Вищий навчальний заклад
Університет економіки та права «КРОК»

М.М. Марусинець, Д.І. Ткач, Д.К. Ткач

ПЕРСОНАЛІЗАЦІЯ МЕДІА ТА АЛГОРИТМІЧНА ЖУРНАЛІСТИКА (ЧАСТИНА 1)

м. Київ – 2026 р.

УДК 070:004.9
М-29

М-29 Марусинець М.М., Ткач Д.І., Ткач Д.К. ПЕРСОНАЛІЗАЦІЯ МЕДІА ТА АЛГОРИТМІЧНА ЖУРНАЛІСТИКА(частина 1): навчально – методичний посібник. Київ: Університет «КРОК», 2026. 438 с.

Навчально-методичний посібник присвячено аналізу персоналізації медіаконтенту та застосуванню алгоритмів у сучасній журналістиці. Розглянуто принципи роботи рекомендаційних систем, використання даних аудиторії, автоматизоване створення і поширення новин, а також вплив алгоритмів на редакційну політику. Окрему увагу приділено етичним, правовим і професійним аспектам алгоритмічної журналістики. Підготовлено з використанням технологій ШІ за науковим контролем авторів. Видання орієнтоване на студентів, викладачів, журналістів і дослідників медіасфери.

ЗМІСТ

Передмова

Від масової аудиторії до аудиторії одного. Уявіть світ, де кожна газета друкується персонально для вас. Де заголовки відображають саме ті теми, що вас турбують, статті написані в стилі, який вам найзручніший для сприйняття, а фотографії підібрані з урахуванням ваших естетичних вподобань. Де вечірній випуск новин на телебаченні має унікальну версію для кожного глядача, а радіостанція грає саме ту музику, яку ви хочете почути цієї миті. Ще двадцять років тому це звучало як утопія або антиутопія — залежно від точки зору. Сьогодні це реальність, у якій ми живемо щодня.

Коли 24 вересня 1833 року Бенджамін Дей випустив перший номер "The Sun" у Нью-Йорку, він створив модель, що домінуватиме наступні півтора століття: одне повідомлення для масової аудиторії. Технологія друку дозволяла тиражувати ідентичний контент для тисяч, потім мільйонів читачів. Радіо та телебачення посилили цю парадигму — one-to-many комунікація, де медіа визначали порядок денний, а аудиторія споживала те, що їй пропонували, у встановлений час та в заданій послідовності.

Ця епоха створила спільний досвід: мільйони людей одночасно дивилися фінал популярного серіалу, обговорювали ту саму передову статтю в ранковій газеті, співали пісню, що звучала на всіх радіостанціях. Медіа не просто інформували — вони формували колективну пам'ять, спільні референсні точки, культурний канон.

Інтернет почав руйнувати цю модель ще в 1990-х, але справжня революція відбулася в 2010-х з появою алгоритмічних новинних стрічок. Facebook замінив хронологічну стрічку на персоналізовану у 2009 році. Netflix запустив персональні рекомендації ще раніше, але справді досконалими вони стали після 2012 року. YouTube перейшов від пошукового сервісу до рекомендаційної машини. А в 2016 році з'явився TikTok — платформа, де персоналізація не була особливістю, а самою сутністю продукту.

Сьогодні ми живемо в світі, де дві людини, відкриваючи одну й ту саму новинну платформу, бачать абсолютно різний контент. Де алгоритми, а не редактори, вирішують, які історії заслуговують вашої уваги. Де кордони між новинами, розвагами та рекламою розмиваються в персоналізованому потоці контенту, безперервно

адаптованого до ваших реакцій у реальному часі.

Ми перейшли від масової аудиторії до аудиторії одного. Але чи це тріумф технології, що нарешті дає нам саме те, чого ми прагнемо? Чи це небезпечна фрагментація, що руйнує спільний простір для громадського діалогу? Чи є це неминуча еволюція медіа, до якої ми маємо адаптуватися? Або технологічна пастка, з якої треба шукати вихід?

Для кого цей навчально – методичний посібник. Персоналізація медіа та алгоритмічна журналістика перетнулися на перехресті технології, бізнесу, етики та суспільних наслідків. Це робить тему надзвичайно міждисциплінарною — і надзвичайно важливою для різних професіоналів.

Цей навчально – методичний посібник створено для студентів журналістики та медіакомунікацій — ви входите в професію, де алгоритми вже є співредакторами, а персоналізація — базовою очікуванням аудиторії. Розуміння того, як працюють рекомендаційні системи, як дані формують контент, як балансувати між релевантністю та журналістською відповідальністю — це не опціональні навички, а необхідність для виживання в індустрії. Цей підручник дасть вам теоретичну базу та практичні інструменти для роботи в алгоритмічну епоху.

Практикуючих журналістів та редакторів — ви спостерігаєте, як змінюються правила гри. Аудиторія фрагментується, традиційні метрики успіху втрачають значення, алгоритми соціальних мереж визначають, чи побачить ваш матеріал хтось окрім вас самих. Ви відчуваєте тиск: персоналізувати чи зберігати редакційну цілісність? Оптимізувати для алгоритмів чи для людей? Тут ви знайдете не готові відповіді (їх не існує), а інструменти для прийняття інформованих рішень.

Data scientists та аналітиків у медіа — ви будуєте системи, що визначають, який контент побачать мільйони людей. Технічна досконалість критична, але недостатня. Розуміння журналістських цінностей, етичних імплікацій алгоритмічних рішень, соціального контексту вашої роботи — це те, що відрізняє відповідального технолога від того, хто просто максимізує метрики. Цей навчально – методичний посібник — міст між технологією та гуманітарними науками.

Product managers медіаплатформ — ви приймаєте стратегічні рішення про те, як персоналізація впроваджується, які метрики оптимізуються, як балансується engagement з благополуччям користувачів. Ви перебуваєте між технічними можливостями, бізнес-цілями та етичною відповідальністю. Підручник надає frameworks для навігації в цих складних компромісах.

Дослідників медіа та комунікацій — персоналізація та алгоритмічна журналістика — одна з найгарячіших тем сучасних media studies. Тут ви знайдете огляд ключових теорій, емпіричних досліджень, дискусій та невирішених питань. Це може бути основою для вашої власної дослідницької програми або корисним ресурсом для викладання.

Громадських активістів та policy makers — алгоритми формують громадську сферу, впливають на демократію, можуть посилювати нерівність або сприяти інклюзії. Регулювання персоналізації та алгоритмічної відповідальності — одне з найважливіших питань цифрової політики. Розуміння того, як технологія працює, допоможе формувати ефективні та справедливі правила.

Медіаграмотних громадян — якщо ви просто хочете зрозуміти, чому ваша новинна стрічка виглядає саме так, чому YouTube рекомендує саме це відео, які дані про вас збираються та як вони використовуються — цей підручник для вас. Критична медіаграмотність в алгоритмічну епоху означає розуміння систем, що опосередковують наше сприйняття світу.

Попередні технічні знання не є обов'язковими, хоча базове розуміння цифрових медіа та статистики зробить деякі розділи легшими для сприйняття. Де потрібні технічні деталі, ми пояснюємо їх доступною мовою з візуалізаціями та прикладами. Водночас, підручник достатньо глибокий для тих, хто вже працює в індустрії та хоче систематизувати знання.

Структура та підхід до вивчення. Цей навчально – методичний посібник побудований навколо центральної ідеї: персоналізація медіа та алгоритмічна журналістика — це не просто технологічні інновації, а фундаментальна трансформація медіасистеми зі складними соціальними, етичними та політичними наслідками. Тому ми рухаємось від технічного розуміння до критичного осмислення.

Чотири части нинавчально – методичного посібника.

Частина I: Фундаментальні концепції (Розділи 1-3) закладає основу. Ми розглядаємо еволюцію медіаспоживання від масової аудиторії до індивідуалізації, визначаємо ключові терміни, досліджуємо економіку уваги та основи того, що робить персоналізацію можливою та необхідною. Тут ми також вводимо концепцію алгоритмічної журналістики — від автоматизованих звітів до ШІ-асистованого створення новин.

Після цих розділів ви зрозумієте: чому персоналізація стала домінуючою парадигмою, які фундаментальні зміни вона приносить, і які питання вона ставить перед медіапрофесіоналами та суспільством.

Частина II: Технології та інструменти (Розділи 4-7) занурюється в механізми. Як працюють рекомендаційні системи? Які дані збираються про аудиторію і як їх аналізувати? Як автоматизувати виробництво контенту? Як персоналізувати дистрибуцію? Ця частина найбільш технічна, але ми балансуємо теорію з практичними прикладами та кейсами. Ви не просто дізнаєтеся, що таке collaborative filtering — ви побачите, як Netflix використовує його для рекомендацій, і спробуєте побудувати просту рекомендаційну систему самостійно. Кожен розділ включає практичні завдання: від налаштування аналітики до A/B тестування персоналізованого контенту.

Частина III: Застосування та кейси (Розділи 8-10) показує персоналізацію в дії. Ми детально розглядаємо три ключові сегменти медіаіндустрії: новинні медіа, розважальний контент та маркетинг/реклама. Для кожного аналізуємо специфічні виклики та можливості персоналізації. Як The New York Times балансує між персоналізацією та редакційною цілісністю? Чому TikTok став майстром персоналізації? Як programmatic advertising змінив медіабізнес? Ця частина насичена реальними кейсами від провідних світових платформ та українських медіа.

Частина IV: Етика, регулювання та майбутнє (Розділи 11-15) переходить до критичного осмислення. Персоналізація — це не просто технологія, це вибір суспільства про те, яким має бути медіапростір. Ми досліджуємо фільтр-бульбашки та поляризацію, маніпуляцію та експлуатацію, algorithmic bias та дискримінацію. Розглядаємо правове регулювання від GDPR до EU AI Act. Аналізуємо психологічні наслідки персоналізованого споживання та соціальні ефекти фрагментації аудиторії. Обговорюємо критику персоналізації

та альтернативні моделі. Нарешті, заглядаємо в майбутнє: які технології на горизонті, як може еволюціонувати індустрія, які сценарії розвитку найбільш ймовірні.

Педагогічний підхід. Кожен розділ структурований однаково для полегшення навчання. Кейс-відкриття: Розділ починається з реальної історії чи прикладу, що ілюструє центральну тему. Це робить абстрактні концепції конкретними та актуальними. Теоретична основа: Пояснення ключових концепцій, теорій, досліджень. Ми прагнемо до балансу між академічною строгістю та доступністю. Практичні інструменти: Огляд технологій, платформ, методів. Де можливо — покрокові інструкції та приклади коду. Критичний аналіз: Обговорення обмежень, етичних дилем, непередбачених наслідків. Кожна технологія розглядається критично, а не як безумовне благо.

Кейс-студії: Детальний розгляд реальних прикладів використання — успішних та невдалих. Практичні завдання: Вправи для застосування знань — від аналітичних завдань до hands-on проєктів. Питання для дискусії: Теми без однозначних відповідей, що заохочують критичне мислення та дебати. Додаткові ресурси: Посилання на дослідження, інструменти, документацію для поглибленого вивчення.

Як працювати з навчально – методичним посібником. Його розроблено як гнучкий ресурс, що підтримує різні стилі навчання. Для семестрового курсу: 15 розділів ідеально відповідають 15-тижневому семестру з одним розділом на тиждень. Лекції можуть базуватися на теоретичних частинах, практичні заняття — на завданнях, а дискусії — на етичних питаннях. Для інтенсивного воркшопу: Якщо ви викладаєте короткий курс, можна вибрати ключові розділи. Наприклад: Розділ 1 (контекст) + Розділ 4 (рекомендації) + Розділ 8 (новини) + Розділ 11 (етика) дадуть концентрований огляд за 4 заняття. Для самостійного навчання: Читайте послідовно або переходьте до розділів, що найбільше цікавлять. Обов'язково виконуйте практичні завдання — теорія без практики швидко забувається. Для дослідницьких проєктів: Бібліографія та посилання на дослідження можуть бути стартовою точкою для власних досліджень. "Питання для дискусії" можуть трансформуватися в дослідницькі питання.

Онлайн-ресурси. Оскільки технологія розвивається швидше, ніж друковані підручники, ми підтримуємо онлайн-супровід. Оновлення: Нові кейси, дослідження, інструменти. Код та датасети: Усі приклади

коду та навчальні датасети. Відео-туторіали: Покрокові демонстрації складніших технічних тем. Спільнота: Форум для обговорення, обміну досвідом, питань. Посилання та QR-коди для доступу до цих ресурсів наведені в кінці кожного розділу.

Етична рамка дослідження. Цей навчально – методичний посібник написаний з чіткою етичною позицією, яку важливо зробити експліцитною з самого початку. Персоналізація медіа та алгоритмічна журналістика — це не нейтральні технології. Вони втілюють цінності, створюють нові форми влади, мають реальні наслідки для людей та суспільства. Як автори та викладачі, ми маємо відповідальність не просто навчити "як", а й допомогти осмислити "чи варто" та "за яких умов".

Принципи нашого підходу.

1. Технологічний реалізм, не утопізм і не катастрофізм. Ми відкидаємо обидві крайності: безкритичний техно-оптимізм, що бачить в персоналізації лише благо, і технофобію, що демонізує будь-яку автоматизацію. Персоналізація може робити медіа більш релевантними, доступними, інклюзивними. Вона також може маніпулювати, поляризувати, створювати нові форми нерівності. Важливо не те, чи персоналізація "добра" чи "погана" — вона неминуча. Важливо, як ми її проєктуємо, регулюємо, використовуємо. Технології не детерміновані — їх наслідки залежать від рішень людей.

2. Пріоритет людського благополуччя над метриками engagement. Сучасна персоналізація часто оптимізується для максимізації часу на платформі, кліків, переглядів. Ці метрики не обов'язково корелюють з інформованістю, благополуччям, задоволенням користувачів. Навпаки, вони можуть заохочувати сенсаційність, обурення, тривогу. Ми вважаємо, що етична персоналізація має ставити благополуччя людей вище короткострокової максимізації залученості. "Time well spent" важливіше за просто "time spent". Інформованість важливіша за вірусність. Різноманітність перспектив важливіша за комфорт підтвердження власних переконань.

3. Прозорість та підзвітність алгоритмів. Алгоритми, що визначають, яку інформацію бачать мільйони людей, не можуть бути "чорними скриньками". Користувачі мають право розуміти, чому їм показують саме цей контент. Журналісти та дослідники мають мати можливість аудитувати системи на предмет bias та маніпуляції. Ре-

гулятори мають інструменти для забезпечення відповідальності. Ми підтримуємо рух за Explainable AI, algorithmic accountability, і вважаємо, що непрозорість — не технічна необхідність, а свідомий вибір, часто мотивований захистом конкурентних переваг над суспільним благом.

4. Справедливість та інклюзивність. Algorithmic bias — не технічна помилка, а дзеркало нерівності в даних, що відображають нерівність у суспільстві. Системи персоналізації можуть відтворювати та посилювати дискримінацію за расою, гендером, класом, географією. Етична розробка вимагає свідомих зусиль для виявлення та мітигації bias. Це означає: різноманітність команд розробників, інклюзивні датасети, fairness metrics, регулярні аудити, готовність жертвувати оптимальністю заради справедливості.

5. Захист приватності як фундаментальне право. Персоналізація вимагає даних, але збір даних не може бути необмеженим. Приватність — не перешкода для інновацій, а фундаментальне право, що має поважатися. Ми підтримуємо принципи data minimization (збирати лише необхідне), purpose limitation (використовувати лише для заявлених цілей), user control (давати людям реальний контроль над їхніми даними). Privacy-preserving персоналізація можлива — вона вимагає більше зусиль, але це етичний імператив.

6. Демократизація, не концентрація влади. Персоналізація може демократизувати доступ до якісного контенту, робити професійні інструменти доступними для всіх. Або вона може концентрувати владу в руках кількох платформ, що контролюють алгоритми та дані. Ми підтримуємо open-source підходи, data portability, interoperability платформ, public interest technology. Потужні інструменти персоналізації не повинні бути монополією корпорацій — вони мають бути доступні громадянському суспільству, незалежним медіа, дослідникам.

7. Збереження спільного публічного простору. Індивідуальна релевантність важлива, але не за рахунок повної атомізації медіадосвіду. Суспільству потрібні спільні референсні точки, shared reality, простір для зустрічі різних перспектив. Повна персоналізація ризикує створити мільйони ізольованих реальностей, де неможливий конструктивний діалог. Етична персоналізація має включати механізми для збереження serendipity, exposure to diversity, shared experiences.

Як ці принципи втілені в цьому навчально – методичному посібнику. Ці етичні принципи — не просто декларації. Вони структурують наш підхід. Кожна технологія розглядається критично: Ми не просто пояснюємо, як щось працює, а й обговорюємо потенційні зловживання, непередбачені наслідки, етичні trade-offs. Рівна увага до альтернатив: Коли є альтернативи домінуючим підходам — від decentralized social media до user-controlled персоналізації — ми їх представляємо, навіть якщо вони менш популярні. Voices from margins: Ми свідомо включаємо перспективи, часто маргіналізовані в технологічних дискусіях: Глобальний Південь, вразливі групи, критики surveillance capitalism. Практичні інструменти для етичної роботи: Не просто "етика погана", а конкретні checklists, frameworks, tools для етичного дизайну та оцінки персоналізаційних систем. Заохочення критичного мислення: Питання для дискусії часто не мають правильних відповідей. Ми хочемо виховувати не технічних виконавців, а рефлексивних практиків, що ставлять складні питання.

Запрошення до етичної рефлексії. Читаючи цейнавчально – методичний посібник, постійно запитуйте себе. Для кого це добре? Хто виграє від цієї технології, а хто може програти? Які альтернативи? Чи це єдиний спосіб вирішити проблему, чи є інші підходи? Які довгострокові наслідки? Що станеться, коли технологія масштабується та стане нормою? Хто приймає рішення? Чи мають ті, кого торкаються рішення, голос у прийнятті цих рішень? Чи це зробить світ кращим? Не для метрик компанії, а для реальних людей та суспільства.

Етика персоналізації — це не фіксований набір правил, а процес постійної рефлексії та діалогу. Ми не претендуємо на остаточні відповіді, але вважаємо за обов'язок ставити правильні питання.

Останнє слово перед початком. Персоналізація медіа та алгоритмічна журналістика — це одні з найбільш трансформативних та контроверсійних феноменів сучасності. Вони впливають на те, як ми дізнаємось про світ, як формуємо думки, як взаємодіємо один з одним, як функціонує демократія.

Цей навчально – методичний посібник— ваш путівник у складному ландшафті, де технологія, етика, бізнес та політика переплетені нерозривно. Ми не обіцяємо простих відповідей — їх немає. Але ми надаємо інструменти для розуміння, критичної оцінки та відпові-

дальної дії. Епоха масової аудиторії закінчилась. Епоха аудиторії одного тільки починається. Її форма ще не визначена. І ви маєте голос у визначенні того, якою вона стане. Ласкаво просимо в подорож від персоналізації як технології до персоналізації як вибору суспільства.

Автори
М.М. Марусинець,
Д.І. Ткач, Д.К. Ткач
Січень 2026

ЧАСТИНА І: ФУНДАМЕНТАЛЬНІ КОНЦЕПЦІЇ

Розділ 1: Еволюція медіаспоживання

1.1. Історична перспектива: Епоха масових медіа

Народження масової комунікації. Франкфурт, 1450 рік. Йоганн Гутенберг завершує роботу над механізмом, що назавжди змінить людську історію — друкарським верстатом з рухомими літерами. Те, що раніше вимагало місяців роботи монахів-переписувачів, тепер можна було виробити за дні. Одна книга перетворювалася на сотні ідентичних копій.

Але справжня революція була не в швидкості, а в самій ідеї масового тиражування. Вперше в історії одне й те саме повідомлення могло досягти тисяч людей одночасно, незалежно від їхнього місцезнаходження. Біблії Гутенберга були абсолютно ідентичними — той самий текст у Парижі та Кракові, в руках селянина та короля. Це був перший крок до парадигми, що домінуватиме наступні п'ять століть: one-to-many комунікація, де один відправник передає однакове повідомлення багатьом одержувачам.

Три хвилі масових медіа. Історію масових медіа можна розділити на три великі хвилі, кожна з яких радикально розширювала охоплення та вплив комунікації, але зберігала базову модель: однакове повідомлення для масової аудиторії.

Перша хвиля: Друковані медіа (1450-1920). Після друкарського верстата потрібно було ще кілька століть, щоб масові медіа перетворилися на справді масовий феномен. Ключовим моментом стала поява penny press у 1830-х роках. Вересень 1833 року, Нью-Йорк. Бенджамін Дей випускає перший номер "The Sun" — газети, що коштувала всього один цент замість традиційних шести. Його інновація була не лише в ціні, а в бізнес-моделі: замість залежності від партійної підтримки чи передплати заможних читачів, газета фінансувалася рекламою. Чим більше читачів — тим більше рекламодавців готові платити. Це створило базову економіку масових медіа: максимізація аудиторії як головна мета. Контент мав привертати якомога більше уваги, щоб продавати цю увагу рекламодавцям. Модель, що здається нам знайомою з соціальних мереж, була винайдена майже два століття тому.

Характеристики епохи друкованих медіа. Одностороння комунікація: Редактори вирішували, що важливо, читачі споживали. Зворотний зв'язок був мінімальним та повільним — листи до редакції, якщо їх публікували, з'являлися через дні чи тижні. Географічна та часова синхронізація: Мільйони людей читали ту саму газету того самого ранку, обговорювали ті самі заголовки під час обіду. Медіа створювали спільний порядок денний для всієї країни або міста. Дефіцит інформації: Звичайна людина мала доступ до обмеженої кількості джерел — місцева газета, можливо кілька національних видань. Інформаційне середовище було скупим, але структурованим. Професійна фільтрація: Журналісти та редактори виконували роль gatekeepers — вони вирішували, які події стають новинами, які точки зору представлені, що заслуговує передової сторінки. Повільний цикл новин: Газети виходили раз на день (або раз на тиждень для регіональних видань). Події мали час "дозріти" до статусу новини, журналісти — перевірити факти та написати зважені матеріали.

Вплив на суспільство. Друковані медіа створили те, що Бенедикт Андерсон назвав "уявленими спільнотами" (imagined communities). Читаючи ту саму газету, люди, які ніколи не зустрінуться особисто, відчували себе частиною єдиної нації. Газети формували національну ідентичність, громадську сферу, політичний дискурс.Водночас, вони були інструментом влади та контролю. Той, хто володів друкарськими пресами, контролював наратив. Уряди цензурували, власники маніпулювали, "жовта преса" перетворювала новини на сенсації заради тиражів.

Друга хвиля: Радіо та телебачення (1920-1990). 2 листопада 1920 року, Піттсбург. Радіостанція KDKA транслює результати президентських виборів — першу комерційну радіотрансляцію новин в історії. Кілька сотень людей з радіоприймачами слухають одночасно. Це початок нової ери. Радіо, а згодом телебачення, посилили масовість медіа до безпрецедентного рівня. Якщо газети досягали мільйонів протягом дня, радіо та ТБ досягали десятків мільйонів одночасно.

Унікальні характеристики електронних медіа. Синхронне споживання в масштабі нації: 21 липня 1969 року приблизно 650 мільйонів людей по всьому світу одночасно дивились, як Ніл Армстронг ступає на Місяць. Такого рівня синхронного глобального досвіду не існува-

ло в історії людства до телебачення. Дефіцит каналів: У більшості країн було лише кілька телеканалів (часто контрольованих державою). У США "золота ера" телебачення означала домінування трьох мереж: ABC, NBC, CBS. Вибір був мінімальним, але саме тому вплив був максимальним.

Програмування як контроль. Глядачі не обирали, що дивитись — вони обирали між тим, що транслювалось у конкретний час. Прайм-тайм означав не просто популярний час перегляду, а концентрацію аудиторії навколо обмеженої кількості програм. Ілюзія близькості: Радіо та особливо телебачення створювали відчуття інтимності. Ведучі новин заходили в ваші вітальні, говорили безпосередньо до вас. Уолтер Кронкайт стає "найбільш довіреною людиною Америки" не через посаду, а через паразоціальні відносини з мільйонами глядачів. Емоційний вплив візуального: Телебачення змінило природу новин. Війна у В'єтнамі стала "першою телевізійною війною" — образи людських страждань у прайм-тайм змінили громадську думку способом, неможливим для друкованих медіа.

Піковий момент масових медіа. Фінальна серія "M*A*S*H" у 1983 році зібрала 125 мільйонів глядачів у США — майже половина всього населення країни. Наступного дня вся нація обговорювала ту саму подію. Це був апогей масових медіа: максимальна концентрація уваги, мінімальна фрагментація аудиторії. Такий рівень культурної синхронізації здається неможливим сьогодні. Навіть найпопулярніші події — фінали Super Bowl, Оскар, королівські весілля — досягають лише частки аудиторії, яку регулярно збирали звичайні телепрограми в епоху трьох каналів.

Економіка уваги в електронну епоху. Бізнес-модель залишалася тією ж, що й у penny press: продавати увагу аудиторії рекламодавцям. Але масштаб зріс експоненційно. Тридцятисекундний рекламний ролик під час прайм-тайму коштував мільйони доларів, тому що досягав десятків мільйонів людей одночасно. Це створило потужний стимул для максимізації охоплення через апелювання до найширшого спільного знаменника. Телебачення прайм-тайму уникало контроверсійних тем, складного контенту, всього, що могло відштовхнути хоча б частину аудиторії. Результат — гомогенізація контенту, домінування середнього смаку, маргіналізація нішевих інтересів.

Третя хвиля: Кабельне та супутникове ТБ (1980-2000). 1 червня 1980 року. Запускається CNN — перший 24-годинний новинний канал. Концепція здається абсурдною: хто дивитиметься новини цілодобово? Виявляється, дуже багато людей. Кабельне та супутникове телебачення стали першою тріщиною в монолітній структурі масових медіа. Якщо раніше були три канали для всіх, тепер з'являлися десятки, потім сотні спеціалізованих каналів.

Початок фрагментації. Нішування: ESPN для спортивних фанатів, MTV для молоді, Discovery для зацікавлених наукою, History Channel для любителів історії. Кожен канал знаходив свою аудиторію. Демографічна сегментація: Рекламодавці більше не платили за недиференційовану масу глядачів. Канали могли обіцяти конкретні демографічні групи: молоді чоловіки 18-34 на ESPN, жінки 25-54 на Lifetime. Політична поляризація: Fox News (1996) та MSNBC займали чіткі ідеологічні позиції, замість фасаду об'єктивності мережевих новин. Глядачі могли обирати новини, що відповідали їхнім політичним переконанням. Часовий зсув: VCR (відеомагнітофони) дозволили записувати програми та дивитись у зручний час. Це був перший крок від диктованого програмування до контролю користувача.

Проте це була все ще масова модель — канали транслювали той самий контент для всіх глядачів. Просто тепер було більше каналів на вибір. Справжня персоналізація ще попереду.

Характеристики епохи масових медіа: узагальнення. Дивлячись на всі три хвилі разом, можемо виділити ключові характеристики парадигми масових медіа. Однаковий контент для всіх. Незалежно від того, хто ви, де ви, коли ви споживаєте медіа — ви отримуєте той самий контент, що й усі інші. Редактори, продюсери, програмні директори вирішували за всіх. Приклад: Вечірні новини о 20:00 були ідентичними для підлітка в Каліфорнії та пенсіонера у Флориді, для вченого та будівельника. Всі отримували той самий набір історій у тому самому порядку з тією самою тривалістю.

Дефіцит як норма. Обмежена кількість каналів, газет, радіостанцій. Обмежена кількість годин мовлення (багато станцій припиняли трансляцію вночі). Обмежений простір на газетній сторінці. Цей дефіцит надавав величезну владу тим, хто контролював канали комунікації. Якщо ваша історія не потрапила у вечірні новини або

на передову газети, вона фактично не існувала для широкої публіки.

Професійні gatekeepers. Журналісти, редактори, продюсери виконували роль фільтрів між подіями та аудиторією. Вони вирішували, що є новиною, а що ні. Які точки зору заслуговують представлення. Скільки часу/простору присвятити темі. Як фреймувати історію. Це давало владу, але й відповідальність. Професійна журналістика розробила етичні кодекси, стандарти верифікації, практики балансу (принаймні теоретично).

Одностороння комунікація. Медіа говорили, аудиторія слухала. Зворотний зв'язок існував (листи до редакції, телефонні опитування), але був мінімальним, повільним та фільтрованим. Діалог між медіа та аудиторією був виключенням, не правилом.

Синхронне споживання. Медіаподії траплялися в конкретний час. Якщо ви пропустили улюблену програму о 20:00 у вівторок — ви її пропустили назавжди (до епохи повторів та відеозапису). Ця синхронність створювала спільний досвід. Наступного дня колеги на роботі, однокласники в школі обговорювали ту саму програму, що бачили напередодні. Медіа були частиною соціального ритуалу.

Економіка масштабу. Створення контенту було дорогим (друкарні, студії, передавачі), але розповсюдження додаткових копій — відносно дешевим. Це створювало стимул до максимізації аудиторії: чим більше глядачів/читачів, тим нижча вартість на одного. Результат — гомогенізація контенту під "середній смак", уникнення контроверсій, фокус на універсально привабливих темах.

Культурний та соціальний вплив масових медіа. Епоха масових медіа мала глибокий вплив на суспільство, багато аспектів якого ми починаємо усвідомлювати лише тепер, коли ця епоха закінчується. Формування національної ідентичності. Масові медіа, особливо телебачення, створили спільну культурну мову. Жителі великих міст і маленьких містечок дивилися ті самі програми, знали тих самих знаменитостей, розуміли ті самі культурні референси.

У багатонаціональних країнах як США чи Індія телебачення відіграло ключову роль у створенні відчуття національної єдності попри регіональні відмінності. Створення спільної реальності. Коли існує обмежена кількість джерел новин, більшість людей оперує спільним набором фактів про світ. Можна не погоджуватись щодо інтерпретації, але базові факти були спільними.

Уолтер Кронкайт завершував вечірні новини фразою "And that's the way it is" ("І так воно є") — і для мільйонів американців це дійсно визначало, "як воно є". Існувала shared reality.

Моменти культурної єдності. Деякі медіаподії ставали точками збігу для всього суспільства. Висадка на Місяць (1969). Виступ Beatles на Ed Sullivan Show (1964). Фінал "M*A*S*H" (1983). Челленджер трагедія (1986). Fall of Berlin Wall (1989). Ці моменти створювали колективну пам'ять — відчуття, що ви були частиною історичного моменту разом з мільйонами інших.

Гегемонія мейнстриму. Обмежена кількість каналів означала, що мейнстримні погляди, домінуючі культурні норми, політичний центр отримували переважне представлення. Маргінальні ідеї, субкультури, альтернативні точки зору мали мінімальний доступ до широкої аудиторії. Це мало як позитивні (стабільність, консенсус), так і негативні (виключення, конформізм) наслідки.

Початок кінця: передвісники змін. Навіть на піку домінування, масові медіа містили зерна власної трансформації. VCR та time-shifting (1980-ті): Можливість записувати програми порушила синхронність споживання. Люди почали дивитись те, що хочуть, коли хочуть. Кабельне ТБ та нішування (1980-90-ті): Сотні каналів фрагментували аудиторію. Загальнонаціональні події ставали рідшими. Пульт дистанційного керування (1980-ті): Можливість легко перемикати канали створила культуру channel surfing та зменшила лояльність до окремих каналів. 24-годинний новинний цикл (1980-90-ті): CNN та інші цілодобові канали прискорили темп новин, зменшили час на верифікацію, посилили конкуренцію за увагу через сенсаційність.

Але справжня революція чекала попереду — з появою інтернету. Ностальгія vs критика: як оцінювати епоху масових медіа? Сьогодні, в епоху фрагментації та персоналізації, існує певна ностальгія за "золотою ерою" масових медіа. Спільний досвід, довірені голоси, shared reality — це те, що багато хто відчуває як втрачене.

Але важливо не ідеалізувати минуле. Що було хорошого. Професійні стандарти журналістики. Спільна основа фактів для публічної дискусії. Культурні моменти єдності. Відносна стабільність інформаційного середовища. Що було проблематичним. Концентрація влади у руках небагатьох. Виключення маргіналізованих голосів. Обмежений вибір та контроль аудиторії. Сприяння конформізму

та цензурі. Маніпуляція громадською думкою власниками медіа та урядами.

Масові медіа не були ні утопією, ні антиутопією — вони були продуктом своїх технологічних можливостей та соціальних структур. Розуміння цієї епохи критично для оцінки того, що втрачено та що здобуто в переході до персоналізованих медіа.

1. КОНТРОЛЬНІ ПИТАННЯ

1. Архітектура: Опишіть принцип "Broadcasting". Чому ця модель вважається неефективною з точки зору задоволення індивідуальних потреб користувача?

2. Роль редактора: Хто такий "Гейткіпер"? Наведіть приклад ситуації з минулого століття, коли рішення одного редактора могло змінити хід історії або суспільну думку.

3. Зворотний зв'язок: Як працював Feedback Loop (петля зворотного зв'язку) в епоху телебачення? Чому він був "відкладеним"?

4. Монополія на істину: Як обмежена кількість каналів (наприклад, лише 3 телеканали в країні) впливала на уніфікацію суспільної думки?

5. Технологічний детермінізм: Як фізичні властивості паперу (неможливість виправити помилку після друку) впливали на стандарти журналістської перевірки фактів? (Це вимагало суворішої пре-модерації, на відміну від сучасної пост-модерації).

2. ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. Симуляція "Редколегія 1980" (The Gatekeeper). Умова: Ви — головний редактор вечірньої газети. У вас є місце лише для 3 новин на першій шпальті. Вхідні події (обрати 3 з 6). Землетрус в Азії (1000 загиблих). Місцевий завод відкрив новий цех (100 робочих місць). Президент країни захворів. Перемога національної збірної з футболу. Зростання цін на хліб на 5 копійок. Науковці знайшли новий вид метеликів. Завдання: Оберіть 3 новини і обґрунтуйте вибір. Що ви "викинули" (відфільтрували)? Як це вплине на те, що люди будуть обговорювати завтра?

Завдання 2. Порівняльний аналіз "Час vs Увага". Дія: Порівняйте програму телепередач за 1995 рік (лінійна сітка) і скріншот головної сторінки Netflix/YouTube сьогодні. Аналіз: Напишіть коротке есе

(100 слів) про те, як змінилося поняття "Прайм-тайм". (Раніше це був час 20:00–22:00, тепер прайм-тайм — це будь-яка секунда, коли користувач відкрив додаток).

Завдання 3. Дослідження монополії. Завдання: Знайдіть інформацію про те, скільки коштував запуск телеканалу в 1990 році, і скільки коштує запуск YouTube-каналу сьогодні. Зробіть висновок про "демократизацію медіа".

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. McLuhan, M. *Understanding Media*. (Розділи про друковане слово та телебачення).

2. Lippmann, W. *Public Opinion*. (Класична праця про те, як медіа створюють "картинки в нашій голові").

3. Postman, N. *Amusing Ourselves to Death*. (Критика телебачення як медіа, що перетворює все на розвагу).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Mass Media (Масові медіа): Засоби комунікації (преса, радіо, ТБ), призначені для передачі інформації великим, розосередженим аудиторіям.

• Gatekeeping (Гейткіпінг/Фільтрація): Процес відбору, обробки та контролю інформації, яка потрапляє до масової аудиторії.

• Agenda Setting (Встановлення порядку денного): Теорія, згідно з якою медіа визначають важливість подій для громадськості, приділяючи їм більше уваги та місця.

• Broadcast (Мовлення): Метод передачі даних, при якому один сигнал одночасно надсилається всім отримувачам у мережі (модель "один до багатьох").

• Linear Media (Лінійні медіа): Медіа, які споживаються у послідовності та в час, визначений мовником (наприклад, прямий ефір ТБ), на відміну від On-Demand (за запитом).

1.2. Перехід до цифрового: інтерактивність та вибір

Народження інтерактивного медіапростору. 6 серпня 1991 року. Тім Бернерс-Лі публікує короткий опис World Wide Web на форумі Usenet. Він запрошує всіх спробувати систему гіпертекстових документів, що з'єднані посиланнями. Кілька сотень людей помічають. Світ ще не розуміє, що щойно народилась технологія, яка за два десятиліття повністю перебудує медіаландшафт.

На відміну від попередніх медіатехнологій, що посилювали модель one-to-many, інтернет був задуманий як децентралізована мережа рівноправних вузлів. Кожен міг бути не лише споживачем, а й виробником контенту. Кожен міг обирати свій шлях через інформаційний простір за допомогою гіперпосилань.

Проте революція відбулася не миттєво. Перехід від масових медіа до цифрової екосистеми зайняв майже три десятиліття та пройшов кілька чітких фаз. Фаза 1: Web 1.0 — Статичний інтернет (1991-2004). Ранній інтернет був схожий на гігантську цифрову бібліотеку. Вебсайти були статичними сторінками, створеними професіоналами або технічно підкованими ентузіастами. Звичайні користувачі могли читати, але не створювати контент.

Характеристики раннього веб. Портали як нові gatekeepers. Yahoo!, Excite, AOL створювали курованні каталоги вебсайтів. Редактори вручну відбирали та категоризували сайти — цифрова версія традиційного редакційного контролю. Потрапити в каталог Yahoo! означало існування у вебі; не потрапити — залишитись невидимим.

Перші пошукові системи. AltaVista (1995), а згодом Google (1998) запропонували радикально інший підхід: алгоритмічний пошук замість людського курування. Вперше машина, а не людина-редактор, вирішувала, яка інформація релевантна для вас. Google PageRank був революційним саме тому, що визначав важливість сторінки на основі того, скільки інших сторінок на неї посилаються — демократичний принцип, де "голосує" вся мережа, а не редактор.

Персональні вебсайти та блоги. GeoCities (1994) дозволив звичайним людям створювати власні вебсторінки. Blogger (1999) зробив публікацію ще простішою. Вперше мільйони людей отримали власні медіаканали без потреби в друкарнях чи телестудіях.

Але аудиторія була фрагментованою. Якщо у телебаченні мільйони дивились одне, в ранньому вебі сотні мікроаудиторій читали тисячі блогів. Email розсилки та RSS. Email став першим масовим інструментом персоналізації. Підписавшись на розсилку, ви самі обирали, який контент отримувати. RSS-фіди (2000-ті) дозволили агрегувати оновлення з улюблених сайтів у персональному "читачі". Це був перший крок від "споживай те, що дають" до "обирай, що споживати".

Медіа у Web 1.0. експериментування та скептицизм. Традиційні медіа ставились до інтернету зі змішаною недовірою та цікавістю. Рання присутність: CNN.com запустився у 1995 році, The New York Times онлайн — у 1996. Спочатку це були просто цифрові версії друкованих чи телевізійних продуктів — той самий контент, перенесений у веб.

Безкоштовність як норма: Культура безкоштовного контенту сформувалась рано. Користувачі очікували, що інформація у вебі має бути вільною. Спроби запровадити платний доступ (paywall) здебільшого провалювались. Медіа ще не знали, як монетизувати онлайн-аудиторію. Обмежена інтерактивність: Коментарі під статтями були рідкістю. Більшість новинних сайтів залишались односторонніми — як друковані газети, тільки на екрані. Революція участі (participation) ще попереду.

Фаза 2. Web 2.0 — Соціальний та інтерактивний інтернет (2004-2010). 4 лютого 2004 року. Марк Цукерберг запускає TheFacebook.com для студентів Гарварду. Через рік платформа доступна у більшості університетів США. Через два роки — для всіх. Facebook не винайшов соціальні мережі (Friendster та MySpace були раніше), але саме він визначив парадигму Web 2.0: інтернет як платформа для користувацького контенту та соціальної взаємодії.

Ключові платформи та їх інновації. Facebook (2004). Реальні імена та ідентичності (на відміну від анонімних форумів). Соціальний граф: з'єднання між людьми як основа платформи. News Feed (2006): алгоритмічно курована стрічка контенту від друзів. "Like" кнопка (2009): найпростіша форма взаємодії, що генерує масивні дані про вподобання.

YouTube (2005). "Broadcast yourself": кожен може бути відеопродюсером. Коментарі та рейтинги: аудиторія оцінює контент.

Рекомендаційний алгоритм, машина підбирає наступне відео. Монетизація для креаторів (2007): можливість заробляти на контенті

Twitter (2006). Мікроблогінг: 140 символів як нова форма висловлювання. Hashtags: користувацька організація контенту навколо тем. Retweet: вірусне поширення без втручання платформи. Real-time: інформація швидше, ніж у традиційних новинах.

Wikipedia (2001, але розквіт у 2004-2010). Колективне створення енциклопедії. Будь-хто може редагувати: радикальна довіра до crowd. Безкоштовна альтернатива професійним енциклопедіям. Зміна ролі аудиторії: від споживачів до "prosumers". Термін "prosumer" (producer + consumer) описує фундаментальну зміну: аудиторія більше не пасивна, вона активно створює контент.

User-Generated Content (UGC) як норма. До соцмереж створення медіаконтенту вимагало спеціальних навичок, обладнання, доступу до каналів дистрибуції. Соцмережі демократизували всі три. Створення: камера у смартфоні, прості інтерфейси. Редагування: вбудовані інструменти. Дистрибуція: публікація одним кліком. Раптом кожен міг мати аудиторію. Блогер з спальні міг мати більше читачів, ніж місцева газета. YouTuber міг конкурувати з телеканалами.

Зміна природи медіаспоживання. Web 2.0 трансформував споживання з пасивного на активне. Коментування: кожна стаття, відео, пост стають точкою для дискусії. Шерінг: поширення контенту стає актом самовираження. Рейтингування, аудиторія колективно оцінює якість. Ремікс, створення нового контенту на базі існуючого (меми!).

Виникнення нових професій. YouTuber, блогер, Instagram-інфлюенсер, TikTok-креатор — професії, що не існували до Web 2.0. Мільйони людей почали заробляти на створенні контенту для цифрових платформ.

Фаза 3: Мобільна революція (2007-2015). 9 січня 2007 року. Стів Джобс представляє iPhone. "Revolutionary mobile phone", "Widescreen iPod", "Breakthrough internet device" — три пристрої в одному. Насправді це був четвертий пристрій: персональний медіацентр, завжди з тобою. Смартфони перетворили медіаспоживання з діяльності, прив'язаної до місця (стілець перед ТБ, стіл з газетою), на постійну активність, інтегровану в кожен момент дня.

Характеристики мобільної ери. Always-on культура. Медіа більше

не те, що ти робиш у певний час. Це постійна фонова активність. Перевіряємо новини в черзі на каву, дивимось відео в метро, скролимо соцмережі перед сном. Середній користувач смартфону перевіряє пристрій 96 разів на день (кожні 10 хвилин у період неспання). Медіаспоживання стало гіперфрагментованим: сотні мікросесій замість кількох тривалих.

Геолокація та контекстуальність. Смартфон знає, де ти. Це дозволяє контент адаптувати до контексту наступне. Локальні новини автоматично. Рекомендації ресторанів поблизу. Погода для твого міста. Події навколо тебе.

Камера в кишені. Кожен стає потенційним журналістом. Важливі події (протести, катастрофи, поліцейське насильство) вперше фіксуються десятками камер одночасно. Традиційні медіа починають покладатись на user-generated відео. App economy. Замість універсального веб-браузера — спеціалізовані аплікації для кожної платформи. Facebook app, Twitter app, Instagram app, YouTube app. Кожна створює окрему екосистему, утримуючи користувача всередині.

Платформи мобільної ери. Instagram (2010): Візуальний мікроблогінг, оптимізований для смартфонів. Фільтри роблять будь-яке фото естетичним. Квадратний формат та простота використання. До 2012 — 100 мільйонів користувачів. Snapchat (2011): Ефемерність: контент зникає через 24 години. Це звільняє від тиску перфекціонізму, заохочує автентичність. Stories формат стане найбільш копійованою інновацією (Instagram Stories, Facebook Stories, YouTube Stories). Messaging apps: WhatsApp, Telegram, WeChat перетворюються на повноцінні медіаплатформи. Новини поширюються через приватні чати швидше, ніж через офіційні канали. "Dark social" — обмін у месенджерах — стає критично важливим, але невидимим для традиційної аналітики.

Фаза 4: Алгоритмічні стрічки та персоналізація (2009-2016). Березень 2009 року. Facebook замінює хронологічну стрічку на алгоритмічну. Замість показу всіх постів від друзів у порядку публікації, алгоритм вирішує, що ви найімовірніше захочете бачити. Це була тиха революція. Більшість користувачів навіть не помітили. Але наслідки були величезними: вперше алгоритм, а не хронологія чи редактор, став головним куратором медіадосвіду мільярдів людей. Як працюють алгоритмічні стрічки. Замість простого правила (нові

пости зверху), складні моделі машинного навчання передбачають, з яким контентом ви найімовірніше взаємодієте:

Сигнали, що аналізуються. Історія взаємодій: що ви лайкали, коментували, ділилися. Час, проведений на пості: чи зупинились ви, чи прогорнули?Тип контенту: відео, фото, текст, посилання. Автор: як часто ви взаємодієте з цією людиною/сторінкою. Популярність: скільки інших людей взаємодіють. Свіжість: нові пости мають бонус.Контекст: час доби, пристрій, локація. Оптимізаційна мета. Максимізувати engagement — час на платформі, кількість взаємодій. Чим довше ви скролите, чим більше лайкаєте /коментуєте, тим успішніший алгоритм з точки зору платформи.

Поширення персоналізації. 2012-2016. Майже всі великі платформи переходять на алгоритмічні стрічки. Instagram (2016): від хронології до алгоритму. Twitter експериментує з "While you were away". YouTube трансформується з пошукового сервісу в рекомендаційну машину. LinkedIn впроваджує алгоритмічну стрічку. Новинні медіа адаптуються. Традиційні медіа оптимізують контент для алгоритмів соцмереж. Заголовки створюються для максимізації кліків (clickbait). Контент тестується через A/B тести. Публікації таймуються за піками активності аудиторії. Формати адаптуються: коротші статті, більше візуалів, відео.

Таблиця 1.1.

Порівняння епох масові медіа, цифрові медіа

Аспект	Масові медіа	Цифрові медіа
Контроль контенту	Професійні редактори	Алгоритми + crowd
Роль аудиторії	Пасивні споживачі	Активні prosumers
Вибір	Обмежений (3-100 каналів)	Практично безмежний
Зворотний зв'язок	Повільний, фільтрований	Миттєвий, публічний
Споживання	Синхронне, за розкладом	Асинхронне, on-demand
Бар'єр входу	Високий (потрібні ресурси)	Низький (смартфон)
Економіка	Масштаб (чим більше, тим дешевше)	Long tail (ніші прибуткові)
Увага	Концентрована (прайм-тайм)	Фрагментована (завжди-on)

Складено автором. Джерело: Вплив соціальних медіа на традиційні ЗМІ // Філологічні трактати (2025). https://www.philol. vernadskyjournals. in.ua/ journals /2025/1_2025/part_2/33.pdf

Парадокс вибору: більше опцій, менше задоволення? Психолог Баррі Шварц у книзі "The Paradox of Choice" стверджує: надмірний вибір не робить нас щасливішими, а навпаки — паралізує та виснажує. У медіаконтексті це проявляється як Decision fatigue: Стільки контенту доступно, що важко вибрати, що дивитись. Netflix scroll — безкінечне гортання каталогу без вибору. FOMO (Fear of Missing Out): Завжди є відчуття, що десь є щось краще, цікавіше. Складно присутньому моменту. Shallow engagement: Замість глибокого занурення в одну книгу чи довгу статтю — поверхневе скакання між десятками джерел.

Персоналізація як відповідь. Алгоритмічні рекомендації частково вирішують парадокс вибору. Замість вибору з 10,000 фільмів на Netflix, алгоритм пропонує 10-20, персоналізованих під вас. Вибір спрощується, але хто контролює, що потрапляє у цей короткий список? Демократизація чи нова нерівність? Цифрові медіа обіцяли демократизацію: кожен може мати голос, талант знайде аудиторію, gatekeepers більше не контролюють доступ.

Реальність виявилась складнішою. Winner-takes-all ефект. Інтернет не створив рівність, а нову форму концентрації. Кілька мегаплатформ (Google, Facebook, Amazon) контролюють величезну частку цифрової економіки. Кілька топ-креаторів отримують левову частку уваги та доходів. Algorithmic gatekeeping. Редактори газет замінені алгоритмами, але останні — теж gatekeepers, просто менш прозорі. Якщо алгоритм YouTube не рекомендує ваше відео, воно не існує для більшості потенційної аудиторії. Цифровий розрив. Доступ до високошвидкісного інтернету, новітніх пристроїв, цифрових навичок нерівномірний. Демократизація працює лише для тих, хто має інфраструктуру та грамотність. Монетизація talent. Створювати контент легко, заробляти на ньому — важко. Більшість креаторів заробляють мало або нічого. Успіх вимагає не лише таланту, а розуміння алгоритмів, маркетингу, постійного виробництва.

Виклики для традиційних медіа. Перехід до цифрового ставить традиційні медіа перед екзистенційними викликами. Економічна модель зруйнована:Класифайди (величезне джерело доходів газет) мігрували на Craigslist. Рекламні бюджети перетекли до Google та Facebook. Аудиторія очікує безкоштовний контент. Увага фрагментована: Замість мільйонів глядачів вечірніх новин —

розпорошення між тисячами джерел. Важко утримувати масову аудиторію. Авторитет підірваний: Експертні журналісти конкурують з блогерами, YouTubers, Twitter-коментаторами. "Mainstream media" стає пейоративним терміном для деяких аудиторій. Швидкість vs якість: 24/7 новинний цикл та соцмережі вимагають публікації миттєво. Менше часу на верифікацію, поглиблений аналіз. Кількість перемагає якість. Платформи як посередники: Більшість людей читають новини не на сайтах медіа, а у соцмережах. Facebook та Google стають головними дистрибуторами, забираючи дані про аудиторію та рекламні доходи.

Адаптація або вимирання. Деякі медіа успішно адаптувалися. The New York Times: digital-first стратегія, платні підписки, інвестиції в якісну журналістику. The Guardian: модель добровільної підтримки читачів. BuzzFeed, Vox Media: нативні цифрові медіа, що розуміють соцмережі. Інші занепали або зникли. Сотні місцевих газет закрились, створюючи "news deserts" — регіони без локальної журналістики.

Нові форми участі: від пасивного споживання до активної участі. Цифрові медіа створили нові форми взаємодії аудиторії з контентом. Коментарі та дискусії:Кожна стаття має секцію коментарів. Іноді дискусія під статтею цікавіша за саму статтю. Але також — токсичність, тролінг, spam.

Crowdsourcing журналістики: The Guardian's "The Counted" проєкт про вбивства поліцією використовував дані від читачів. ProPublica залучає аудиторію до розслідувань. Фактчекінг спільноти: Community Notes у Twitter (колишній Birdwatch) дозволяє користувачам додавати контекст до potentially misleading tweets. Фінансова підтримка: Patreon, Substack, OnlyFans — платформи, де фанати безпосередньо підтримують креаторів, без посередників.

Культурні зміни: як інтернет змінив нас. Скорочення span of attention:

Дослідження показують, що середній час фокусування на одному завданні скоротився з 2.5 хвилин (2004) до 47 секунд (2021). Постійні нотифікації, мультитаскінг, безкінечні стрічки тренують мозок на поверхневість. Culture of immediacy. Очікуємо миттєвих відповідей на повідомлення, новин у реальному часі, сервісів on-demand. Терпіння стало дефіцитним ресурсом.

Performative living: Соцмережі заохочують жити для аудиторії.

Подорожі, їжа, події документуються не для пам'яті, а для Instagram. Життя стає перформансом. Echo chambers ранньої версії: Ще до алгоритмічних стрічок, можливість обирати лише джерела, що підтверджують наші погляди, створювала ризик ізоляції в інформаційних бульбашках.

Висновки: невідворотність трансформації. До середини 2010-х стало очевидно: епоха масових медіа закінчилась безповоротно. Цифрові платформи не просто доповнюють традиційні медіа — вони їх замінюють для більшості аудиторії, особливо молодшої.

Ключові зміни перехідного періоду. Від дефіциту до надлишку: Обмеженість контенту замінена інформаційним перевантаженням. Від пасивності до участі: Аудиторія трансформувалась у prosumers. Від синхронності до on-demand: Контроль над часом та місцем споживання перейшов до користувачів. Від централізованих gatekeepers до розподіленого курування: Редактори поступилися алгоритмам та crowd wisdom. Від масової до персоналізованої аудиторії: Кожен отримує унікальний медіадосвід. Але справжня революція персоналізації тільки почалась. Наступна фаза — машинне навчання та ШІ — підніме індивідуалізацію медіа на абсолютно новий рівень.

1. КОНТРОЛЬНІ ПИТАННЯ

1. Зміна формату: У чому різниця між "оцифруванням" (digitization) і "цифровізацією" (digitalization)? Чому просто викласти PDF-версію газети на сайт — це погана стратегія в епоху інтерактивності?

2. Економіка: Поясніть теорію "Довгого хвоста" на прикладі Netflix або Spotify. Чому їм вигідно тримати фільми, які дивляться лише 100 людей на рік?

3 Влада користувача: Як можливість коментувати новини змінила журналістські стандарти? (Журналісти стали більш підзвітними, але й більш вразливими до булінгу).

4. Гіпертекст: Як посилання (hyperlinks) змінили структуру написання новин? (Принцип перевернутої піраміди став ще агресивнішим; необхідність утримувати увагу, щоб користувач не клікнув на посилання і не пішов).

5. Web 2.0: Хто ввів цей термін і яка його головна характеристика? (Тім О'Райлі; перехід до платформ, де контент створюють

користувачі).

2. ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. "Цифрові розкопки" (Internet Archive). Інструмент: Wayback Machine (web.archive.org). Дія: Знайдіть, як виглядала головна сторінка сайту *BBC* або *New York Times* у 1998 році. Знайдіть, коли на сайті з'явилася можливість залишати коментарі. Аналіз: Опишіть еволюцію від "електронної дошки оголошень" до інтерактивного порталу.

Завдання 2. Пошук ніші (Long Tail in Action). Умова: Ви хочете запустити медіа, яке не могло б існувати в епоху друкованої преси. Завдання: Придумайте тему для блогу/каналу, яка є занадто вузькою для газети, але успішною для інтернету. *Приклад:* "Відновлення радянських синтезаторів" або "Новини для власників ліворульних авто в Британії". Обґрунтування: Чому глобальний доступ робить цю нішу прибутковою?

Завдання 3. UX-порівняння. Об'єкти: Стаття у Вікіпедії vs Стаття у Британській енциклопедії (паперовій або сканованій). Завдання: Прослідкуйте свій шлях користувача. Скільки разів ви перейшли за посиланнями у Вікіпедії? Чи повернулися ви до початкової теми? Це ілюстрація нелінійності.

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. *Negroponte, N. Being Digital.* (Пророча книга 1995 року про перехід від атомів до бітів).

2. *Anderson, C. The Long Tail: Why the Future of Business is Selling Less of More.* (Обов'язково для розуміння економіки інтернету).

3. *O'Reilly, T. What Is Web 2.0.* (Стаття, що визначила епоху соцмереж).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Digitization (Оцифрування): Технічний процес переведення аналогової інформації (текст, фото, звук) у цифровий формат (код).

• Hypertext (Гіпертекст): Текст, що містить посилання на інші тексти, дозволяючи читачеві переміщатися між ними нелінійно.

• Web 1.0: Перший етап розвитку інтернету (90-ті), що характеризується статичними сайтами, де користувачі могли лише спожива-

ти інформацію ("Read-only").

• Web 2.0: Другий етап (2000-ні), орієнтований на інтерактивність, соціальні мережі та контент, створений користувачами ("Read-Write").

• The Long Tail (Довгий хвіст): Економічна модель, в якій продаж великої кількості унікальних нішевих товарів сумарно дає такий же або більший прибуток, ніж продаж невеликої кількості хітів.

1.3. Алгоритмічна революція: від програмування до персоналізації

Перехід від традиційного програмування медіа до алгоритмічної персоналізації представляє одну з найбільш фундаментальних трансформацій в історії комунікацій. Ця революція не сталася в один момент, а розгорталася поступово протягом кількох десятиліть, змінюючи не лише технічні аспекти доставки контенту, але й саму природу відносин між виробниками медіа, споживачами та інформаційним середовищем. Розуміння цієї трансформації вимагає дослідження її технологічних, економічних, соціальних та культурних вимірів.

У традиційній моделі мовлення, що домінувала протягом більшої частини двадцятого століття, концепція "програмування" була центральною. Редактори новин, програмні директори телебачення та радіо, кінокуратори приймали рішення про те, який контент виробляти, коли його транслювати, та в якій послідовності його представляти. Ці професіонали діяли як "привратники" (gatekeepers) інформації, застосовуючи своє професійне судження, розуміння аудиторії та редакторські стандарти для створення програмних сіток, що, як вважалося, служили інтересам широкої публіки. Модель була по суті централізованою та односпрямованою: від кількох виробників до багатьох споживачів, з обмеженими можливостями для зворотного зв'язку або індивідуалізації.

Ця система мала певні переваги. Професійне судження редакторів та кураторів часто забезпечувало якість та різноманітність контенту. Спільний досвід перегляду створював культурні точки згуртування – коли мільйони людей дивилися одну й ту саму програму одночасно, це створювало спільну культурну пам'ять та базу для соціальних розмов. Редакторська відповідальність означала, що хтось нес відповідальність за точність, справедливість та етичність контенту. Водночас ця система також мала значні обмеження. Вона була негнучкою, пропонуючи обмежений вибір у конкретний час. Вона часто ігнорувала нішеві інтереси та потреби меншин, оптимізуючи для "найменшого спільного знаменника" масової аудиторії. І вона надавала величезну владу відносно невеликій групі привратників, які визначали, що є гідним уваги громадськості.

Перші проблиски алгоритмічної революції можна простежити до появи кабельного телебачення в 1970-х та 1980-х роках. Хоча кабельне телебачення все ще покладалося на традиційне програмування, воно драматично збільшило кількість доступних каналів, дозволяючи більшу спеціалізацію та сегментацію аудиторії. Канали могли орієнтуватися на конкретні демографічні групи чи інтереси – новини 24/7, спорт, фільми, музика, дитячі програми – способами, які були економічно нежиттєздатними в еру обмеженого ефірного мовлення. Це представляло рух від моделі "одне для всіх" до більш сегментованої моделі "щось для кожного", хоча вибір все ще визначався програмістами, а не індивідуальними споживачами.

Справжній каталізатор алгоритмічної революції прийшов з цифровізацією та інтернетом. Коли медіа-контент став цифровим, він отримав фундаментально нові властивості: він міг бути легко копійованим, передаваним, індексованим, пошуковим та, критично важливо, маніпульованим алгоритмами. Інтернет створив інфраструктуру для двостороннього зв'язку та миттєвої доставки контенту на вимогу. Ці технологічні зміни відкрили можливості для радикально різних моделей медіа-споживання.

Рання фаза цифрової революції, приблизно з середини 1990-х до середини 2000-х, характеризувалася тим, що можна назвати "моделлю на вимогу". Послуги, такі як ранні веб-сайти новин, Netflix (спочатку як сервіс доставки DVD), iTunes та подкасти, дозволяли користувачам вибирати контент, коли та де вони хотіли його споживати, а не бути обмеженими програмними сітками. Це був важливий крок до персоналізації, оскільки він переміщував контроль від програмістів до споживачів. Однак на цьому етапі вибір все ще робився переважно людиною-користувачем на основі їхніх свідомих переваг та активного пошуку. Роль алгоритмів була обмеженою – в основному для індексації, пошуку та організації контенту, а не для активної рекомендації або курації.

Справжній проривний момент настав з появою складних систем рекомендацій у середині-кінці 2000-х років. Amazon була піонером у використанні колаборативної фільтрації для рекомендації продуктів на основі поведінки інших користувачів з подібними уподобаннями. Netflix запустив знаменитий конкурс Netflix Prize у 2006 році, пропонуючи мільйон доларів команді, яка могла б найбільше

покращити точність їхнього алгоритму рекомендацій. Цей конкурс не лише призвів до технологічних інновацій, але й сигналізував про стратегічну важливість персоналізації для цифрових медіа-компаній. Стало зрозуміло, що в світі практично необмеженого вибору здатність допомогти користувачам знайти контент, який їм сподобається, стане ключовою конкурентною перевагою.

Паралельно зростання соціальних мереж, особливо з запуском Facebook News Feed у 2006 році, представило новий тип алгоритмічної курації. На відміну від традиційних медіа, де професійні редактори вибирали контент, або ранніх цифрових платформ, де користувачі активно вибирали, що споживати, соціальні мережі впровадили модель, де алгоритми вибирали та ранжували контент з величезного потоку оновлень від друзів, сімейних членів та наступних сторінок. News Feed не був просто хронологічним переліком; він був персоналізованим потоком, оптимізованим алгоритмами для максимізації релевантності та залученості для кожного індивідуального користувача.

Цей розвиток представляв фундаментальну зміну парадигми. Вперше в історії мас-медіа кожна людина потенційно отримувала унікальну версію "програми". Два користувачі Facebook могли мати радикально різні досвіди платформи навіть якщо вони підписані на ті ж самі сторінки, оскільки алгоритм визначав, які пости показувати, в якому порядку та з якою пріоритетністю, на основі індивідуальної поведінки та характеристик. Це було народженням справді персоналізованих медіа в масштабі.

Технологічні основи цієї трансформації спиралися на кілька ключових інновацій у галузі машинного навчання та обробки даних. Колаборативна фільтрація, згадана раніше, дозволяла системам знаходити закономірності в поведінці мільйонів користувачів та робити передбачення на основі подібностей. Якщо користувачі А та В демонструють подібні переваги в минулому, то речі, які подобаються користувачеві А, ймовірно, сподобаються користувачеві В. Контент-базована фільтрація аналізувала характеристики самого контенту – жанр, тема, автор, ключові слова – щоб рекомендувати подібні елементи. Гібридні підходи поєднували обидва методи для більш надійних рекомендацій.

З часом алгоритми стали більш складними, інкорпоруючи глибо-

ке навчання та нейронні мережі. Ці технології дозволяли системам вчитися з величезних обсягів даних, виявляючи нелінійні та складні закономірності, які були б неможливими для людей або простіших алгоритмів. Вони могли враховувати сотні або тисячі факторів одночасно – від очевидних, таких як історія переглядів, до тонких, таких як час дня, тип пристрою, швидкість прокрутки або навіть мікровирази, вловлені через камеру (хоча останнє залишається в основному в сфері досліджень та викликає значні етичні занепокоєння).

Критично важливим для цього розвитку була також експоненційна зростання обчислювальної потужності та здатності зберігання даних. Закон Мура, який спостерігав подвоєння обчислювальної потужності приблизно кожні два роки, зробив можливою обробку величезних масивів даних в реальному часі. Хмарні обчислення дозволили компаніям масштабувати свою інфраструктуру для обслуговування мільярдів користувачів. Великі дані (Big Data) – терм, який увійшов у широке використання в кінці 2000-х – відносився не лише до обсягу даних, але й до нових методологій для їх зберігання, обробки та аналізу.

Економічна логіка, що керувала алгоритмічною революцією, була потужною. У цифровому світі, де копіювання та розповсюдження контенту майже безкоштовні, традиційні економічні моделі, засновані на дефіциті – обмежена кількість телевізійних каналів, обмежений простір на газетних шпальтах – більше не працювали. Натомість виникла "економіка уваги", де справжнім дефіцитним ресурсом була увага користувачів, а не контент. У світі, де будь-хто може публікувати і де доступний практично необмежений контент, виклик полягав не у виробництві контенту, а в захопленні та утриманні уваги аудиторії.

Персоналізація стала ключовою стратегією в цій конкуренції за увагу. Адаптуючи контент до індивідуальних переваг, платформи могли збільшити релевантність, а отже, залученість та час, проведений на платформі. Це мало пряме економічне значення в рекламно-підтримуваних моделях, де прибуток залежав від показів реклами та кліків. Чим більше часу користувачі проводять на платформі та чим більше вони взаємодіють з контентом, тим більше можливостей для монетизації через рекламу.

Більше того, дані, зібрані через відстеження поведінки користувачів, самі стали цінним активом. Ці дані не лише покращували

рекомендації контенту, але й дозволяли надзвичайно точне таргетування реклами. Рекламодавці могли досягти не просто широких демографічних категорій, але конкретних індивідів на основі їхніх інтересів, поведінки, місцезнаходження та навіть емоційного стану. Це створило "цикл позитивного зворотного зв'язку": більше користувачів означало більше даних, що призводило до кращих алгоритмів, що приваблювало ще більше користувачів, генеруючи ще більше даних. Це динаміка, часто називана "ефектами мережі даних", допомогла кільком платформам швидко домінувати на своїх ринках.

Соціальні та культурні наслідки алгоритмічної революції були глибокими та далекосяжними. На найбазовішому рівні персоналізація змінила саму природу медіа-споживання від в основному спільного досвіду до в основному індивідуалізованого. У еру мовлення сім'ї збиралися навколо телевізора, щоб дивитися одні й ті ж програми, колеги обговорювали вечірні новини, нації об'єднувалися навколо великих телевізійних подій. Персоналізовані цифрові медіа, навпаки, створили ситуацію, де кожна людина споживає унікальну суміш контенту, часто на особистих пристроях, у власному темпі та за власним графіком.

Ця індивідуалізація мала як позитивні, так і негативні аспекти. З позитивного боку, вона дала людям безпрецедентну свободу та контроль над їхнім медіа-досвідом. Більше не обмежені тим, що транслюють в певний час, користувачі могли досліджувати нішеві інтереси, знаходити спільноти навколо специфічних захоплень та споживати контент, який справді резонував з їхніми особистими смаками та потребами. Для меншин та маргіналізованих груп, чиї інтереси та перспективи часто ігнорувалися масовими медіа, цифрова персоналізація відкрила можливості для представлення та спільноти.

З негативного боку, фрагментація спільного медіа-досвіду підняла занепокоєння щодо соціальної згуртованості та можливості конструктивного громадського діалогу. Якщо люди живуть у фундаментально різних інформаційних всесвітах, споживаючи різний контент, експонуючись до різних фактів, і навіть використовуючи різні концептуальні рамки для розуміння подій, як вони можуть знайти спільну основу для обговорення та вирішення соціальних проблем? Ця занепокоїність посилилася з появою концепції "фільтр-бульбашок" та "ехо-камер", термінів, популяризованих активістом

інтернету Еліем Паризером та іншими наприкінці 2000-х та на початку 2010-х.

Фільтр-бульбашка відноситься до інтелектуальної ізоляції, яка може виникнути, коли алгоритми персоналізації вибірково представляють інформацію на основі минулої поведінки користувача, потенційно обмежуючи експозицію до різноманітних перспектив. Якщо алгоритм визначає, що ви зацікавлені в певній політичній перспективі, він може показувати вам більше контенту, що узгоджується з цією перспективою, та менше контенту, що кидає їй виклик, створюючи самопідсилювальний цикл підтвердження. Ехо-камера є схожою концепцією, що відноситься до ситуації, коли переконання посилюються або підсилюються через комунікацію та повторення всередині закритої системи, де альтернативні погляди відфільтровані або недостатньо представлені.

Дебати про фільтр-бульбашки виявилися складними та нюансованими. Деякі дослідження показали докази персоналізованої фільтрації та ідеологічної сегрегації в онлайн-середовищах, особливо навколо політичного контенту. Інші дослідження припустили, що ефект переоцінений, вказуючи на те, що багато людей все ще експонуються до різноманітних поглядів онлайн, часто більше, ніж вони були б у доцифрову еру, коли географічне та соціальне розділення створювало фізичні фільтр-бульбашки. Більше того, дослідження показали, що власний вибір користувачів – люди, що вибірково шукають інформацію, що підтверджує їхні існуючі переконання – може бути важливішим драйвером ідеологічної сегрегації, ніж алгоритмічна фільтрація.

Незалежно від емпіричних дебатів, занепокоєння про фільтр-бульбашки відображала ширшу тривогу щодо ролі алгоритмів у формуванні громадського дискурсу. На відміну від традиційних редакторів, чиї критерії та процеси прийняття рішень були відносно прозорими та підлягали професійним стандартам та громадському контролю, алгоритми часто діяли як непрозорі "чорні скриньки". Користувачі не знали, чому їм показували конкретний контент або що не показувалося. Навіть якщо алгоритми не мали явної політичної упередженості, їхня оптимізація для залученості могла мати ненавмисні соціальні наслідки, такі як просування сенсаційного або поляризаційного контенту, оскільки такий контент часто генерував

сильні емоційні реакції та високу залученість.

Ця занепокоїність досягла кульмінації після президентських виборів у США 2016 року, коли питання про роль соціальних мереж у розповсюдженні дезінформації, втручання іноземних акторів та можливий вплив на результати виборів стали предметом інтенсивного громадського та наукового дослідження. Хоча причинні зв'язки залишаються оспорюваними, широко визнається, що алгоритмічні системи персоналізації зіграли значну роль у формуванні інформаційного середовища виборів способами, які підняли важливі питання про демократичне управління, цілісність інформації та владу технологічних платформ.

У відповідь на зростаючу критику та регуляторний тиск технологічні компанії почали робити зміни у своїх алгоритмічних системах. Facebook, наприклад, оголосив у 2018 році, що змінює алгоритм News Feed для пріоритезації "значущих соціальних взаємодій" над пасивним споживанням контенту, стверджуючи, що це покращить благополуччя користувачів. YouTube заявив про зусилля зменшити рекомендації "прикордонного контенту" та поганої дезінформації. Twitter експериментував з різними підходами до організації таймлайну та позначення потенційно оманливого контенту.

Водночас почалася хвиля регуляторної активності, особливо в Європі. Загальний регламент про захист даних (GDPR), що вступив у силу в 2018 році, встановив нові стандарти для конфіденційності даних та дав користувачам більше контролю над їхніми особистими даними. Закон про цифрові послуги (Digital Services Act) та Закон про цифрові ринки (Digital Markets Act), запропоновані Європейською Комісією в 2020 році, прагнули встановити нові правила для цифрових платформ, включаючи вимоги прозорості для алгоритмічних систем та обмеження на використання персоналізованої реклами.

Академічна та дослідницька спільнота також інтенсифікувала свою увагу до алгоритмічних систем. Нова область "досліджень критичних алгоритмів" виникла, об'єднуючи дослідників з комп'ютерних наук, соціальних наук, гуманітарних наук та права для вивчення соціальних, етичних та політичних вимірів алгоритмічних систем. Важливі роботи, такі як "Weapons of Math Destruction" Кеті О'Ніл, "Algorithms of Oppression" Сафії Ноубл та "The Age of

Surveillance Capitalism" Шошани Зубофф, привернули громадську увагу до темних боків алгоритмічної персоналізації, від дискримінації та упередження до комерціалізації приватності та маніпулятивних бізнес-практик.

Технологічні розробки також продовжували швидко розвиватися. Штучний інтелект та машинне навчання ставали все більш складними, з проривами в обробці природної мови (наприклад, GPT-3 та пізніше моделі), комп'ютерному зорі та рекомендаційних системах. Ці досягнення відкрили нові можливості для персоналізації, від автоматичного генерування персоналізованого контенту до передбачення потреб користувачів перед тим, як вони їх висловлюють. Водночас вони також підняли нові етичні та соціальні питання про автентичність, маніпуляцію та межі автоматизації в медіа та комунікації.

Поява генеративного штучного інтелекту, здатного створювати реалістичні тексти, зображення, аудіо та відео, представляє останню фазу алгоритмічної революції. Ці технології відкривають можливість не лише персоналізації курації існуючого контенту, але й створення нового, унікального контенту для кожного користувача. Медіа-досвід може стати не просто персоналізованим, а індивідуально згенерованим – де фільми, новини, музика та навіть літературні твори створюються в реальному часі на основі переваг, контексту та емоційного стану конкретного користувача.

Ця перспектива підкреслює, наскільки далеко ми пройшли від традиційної моделі програмування. У той час як ера мовлення характеризувалася централізованим контролем, обмеженим вибором та спільним досвідом, алгоритмічна ера характеризується децентралізованою кураці єю (хоча й контрольованою кількома потужними платформами), практично необмеженим вибором та глибоко індивідуалізованим досвідом. Перехід не був лише технологічним, але й фундаментально змінив владні відносини в медіа-екосистемі, економічні моделі, що підтримують виробництво та розповсюдження контенту, та саму природу громадської сфери.

Розуміння цієї траєкторії – від програмування до персоналізації – є критично важливим для осмислення того, де ми знаходимося сьогодні та куди ми можемо рухатися в майбутньому. Алгоритмічна революція не є завершеною; вона продовжує розгортатися, з

новими технологіями, бізнес-моделями та соціальними практиками, що постійно з'являються. Виклики, які вона представляє – для приватності, автономії, демократії, соціальної згуртованості та людської гідності – вимагають постійної уваги, критичного аналізу та продуманого регулювання. Водночас можливості, які вона відкриває для розширення доступу до інформації, підтримки різноманітності голосів та адаптації медіа-досвіду до індивідуальних потреб, не повинні бути втрачені. Навігація цієї складності вимагає як технічного розуміння, так і етичної мудрості – саме те, що цей підручник прагне розвивати.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допомагають зрозуміти зміну парадигми:

1. Зміна ролі "Гейткіпера" (Gatekeeper): Раніше редактор вирішував, що важливо знати суспільству (Human Curation). Тепер це вирішує алгоритм. Яка головна мета алгоритму (утримання уваги/ engagement) і як вона конфліктує з класичною метою журналістики (інформування громадян)?

2. Інформаційна бульбашка (Filter Bubble): Концепція Ілая Парізера. Чому персоналізація, яка здається зручною (ми бачимо те, що нам подобається), стає загрозою для демократії? (Ізоляція від альтернативних точок зору).

3. Ехо-камера (Echo Chamber): Чим "ехо-камера" відрізняється від "бульбашки"? (Бульбашка — це пасивний фільтр, а ехо-камера — це активне середовище, де ваші переконання постійно підсилюються та радикалізуються однодумцями).

4. Економіка уваги (Attention Economy): Чому платформи оптимізують стрічку під емоційні реакції (гнів, страх, сміх)? Як це впливає на тип контенту, який стає вірусним?

5. Serendipity (Ефект випадкового відкриття): У друкованій газеті ви могли випадково прочитати статтю про балет, шукаючи статтю про футбол, бо вони були поруч. Чому алгоритмічна стрічка вбиває цю випадковість і звужує світогляд?

2. ПРАКТИЧНІ ЗАВДАННЯ. Ці завдання демонструють роботу "персональної реальності":

Завдання 1. Експеримент "Два гугла" (Google Bubble Test)

• Учасники: Два студенти з різними інтересами (наприклад, політично активний та аполітичний).

• Дія: Одночасно введіть у пошук Google (не в режимі Інкогніто) одне й те саме нейтральне, але багатозначне слово. Наприклад: "Єгипет" (туроператори чи революція?), "Apple" (фрукт чи компанія?), "Бюджет" (сімейний чи державний?).

• Аналіз: Порівняйте перші 5 результатів. Наскільки вони відрізняються?

Завдання 2. Аналіз "Стрічки друга"

• Завдання: Попросіть друга/колегу дати вам погортати його TikTok або Instagram Reels протягом 5 хвилин.

• Спостереження: Опишіть свої відчуття. Чи здається вам цей контент дивним, нецікавим або навіть дратівливим?

• Висновок: Це ілюстрація того, наскільки глибоко алгоритм знає *конкретного* користувача і наскільки різні реальності ми населяємо.

Завдання 3. Дебати: "Редактор проти Робота"

• Теза: "Алгоритми знають, що нам *потрібно*, краще, ніж редактори".

• Аргументи ЗА: Редактори суб'єктивні, елітарні, ігнорують нішеві інтереси. Алгоритм демократичний — він дає те, на що люди клікають.

• Аргументи ПРОТИ: Алгоритм не має етики, він просуває фейки та ненависть заради кліків. Суспільству потрібен "спільний порядок денний".

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You.* (Книга, що ввела цей термін у 2011 році. Обов'язкова класика).

2. Cathy O'Neil. *Weapons of Math Destruction.* (Про темний бік алгоритмів і те, як вони посилюють нерівність).

3. Shoshana Zuboff. *The Age of Surveillance Capitalism.* (Фундаментальна праця про те, як наш досвід став сировиною для корпорацій).

4. Yuval Noah Harari. *Homo Deus.* (Розділ про "Датаїзм" — ідеологію, де потік даних є вищою цінністю).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• Персоналізація (Personalization): Процес адаптації контенту під

індивідуальні характеристики користувача (історію переглядів, демографію, поведінку) за допомогою машинного навчання.

• Інформаційна бульбашка (Filter Bubble): Стан інтелектуальної ізоляції, що виникає, коли вебсайти використовують алгоритми для вгадування того, що користувач хоче побачити, і приховують інформацію, яка суперечить його поглядам.

• Економіка уваги (Attention Economy): Підхід до управління інформацією, який розглядає людську увагу як дефіцитний товар. Медіа та платформи конкурують за кожну секунду вашого часу.

• Engagement (Залучення): Ключова метрика для алгоритмів (лайки, шери, час перегляду). Часто критикується за те, що пріоритезує емоційний та поляризуючий контент над якісним.

• Алгоритмічна упередженість (Algorithmic Bias): Ситуація, коли алгоритм видає несправедливі або дискримінаційні результати через недоліки в коді або даних, на яких він навчався.

1.4. Мобільна ера: контекстуальне споживання медіа

Поява мобільних пристроїв та їх швидке поширення протягом останніх двох десятиліть представляє ще одну фундаментальну трансформацію в еволюції медіаспоживання. Якщо алгоритмічна революція змінила те, який контент ми бачимо, то мобільна революція трансформувала де, коли та як ми споживаємо медіа. Більше того, мобільні пристрої додали новий вимір персоналізації – контекстуальну усвідомленість – що дозволило медіа адаптуватися не лише до того, хто ви є, але й до того, де ви знаходитесь, що ви робите, та навіть який час дня. Ця трансформація мала глибокі наслідки для медіа-індустрії, користувацької поведінки, соціальних взаємодій та самої тканини повсякденного життя.

Історія мобільних медіа може бути прослідкована через кілька ключових фаз розвитку. Рання епоха мобільних комунікацій, що розпочалася в 1980-х з появою перших комерційних стільникових телефонів, була в першу чергу про голосовий зв'язок. Ці пристрої були громіздкими, дорогими та обмеженими в функціональності, служачи в основному як портативні телефони для бізнес-професіоналів та заможних споживачів. Медіа в традиційному сенсі – новини, розваги, інформація – все ще споживалися через стаціонарні платформи, такі як телебачення, радіо та друковані видання.

Перша значна зміна прийшла з введенням SMS (служби коротких повідомлень) в 1990-х роках. Хоча спочатку задумане як технічна функція для інженерів, текстові повідомлення швидко були прийняті споживачами, особливо молоддю, як новий спосіб комунікації. SMS представив концепцію асинхронної, текстової, мобільної комунікації, що передбачила багато аспектів пізніших мобільних медіа. До початку 2000-х років мільярди текстових повідомлень надсилалися щодня, а мобільні телефони перетворювалися з голосових пристроїв на мультимедійні комунікаційні центри.

Наступна фаза почалася з появою WAP (протоколу застосування бездротових пристроїв) та ранніх мобільних інтернет-сервісів. Ці технології дозволяли мобільним телефонам отримувати доступ до спрощених версій веб-сайтів та базових інтернет-послуг. Досвід був примітивним за сучасними стандартами – повільні швидкості з'єднання, крихітні монохромні екрани, незручні інтерфейси – але

він продемонстрував потенціал мобільного доступу до інформації. У деяких країнах, особливо в Японії з її передовою системою i-mode, мобільний інтернет досяг раннього успіху, дозволяючи користувачам отримувати доступ до новин, погоди, електронної пошти та навіть проводити транзакції зі своїх телефонів.

Справжнім переломним моментом, що визначив сучасну мобільну еру, став запуск iPhone від Apple в 2007 році. Хоча смартфони існували до iPhone – найбільш помітно серія BlackBerry та пристрої на базі Windows Mobile – iPhone представив радикально нову концепцію того, чим може бути мобільний пристрій. З його великим сенсорним екраном, інтуїтивним інтерфейсом, потужними обчислювальними можливостями та, критично важливо, App Store (запущеним у 2008 році), iPhone перетворив мобільний телефон з комунікаційного пристрою на універсальну обчислювальну платформу. Швидко з'явилися конкуруючі платформи, особливо Android від Google, і протягом кількох років смартфони стали повсюдними в розвинених країнах та все більш поширеними в країнах, що розвиваються.

Ця трансформація мала глибокі наслідки для медіаспоживання. Вперше люди мали постійний, завжди ввімкнений доступ до інтернету та цифрових медіа. Смартфон став пристроєм, який люди носили з собою весь час, консультувалися десятки або сотні разів на день, і використовували практично для всього – від комунікації до навігації, від розваг до роботи, від покупок до датування. Медіа більше не були чимось, що ви споживали у певних місцях (вдома перед телевізором, в офісі за комп'ютером) або в певні часи (вечірні новини, ранкова газета), а стали постійною присутністю, що пронизує кожен момент та місце повсякденного життя.

Одним з найбільш важливих нововведень мобільної ери було введення контекстуальної усвідомленості в медіаспоживання. Смартфони обладнані безліччю сенсорів – GPS для визначення місцезнаходження, акселерометри для виявлення руху, гіроскопи для орієнтації, датчики наближення, датчики освітлення, мікрофони, камери – які забезпечують постійний потік даних про фізичний контекст користувача. Додайте до цього цифровий контекст – який додаток використовується, яка інформація переглядається, які комунікації відбуваються – і мобільні пристрої стають неймовірно багатими джерелами контекстуальної інформації.

Ця контекстуальна інформація відкрила нові можливості для персоналізації медіа. Локаційно-базовані сервіси дозволили медіа адаптуватися до того, де знаходиться користувач. Додатки новин могли показувати локальні історії, релевантні до поточного місця користувача. Рекламодавці могли доставляти оголошення на основі географічного місцезнаходження, наприклад, рекламуючи ресторан людям, які знаходяться поблизу. Сервіси, такі як Foursquare, побудували цілі платформи навколо концепції «чекінів» та локаційно-базованих соціальних взаємодій.

Часовий контекст також став важливим. Алгоритми могли враховувати не лише що ви зазвичай робите, але й що ви ймовірно робите в даний момент на основі часу дня. Вранці ви могли отримувати новини та інформацію про погоду. Під час вечірньої поїздки додому – музику або подкасти. Пізно ввечері – розважальний контент. Ці темпоральні закономірності, поєднані з іншими сигналами контексту, дозволили більш нюансовану персоналізацію, яка адаптувалася до ритмів повсякденного життя.

Активність та стан користувача також могли бути визначені та враховані. Акселерометри та гіроскопи могли виявити, чи користувач стоїть, сидить, йде, біжить або їде в транспорті. Це інформувало про те, який тип контенту був доречним. Довгі статті для читання могли бути відкладені на час, коли користувач був нерухомим, в той час як короткі аудіо-фрагменти могли бути запропоновані під час транзиту. Деякі додатки експериментували навіть з більш інтимними формами контекстуальної усвідомленості, такими як виявлення емоційного стану через аналіз голосу, набору тексту або вибору контенту, хоча це підняло значні занепокоєння щодо приватності.

Мобільна ера також драматично змінила темпоральну структуру медіаспоживання. У традиційній моделі мовлення медіа споживалися відповідно до фіксованих графіків, визначених програмістами. Навіть рання цифрова ера, хоча вона дозволяла споживання на вимогу, все ще передбачала відносно тривалі сеанси заангажованої уваги – сідати за комп'ютер, щоб прочитати новини або подивитися відео. Мобільні пристрої, навпаки, зробили можливим новий режим медіаспоживання, характеризований частими, короткими, розподіленими сеансами взаємодії.

Концепція «мікромоментів», популяризована Google, відобразила

цю зміну. Мікромоменти – це короткі випадки, коли люди звертаються до своїх смартфонів з конкретним наміром – дізнатися щось, зробити щось, купити щось, піти кудись. Ці моменти часто відбуваються в «проміжний час» – очікування в черзі, поїздка в транспорті, короткі перерви між діяльностями – який раніше міг би бути заповнений роздумами, спостереженням або простою бездіяльністю. Мобільні медіа колонізували цей проміжний час, перетворюючи кожен вільний момент на потенційну можливість для споживання контенту або взаємодії з платформою.

Це мало важливі наслідки для форматів та стратегій медіа. Контент повинен був адаптуватися до фрагментованої уваги та коротких сеансів взаємодії. Новинні організації розробили «мобільні» формати – коротші статті, візуальні історії, швидкі оновлення – оптимізовані для споживання на маленьких екранах та в ситуаціях часткової уваги. Відео-платформи експериментували з коротшими форматами; Vine (хоча пізніше закритий) піонерував шестисекундні відео, Instagram Stories популяризував ефемерний, вертикальний відео-контент, а TikTok побудував імперію на ультра-коротких, захоплюючих відео-кліпах, оптимізованих для мобільного перегляду.

Соціальні мережі стали нерозривно пов'язані з мобільним досвідом. Хоча платформи, такі як Facebook та Twitter, почалися як веб-сайти на основі десктопів, вони швидко перейшли до стратегії «мобільний-спочатку», визнаючи, що більшість їхніх користувачів отримували доступ до сервісів через смартфони. Дійсно, деякі з найуспішніших соціальних платформ – Instagram, Snapchat, WhatsApp, TikTok – були розроблені з нуля для мобільних пристроїв. Ці платформи розуміли унікальні можливості та обмеження мобільного контексту та проектували свої інтерфейси, функції та контентні формати відповідно.

Мобільні медіа також ввели нові режими взаємодії та залученості. Сенсорні інтерфейси зробили взаємодію більш тактильною та безпосередньою – свайп, пінч, тап замість клацання та набору тексту. Камери перетворили користувачів з пасивних споживачів на активних творців, з мільярдами зображень та відео, що завантажуються щодня. Push-повідомлення створили новий канал для залучення уваги користувачів, дозволяючи додаткам та сервісам переривати та спонукати до взаємодії навіть коли вони не використовувалися

активно.

Ця «економіка уваги» на стероїдах призвела до розробки все більш складних стратегій для захоплення та утримання користувацької залученості. Додатки використовували техніки з ігрового дизайну – стріки (поточні серії), значки досягнень, геміфікацію – щоб заохочувати регулярне використання. Безкінечна прокрутка та автоматичне відтворення елімінували природні точки зупинки, заохочуючи триваліші сеанси взаємодії. Алгоритми оптимізувалися для максимізації «часу взаємодії» та «залипання», часто за рахунок благополуччя користувачів.

Психологічні дослідження почали документувати наслідки цього постійного мобільного з'єднання. Термін «номофобія» (no-mobile-phone phobia) був викований для опису тривоги, яку люди відчувають, коли вони відокремлені від своїх смартфонів. Дослідження показали, що середній користувач перевіряв свій телефон від 50 до 150 разів на день, часто в рамках майже автоматичної, компульсивної поведінки. Феномен «фантомних вібрацій» – відчуття, що ваш телефон вібрує, коли він цього не робить – став широко визнаним. Більш серйозні занепокоєння виникли щодо впливу постійного з'єднання на увагу, пам'ять, соціальні навички та психічне здоров'я, особливо серед молоді.

Водночас мобільні пристрої також демократизували доступ до інформації та медіа способами, які були неможливі раніше. У країнах, що розвиваються, де стаціонарна інтернет-інфраструктура була обмеженою, мобільні телефони часто надавали перший та єдиний доступ до цифрового світу. За даними Міжнародного союзу електрозв'язку, станом на 2020 рік більше людей мали доступ до мобільних телефонів, ніж до чистої питної води чи електрики. Це мало трансформаційні наслідки для освіти, охорони здоров'я, економічної участі та громадянського залучення в багатьох частинах світу.

Мобільні медіа також змінили характер журналістики та виробництва новин. «Громадянська журналістика» – звичайні люди, що документують та поширюють новини через мобільні пристрої – стала значною силою. Під час великих подій, від Арабської весни до руху Black Lives Matter, мобільне відео та соціальні мережі дозволили свідкам та учасникам обходити традиційних медійних посередників

та безпосередньо поширювати інформацію до глобальної аудиторії. Професійні журналісти також адаптували свої практики, використовуючи смартфони не лише як інструменти спілкування, але й як засоби виробництва, здатні захоплювати, редагувати та публікувати контент безпосередньо з місця подій.

Бізнес-моделі медіа-індустрії були глибоко порушені мобільною трансформацією. Традиційні моделі на основі реклами стикнулися з викликами, оскільки маленькі екрани та споживання на ходу зробили деякі формати реклами менш ефективними. Водночас мобільні пристрої відкрили нові можливості для таргетованої, контекстуально релевантної реклами на основі місцезнаходження, часу та активності користувача. Мобільна комерція виникла як значний сектор, з користувачами, що все частіше здійснювали покупки безпосередньо зі своїх телефонів.

Моделі на основі підписки також знайшли новий успіх у мобільну еру. Послуги, такі як Spotify для музики, Netflix для відео та різноманітні новинні видання, пропонували модель «все, що ви можете споживати» через мобільні додатки, забезпечуючи постійний доступ до величезних бібліотек контенту за фіксовану місячну плату. Зручність мобільного доступу та персоналізовані рекомендації зробили ці послуги привабливими для мільйонів передплатників.

Роль контексту в мобільній персоналізації також підняла нові етичні та соціальні занепокоєння. Збір детальних локаційних даних створив ризики приватності, оскільки ці дані могли розкривати надзвичайно чутливу інформацію про звички, стосунки та діяльність людей. Де ви живете, працюєте, молитеся, отримуєте медичну допомогу; кого ви відвідуєте та як часто; яких демонстрацій ви відвідуєте або які бари ви відвідуєте – все це може бути виведено з локаційних даних. Випадки неправомірного використання таких даних, від продажу третім сторонам до використання правоохоронними органами без ордерів, підкреслили ризики.

Контекстуальна персоналізація також підняла питання про маніпуляцію та вразливість. Якщо системи можуть виявити, що ви втомлені, стресовані або емоційно вразливі на основі ваших шаблонів використання та контекстуальних сигналів, чи етично використовувати цю інформацію для таргетування вас конкретними повідомленнями чи рекламою? Концепція «темних патернів» –

дизайнерських виборів, які маніпулюють користувачами до дій, які не відповідають їхнім інтересам – стала особливо помітною в мобільному контексті, де обмежений розмір екрана та розподілена увага робили користувачів більш вразливими до маніпулятивного дизайну.

Питання «цифрової нерівності» також еволюціонували в мобільну еру. У той час як мобільні пристрої розширили доступ у багатьох відношеннях, також з'явилися нові форми нерівності. «Мобільно-тільки» користувачі – ті, хто отримує доступ до інтернету виключно через мобільні телефони – часто мали обмежений досвід порівняно з тими, хто мав доступ до комп'ютерів та широкосмугового інтернету. Деякі веб-сайти та онлайн-послуги були погано оптимізовані для мобільних пристроїв, обмеження даних та витрати могли обмежувати те, що люди могли робити, а складні завдання, такі як створення контенту, заповнення форм або професійна робота, залишалися складнішими на мобільних пристроях.

Регуляторні відповіді на мобільну персоналізацію розвивалися в різних юрисдикціях. GDPR у Європі встановив нові стандарти для мобільного відстеження та вимагав чітку згоду для збору даних, включаючи локаційну інформацію. Каліфорнійський акт про приватність споживачів (CCPA) встановив подібні правила в США. Apple та Google, як оператори двох домінуючих мобільних операційних систем, також впровадили зміни в приватності, такі як вимога дозволів для відстеження та надання користувачам більше контролю над тим, які дані додатки можуть отримувати.

Технологічні досягнення продовжували формувати мобільну еру. Швидкість мереж різко зросла від 2G до 3G, 4G та зараз 5G, дозволяючи більш насичений медіа-досвід – відео високої якості, доповнену реальність, хмарне ігри – на мобільних пристроях. Обчислювальна потужність смартфонів зросла до точки, де вони стали більш потужними, ніж комп'ютери-десктопи попереднього десятиліття. Батарейне життя покращилося, хоча залишалося обмеженням, часто визначаючи, скільки та як люди могли використовувати свої пристрої.

Поява «носимих» пристроїв – розумних годинників, фітнес-трекерів, навіть розумних окулярів – представила ще один рівень контекстуальної усвідомленості та потенціал для ще більш безшовної інтеграції медіа в повсякденне життя. Ці пристрої могли збирати

ще більш інтимні дані – частоту серцевих скорочень, якість сну, фізичну активність, навіть електрокардіограми – що відкривало нові можливості для персоналізації на основі здоров'я та благополуччя, але також підняло ще більш глибокі питання приватності.

Культурні наслідки мобільної ери були глибокими та всепроникаючими. Смартфон став центральним артефактом сучасного життя, об'єктом, який формує наші розклади, відносини та навіть самосприйняття. Феномен «фоторобки для Instagram» – документування досвіду в першу чергу для обміну в соціальних мережах – відображав, як мобільні медіа змінили наші відносини з досвідом та пам'яттю. Постійне з'єднання стирало межі між роботою та особистим життям, між публічним та приватним, між онлайн та офлайн.

Соціальні норми та етикет еволюціонували навколо мобільного використання. «Фабінг» (phone snubbing – ігнорування людини на користь телефону) став визнаною соціальною проблемою. Дебати виникли щодо відповідності використання телефонів у різних контекстах – за обіднім столом, на зустрічах, під час розмов. Деякі місця та події почали встановлювати «зони без телефонів», а рухи, такі як «цифровий детокс» та практики мінімалістичного використання технологій, виникли як форми опору постійному з'єднанню.

Наукова спільнота інтенсифікувала дослідження психологічних, соціальних та когнітивних ефектів мобільного використання. Дослідження запропонували складну картину. З одного боку, смартфони забезпечували безпрецедентний доступ до інформації, підтримували соціальні зв'язки, особливо для ізольованих індивідів, та пропонували інструменти для продуктивності, творчості та навчання. З іншого боку, з'явилися занепокоєння щодо залежності, зменшення тривалості уваги, порушення сну (частково через блакитне світло екранів), впливу на очні соціальні взаємодії та потенційних зв'язків з депресією та тривожністю, особливо серед підлітків.

Поняття «когнітивної розвантаження» – покладатися на зовнішні пристрої для збереження та обробки інформації замість нашого власного мозку – стало предметом дослідження. У той час як це може звільнити когнітивні ресурси для інших завдань, деякі дослідники занепокоїлися, що це може також атрофувати певні когнітивні навички, такі як навігація без GPS або запам'ятовування без цифрових

помічників.

Майбутнє мобільних медіа, ймовірно, буде характеризуватися ще більшою інтеграцією в тканину повсякденного життя. Концепція «навколишніх обчислень» – де обчислювальні можливості вбудовані в середовище, а не обмежені дискретними пристроями – може зробити розрізнення між «мобільним» та іншими формами медіаспоживання застарілим. Інтернет речей, де побутові об'єкти від холодильників до дзеркал стають підключеними та інтелектуальними, може створити ще більш контекстуально усвідомлене середовище.

Доповнена реальність (AR), яка накладає цифрову інформацію на фізичний світ, представляє потенційну наступну фазу мобільних медіа. Замість дивитися на екран, щоб отримати доступ до цифрової інформації, AR може проектувати цю інформацію безпосередньо у ваше поле зору через розумні окуляри або контактні лінзи. Це може дозволити ще більш безшовну, контекстуально релевантну персоналізацію, де інформація з'являється точно тоді та там, де вона потрібна. Водночас це також підняло б глибокі питання про увагу, приватність та межу між цифровим та фізичним.

Мобільна ера фундаментально трансформувала медіаспоживання, зробивши його контекстуальним, постійним та глибоко інтегрованим у повсякденне життя. Розуміння цієї трансформації – її технологічних основ, соціальних наслідків, психологічних ефектів та етичних викликів – є критично важливим для будь-кого, хто прагне орієнтуватися в сучасному медіа-ландшафті, чи то як професіонал, дослідник або усвідомлений громадянин. Мобільна революція не завершена; вона продовжує розгортатися, з новими технологіями та практиками, що постійно з'являються, та наслідками, які будуть відчуватися протягом десятиліть.

1. КОНТРОЛЬНІ ПИТАННЯ

1. Зміна ролей: Чому алгоритм називають "новим воротарем" (New Gatekeeper)? Чим його критерії відбору новин відрізняються від критеріїв журналіста-редактора? (Журналіст: суспільна вага; Алгоритм: залученість).

2. Неявні дані: Чому TikTok став успішнішим за Instagram у персоналізації? (Орієнтація на *час перегляду*, а не на соціальні зв'язки чи лайки).

3. Ехо-камера: Чим "ехо-камера" відрізняється від "бульбашки фільтрів"? (Бульбашка — це технологічне обмеження вибору, ехо-камера — це соціальне середовище, де посилюються однотипні думки).

4. Етична дилема: Якщо алгоритм показує людині фейкові новини, бо вона їх любить читати, хто винен: платформа, алгоритм чи користувач?

5. Прозорість: Що таке "Algorithm Transparency" і чому європейські закони (DSA) вимагають від платформ розкривати принципи роботи рекомендацій?

2. ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. Експеримент "Дресирування алгоритму". Платформа: YouTube Shorts або TikTok (новий акаунт або режим Інкогніто). Дія. Проскрольте 20 відео, не виявляючи реакції. Запишіть теми. Почніть лайкати та додивлятися до кінця відео *тільки* на одну специфічну тему (наприклад, "гончарство" або "ремонт авто"). Продовжуйте 10-15 хвилин. Аналіз: Як швидко змінилася стрічка? Скільки відсотків контенту тепер складає ваша тема? Це демонструє швидкість петлі зворотного зв'язку.

Завдання 2. "Тіньовий профіль" (Digital Shadow). Інструмент: Google My Activity (myactivity.google.com) або Facebook Ad Preferences. Дія: Зайдіть у налаштування рекламних вподобань. Завдання: Подивіться, які "ярлики" навісив на вас алгоритм. (Наприклад: "Любитель розкоші", "Батьки підлітків", "Політично помірковані"). Рефлексія: Наскільки точним є цей портрет? Чи є там помилки?

Завдання 3. Дебати "Serendipity vs Relevance". Тема: Що важливіше для медіаспоживача: отримувати лише те, що він хоче (Relevance), чи випадково натрапляти на щось нове і несподіване (Serendipity)? Аргументація: Наведіть по 2 аргументи за кожну позицію.

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You.* (Книга, яка ввела поняття бульбашки фільтрів).

2. O'Neil, C. *Weapons of Math Destruction.* (Про те, як алгоритми посилюють нерівність).

3. Harari, Y. N. *Homo Deus.* (Розділ про "Датаїзм" — ідеологію, де

потік даних є найвищою цінністю).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Algorithm (Алгоритм): Набір математичних інструкцій, які визначають, який контент, коли і кому показувати.

• Black Box (Чорна скринька): Система, внутрішні процеси якої непрозорі для користувача. Ми бачимо вхідні дані (наш лайк) і результат (нова рекомендація), але не знаємо, як саме прийнято рішення.

• Engagement (Залученість): Ключова метрика для алгоритмів. Сума реакцій користувача (лайки, коментарі, час перегляду).

• Echo Chamber (Ехо-камера): Метафоричний опис ситуації, коли переконання посилюються або зміцнюються через спілкування та повторення всередині закритої системи.

• Implicit Signals (Неявні сигнали): Пасивні дані про поведінку (зупинка скролу, час читання), які є більш чесними індикаторами інтересу, ніж активні дії (лайки).

1.5. Економіка уваги

Економіка уваги сьогодні є не просто однією з багатьох концепцій у сфері медіа чи маркетингу; це фундаментальна парадигма, що визначає логіку функціонування всієї цифрової екосистеми та сучасного капіталізму в цілому. Її суть полягає в тому, що в умовах радикального надлишку інформації та практично нескінченного вибору контенту дефіцитним і, отже, найціннішим ресурсом стала людська увага — обмежена, нероздільна та невідновлювана психологічна валюта. Саме навколо боротьби за цей ресурс будуються новітні бізнес-моделі, технологічні інновації, культурні тренди та навіть політичні комунікації. Це система, де переможцями стають ті, хто може не лише ефективно атрактувати фокус споживача, але й утримувати його якомога довше, трансформуючи цей час у різноманітні форми капіталу — від даних і рекламних доходів до соціального впливу та політичного капіталу.

Зародження цієї концепції, термін для якої популяризував економіст Герберт Александер Саймон ще наприкінці XX століття, стало неминучим наслідком інформаційної революції. Якщо індустріальна епоха була епохою дефіциту матеріальних товарів, а ранній капіталізм — дефіциту капіталу, то сьогодні ми стикаємося з парадоксальною ситуацією: інформація та контент є в надлишку, майже безкоштовні у виробництві та копіюванні, проте пропускна спроможність людської свідомості залишається жорстко обмеженою. Цей розрив між нескінченним виробництвом стимулів і обмеженим споживанням і створив новий ринок — ринок уваги, де є продавці (медіа, платформи, бренди, творці), покупці (рекламодавці, політичні сили) і товар — психічний простір користувача.

Основним архітектором і головним бенефіціаром цієї нової економіки стали масштабні цифрові платформи — такі як Meta, Google, ByteDance, які розробили досконалі технологічні механізми для видобутку, каналізації та монетизації уваги. Їхня бізнес-модель, заснована на зборі детальних даних про поведінку користувача, ґрунтується на простому принципі: платформа надає безкоштовні послуги (пошук, спілкування, розваги), користувач сплачує за них своєю увагою та персональними даними, а платформа продає доступ до цієї уваги та даних рекламодавцям. Ключовим інструментом тут

є рекомендаційний алгоритм — складний математичний механізм, який у реальному часі аналізує поведінку, передбачає уподобання та пропонує наступний фрагмент контенту, щоб максимізувати показник «часу перебування на платформі» або «залучення». Ці алгоритми оптимізовані не на якість інформації в її традиційному розумінні (правдивість, глибина, суспільна корисність), а на здатність контенту викликати сильну емоційну або поведінкову реакцію — цікавість, обурення, радість, тривогу. Таким чином, вся архітектура цифрового середовища, від дизайну інтерфейсу до логіки стрічки новин, підпорядкована меті захопити та утримати фокус.

Це породжує низку ключових проявів економіки уваги у повсякденній практиці. В медіа ми спостерігаємо домінування клікбейт-заголовків, візуально агресивних форматів і стрімкої швидкості оновлення новин, де важливість часто поступається місцем «вірусності». В соціальних мережах поширені нескінченні стрічки, автоматичне відтворення відео та ефекти, що базуються на FOMO (страх упустити щось важливе). Маркетинг все частіше використовує мікротаргетинг, персоналізовані рекламні послання та формати native content та influencer marketing, які минають традиційні психологічні захисні бар'єри споживача. Навіть політична комунікація перетворюється на постійну боротьбу за медіапростір, де скандали та поляризуючі висловлювання часто ефективніші за раціональні програми, оскільки генерують більше висвітлення та уваги.

Однак найглибший вплив економіки уваги проявляється на психологічному та соціокультурному рівнях. Індивідуально вона формує новий тип уваги — фрагментовану, розсіяну, постійно перемикається між завданнями. Здатність до тривалого глибокого концентрації, необхідної для навчання чи складної роботи, витісняється звичкою до поверхневого скролінгу. Це безпосередньо впливає на когнітивне здоров'я, сприяючи синдрому втомленої уваги, цифровій залежності та підвищеному рівню тривожності. Соціально економіка уваги стає потужним двигуном поляризації. Алгоритми, що прагнуть максимізувати залучення, природним чином пропонують користувачам контент, який підтверджує їхні існуючі переконання та часто викликає сильні емоції, наприклад, обурення. Це призводить до формування ізольованих інформаційних капітул, посилення

радикальних позицій та ерозії спільного інформаційного простору, де можливий конструктивний діалог.

Крім того, культурним наслідком стає домінування формату над змістом, та тенденція до стандартизації та «платформізації» творчості. Контент, який легко алгоритмізується, пакується та поширюється, отримує перевагу перед складними, повільними або експериментальними формами мистецтва. Виникає парадокс вибору: при формально нескінченному асортименті контенту алгоритми часто ведуть мільйони користувачів до обмеженого набору вірусних трендів, дещо нівелюючи саме розмаїття, яке мало забезпечити цифрова ера.

Проте, як будь-яка потужна система, економіка уваги породжує і контртенденції, і шляхи опору. Зростає усвідомлення цінності цифрового детоксу, практик медіагігієни та свідомого споживання інформації. Формуються нові бізнес-моделі, що апелюють не до найширшої, а до найякіснішої уваги: підписки без реклами, платні медіа, що інвестують у глибоку журналістику, платформи, які обмежують час користування або відмовляються від алгоритмічних стрічок на користь хронологічних. З'являються етичні дискусії про дизайн-інтервенції, що покликані захищати автономію користувача, наприклад, функції відліку часу або нагадування про перерву. Відповідь на виклики економіки уваги поступово формується на перетині регуляторних ініціатив (як GDPR в ЄС чи Digital Services Act), технологічних інновацій, спрямованих на децентралізацію та захист приватності, та зміни суспільних цінностей, де цифрова аскетичність та контроль над власним психічним простором стають новими символами статусу.

Таким чином, економіка уваги — це комплексна і всеосяжна система координат, що визначає не тільки те, що ми дивимось і читаємо, але й те, як ми мислимо, спілкуємося та будуємо соціальні зв'язки. Вона перетворює людський досвід на сировину, а психічні процеси — на ринок. Розуміння її механізмів, наслідків і антитез є критично важливим не лише для фахівців у сферах комунікацій, технологій чи політики, але й для будь-якої особи, яка прагне зберегти суверенітет власної свідомості в світі, де кожна мить нашої уваги стала об'єктом запеклої конкурентної боротьби, часто невидимої, але надзвичайно впливовою. Майбутнє цієї економіки буде визначатися

тим, чи вдасться знайти баланс між технологічною ефективністю видобутку уваги та етичною необхідністю захисту людської психіки та соціальної тканини від її найбільш деструктивних ефектів.

1. КОНТРОЛЬНІ ПИТАННЯ

1. Трансформація влади: Як змінилася роль "гейткіпера" (воротаря) інформації при переході від друкованої газети до стрічки TikTok, а тепер — до ChatGPT? Хто тепер фільтрує інформацію?

2. Економіка уваги: Що означає фраза Герберта Саймона: "Багатство інформації створює бідність уваги"? Як ШІ намагається вирішити цю проблему (через самарізацію) і як він її погіршує (через генерацію спаму)?

3. Prosumer: Хто такий "проз'юмер" (виробник-споживач)? Як генеративний ШІ знижує поріг входу для створення контенту, перетворюючи кожного читача на потенційного конкурента професійного медіа?

4. Смерть пошуку: Чому молоде покоління (Gen Z / Alpha) все частіше шукає інформацію в TikTok або ChatGPT, а не в Google? Як це змінює вимоги до створення контенту (SEO vs GEO - Generative Engine Optimization)?

5. FOMO vs JOMO: Як еволюціонувало психологічне сприйняття новин: від страху пропустити важливе (FOMO) до радості від відключення (JOMO), і яку роль тут відіграють ШІ-дайджести?

2. ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. "Медіа-археологія"

• Мета: Відчути різницю в швидкості та глибині.

• Завдання: Оберіть одну історичну подію (наприклад, аварія на ЧАЕС або висадка на Місяць).

1. Знайдіть, як про неї писали газети того часу (Web 1.0 / Pre-web).

2. Знайдіть, як про неї розповідають у YouTube/TikTok блогери (Web 2.0).

3. Запитайте у ChatGPT/Perplexity про цю подію (AI Era).

• Аналіз: Порівняйте джерела за трьома критеріями: *Достовірність, Емоційність, Час на отримання інформації.*

Завдання 2. "Аудит інформаційної бульбашки"

• Умова: Використайте свій смартфон.

• Дія: Відкрийте стрічку рекомендацій (TikTok, Instagram Reels або Google Discover). Проскрольте 20 одиниць контенту.

• Аналіз:

- Скільки з них є новинами, а скільки — розвагами?

- Чому алгоритм показав це саме вам? (Наприклад: "Я лайкнув відео про котів, тепер бачу рекламу корму").

- Уявіть, як виглядатиме ця стрічка в епоху ШІ: чи буде там відео, згенероване спеціально для вас?

Завдання 3. Прогнозування (Essay)

• Тема: "Мій ранок у 2030 році".

• Завдання: Опишіть процес споживання новин через 5 років. Чи будуть існувати сайти? Чи будуть новини читати голосом у навушник? Чи буде ШІ фільтрувати негатив?

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. McLuhan, M. *Understanding Media: The Extensions of Man.* (Класика про те, що "засіб комунікації є повідомленням". Актуально для розуміння природи ШІ).

2. Toffler, A. *The Third Wave.* (Тут вперше введено поняття "Prosumer").

3. Manovich, L. *AI Aesthetics*. (Як ШІ змінює наше сприйняття культури та медіа).

4. Reuters Institute Digital News Report. (Щорічний звіт про те, як змінюється медіаспоживання у світі. Обов'язково до перегляду актуального випуску).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Gatekeeper (Гейткіпер/Воротар): Особа або технологія, що контролює доступ до інформації та вирішує, який контент потрапить до аудиторії. Раніше це були головні редактори, зараз — алгоритми рекомендацій та моделі ШІ.

• Attention Economy (Економіка уваги): Концепція, яка розглядає людську увагу як дефіцитний ресурс. Медіа та платформи конкурують не за гроші (напряму), а за час користувача.

• Filter Bubble (Інформаційна бульбашка): Стан інтелектуальної ізоляції, коли алгоритми показують користувачу лише ту інформа-

цію, яка підтверджує його погляди, відсікаючи альтернативні думки.

• UGC (User Generated Content): Контент, створений кінцевими користувачами (блоги, коментарі, відео в TikTok), а не професійними редакціями.

• Echo Chamber (Ехо-камера): Ситуація, в якій переконання користувача посилюються через багаторазове повторення всередині закритої системи.

Висновок до розділу 1

Еволюція медіаспоживання, розглянута в розділі, є динамічним процесом, що простежується від традиційних централізованих каналів до сучасного цифрового, персоналізованого та фрагментованого ландшафту. Її основними драйверами були технологічні інновації, що радикально змінили способи створення, поширення та споживання інформації та розваг, а також соціокультурні зміни та зовнішні обставини, такі як конфлікти, які формують суспільні потреби. Сьогодні ця еволюція характеризується домінуванням мобільних пристроїв, зокрема смартфонів, які стали основним порталом до цифрового світу для переважної більшості споживачів, забезпечуючи доступ до новин, соціальних мереж та стрімінгових платформ. При цьому спостерігається суттєвий перехід від пасивного споживання контенту за розкладом, як у лінійному телебаченні, до активної взаємодії з контентом на вимогу (VOD) та на стрімінгових сервісах, що пропонують індивідуалізовані підписки, хоча й стикаються з викликом "втоми від підписок" через перенасичення ринку.

Окремою значущою рисою сучасної медіаринку є криза довіри, що впливає на всі гравців — від медіакомпаній до рекламодавців. У відповідь на це відбувається переоцінка цінності та ролі різних каналів. Наприклад, лінійне телебачення зберігає свою силу для масштабних подій, таких як спортивні трансляції чи загальнонаціональні інформаційні марафони в умовах кризи, демонструючи високу консолідуючу спроможність та довіру аудиторії, що підтверджується дослідженнями в Україні. Водночас підключене телебачення (CTV) та програматик реклама стрімко розвиваються, пропонуючи небачену раніше точність таргетування. Парадоксально, але друковані медіа, втрачаючи масовість, набувають нової цінності як джерела глибини, аналізу та довіри для вузької, часто преміум-аудиторії, яка втомилася від цифрового шуму. Цей тренд перегукується із загальносвітовим

зростанням інтересу до якісного локального та патріотичного контенту в умовах викликів, що спостерігається, зокрема, серед української аудиторії старшого віку.

Технологічним фронтом еволюції стає поширення штучного інтелекту та синтетичних медіа, які трансформують ландшафт створення контенту — від автоматизованої текстівки до генерації реалістичних відео, — ставлячи нові виклики щодо аутентичності, авторства та інформаційної безпеки. На цьому тлі ключовим імперативом для всіх учасників медіаринку стає гнучкість та здатність адаптувати бізнес-моделі. Це проявляється в поєднанні різних джерел доходу, таких як гібридні підписки з рекламою (AVOD), співпраця між платформами, інтеграція традиційного та цифрового контенту (конвергентне телебачення) та пошук нових форматів взаємодії з аудиторією. Таким чином, еволюція медіаспоживання веде нас до екосистеми, де технологічна динаміка, якісна журналістика, консолідуюча роль медіа в кризових ситуаціях та індивідуальні уподобання споживачів знаходяться у постійній взаємодії, формуючи складний, багаторівневий, але надзвичайно життєздатний простір сучасної комунікації.

Розділ 2: Що таке персоналізація медіа
2.1. Визначення та типологія

Персоналізація медіа — це процес адаптації контенту, його презентації та способу доставки до індивідуальних характеристик, уподобань та контексту конкретного користувача з метою підвищення релевантності, залученості та задоволеності від медіаспоживання. Якщо традиційні медіа працювали за принципом "один для всіх" (one-to-many broadcasting), то персоналізовані медіа функціонують за логікою "унікальний для кожного" (one-to-one narrowcasting). Газета друкувалась в мільйонах ідентичних примірників. Ваша стрічка Facebook унікальна — жодна інша людина не бачить точно таку саму послідовність постів.

Персоналізація базується на даних про користувача. Але ці дані можна отримати двома принципово різними способами — через явну та неявну персоналізацію. Явна персоналізація передбачає, що користувач свідомо та активно надає інформацію про свої преференції. Це відбувається через рейтинги та оцінки, коли Netflix просить оцінити фільми від одного до п'яти зірок або "подобається/не подобається", а Spotify пропонує лайкнути або дизлайкнути трек. Ці оцінки прямо сигналізують алгоритму про ваші смаки. При реєстрації на платформі ви вказуєте вік, стать, місто, інтереси. Medium запитує, які теми вас цікавлять — технології, бізнес, дизайн. YouTube пропонує підписатись на канали при першому заході. Кожна підписка на Instagram-акаунт, YouTube-канал чи Telegram-групу стає явним сигналом про інтереси. Алгоритм знає: якщо ви підписані на двадцять каналів про кулінарію, ви любите готувати. Списки бажань та збережене — Pinterest boards, Spotify playlists, "Прочитати пізніше" в Medium, "Збережені" в Instagram — все це явні індикатори того, що вас приваблює. Що ви вводите в пошук Google, YouTube, Netflix — пряме вираження намірів та інтересів.

2.1.1. Явна vs неявна персоналізація.

Явна персоналізація має свої переваги: висока точність, адже людина прямо каже, що їй подобається; прозорість, бо користувач розуміє, чому отримує певні рекомендації; контроль, оскільки можна змінити преференції, якщо рекомендації не влаштовують; менше етичних проблем, адже людина свідомо ділиться інформацією. Проте

є й недоліки: потребує зусиль від користувача, багато хто не хоче витрачати час на оцінювання; "декларативні" преференції не завжди збігаються з реальними — люди кажуть, що люблять документалки, а насправді дивляться реаліті-шоу; cold start problem для нових користувачів, яким треба спочатку багато чого оцінити; статичність, бо інтереси змінюються, а профіль залишається старим, якщо не оновлювати.

Неявна персоналізація працює інакше — система автоматично виводить преференції з поведінки користувача, без його усвідомленої участі. Це відбувається через збір поведінкових даних: що дивитесь навіть якщо не оцінюєте, скільки часу витрачаєте на кожен контент, коли зупиняєте, перемотуєте, переглядаєте повторно, що ігноруєте або скіпаєте, швидкість скролінгу. Швидко проскролили означає нецікаво, зупинились — зацікавило. Контекстуальні дані також важливі: час доби, коли вранці читаєте новини, а ввечері дивитесь серіали; день тижня, коли п'ятниця асоціюється з розважальним контентом, а понеділок з мотиваційним; пристрій, адже смартфон підходить для коротких відео, а телевізор для довгих фільмів; місцезнаходження, дома чи у транспорті. Кожен клік на посилання, кожне відкриття статті, кожен перегляд превью стає сигналом про інтерес, навіть якщо ви не лайкнули. Соціальні зв'язки теж мають значення: з ким ви спілкуєтесь, чиї пости коментуєте, кого часто згадуєте. "Покажи мені твоїх друзів, і я скажу, хто ти" — принцип соціальної персоналізації. Деякі платформи навіть відстежують, де затримується погляд через веб-камеру на smart TV, куди рухається курсор, на що ви наводите, але не клікаєте.

TikTok є найдосконалішим прикладом неявної персоналізації. Алгоритм аналізує кожну мілісекунду взаємодії: скільки секунд дивились відео перед свайпом, якщо дві з шістдесяти — нецікаво, якщо всі шістдесят до кінця — дуже цікаво; чи переглянули повторно; чи зайшли на профіль автора після відео; чи поставили лайк, коментар чи шер; чи додали звук у вибране; навіть чи трохи затрималися перед свайпом, фіксуючи мікро-сигнали вагання. Результат вражає: TikTok "розкусує" ваші смаки за тридцять-сорок хвилин використання. Ви не вказуєте жодних інтересів, але алгоритм знає, що вам подобається. YouTube аналізує не просто що ви переглянули, а скільки відсотків відео. Якщо регулярно дивитесь десятихвилинні відео певного типу

до кінця, а тридцятихвилинні скіпаєте через дві хвилини, алгоритм зробить висновки про ваші преференції щодо формату. Netflix фіксує, де ви поставили на паузу, можливо щоб роздивитись деталь сцени — сигнал уваги; де перемотали назад, бо щось важливе пропустили; чи бросили серіал на третій серії, що означає нецікаво, чи передивились весь сезон за ніч, що свідчить про захоплення.

Неявна персоналізація не потребує зусиль користувача, працює автоматично; відображає реальну поведінку, а не декларовані преференції; генерує величезні обсяги даних, де кожна дія стає data point; адаптується в реальному часі до зміни інтересів; працює навіть для пасивних користувачів. Але є недоліки: непрозорість, коли користувач не розуміє, чому бачить певний контент; відсутність контролю, бо важко "виправити" алгоритм; питання приватності та відчуття, що "за тобою стежать"; помилкові висновки, наприклад переглянули весь фільм не тому, що подобається, а тому що заснули з увімкненим екраном; reinforcement loops, коли алгоритм показує більше того, що вже дивились, звужуючи діапазон контенту.

Служба Spotify давно перетворилася з простого музичного сховища на складну екосистему персональних відкриттів, серцем якої є її гібридна система рекомендацій. Її феноменальна точність, що виражається в персоналізованих плейлистах на кшталт Discover Weekly, є прямою наслідком майстерного поєднання двох фундаментальних підходів до машинного навчання: фільтрації за вмістом та спільної фільтрації, які живляться мільярдами явних та неявних сигналів від користувачів. Ця архітектура дозволяє системі не лише розуміти музику на нейронному рівні, але й розуміти слухача через його поведінку, створюючи динамічну та еволюціонуючу картину смаку.

На першому рівні цієї системи працює фільтрація за вмістом, мета якої — об'єктивно зрозуміти, що являє собою кожен із понад 100 мільйонів треків у каталозі. Spotify аналізує як метадані (жанр, темп, тональність), так і, що важливіше, сам сирий аудіосигнал. Складні алгоритми розбивають музику на складові, визначаючи такі параметри, як танцювальність, енергійність, валентність (емоційний тон) та інструментальність. Це дозволяє оцифрувати саундтрек: наприклад, класифікувати пісню як мрійливу, з низькою енергією та високою танцювальністю. Однак у 2025 році аналіз вийшов

далеко за межі базових аудіохарактеристик. Завдяки впровадженню великих мовних моделей (LLM), Spotify тепер може вбудовувати в алгоритмічний простір весь культурний контекст навколо релізу: тексти пісень, опис від артиста, обкладинки, навіть обговорення в соціальних мережах та музичних блогах. Таким чином, трек визначається не лише тим, як він звучить, але й тим, про що він, які емоції викликає та з якими явищами культури асоціюється.

Однак знати музику недостатньо — потрібно знати слухача. Тут на сцену виходить спільна фільтрація, яка є другим стовпом системи. Її логіка ґрунтується на простому принципі: «користувачі, які слухали подібне, мають подібні смаки». Алгоритм будує гігантську матрицю взаємодій, де рядки — це мільйони користувачів, а стовпці — мільйони пісень. Якщо ви і мільйон інших людей регулярно слухаєте певних виконавців, система зробить висновок, що ваші смаки схожі, і запропонує вам музику, яку обожнюють ті «алгоритмічні сусіди», а ви ще не чули. Цей метод неймовірно ефективний для відкриттів у межах ваших жанрових уподобань.

Справжня магія відбувається в гібридизації цих двох методів, і паливо для неї — це дані, які Spotify збирає про кожну вашу взаємодію з платформою. Ці дані поділяються на явні (експліцитні) та неявні (імпліцитні) сигнали. До явних відносяться свідомі дії: лайк (лайк) пісні, додавання треку до плейлисту, підписка на артиста чи створення власних плейлистів. Це прямий і зрозумілий відгук. Однак значно багатшим джерелом інформації є неявні сигнали: чи пропускаєте ви пісню в перші 30 секунд, дослуховуєте до кінця, додаєте її в «Вподобане» після кількох прослуховувань, у який час доби чи на якому пристрої ви її слухаєте. Пропуск треку є, мабуть, найсильнішим негативним сигналом, тоді як повне прослуховування та додавання до плейлисту — вагомим позитивним. Усі ці мільйони мікро-дій постійно оновлюють ваш «профіль смаку» — динамічне внутрішнє представлення ваших уподобань у Spotify.

Discover Weekly є апофеозом цієї гібридної системи. Щотижня алгоритм, використовуючи спільну фільтрацію, знаходить користувачів з надзвичайно схожими на ваш профілями смаку. Потім він аналізує, що слухають ці «музичні двійники», а ви — ні. Але замість того, щоб сліпо пропонувати ці треки, система пропускає їх через фільтр фільтрації за вмістом, перевіряючи, чи відповідають

їхні аудіохарактеристики (темп, енергія, валентність) та семантичний контекст тому, що ви імпліцитно любите. Наприклад, вона може запропонувати не просто ще один трек у жанрі інди-поп, а саме той, який має аналогічну мрійливу атмосферу та ліричні теми, що й ваші улюблені пісні, і який, до того ж, зберігають у своїх плейлистах люди з ідентичним вашому музичним смаком. Важливо, що ця система не статична. Вона працює в режимі постійного зворотного зв'язку: якщо ви пропускаєте рекомендований трек у Discover Weekly, алгоритм відзначає це та коригує подальші пропозиції, тим самим постійно вчащаючись і адаптуючись до змін у ваших уподобаннях.

Еволюцією цього гібридного підходу стали інновації на кшталт Prompted Playlists (запропоновані плейлисти). Ця функція, анонсована наприкінці 2025 року, поєднує глибоке розуміння системою вашого профілю с новим рівнем явного контролю. Ви можете словесно описати бажаний плейлист природною мовою, наприклад: «музика моїх улюблених артистів за останні п'ять років, але невідомі рідкісні треки» або «енергійний поп та хіп-хоп для 5-кілометрового бігу, що переходить у розслабляючі пісні для заминки». Алгоритм, використовуючи гібридну модель (ваш історичний профіль смаку + аналіз вмісту для розуміння запиту), генерує точно відповідний плейлист. Це радикально новий етап, де людина та алгоритм входять у творчу співпрацю: користувач стає куратором, вказуючи напрямок, а AI — неймовірно еруйованим та особистим асистентом, який цей намір виконує.

Отже, ефективність Spotify полягає саме в цій синергії. Система не покладається виключно на схожість між користувачами (спільна фільтрація) або виключно на схожість між піснями (фільтрація за вмістом). Вона будує наскрізний ланцюжок розуміння: від фізичних та культурних характеристик музики через колективну поведінку мільйонів слухачів до особистих, часто неусвідомлених, звичок конкретної людини.

2.1.2. Колаборативна фільтрація.

Колаборативна фільтрація працює за принципом "Люди, схожі на вас, також полюбили...". Вона не аналізує сам контент, а аналізує патерни вподобань різних користувачів. Ідея проста: якщо ви та інша людина оцінюєте сто фільмів дуже схоже, то сто перший фільм, який їй сподобався, ймовірно сподобається і вам. User-

Based Collaborative Filtering працює так: знаходимо користувачів зі схожими вподобаннями, дивимось що вони полюбили а ви ще не бачили, рекомендуємо вам це. Наприклад, ви оцінили позитивно "Breaking Bad", "The Wire", "True Detective". Користувач Б оцінив позитивно ті самі серіали плюс "Fargo". Система робить висновок: ви схожі на Користувача Б. Рекомендація: "Подивіться Fargo". Item-Based Collaborative Filtering замість пошуку схожих людей шукає схожі об'єкти — фільми, пісні, статті. Система аналізує, які пари об'єктів часто споживаються разом. Якщо ви подивились об'єкт А, вона рекомендує об'єкт Б, який часто споживається разом з А. Amazon використовує це у формулі "Customers who bought this also bought...". Вісімдесят відсотків людей, які купили книгу "Sapiens", також купили "Homo Deus". Ви купили "Sapiens". Рекомендація: "Homo Deus".

Matrix Factorization — більш просунута математична техніка, що лежить в основі Netflix Prize, конкурсу Netflix 2006-2009 років на кращий рекомендаційний алгоритм з призом один мільйон доларів. Ідея полягає в тому, щоб представити всіх користувачів та весь контент у багатовимірному просторі "прихованих факторів". Для фільмів такими факторами можуть бути сила екшн-компоненти, рівень романтики, інтелектуальна складність, темний настрій. Кожен фільм та кожен користувач отримує числові значення по кожному фактору. Потім система рекомендує фільми, які "близькі" до користувача в цьому багатовимірному просторі.

Колаборативна фільтрація має переваги: не потребує розуміння самого контенту, працює з будь-яким типом об'єктів; може знаходити несподівані зв'язки, створюючи ефект serendipity — приємні неочікувані знахідки; має ефект мережі, коли чим більше користувачів, тим точніші рекомендації; здатна робити cross-domain discoveries, рекомендуючи речі з категорій, які ви самі б не досліджували. Але є й недоліки: cold start problem для нових користувачів без історії або нового контенту, який ніхто ще не оцінив; popularity bias, коли популярні об'єкти отримують ще більше рекомендацій, а нішеві залишаються невидимими; filter bubble, коли рекомендації базуються на минулому і важко вийти за межі звичного; sparsity, бо матриця "користувачі × об'єкти" дуже розріджена, Netflix має мільйони фільмів, а кожен користувач бачив лише десятки; shilling attacks, коли можна маніпулювати системою, створюючи фейкові акаунти з

штучними оцінками.

2.1.3. Контент-орієнтована персоналізація.

Контент-орієнтована персоналізація працює за принципом "Якщо вам сподобалось це, вам сподобається схоже". На відміну від колаборативної фільтрації, тут аналізується сам контент — його характеристики, атрибути, властивості. Спочатку створюється "профіль" контенту з описом його характеристик, потім "профіль" користувача з характеристиками контенту, який йому подобається, і нарешті рекомендується контент, чий профіль найкраще відповідає профілю користувача. Spotify Music Genome аналізує кожен трек по сотнях параметрів: акустичні характеристики як темп, тональність, гучність, енергійність, танцювальність; інструментальний склад з наявністю вокалу, гітар, синтезаторів, ударних; емоційна валентність від веселого до сумного, від агресивного до спокійного; жанрові маркери від року та попу до електроніки та джазу; епоха від шістдесятих та вісімдесятих до сучасності. Якщо ви регулярно слухаєте треки з характеристиками високий темп понад сто сорок ударів на хвилину, електронні синтезатори, мінор, енергійно, без вокалу, система рекомендує інші треки з подібними параметрами.

Для текстового контенту платформи як Medium, Feedly та новинні агрегатори використовують TF-IDF, Term Frequency-Inverse Document Frequency, аналіз того, які слова та фрази характерні для статей, що вам подобаються. Ви читаєте багато статей зі словами "machine learning", "neural networks", "AI", "deep learning". Система робить висновок: вас цікавить штучний інтелект. Вона рекомендує статті з високою концентрацією цих термінів. Topic Modeling через LDA, Latent Dirichlet Allocation, дозволяє автоматично виявляти "теми" у документах. Статті групуються за прихованими темами без ручної розмітки: Тема 1 може бути про стартапи, венчур, фандрейзинг, інвестиції; Тема 2 про дизайн, UX, інтерфейси, прототипи; Тема 3 про блокчейн, криптовалюту, децентралізацію. Якщо ви часто читаєте статті з Теми 2, рекомендуються інші статті цієї теми.

YouTube здійснює аналіз відео-контенту на кількох рівнях: візуальний аналіз об'єктів на екрані — люди, тварини, пейзажі, колірна гама, динамічність кадрів; аудіо-аналіз мови, англійська чи українська, музики з жанром та настроєм, звукових ефектів; метадані як заголовок, опис, теги, категорія; субтитри та транскрипція для

NLP-аналізу тексту, про що говорять. Якщо ви дивитесь багато відео з характеристиками англійська мова плюс код на екрані плюс слова "tutorial", "programming" у заголовку, рекомендуються інші програмістські туторіали.

Контент-орієнтована персоналізація має переваги: немає cold start для контенту, новий фільм чи пісню можна рекомендувати одразу, адже є опис характеристик; прозорість, бо можна пояснити чому рекомендується через схожий жанр, акторів, тематику; незалежність від інших користувачів, працює навіть для унікальних смаків; здатність рекомендувати нішевий контент, малопопулярні об'єкти з потрібними характеристиками. Недоліки включають обмеженість новизною, коли рекомендується лише схоже і важко виходити за межі звичного; over-specialization, що може загнати в дуже вузьку нішу; потребу в якісному описі контенту, не весь контент легко описати, як наприклад пояснити чому цей фільм цікавий; відсутність serendipity з малими шансами відкрити щось несподіване; content acquisition costs, коли створення профілів контенту може бути дорогим, особливо для мультимедіа.

2.1.4. Гібридні системи.

Найефективніші сучасні рекомендаційні системи комбінують множинні підходи, компенсуючи слабкості одних сильними сторонами інших. Weighted Hybrid, зважена гібридизація, комбінує результати різних алгоритмів з певними вагами. Наприклад, п'ятдесят відсотків колаборативна фільтрація, тридцять відсотків контент-орієнтована, двадцять відсотків популярність чи trending. Фінальний рейтинг дорівнює нуль цілих п'ять десятих помножити на CF score плюс нуль цілих три десятих помножити на CB score плюс нуль цілих дві десятих помножити на popularity score. Switching Hybrid, перемикальна гібридизація, означає що система вибирає різні алгоритми залежно від ситуації. Netflix для нового користувача використовує popularity-based топ-десять фільмів плюс explicit preferences, запитує жанри; після десяти оцінених фільмів додає item-based collaborative filtering; після п'ятдесяти оцінених переходить на matrix factorization; для нового фільму використовує content-based через схожість з іншими фільмами.

Cascade Hybrid, каскадна гібридизація, застосовує алгоритми послідовно, кожен наступний уточнює результати попереднього.

Спочатку контент-орієнтована фільтрація відбирає тисячу кандидатів зі схожим жанром та тематикою. Потім колаборативна фільтрація ранжує ці тисячу за схожістю користувачів. Нарешті контекстуальні фактори як час доби та пристрій роблять фінальне ранжування топ-двадцяти. Feature Augmentation, доповнення ознак, означає що результати одного методу стають вхідними даними для іншого. Контент-аналіз виявляє що фільм належить до кластеру "psychological thrillers". Ця інформація додається як feature для колаборативної фільтрації. Тепер CF може знаходити користувачів, які люблять саме цей кластер. Meta-Level Hybrid, мета-рівнева гібридизація, передбачає що один алгоритм навчається на виході іншого. Контент-орієнтована система створює "профіль користувача", опис його смаків через характеристики контенту. Колаборативна система використовує цей профіль для пошуку схожих користувачів. Знаходяться люди зі схожими профілями, але які споживали інший контент.

YouTube комбінує десятки сигналів у своїй гібридній системі. На етапі Candidate Generation, генерації кандидатів, працюють колаборативна фільтрація для відео які дивились люди схожі на вас, контентна для відео зі схожими тегами та категоріями, соціальна для відео від каналів на які ви підписані, trending для популярного зараз контенту. Результат — список з кількох тисяч потенційних відео. На етапі Ranking, ранжування, нейронна мережа оцінює кожне відео по сотням факторів: передбачуваний watch time, як довго ви будете дивитись; click-through rate, ймовірність кліку; engagement, ймовірність лайку чи коменту; свіжість, коли нові відео отримують boost; різноманітність, щоб не було десяти відео підряд від одного каналу. На етапі Re-Ranking, пере-ранжування, відбувається фінальна корекція: фільтрування неприйнятного контенту, забезпечення різноманітності жанрів, каналів, форматів, контекстуальна адаптація до часу доби та пристрою.

2.1.5. Контекстуальна персоналізація.

Контекстуальна персоналізація працює за принципом, що персоналізація залежить не лише від того хто ви, а й від де, коли та в якій ситуації ви зараз перебуваєте. Один і той самий користувач має різні потреби в різних контекстах. Часовий контекст включає час доби: ранок від шостої до дев'ятої години асоціюється з новинами, мотиваційним контентом, енергійною музикою; день від дванадцятої

до чотирнадцятої з легкими розважальними відео та подкастами; вечір від вісімнадцятої до двадцять другої з фільмами, серіалами, довгочитанням; ніч після двадцять другої з relaxing музикою, ASMR, легким контентом. Spotify помітив патерн: люди слухають різну музику в різний час. Ранком мотиваційну, енергійну музику для тренувань, workout playlists. Ввечері спокійну, розслаблюючу, chill playlists. Алгоритм адаптує рекомендації до часу. День тижня також має значення: понеділок асоціюється з мотиваційним, продуктивним контентом; п'ятниця з розважальним, вечірковим; неділя з релаксуючим, сімейним. Сезонність впливає теж: грудень приносить новорічний контент, літо музику для відпочинку та тревел-контент.

Локаційний контекст визначає де ви перебуваєте: вдома споживаєте довгий контент, фільми та серіали; у транспорті подкасти, короткі відео, музику; на роботі фонову музику, quick-read статті; у спортзалі енергійну музику, workout videos; в кафе середньотемпову музику, легке читання. Якщо ваша локація за даними Google Maps показує що ви в автомобілі через швидкість руху, YouTube може рекомендувати аудіо-контент, подкасти та музику замість відео. Proximity marketing показує рекламу кафе коли ви в радіусі двохсот метрів. Рекомендація статті про місцеві визначні пам'ятки з'являється коли ви в новому місті. Девайс-контекст має значення: смартфон підходить для коротких форматів, вертикального відео як Stories, Reels, TikTok; планшет для середніх форматів, інтерактивного контенту; ПК чи ноутбук для довгих статей, багатозадачності з роботою плюс YouTube; Smart TV для фільмів, серіалів, довгих відео в HD. Якість з'єднання також впливає: Wi-Fi дозволяє HD та 4K відео, 4G обмежується SD відео, 3G підходить для аудіо-контенту та тексту. YouTube автоматично знижує якість відео при повільному з'єднанні.

Соціальний контекст враховує з ким ви: наодинці дозволяє персональні преференції, з сім'єю передбачає family-friendly контент, з друзями популярний розважальний контент, на побаченні романтичну музику та рекомендації ресторанів. Netflix створює різні профілі для різних членів сім'ї. Але навіть в одному профілі система може визначати контекст: якщо увімкнено "Kids Mode", рекомендації змінюються. Активність-контекст визначає що ви робите: під час тренування пропонується енергійна музика, Spotify Running feature автоматично підбирає темп під ваш біг; під час роботи чи

навчання фокус-музика, lo-fi beats, instrumental; під час приготування їжі кулінарні відео, легка музика; під час прибирання енергійні плейлисти; під час засинання медитації, relaxing soundscapes. Якщо ваш Apple Watch визначає що ви тренуєтесь через високий пульс та активність, Apple Music автоматично пропонує workout playlists.

Емоційний контекст намагається визначити настрій через поведінку: швидкий скролінг, багато скіпів означає знуджений чи дратований стан, потрібно змінити тип контенту; повторне прослуховування сумних пісень свідчить про сумний настрій, можна рекомендувати більше або навпаки спробувати підняти настрій; активна взаємодія, багато лайків показує позитивний настрій, варто продовжувати в тому ж дусі. Spotify створює Mood Playlists як "Happy Hits", "Sad Songs", "Chill Vibes" під різні настрої. Алгоритм може визначати ваш настрій через музичний вибір та адаптувати рекомендації. Погодний контекст теж враховується: під час дощу пропонується спокійна меланхолійна музика, фільми для перегляду вдома; в сонячну погоду енергійна музика, outdoor активності; в холод comfort food рецепти, cozy контент.

Технології контекстуальної персоналізації базуються на даних сенсорів: GPS для локації, акселерометр для руху та активності, гіроскоп для орієнтації пристрою, мікрофон для ambient sound та визначення середовища, камера для розпізнавання обличчя та настрою через face recognition. Contextual Bandits — це Machine Learning підхід де алгоритм вибирає дії, рекомендації, на основі контексту, балансуючи exploration, пробувати нове, та exploitation, використовувати вже відоме. Deep Learning для контексту використовує нейронні мережі які приймають на вхід множинні контекстуальні сигнали одночасно і видають персоналізовані рекомендації.

Uber Eats демонструє комплексну контекстуальну систему. Рекомендації їжі залежать від часу: сніданок від восьмої до одинадцятої, обід від дванадцятої до чотирнадцятої, вечеря від вісімнадцятої до двадцять першої, нічна їжа після двадцять другої; локації, вдома означає доставка повноцінної їжі, офіс швидкі ланчі; погоди, холодно означає супи та гаряче, спека салати та холодне; дня тижня, вихідні brunch, будні quick meals; історії замовлень але з адаптацією до контексту, улюблена піца але вранці не рекомендується.

Контекстуальна персоналізація несе етичні виклики. Privacy

Invasion ставить питання наскільки відстеження локації, активності, навіть настрою через сенсори є прийнятним, скільки це занадто багато. Predictive Overreach виникає коли алгоритм "знає" про ваші потреби раніше ніж ви самі, наприклад Target передбачив вагітність дівчини раніше ніж її батько дізнався про це. Manipulation відбувається через використання контексту для маніпуляції, показ реклами фастфуду коли алгоритм визначив що ви голодні та поруч є McDonalds.

Персоналізація — це не монолітна технологія, а складна екосистема методів, кожен з яких має свої сильні та слабкі сторони. Явна персоналізація дає контроль користувачу, дозволяючи йому вручну налаштовувати свої уподобання, інтереси та параметри фільтрації контенту. Цей підхід будується на принципах прозорості та згоди, підвищуючи довіру, оскільки користувач точно розуміє, на підставі чого система формує його досвід. Однак його ефективність обмежена: більшість користувачів не вказують усіх своїх інтересів, а їхні переваги можуть змінюватися з часом, що призводить до неповної або застарілої картини.

Неявна персоналізація забезпечує автоматизовану та глибоку адаптацію, аналізуючи поведінку користувача в реальному часі: що він переглядає, як довго, що лайкає та коментує. Це дозволяє системі передбачати потреби та пропонувати релевантний контент, навіть про який користувач не замислювався. Але цей підхід несе ризики: формування «фільтрувальних бульбашок», де користувач ізолюється в інформаційному середовищі, що лише підтверджує його існуючі погляди; занепокоєння щодо конфіденційності через прихований збір даних; та відчуття втрати контролю над тим, яку інформацію ми отримуємо.

Ідеальна система персоналізації в сучасних медіа поєднує обидва підходи, створюючи синергію між контролем користувача та розумінням алгоритму.

Висновок для медіапрофесіонала: Сучасна стратегія персоналізації повинна балансувати між цими двома полюсами. Важливо розвивати системи, які:

1. Використовують неявну персоналізацію для ефективного виявлення контенту та підвищення залучення.

2. Надають користувачам явні інструменти контролю (наприклад, кнопки «Не цікаво», налаштування інтересів, зрозумілі пояс-

нення «Чому я це бачу?») для підвищення довіри та подолання ефекту бульбашки.

3. Свідомо інтегрують в алгоритми елементи серендипітету — можливість знайти щось неочікуване та нове, розширюючи кругозір аудиторії замість його звуження.

Таблиця 2.1.

Порівняння двох моделей явної та неявної персоналізації

Критерій	Явна персоналізація	Неявна персоналізація
Тип даних	Прямі вказівки користувача (налаштування, вибір категорій).	Поведінкові дані (кліки, перегляди, тривалість).
Контроль	Високий. Користувач активно керує процесом.	Низький. Процес автоматизований і часто непрозорий.
Точність	Обмежена свідомістю та ініціативою користувача.	Висока, за умови якісних алгоритмів та достатньо даних.
Гнучкість	Низька. Інтереси оновлюються тільки при ручному втручанні.	Висока. Система постійно адаптується до нової поведінки.
Основні ризики	Неповна картина інтересів, «лінь» користувача.	Фільтрувальна бульбашка, порушення приватності, маніпуляція.
Приклад	Підписка на конкретні теми в Google News, налаштування стрічки в соцмережах.	Рекомендації YouTube чи Netflix, стрічка Facebook/ Instagram.

Складено автором. Джерело: Personalization 101: Implicit vs Explicit // Hannon Hill (2019). https://www.hannonhill.com/blog/2019/personalization-101-implicit-versus-explicit-personalization.html

Таким чином, персоналізація перетворюється з технічного завдання на етичний вибір, де поєднуються потужність алгоритмів, права користувача та соціальна відповідальність медіа.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання перевіряють розуміння природи даних, з якими працюють сучасні медіа:

1. Концепція "5V": Традиційно Big Data описують через три "V" (Volume, Velocity, Variety). Які ще дві "V" (Veracity, Value) додають

сучасні дослідники і чому Veracity (Достовірність) є критичною саме для алгоритмічної журналістики?

2. Структуровані vs Неструктуровані дані: 80% даних у світі — неструктуровані (відео, текст постів, аудіо). Чому алгоритмам значно важче працювати з ними, ніж зі структурованими даними (таблицями, CRM), і які технології (NLP, Computer Vision) це вирішують?

3. Метадані: Поясніть вислів "Метадані розповідають про вас більше, ніж самі дані". Як час публікації, модель телефону та геотег допомагають створити психологічний портрет користувача без читання його листування?

4. Типи даних за джерелом: У чому різниця між First-party data (дані, які медіа збирає саме), Second-party data (дані партнерів) та Third-party data (куплені дані)? Чому світ відмовляється від останнього типу (Cookiepocalypse)?

5. Цифровий слід (Digital Footprint): Чим відрізняється *активний* цифровий слід (пости, які ви написали свідомо) від *пасивного* (історія пошуку, рухи мишкою, IP-адреса)? Який з них цінніший для маркетингу?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання спрямовані на навичку класифікації інформаційних потоків.

Завдання 1. Класифікація даних медіахолдингу. Ситуація: Ви керуєте аналітикою онлайн-видання. Завдання: Розподіліть наведені нижче дані на три колонки: *Структуровані, Напівструктуровані, Неструктуровані.* Список: Текст статті, ID користувача (число), коментар під статтею ("Круто!"), час перебування на сторінці (сек), відеорепортаж, HTML-код сторінки, хештеги (#новини), геолокація (координати). Мета: Зрозуміти, що комп'ютер легко "розуміє" координати, але для розуміння коментаря "Круто!" (чи це сарказм?) йому потрібен складний ШІ.

Завдання 2. Кейс "Метадані фотографії". Об'єкт: Зробіть фото на смартфон і завантажте його на сервіс перевірки EXIF-даних (або просто відкрийте "Властивості" файлу). Аналіз: Випишіть, що алгоритм знає про вас, просто отримавши цей файл. Модель камери (Рівень доходу?). GPS-координати (Де ви живете/працюєте?). Час зйомки (Ваш режим дня?). Налаштування світла (Ви профі чи

аматор?).

Завдання 3. Аналіз First-party data strategy. Ситуація: Google планує скасувати сторонні cookies (Third-party cookies). Завдання: Запропонуйте для новинарного сайту 3 способи легального збору *First-party data* (власних даних) про читачів. *Приклад:* "Впровадження реєстрації через пошту для доступу до ексклюзивних лонгрідів".

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Віктор Маєр-Шенбергер, Кеннет Кук'єр. *Великі дані. Революція, яка змінить наше мислення, працю і життя.* (Класичний вступ до філософії Big Data).

2. Лев Манович. *Cultural Analytics.* (Як використовувати Big Data для аналізу культури та медіаконтенту).

3. Martin Hilbert. *Big Data for Development.* (Огляд типів та джерел даних у цифровому суспільстві).

4. ГЛОСАРІЙ ТЕРМІНІВ.

- Big Data (Великі дані).
- Структуровані дані (Structured Data).
- Неструктуровані дані (Unstructured Data)
- Метадані (Metadata): "Дані про дані".
- Cookies (Куки)

2.2. Рівні персоналізації

Персоналізація як феномен має чітку ієрархію складності та глибини втручання в індивідуальний досвід користувача. Її розвиток можна розглядати як еволюцію від загального до приватного, від реактивного до прогностичного, і, нарешті, до творчого. Цей шлях пролягає через п'ять ключових рівнів, кожен з яких ґрунтується на попередньому і розширює можливості впливу.

2.2.1. Сегментація аудиторії (групова персоналізація).

Сегментація аудиторії, або групова персоналізація, становить концептуальну та практичну основу всього спектру персоналізованих комунікацій, діючи за принципом «один до багатьох». На відміну від індивідуального підходу, вона не прагне охопити неповторність окремої особистості, а виконує стратегічне завдання структурування гетерогенної маси клієнтів або користувачів на керовані, відносно однорідні групи (сегменти). Мета полягає у виявленні таких сегментів, внутрішні відмінності в межах яких щодо певної потреби, поведінки або характеристики є мінімальними, а відмінності між самими сегментами — максимальними. Це дозволяє організаціям оптимізувати ресурси, адаптуючи продукти, маркетингові повідомлення та канали комунікації не під кожного окремо, а під типових представників ключових груп, тим самим значно підвищуючи релевантність порівняно з єдиним масовим підходом.

Теоретичні та прикладні основи сегментації ґрунтуються на виборі критеріїв, які умовно поділяються на кілька категорій. Демографічні критерії (вік, стать, розмір сім'ї, рівень доходу, освіта, професія) є найбільш поширеними через простоту вимірювання та зв'язок з попитом на багато товарів і послуг. Однак їхня простота часто виявляється обмеженням, оскільки вони не розкривають мотивацію. Географічні критерії (країна, регіон, місто, тип місцевості, клімат) допомагають адаптувати пропозиції до місцевих умов, культурних особливостей та законодавства. Значно глибший рівень аналізу пропонують психографічні критерії, що розкривають внутрішній світ споживача: стиль життя, цінності, інтереси, ставлення до певних явищ (наприклад, до здорового харчування або технологій). Найбільш дієвими з точки зору прогнозування поведінки є критерії, засновані на поведінці. Сюди відносяться: статус користувача (новий, постійний,

колишній), користувацька активність (частота, тривалість сесій), лояльність, реакція на маркетингові стимули, етап в customer journey (знайомство, розгляд, покупка, післяпродажне обслуговування) та RFM-аналіз (Recency — як давно була остання покупка, Frequency — частота покупок, Monetary — грошова цінність). В B2B-сфері додаються такі критерії, як галузь компанії, її розмір, технологічний стек та структура прийняття рішень.

Технологічна реалізація сегментації еволюціонувала від ручного аналізу в ексель-таблицях до високоавтоматизованих процесів. Ключовим інструментом є кластерний аналіз — група методів машинного навчання без учителя (наприклад, k-середніх, ієрархічна кластеризація), які автоматично групують об'єкти (користувачів) на основі схожості за заданими ознаками. Дані для аналізу надходять із різноманітних джерел: CRM-системи, що зберігають історію взаємодій; веб-аналітика (Google Analytics), що фіксує поведінку на сайті; трекери соціальних мереж та медіапланінгові платформи; а також зовнішні бази даних (соціо-демографічні панелі). Сучасні Customer Data Platforms (CDP) інтегрують всі ці потоки даних, створюючи єдиний профіль на кожного користувача, що значно полегшує подальшу динамічну і точну сегментацію в реальному часі.

Класичними прикладами застосування групової персоналізації є: таргетована email-розсилка (окремий лист для чоловіків 30-45 років, які цікавляться авто, та інший — для жінок 20-30 років, які шукали косметику); стрічки новин або товарів у медіа та e-commerce (розділи «Для бізнесу», «Для розваг»); рекламні кампанії в соціальних мережах, націлені на аудиторії за інтересами; а також різні пропозиції та програми лояльності для різних сегментів клієнтів (наприклад, «Стандарт», «Преміум»).

Проте цей рівень персоналізації має суттєві обмеження. Надмірне узагальнення — головний недолік. Усередині сегмента «жінки 25-35 років, Київ» можуть опинитися як кар'єристки з високим доходом, так і молоді мами з іншими пріоритетами, що призводить до низької релевантності комунікації для частини сегмента. Статичність — багато традиційних сегментів оновлюються недостатньо часто і не встигають за динамікою змін у поведінці та уподобаннях людей. Симптоматичний підхід — сегментація часто оперує зовнішніми проявами (що купили), а не глибинними причинами (чому купили),

що обмежує прогностичну силу. Для подолання цих недоліків сучасна практика прямує до мікросегментації та динамічної сегментації в реальному часі на основі потокових даних, а також до поєднання класичної сегментації з іншими, більш глибинними рівнями персоналізації, створюючи гібридні моделі. Таким чином, сегментація залишається незамінним стратегічним інструментом, але її максимальна ефективність досягається не ізольовано, а як перший крок на шляху до більш індивідуалізованої взаємодії.

2.2.2. Персоналізація на рівні користувача.

Персоналізація на рівні користувача представляє собою фундаментальну зміну парадигми, перехід від роботи з групами до побудови унікального досвіду для кожної окремої особи, що реалізує ідеальну модель маркетингу «один до одного». Фокус повністю зміщується з узагальнених характеристик сегмента на динамічний цифровий слід, який залишає користувач у цифровому середовищі. Цей профіль будується не на припущеннях або демографії, а на фактичній поведінці, що розкриває справжні інтереси та наміри. Основним джерелом даних слугують явні дії, такі як здійснені покупки, проставлені лайки, явні пошукові запити, завантаження файлів або оформлення підписок. Однак справжню глибину та точність профілю забезпечують саме неявні сигнали, які користувач залишає непомітно для себе: тривалість перегляду сторінки або відео, швидкість та характер прокручування, рухи курсора миші, які фіксуються тепловими картами, послідовність переходів між розділами сайту та навіть моменти паузи чи швидкого гортання. Сучасні системи ідуть ще далі, інтегруючи контекстуальні дані про час доби, тип пристрою, місцезнаходження та навіть погоду, щоб зрозуміти стан та ситуацію користувача в момент взаємодії.

Технологічним серцем такої персоналізації є рекомендаційні системи, які досягають ефективності через поєднання різних методів. Колаборативна фільтрація працює за принципом колективної мудрості, знаходячи приховані зв'язки між користувачами. Вона аналізує масштабні масиви поведінкових даних, щоб виявити закономірності: якщо користувачі А і Б мають історію дуже схожих вподобань, і користувач Б щойно придбав товар Х, якого ще немає в історії користувача А, то система з високою ймовірністю порекомендує цей товар Х. Цей метод не потребує глибокого аналізу контенту, але

стикається з проблемою «холодного старту» для нових користувачів або товарів, про які ще немає достатньо даних. Контентна фільтрація вирішує цю проблему, орієнтуючись не на схожість між людьми, а на схожість між об'єктами. Вона аналізує атрибути товарів, статей або медіафайлів (жанр, автор, технічні характеристики, ключові слова) та порівнює їх з профілем переваг користувача, сформованим на основі його минулих взаємодій. Найпотужніші сучасні системи, такі як ті, що використовуються Netflix або Amazon, є гібридними. Вони поєднують обидва підходи, а також залучають алгоритми глибинного навчання для обробки нейронних сигналів та контексту, що дозволяє передбачати потреби з високою точністю, пропонуючи не просто схожі, а й доповнюючі або наступні логічним кроком продукти чи контент.

Однак така глибина індивідуалізації породжує значні суспільні та етичні ризики, найвідомішим з яких є формування так званої «фільтр-бульбашки». Це явище, коли алгоритм, намагаючись максимізувати залучення та задоволення, потрапляє в петлю позитивного зворотного зв'язку, постійно пропонуючи користувачеві контент, який лише підтверджує його існуючі переконання та смаки. В результаті інформаційний простір користувача звужується, альтернативні точки зору відсіюються, а можливість випадкового, серендипічного знайомства з новими ідеями зникає. Це веде до посилення соціальної поляризації, радикалізації поглядів та обмеження кругозору. Крім того, персоналізація на цьому рівні нерозривно пов'язана з колосальним збором персональних даних, що створює фундаментальну загрозу приватності та потенційно веде до маніпуляцій на основі вразливостей, виявлених у поведінці. Таким чином, персоналізація на рівні користувача є потужним інструментом, який водночас є викликом, вимагаючи балансу між технологічною ефективністю, етичною відповідальністю та повагою до автономії та права на інформаційну різноманітність кожної людини.

2.2.3. Контекстуальна адаптація (час, місце, пристрій).

Контекстуальна адаптація являє собою критично важкий рівень персоналізації, на якому система перестає сприймати користувача як статичний набір уподобань і починає інтерпретувати його як динамічну сутність, чиї потреби та інтенції кардинально змінюються в залежності від ситуації, що оточує його в конкретний момент часу.

Цей підхід ґрунтується на розумінні, що людина вдома ввечері та людина в дорозі вдень — це, з точки зору контексту, майже різні користувачі. Адаптація відбувається шляхом аналізу та синтезу низки контекстних факторів у реальному часі, що дозволяє не просто вгадувати, а розуміти поточні наміри.

Найочевиднішим фактором є часовий контекст. Сюди входять не лише час доби (ранок, день, вечір), а й день тижня, пори року, свята та навіть історичні періоди життя користувача. Алгоритми можуть розпізнавати паттерни: наприклад, у будні о 9:00 користувач регулярно відкриває стрічку бізнес-новин, а після 21:00 переглядає розважальні відео або серіали. Система адаптується, виводячи відповідний контент на перший план. Більш просунуті системи використовують час для тригерування специфічних дій: автоматичне замовлення кави через додаток доставки о 7:30 у будній день або рекомендація путівок у теплі країни взимку. Основним завданням є передбачення режиму дня та циклічності активності користувача.

Другим ключовим фактором є просторовий або геолокаційний контекст. Точні координати, отримані через GPS, визначення по WiFi або стільниковим вишкам, переводять персоналізацію з віртуальної площини у фізичний світ. Це дозволяє реалізувати механізми гео-таргетингу та гео-фенсінгу. Наприклад, коли користувач наближається до певного торгового центру, його додаток магазину може надіслати пуш-сповіщення про акцію саме в цій точці. Картографічні сервіси (Google Maps, Maps.me) адаптують інтерфейс і рекомендації закладів (кав'ярні, АЗС) на основі поточного місця знаходження та напрямку руху. Геолокація також використовується для автоматичного визначення мови інтерфейсу, валюти платежу та пропозицій, релевантних саме для даного регіону або міста.

Третім стовпом є контекст пристрою. Розуміння того, чи користувач взаємодіє з системою через смартфон, планшет, ноутбук, smart-TV або голосовий помічник, є критичним для адаптації форми та змісту. Це виходить далеко за рамки адаптивного веб-дизайну (Responsive Web Design). Наприклад, на смартфоні з невеликим екраном, особливо в умовах одноручного використання (one-hand use), система може пріоритезувати великі кнопки, короткі вертикальні відео (як в TikTok), спрощений текст та голосове введення. Натомість на десктопі з широким монітором та клавіатурою та ж система

може запропонувати багатовіконний інтерфейс, детальні таблиці даних, складні інструменти редагування та довгі текстові матеріали. Тип пристрою також вказує на ймовірний контекст використання: смартфон часто асоціюється з «мобільністю» та короткими сесіями, тоді як ноутбук — з роботою або навчанням.

Окрім цих трьох основних факторів, сучасні системи починають інтегрувати дані з сенсорів пристрою та інших контекстуальних джерел. Дані про рівень заряду батареї можуть спонукати систему до скорочення анімацій або пропозиції режиму енергозбереження. Дані про швидкість інтернет-з'єднання (3G, 4G, 5G, WiFi) дозволяють автоматично регулювати якість потокового відео або завантажувати контент заздалегідь. Метеодані (дощ, сонце, температура) можуть змінювати рекомендації: пропозиції таксі під час дощу, реклама кави в холодну погоду або сонцезахисних кремів у спекотний день. На передовій знаходиться інтеграція з Internet of Things (IoT): ваш розумний годинник може повідомити додатку про те, що ви почали пробіжку, і музичний сервіс автоматично запустить динамічний плейлист для бігу.

Технологічна реалізація контекстуальної адаптації вимагає архітектури, здатної до обробки потокових даних у реальному часі. Це передбачає використання подієво-орієнтованих архітектур (event-driven architecture), складної обробки сигналів (signal processing), де дані з різних джерел (геодатчиків, годинника системи, параметрів пристрою) агрегуються, фільтруються та аналізуються для миттєвого формування контекстного профілю. Машинне навчання, особливо моделі, що працюють з часовими рядами (time-series models), використовуються для виявлення складних залежностей між контекстом та поведінкою.

Однак цей рівень персоналізації несе в собі підвищені ризики для приватності. Постійний моніторинг місця знаходження, активності та стану пристрою створює вкрай детальну картину життя користувача. Крім того, існує небезпека контекстуальної помилки — коли система неправильно інтерпретує ситуацію. Наприклад, пропозиція весільного послуг після відвідування весілля друга може бути сприйнята як нав'язлива та некоректна. Таким чином, контекстуальна адаптація, будучи вершиною технічної релевантності, одночасно є і вершиною технічної нав'язливості, вимагаючи від розробників

не лише високої майстерності, але й глибокого етичного чуття та надання користувачеві прозорого контролю над тим, який контекст і для яких цілей використовується.

2.2.4. Передбачувальна (предиктивна) персоналізація.

Передбачувальна персоналізація є еволюційним кроком, що перетворює алгоритм з інструменту аналізу минулої та поточної поведінки на активного учасника, здатного моделювати ймовірне майбутнє користувача. На цьому рівні система перестає бути реактивною та навіть контекстуально-залежною; вона стає проактивною, прагнучи не просто задовольнити сформовану потребу, а визначити та сформувати її задовго до моменту усвідомлення самим користувачем. Замість відповіді на запит "що зараз?", вона намагається відповісти на питання "що далі?" і "що потрібно буде після цього?". Методологічною основою тут є складні моделі машинного навчання, зокрема алгоритми роботи з часовими рядами, які виявляють циклічність і тренди в поведінці, та графові нейронні мережі, що моделюють складні взаємозв'язки між користувачами, об'єктами та контекстами. Ці моделі аналізують не лише індивідуальну історію, а й макротренди в поведінці мільйонів інших користувачів, знаходячи приховані кореляції та шаблони, за допомогою яких можна з високою ймовірністю екстраполювати наступний крок конкретної особи.

Технічна реалізація передбачувальної персоналізації спирається на концепцію створення динамічної "моделі намірів" користувача. Алгоритми прогнозують не тільки наступну очевидну дію (наприклад, покупку того ж товару), а й потенційні життєві події та зміни контексту, що можуть породити нові потреби. Наприклад, стрімінговий сервіс, аналізуючи вашу активність, може виявити, що після тривалого перегляду серіалів певного жанру ви зазвичай шукаєте документальні фільми на близьку тему, і запропонує такий фільм заздалегідь. Банківська система, інтегруючи дані про транзакції, геолокацію (наприклад, відвідування автосалонів або агентств нерухомості) та етап життєвого циклу, може запропонувати індивідуальну іпотечну або автомобільну кредитну пропозицію ще до того, як клієнт офіційно звернеться по неї. У сфері логістики та електронної комерції предиктивні моделі дозволяють здійснювати "передбачуване розміщення" товарів на складах, переміщаючи популярні позиції ближче до регіонів, де попит на них очікується

найвищим у найближчі дні, що радикально скорочує терміни доставки.

Однак ця надзвичайна потужність породжує найбільш глибокі етичні дилеми серед усіх рівнів персоналізації. Перша проблема — це онтологічна помилка, коли ймовірнісний прогноз, зроблений алгоритмом, починає сприйматися системою (а згодом, можливо, і суспільством) як об'єктивна істина або неминучість. Це може призвести до того, що користувачеві починають пропонувати лише ті опції, які узгоджуються з прогнозом, фактично обмежуючи його вибір і створюючи ефект "пророцтва, що самовиконується". Друга загроза — дискримінація на основі предиктивних моделей. Алгоритм, навчений на історичних данах, може систематично занижувати оцінку кредитоспроможності або пропонувати менш вигідні умови певним соціальним групам, лише тому що в минулому їхні представники статистично частіше допускали прострочення, тим самим законсервувавши соціальну нерівність. Третя ключова проблема — це ерозія автономії та свободи волі. Коли система знає про наші майбутні дії, ймовірно, краще за нас самих, вона отримує безпрецедентну силу для мікротаргетованої маніпуляції — від політичної агітації до динамічного ціноутворення, що експлуатує передбачувану терміновість нашої потреби.

Таким чином, передбачувальна персоналізація відкриває нову парадигму взаємодії "людина-машина", в якій алгоритм виконує роль не слуги, а активного, а часом і надмірно втрутливого, порадника. Вона є логічним завершенням аналітичної гілки розвитку персоналізації, де весь минулий і поточний досвід користувача переробляється в ймовірнісну модель його майбутнього. Подальший розвиток у цьому напрямку вимагає вже не лише технічних інновацій, але й формування нового соціального та правового консенсусу щодо меж, у яких алгоритмам дозволено "думати за нас" і впливати на наші життєві траєкторії на основі власних передбачень.

2.2.5. Генеративна персоналізація (ШІ-створений унікальний контент)

Генеративна персоналізація є найвищим та найбільш трансформаційним рівнем розвитку персоналізованих систем, що становить радикальний відхід від усіх попередніх моделей. Якщо попередні рівні були орієнтовані на відбір, адаптацію або

прогнозування на основі існуючого контенту, то генеративна персоналізація означає перехід до синтезу абсолютно нового, унікального контенту в реальному часі для кожного окремого користувача. Ця можливість з'явилася завдяки стрімкому розвитку генеративних моделей штучного інтелекту, таких як великі мовні моделі (ChatGPT, Gemini, Claude) та системи генерування зображень і відео (DALL-E, Stable Diffusion, Midjourney, Sora). Фундаментальна відмінність полягає в тому, що система більше не є куратором чи аналітиком; вона стає творцем, що виробляє контент, якого раніше не існувало, ґрунтуючись на глибокому розумінні профілю, контексту та запитів користувача.

Технологічною основою цього рівня є генеративний штучний інтелект (GenAI) — різновид ШІ, який навчається на величезних масивах даних для розпізнавання складних шаблонів, а потім використовує ці знання для створення нових текстових, графічних, аудіо- чи відеоматеріалів із схожими характеристиками. У контексті персоналізації ці можливості виходять на новий рівень завдяки поєднанню з іншими технологіями: предиктивною аналітикою для передбачення потреб, обробкою даних у реальному часі для врахування контексту та агентними системами для автономного виконання складних завдань. Наприклад, платформи на кшталт Amazon Personalize вже зараз спрощують розробку складних систем рекомендацій, які можуть інтегрувати генеративні моделі для створення унікальних продуктів описів або маркетингових матеріалів.

Сфери практичного застосування генеративної персоналізації надзвичайно широкі та продовжують швидко розширюватися. В освіті (EdTech) ШІ може генерувати індивідуальні навчальні матеріали, наприклад, задачі з математики, що використовують улюблених персонажів дитини, або створювати адаптивні сценарії для навчальних симуляторів. У сфері розваг та ігрової індустрії (GameTech) відбувається революція: розробники експериментують з генерацією сюжетів, діалогів, ігрових ландшафтів і навіть цілих персонажів зі штучним інтелектом, які адаптуються до стилю гри та вибору гравця. Зокрема, такі технології, як NVIDIA Ace, прагнуть створювати ШІ-напарників для онлайн-ігор, здатних спілкуватися та діяти як реальні гравці. У креативних індустріях генеративний ШІ

дозволяє створювати персоналізований дизайн — від унікальних варіантів логотипів і предметів одягу до цілісних концепцій інтер'єрів на основі текстового опису вподобань користувача. Окремо варто відзначити появу сервісів, які адаптують інтерфейси програмного забезпечення під конкретні робочі потоки користувача або навіть генерують підсумки серій фільмів, як це робить Prime Video, що кардинально підвищує зручність.

Механізм генеративної персоналізації ґрунтується на постійному циклі «спостереження-синтезу-взаємодії». Система спостерігає за користувачем, аналізуючи його явні дії (пошук, покупки) та неявні сигнали (тривалість перегляду, контекст), формує динамічний профіль. Потім, отримавши запит або передбачивши потребу, вона не шукає готове рішення, а генерує новий контент, використовуючи підказку (prompt), яка поєднує запит користувача з даними його профілю. Наприклад, замість того щоб пропонувати існуючий рецепт, система може створити унікальний рецепт страви, зображеної у серіалі, який дивиться користувач, як це робить функція Smart Food на телевізорах Samsung. Подальше вдосконалення цього механізму пов'язане з розвитком мультимодальних моделей, здатних одночасно обробляти текст, зображення, звук і відео для створення ще більш комплексного та персоналізованого досвіду.

Однак така безпрецедентна могутність творення породжує низку серйозних викликів та ризиків. Перша група проблем стосується якості та контролю: генерований контент може містити фактичні помилки, упередження, унаслідовані з тренувальних даних, або бути «бездушним», що призводить до втрати довіри користувачів. Питання авторського права стають надзвичайно гострими — важко визначити право власності на контент, створений ШІ на основі мільйонів творів людей. Друга група ризиків пов'язана з безпекою та етикою: технології генерації глибоких підробок (deepfakes) можуть використовуватися для поширення дезінформації та маніпуляцій, а системи, що базуються на великих даних, становлять загрозу конфіденційності особистих даних. Найглибший філософський виклик полягає в потенційному звуженні культурного досвіду та формуванні «ідеальної ізольованої реальності». Якщо алгоритм постійно генерує лише той контент, який ідеально відповідає смакам та переконанням користувача, це може призвести до ще більшого

посилення ефекту «фільтр-бульбашки», обмеження критичного мислення та позбавлення людини можливості випадкових, але важливих відкриттів. Крім того, масове впровадження таких технологій може спричинити значні зміни на ринку праці, особливо в креативних професіях.

Таким чином, генеративна персоналізація представляє собою нову парадигму взаємодії людини з технологіями, переходять від інструменту задоволення запитів до співтворця індивідуальної реальності. Вона відкриває вражаючі можливості для креативності, адаптації та підвищення ефективності в навчанні, роботі та розвагах. Проте її розвиток вимагає виваженої, етично обґрунтованої дорожньої карти, що включає створення надійних систем перевірки якості, чітких правових рамок для захисту авторства та приватності, а також розвиток цифрової грамотності в суспільстві. Майбутнє полягає не у відмові від цієї потужної технології, а в пошуку балансу, де генеративний ШІ буде слугувати розширенню людських можливостей, а не їх обмеженню, допомагаючи створювати унікальний, але не ізольований досвід.

Висновок. Еволюція рівнів персоналізації демонструє чіткий тренд: від групи до індивіда, від минулого до майбутнього, від відбору до творення. Кожен наступний рівень підвищує релевантність і залученість, але експоненційно збільшує складність, вимоги до даних та масштаб етичних дилем. Ефективна та відповідальна персоналізація майбутнього вимагатиме балансу між усіма цими рівнями, де технологічна могутність буде обмежена рамками прозорості, контролю користувача та соціальної відповідальності.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допоможуть розрізнити етапи розвитку персоналізації:

1. Від One-to-All до One-to-One: У чому полягає фундаментальна різниця між моделлю масової комунікації (ТБ, газета) та моделлю індивідуальної комунікації? Чому модель *One-to-One* була економічно неможливою до ери Big Data?

2. Сегментація vs. Персоналізація: Маркетологи часто плутають ці поняття. Чому розсилка листа всім "жінкам 25-35 років" — це сегментація, а розсилка листа "Маріє, ось кросівки до вашої вчорашньої сукні" — це персоналізація?

3. Динамічний контент (DCO): Як технологія *Dynamic Creative Optimization* дозволяє змінювати елементи банера (колір, текст, картинку) в реальному часі залежно від того, хто на нього дивиться?

4. Контекстуальна (Гіпер-)персоналізація: Як дані про *поточний момент* (погода, геопозиція, швидкість руху, заряд батареї) змінюють контент? (Приклад: Spotify пропонує різну музику для ранкової пробіжки та вечірнього відпочинку, навіть для однієї людини).

5 ШІ-асістоване створення (Generative Personalization): Це найвищий рівень. Як генеративний ШІ (GenAI) змінює парадигму з "рекомендації *готового* контенту" на "створення *нового* контенту" під користувача (наприклад, персоналізований переказ новин)?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання допоможуть на практиці побачити різницю між рівнями:

Завдання 1. Побудова "Піраміди персоналізації"

• Об'єкт: Оберіть відомий бренд (наприклад, Nike або Netflix).

• Завдання: Розпишіть його комунікацію на 4 рівнях:

1. *Масовий (Broadcast):* (Реклама на Супербоулі).

2. *Сегментований:* (Розсилка для бігунів).

3. *Персоналізований:* (Рекомендація конкретної моделі взуття на основі історії покупок).

4. *Генеративний / Предиктивний:* (Створення унікального відео привітання з річницею тренувань, згенерованого ШІ).

Завдання 2. Експеримент з ChatGPT: "Адаптивна журналістика"

• Ситуація: У вас є суха економічна новина про підвищення податків.

Завдання: Використовуючи ШІ, згенеруйте три варіанти заголовку та ліду (вступу) для різних персон:

- *Персона А:* Студент, який підробляє кур'єром. (Акцент на цінах).

- *Персона Б:* Власник малого бізнесу. (Акцент на звітності).

- *Персона В:* Пенсіонер. (Акцент на субсидіях).

• Мета: Зрозуміти, як медіа майбутнього можуть показувати одну й ту ж новину по-різному різним людям.

Завдання 3. Аналіз Netflix Artwork Personalization

• Кейс: Netflix змінює обкладинки (thumbnails) фільмів.

• Завдання: Поясніть логіку алгоритму:

- Чому користувачу, який любить мелодрами, на обкладинці фільму "Кримінальне чтиво" покажуть Уму Турман і Траволту в танці?

- Чому фанату бойовиків на тому ж фільмі покажуть Семюела Л. Джексона з пістолетом?

• Висновок: Як це впливає на очікування від фільму?

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Don Peppers, Martha Rogers. *The One to One Future.* (Класична книга, яка передбачила цей перехід ще у 90-х).

2. McKinsey & Company. *The value of getting personalization right—or wrong is multiplying.* (Звіт про економічну вигоду гіперперсоналізації).

3. Netflix Tech Blog. *Artwork Personalization at Netflix.* (Технічний опис того, як працює підбір картинок).

4. HBR (Harvard Business Review). Customer Loyalty in the Age of AI.

4. ГЛОСАРІЙ ТЕРМІНІВ

• Масова кастомізація (Mass Customization): Виробництво товарів або послуг, що задовольняють індивідуальні потреби клієнтів, але з ефективністю масового виробництва (наприклад, ім'я на стаканчику Starbucks).

• Сегментація (Segmentation): Поділ аудиторії на групи (кластери) за спільними ознаками (вік, стать, географія). Це метод "від одного до багатьох".

• Гіперперсоналізація (Hyper-personalization): Найбільш просунута форма персоналізації, яка використовує дані в реальному часі та ШІ для надання максимально релевантного контенту конкретному користувачеві в конкретний момент.

• DCO (Dynamic Creative Optimization): Рекламна технологія, яка автоматично створює тисячі варіантів одного оголошення (комбінуючи картинки, текст, заклики до дії), тестує їх і показує найефективніший для конкретного користувача.

• Next Best Action (NBA): Предиктивна модель, яка визначає, яку дію компанія має запропонувати клієнту наступною (подзвонити, надіслати email, показати банер), щоб максимізувати результат.

2.3. Технологічна інфраструктура

2.3.1. Збір даних: cookies, піксель трекінг, поведінкова аналітика.

Щоб створити персоналізований контент, алгоритми штучного інтелекту повинні мати доступ до величезних обсягів інформації про користувачів. Технологічна інфраструктура збору цих даних — це «нервова система» сучасних медіа. Вона включає як знайомі інструменти (кукі), так і менш очевидні, але потужні механізми, які дозволяють не просто рахувати відвідувачів, але й глибоко аналізувати їхню поведінку на рівні кожного кліка та руху миші. Розуміння цієї інфраструктури є критично важливим для повного усвідомлення того, як існує сучасна цифрова медіа-екосистема.

1. Cookies: Пам'ять інтернету та епоха «Cookieless Future». Cookies (куки) — це невеликі текстові файли, які веб-сайт зберігає у браузері користувача. Вони виконують роль пам'яті інтернету, але не всі куки однакові за своєю природою та призначенням.

• First-Party Cookies (Власні куки) належать домену сайту, на якому перебуває користувач. Це, у певному сенсі, «хороші» куки, оскільки без них неможливий базовий функціонал: вони зберігають дані для авторизації (щоб не вводити логін і пароль щоразу), запам'ятовують вміст кошика в інтернет-магазині або мовні налаштування. Для медіа та ШІ вони дозволяють зв'язувати дії одного користувача в межах одного сайту між різними візитами.

• Third-Party Cookies (Сторонні куки) належать стороннім доменам, переважно рекламним мережам (Google, Criteo). Вони зазвичай вбудовані в рекламні банери або соціальні плагіни та можуть відстежувати користувача при переході між десятками різних сайтів. Саме вони дозволили побудувати систему глобального профілювання інтересів: відвідавши сайт про туризм, ви могли побачити рекламу готелів на зовсім іншому ресурсі.

Сьогодні індустрія переживає революцію, відому як Cookieless Future. Провідні браузери (Apple Safari, Mozilla Firefox, а з 2024 року — і Google Chrome) поступово повністю блокують сторонні куки. Це пряма відповідь на зростаючу турботу суспільства про приватність. Для медіа це означає необхідність корінної зміни стратегії: зосередження на зборі та обробці First-Party Data (власних даних), отриманих через прямі взаємодії з аудиторією — підписки,

реєстрації, опитування, переваги в налаштуваннях.

2. Піксель трекінг (Tracking Pixels): Невидимі «маяки». Піксель трекінгу — це надзвичайно простий, але ефективний інструмент відстеження. Це невидиме зображення розміром 1x1 піксель, вбудоване в код HTML-сторінки або електронного листа.

• Принцип роботи: Коли користувач завантажує сторінку або відкриває листа, його браузер автоматично завантажує цей крихітний піксель з сервера компанії-трекера (наприклад, Facebook або Google). Разом із запитом на завантаження передається набір даних: IP-адреса, інформація про пристрій і браузер, а також точні відомості про те, яку саме сторінку або дію відстежували.

• Значення для ШІ: Піксель дозволяє безшовно інтегрувати зовнішні платформи. Наприклад, Meta Pixel дає змогу алгоритмам Facebook зв'язати анонімну активність користувача на сторонньому сайті (перегляд товару, додавання до кошика) з його реальним профілем у соціальній мережі. Це основа для механізму ретаргетингу: побачивши кросівки на сайті магазину, ви через кілька хвилин знаходите їх у рекламній стрічці Instagram. Подібні пікселі є у всіх великих рекламних і аналітичних системах, формуючи розгалужену мережу обміну даними.

3. Поведінкова аналітика (Behavioral Analytics): Від кількості до якості взаємодії. Якщо класичні системи аналітики (наприклад, застарілий Google Analytics Universal) зосереджувалися на агрегованих метриках на кшталт «переглядів сторінок», то сучасна поведінкова аналітика робить крок далі. Її мета — зрозуміти не *скільки*, а *як* користувачі взаємодіють з контентом. Для цього використовуються новітні інструменти, такі як Google Analytics 4 (GA4), Hotjar чи Microsoft Clarity, які аналізують не сторінки, а події (*events*).

• Теплові карти (Heatmaps): Візуалізують області сторінки, які привертають найбільшу увагу (часті кліки, зависання курсора) або, навпаки, ігноруються («сліпі зони»). Для редактора це прямий інсайт: який заголовок чи зображення працює, а який блок тексту ніхто не читає.

• Аналіз глибини скролу та часу перебування (Dwell Time): Ці метрики дозволяють відрізнити «клікбейтний» візит, коли користувач швидко закрив сторінку, від якісного занурення в контент, коли він прокрутив статтю до кінця і провів на ній значний час. ШІ може

використовувати ці дані, щоб оцінювати якість контенту та прогнозувати, які теми чи формати найкраще утримують аудиторію.

• Записи сесій (Session Recordings): Дозволяють буквально побачити, як користувач навігує сайтом, куди клікає, де зупиняється. Це незамінний інструмент для виявлення технічних проблем або незрозумілих інтерфейсів.

Для ШІ поведінкова аналітика стає джерелом найцінніших контекстуальних даних. Алгоритм бачить не просто «користувач X відвідав статтю Y», а «користувач X уважно прочитав 80% статті Y, двічі повернувся до другого абзацу, потім перейшов на статтю Z на схожу тему». Така деталізація дозволяє будувати набагато точніші прогнози та рекомендації.

Поведінкова аналітика — лише одна з технологій у ланцюжку збору даних. Порівняємо її з іншими ключовими методами.

Таблиця 2.2.

Порівняння поведінкова аналітика First-Party Cookies, Third-Party Cookies, Пікселі трекінгу

Метод	Що це?	Як живить ШІ	Переваги для медіа	Ризики/ обмеження
First-Party Cookies (Власні)	Текстові файли, що зберігають дані (логін, налаштування) на одному сайті.	Дозволяють зв'язувати дії одного користувача в рамках одного сеансу або між візитами.	Безпечні, підтримують базовий функціонал. Ключові в стратегії First-Party Data.	Мають обмежений термін дії.
Third-Party Cookies (Сторонні)	Файли, що належать стороннім доменам (рекламним мережам) та відстежують користувача між різними сайтами.	Створюють детальні профілі інтересів для таргетингу.	Історично дозволяли точну рекламу та збір даних про аудиторію.	Поступово блокується браузерами (Safari, Firefox, Chrome). Ризик для приватності.

Метод	Що це?	Як живить ШІ	Переваги для медіа	Ризики/ обмеження
Пікселі трекінгу	Невидимі зображення 1x1 піксель у коді сторінки/ листа, що сповіщають сервер про дію користувача.	Пряма інтеграція з платформами (наприклад, Meta Pixel). Дозволяє ШІ зв'язати активність на сайті з профілем у соцмережі.	Потужний інструмент ретаргетингу та атрибуції конверсій.	Також під загрозою через блокування трекерів. Потенційне порушення приватності.
Поведінкова аналітика	Системи (GA4, Hotjar, Clarity), що аналізують детальні взаємодії користувача з сайтом (події, теплові карти).	Надає ШІ контекстуальні дані про *якість* взаємодії, а не лише про її факт. Допомагає відрізнити цінну поведінку від випадкових кліків.	Прямий інсайт в UX, оптимізація контенту, підвищення залученості.	

Складено автором. Джерело. First-Party vs Third-Party Cookies // Tune (2025). https://www.tune.com/blog/first-party-vs-third-party-tracking-cookies-what-they-are-why-you-should-drop-them/

Висновок. Інфраструктура збору даних — це не пасивний механізм логування, а активний стратегічний актив. Відстеження за допомогою куків і пікселів забезпечує «горизонтальне» профілювання користувача в мережі, тоді як поведінкова аналітика дає «вертикальне», глибинне розуміння його взаємодії з конкретним контентом. Разом вони формують ту самісну базу знань, на якій навчаються та працюють усі сучасні системи персоналізації. Умови гри постійно змінюються (заборони третіх осіб, посилення регулювання приватності), що змушує медіакомпанії постійно адаптуватися, інвестувати в власні дані та будувати довіру з аудиторією через прозорість у їх використанні.

2.3.2. Хмарні обчислення, API та апаратне забезпечення (GPU): Невидима механіка ШІ.

Сучасний ШІ, особливо його найпотужніші моделі, не функціонує на персональних пристроях. Його життєвий простір — це гігант-

ські хмарні дата-центри, а взаємодія з ним відбувається через спеціалізовані інтерфейси. Розуміння цієї логіки критично важливе для медіапрофесіоналів, оскільки воно впливає на операційні бюджети, технічну автономію та навіть екологічну політику редакції.

1. Економіка та логіка API (Application Programming Interface)

API можна уявити як цифрового офіціанта, що спрощує складну комунікацію. Коли журналіст натискає кнопку «Згенерувати заголовок» у системі управління контентом (CMS, наприклад, WordPress), він не звертається безпосередньо до серверів OpenAI чи Google. Натомість CMS відправляє стандартизований запит через API. Цей «офіціант» доставляє запит на «кухню» — у хмарний дата-центр, де модель ШІ (як шеф-кухар) обробляє його і створює результат. Потім API так само непомітно доставляє відповідь назад у CMS, де вона відображається користувачеві.

Цей спосіб інтеграції робить передові технології доступними без необхідності тримати власні суперкомп'ютери. Однак він формує особливу економіку токенів (Tokenomics). Більшість провайдерів стягують плату не за підписку, а за фактичне використання, де одиницею розрахунку є токен (зазвичай, частина слова). Така модель «плати за обсяг» (Pay-as-you-go) є економічно ефективною для нерегулярних завдань, але може призвести до значних непередбачуваних витрат при масштабуванні, що вимагає від редакції уважного моніторингу та бюджетування.

2. Апаратна революція: від CPU до GPU. Могутність сучасного ШІ стала можливою завдяки фундаментальній зміні в апаратному забезпеченні. Традиційні центральні процесори (CPU) — це «універсальні професори». Вони блискуче виконують складні, але послідовні завдання, як-от запуск операційної системи чи обробка таблиць. Однак навчання нейронної мережі чи генерація мультимедіа — це не одне складне обчислення, а мільярди одночасних простих операцій з матрицями чисел.

Саме для цього оптимальні графічні процесори (GPU). Спочатку створені для обробки мільйонів пікселів у відеоіграх одночасно, GPU архітектурно є «армією старанних п'ятикласників». Вони мають тисячі невеликих ядер, здатних паралельно виконувати величезні обсяги однотипної роботи. Ця здатність робить GPU незамінними для ШІ. Сьогодні ринок апаратного забезпечення для навчання

найпотужніших моделей визначається дефіцитом спеціалізованих чіпів, як-от NVIDIA H100, фактично роблячи тих, хто має доступ до таких потужностей, технологічними гегемонами.

3. Екологічний слід: прихована ціна цифрового інтелекту. Потужна інфраструктура має конкретну фізичну та екологічну вартість. Енергоспоживання систем ШІ є колосальним. За даними досліджень, один запит до великої мовної моделі, на кшталт ChatGPT, може споживати в 10-100 разів більше енергії, ніж стандартний пошуковий запит у Google. Це пов'язано з титанічними обчисленнями, що виконуються в дата-центрах для генерації кожної відповіді.

Крім електроенергії, критичним ресурсом є вода, яка використовується для охолодження серверів, що працюють на межі потужностей. Великі дата-центри можуть споживати мільйони літрів питної води щороку для своїх систем охолодження. Це змушує прогресивні медіа та технологічні компанії починати враховувати вуглецевий та водний слід своїх ШІ-проєктів, шукаючи способи використовувати «зелені» джерела енергії та оптимізувати обчислювальні процеси для зменшення впливу на навколишнє середовище.

Таблиця 2.3.

Порівняння функцій CPU (Центральний процесор) та GPU (Графічний процесор)

Аспект	CPU (Центральний процесор)	GPU (Графічний процесор)
Архітектура	Небагато потужних ядер, оптимізованих для послідовних задач.	Тисячи менш потужних ядер, оптимізованих для паралельних обчислень.
Аналогія	Професор, що вирішує одну складну задачу.	Армія студентів, що одночасно вирішують тисячі простих підзадач.
Основна сфера	Універсальні завдання: ОС, офісні програми, веб-серфінг.	Графіка, ШІ, машинне навчання, криптомайнінг.
Роль у ШІ	Управління процесом, обробка логіки програми.	Безпосереднє навчання нейромереж та генерація контенту (текст, зображення, відео).

Складено автором. Джерело. RedSwitches. "CPU Vs GPU In 2026: Key Differences, Use Cases" (2025). https://www.redswitches.com/blog/cpu-vs-gpu-in-2026/

Таким чином, технологічна інфраструктура ШІ — це складний

симбіоз програмних інтерфейсів (API), спеціалізованого апаратного забезпечення (GPU) та масштабних хмарних ресурсів, що має не лише технологічні, а й економічні та екологічні наслідки. Для медіа це означає залежність від екосистем технологічних гігантів, необхідність гнучкого бюджетування на основі використання та зростаючу соціальну відповідальність за екологічні наслідки цифрової діяльності.

2.3.3. Рекомендаційні механізми: Алгоритмічний редактор.

Рекомендаційні системи стали центральною логікою функціонування сучасних цифрових платформ, виконуючи роль безперервного та алгоритмічного редактора, що формує інформаційний раціон користувача. Їх мета — передбачити та запропонувати контент (статтю, відео, товар) з максимальною ймовірністю взаємодії в даний момент. Для медіапрофесіонала розуміння внутрішньої механіки цих систем є критичним для ефективного досягнення цільової аудиторії.

Основи класифікації: як система навчається розуміти уподобання. Існують два фундаментальні підходи до побудови рекомендацій, кожен з яких має різні механізми навчання та обмеження.

Content-Based Filtering (Фільтрація на основі змісту) працює за принципом логічного продовження: "Якщо вам сподобався цей об'єкт, то, швидше за все, сподобаються й інші, схожі за характеристиками". Алгоритм аналізує атрибути контенту: ключові слова, теги, жанри, метадані, стилістику. Технічно це часто реалізується через векторизацію текстових даних (наприклад, моделі TF-IDF або word embeddings) та обчислення їхньої косинусної подібності. Основний плюс цього методу — прозорість і відсутність проблеми "холодного старту" для нових елементів. Однак він має два ключові недоліки: по-перше, схильність до надмірної спеціалізації, що веде до формування "бульбашки" одноманітного контенту; по-друге, нездатність знаходити міжпредметні та неочікувані асоціації, обмежуючи користувача лише тим, що він вже знає.

Collaborative Filtering (Колаборативна фільтрація) використовує колективну мудрість спільноти. Її логіка: "Користувачі, які в минулому демонстрували схожі смаки, мають високі шанси полюбити й ті ж самі об'єкти в майбутньому". Системі байдуже на внутрішній зміст статті чи відео; вона працює виключно з матрицею взаємодій "користувач-об'єкт". Найпоширеніші методи, такі як матрична факторизація (SVD),

шукають приховані (латентні) фактори, що пояснюють ці взаємодії (наприклад, зацікавленість у політиці, любов до документального кіно). Цей підхід дозволяє робити серендіпітні (несподівані) відкриття, рекомендуючи контент, що не має прямого зв'язку з минулою історією, але який оцінили схожі користувачі. Головним викликом є проблема "холодного старту" для нового користувача або нового елемента контенту, для яких ще немає достатньої кількості взаємодій. Тому на практиці найчастіше використовують гібридні моделі, які поєднують переваги обох підходів для отримання більш збалансованих та точних рекомендацій.

Сигнали: паливо для алгоритмічної машини. Точність рекомендацій прямо залежить від якості та різноманітності вхідних даних. Алгоритми аналізують як пряму, так і опосередковану зворотній зв'язок, які поділяються на дві категорії.

• Явні сигнали — це свідомі та прямі оцінки користувача: лайк, дизлайк, шкала рейтингу (наприклад, "5 зірок"), підписка на канал, текстовий коментар. Це сильні та недвозначні сигнали, проте отримати їх від широкої аудиторії складно — лише невеликий відсоток користувачів регулярно ставить оцінки або коментує.

• Неявні сигнали — це поведінкові дані, що збираються автоматично в процесі перегляду. Вони становлять основну частину вхідних даних для сучасних систем і часто є точнішими індикаторами реального інтересу. До них належать Тривалість перегляду / дочитування до кінця (Completion Rate): Найважливіший метрика для відеоплатформ та новинних агрегаторів. Швидкість прокручування та час затримки на елементі. Дії: шеринг, збереження, повторний перегляд, натискання паузи, пропуск вступу. Контекст споживання: час доби, тип пристрою, геолокація.

Інсайт: Для таких платформ, як TikTok, утримання уваги (Retention) є значно важливішим сигналом, ніж лайк. Якщо користувач дивиться відео до кінця, але не ставить лайк, алгоритм інтерпретує це як успіх і продовжить рекомендувати подібний контент.

Проблема "холодного старту" та шляхи її вирішення. "Холодний старт" — це фундаментальна проблема, коли система не має достатньо даних для формування осмислених рекомендацій. Це стосується двох випадків: новий користувач (немає історії поведінки) та новий об'єкт контенту (немає взаємодій).

Алгоритмічне рішення цієї проблеми зазвичай проходить через кілька етапів:

1. Етап популярності (Generic Popularity): Новому користувачеві спочатку показують контент, популярний серед усієї аудиторії або в його георегіоні. Це дозволяє негайно зафіксувати перші реакції.

2. Швидка кластеризація: Після перших 5-10 взаємодій система вже може віднести користувача до певного широкого кластера на основі демографії або найпростіших переваг.

3. Залучення гібридних моделей: Для нового контенту на перших порах активно використовується контентна фільтрація, порівнюючи його атрибути з вподобаннями користувачів.

Для медіа ця проблема має пряме практичне значення. Нові проєкти або статті потребують стратегічного просування для подолання початкового вакууму уваги. Можна використовувати A/B-тестування заголовків та зображень, публікацію в "пікові" години або сприяння шерингу для швидкого набору первинних сигналів. Розуміння цієї логіки підказує, чому перші години життя публікації в соціальних мережах так критично важливі — саме в цей період алгоритм вирішує, чи варто надати їй широке рекомендаційне покриття.

Таким чином, рекомендаційні системи — це не чорна скриня, а складний, але зрозумілий механізм, що працює на перетині даних, статистики та машинного навчання. Для медіапрофесіонала глибоке розуміння принципів content-based та collaborative фільтрації, механіки збору сигналів та стратегій подолання "холодного старту" перетворюється з технічної знання на конкурентну перевагу, що дозволяє ефективно вести діалог з основним редактором сучасності — алгоритмом.

2.3.4. A/B-тестування та багатоваріантне тестування: Науковий метод у медіа.

Сучасні медіа все рідше покладаються на редакторську інтуїцію та все частіше — на дані, отримані в результаті контрольованих експериментів. A/B-тестування та його розширені версії стали стандартом наукового підходу до вдосконалення контенту, дизайну та взаємодії з аудиторією, дозволяючи замінити суб'єктивне «я думаю» на об'єктивне «дані показують».

Класичне A/B-тестування: Фундамент дата-орієнтованого під-

ходу. A/B-тестування (спліт-тестування) — це методика порівняння двох. версій одного елемента (A — контрольна версія, B — варіант) для визначення ефективнішої. Суть полягає у випадковому поділі трафіку користувачів на дві групи, кожній з яких демонструється один із варіантів. Після збору достатнього обсягу даних аналізуються ключові метрики для вибору переможця.

Об'єкти тестування в медіа. Заголовки та підзаголовки: «Чому сон важливий» проти «Науково доведено: як сон впливає на ваш успіх». Зображення та обкладинки (thumbnails): Емоційне фото проти схеми або інфографіки. Заклики до дії (CTA): «Підписатися» проти «Отримати доступ» чи «Дізнатися більше». Розклад та форма публікації: Оптимальний час доби чи дня тижня для публікації контенту. Варіанти розміщення реклами та монетизації.

Критично важливе поняття — статистична значущість. Припинення тесту надто рано, на основі перших ознак переваги, може призвести до хибних висновків через вплив випадковості. Для надійності результатів необхідно заздалегідь розрахувати та досягти достатнього розміру вибірки.

Багатоваріантне тестування: Складність та масштаб. Багатоваріантне тестування (Multivariate Testing, MVT) є розширенням логіки A/B-тестування. Замість порівняння двох цілісних варіантів сторінки, MVT дозволяє одночасно тестувати множину комбінацій різних елементів на одній сторінці.

Механіка та приклад: Якщо ви тестуєте два заголовки, два зображення та два варіанти кнопки CTA, система автоматично створює та показує аудиторії всі можливі комбінації ($2 \times 2 \times 2 = 8$ різних версій). Потім аналізується, яка саме *комбінація* елементів, а не окремий елемент, працює найкраще. Такий підхід розкриває синергетичні ефекти між компонентами.

Вимоги та застосування: Для отримання статистично валідних результатів MVT потребує значно більшого трафіку та тривалого часу, ніж просте A/B-тестування. Тому вона традиційно доступна великим медіахолдингам, інтернет-виданням з мільйонною аудиторією та агрегаторам, які можуть собі дозволити подібні масштабні експерименти.

Еволюція за допомогою ШІ: Від статичних тестів до адаптивних алгоритмів. Традиційне A/B-тестування має суттєвий недолік:

поки триває фіксований експеримент (наприклад, тиждень), значна частина аудиторії продовжує бачити менш ефективний варіант, що веде до втрачених можливостей. Штучний інтелект усуває цю проблему завдяки алгоритмам динамічної оптимізації.

Алгоритм «Багаторукий бандит» (Multi-Armed Bandit): Ця модель натхненна азартною грою з кількома ігровими автоматами («бандитами»). На відміну від класичного A/B-тесту, яке спочатку збирає дані, а потім робить висновок, алгоритм постійно балансує між двома стратегіями:

1. Експлуатація (Exploitation): Направляти трафік на варіант, який на даний момент показує найкращі результати.

2. Дослідження (Exploration): Продовжувати тестувати інші варіанти на невеликій частці трафіку, щоб виявити потенційних майбутніх лідерів.

Таким чином, система динамічно перерозподіляє трафік: якщо варіант B демонструє перевагу вже через кілька годин, ШІ автоматично збільшує його частку з 50% до 70%, 80% і далі, мінімізуючи втрати на неоптимальному варіанті.

Персоналізація на льоту, як найдосконаліший вираз сучасної цифрової адаптивності, представляє собою якісний стрибок від статичного таргетування до динамічної, самонавчальної системи комунікації. Її суть полягає не просто у виборі найкращого контенту в цілому, а у миттєвому визначенні та демонстрації оптимального варіанта для кожного мікросегмента або навіть окремого користувача в конкретному контексті, створюючи унікальний досвід у реальному часі. Ця здатність ґрунтується на потужних алгоритмах машинного навчання, які проводять безперервні експерименти — так зване А/Б-тестування нового покоління, або багатовимірне тестування, де одночасно перевіряються сотні комбінацій заголовків, зображень, формату, часу публікації та каналу поширення. Система не чекає фінального результату кампанії, щоб зробити висновки; вона аналізує потік взаємодій у реальному часі, визначаючи не лише що працює, але й для кого саме, за яких умов і чому, динамічно перенаправляючи ресурси на найефективніші комбінації.

Проте сфера застосування персоналізації на льоту значно ширше за рамки веб-сайтів. У медіа та контент-маркетингу вона перетворює єдиний масив контенту на безліч адаптивних шляхів. Наприклад,

одна й та сама новина може автоматично генерувати різні варіанти вступу: для фінансової платформи — з акцентом на біржові наслідки, для молодіжного видання — з фокусом на культурний вплив, для регіонального ЗМІ — з посиланням на місцевий контекст. У цифровій рекламі платформи використовують динамічне творче оптимізацію, де один рекламний креатив має сотні заздалегідь підготовлених варіантів (різні відеофрагменти, кольори кнопок, тексти закликів до дії), а система показує саме ту комбінацію, яка дає найвищий показник конверсії для даного конкретного користувача в цю секунду. У електронній комерції ця технологія проявляється у динамічному ціноутворенні, персоналізації каталогу товарів на головній сторінці та навіть у адаптації процесу оформлення замовлення, де кількість полів або спосіб оплати можуть змінюватися залежно від очікуваної ймовірності скасування.

Отже, персоналізація на льоту — це квінтесенція інтернету, що живе і дихає в ритмі своїх користувачів. Вона пропонує неймовірну ефективність і зручність, створюючи середовище, яке буквально «думає» разом з вами. Але водночас вона ставить нас перед глибокими питаннями про природу об'єктивної реальності, свободи вибору та цифрової рівності. Це вже не просто інструмент маркетингу, а соціотехнологічна сила, що формує те, як мільярди людей сприймають інформацію, приймають рішення та взаємодіють зі світом, що перетворює монолог брендів на безкінечну множинність діалогів, кожен з яких унікальний і миттєвий.

Таблиця 2.4.

Порівняльна таблиця підходів класичне A/B-тестування, багатоваріантне тестування (MVT), алгоритм «Багаторукий бандит»

Критерій	Класичне A/B-тестування	Багатоваріантне тестування (MVT)	Алгоритм «Багаторукий бандит»
Суть	Порівняння двох цілісних варіантів.	Одночасне тестування всіх комбінацій кількох елементів.	Динамічне перерозподілення трафіку між варіантами під час тесту.
Перевага	Простота, не вимагає великого трафіку.	Виявляє синергію між елементами.	Мінімізує втрати під час тесту, швидше знаходить переможця.

Критерій	Класичне A/B-тестування	Багатоваріантне тестування (MVT)	Алгоритм «Багаторукий бандит»
Недолік	Фіксовані втрати на гіршому варіанті; тестує лише один елемент.	Потребує колосального трафіку і часу.	Складніша настройка та інтерпретація.
Ідеальний випадок	Оптимізація одного ключового елемента (заголовок, кнопка).	Редизайн головної сторінки або статті з багатьма змінними.	Постійна оптимізація високопріоритетних цілей (конверсія, залученість).

Складено автором. Джерело. Перевірено на людях: A/B тестування" // DOU.ua (2020). https://dou.ua/lenta/articles/what-is-a-b-testing/

Висновок. Інтеграція методів тестування в робочий процес перетворює медіапроєкт на лабораторію, де кожне рішення може бути перевірено. Від простого A/B-тестування до складних адаптивних алгоритмів ШІ — ці інструменти дозволяють не вгадувати, а точно знати, що резонує з аудиторією. Для сучасного медіапрофесіонала володіння цим «науковим методом» стає обов'язковою компетенцією, що забезпечує конкурентоспроможність в умовах боротьби за увагу.

2.3.5. Real-time персоналізація: Інтернет "Тут і Зараз".

Реальна персоналізація є логічним апогеєм розвитку інтернету, трансформуючи його з пасивного сховища інформації в активне, адаптивне середовище, що реагує на кожну нашу дію в режимі «тут і зараз». Це явище виходить далеко за межі простих рекомендацій на основі історії пошуку; це створення унікального, динамічного цифрового дзеркала для кожного користувача, яке змінюється в реальному часі, передбачаючи наміри та підлаштовуючи весь цифровий досвід під миттєвий контекст. В основі цього процесу лежать потужні алгоритми машинного навчання, які безперервно аналізують потік даних: нашу геолокацію, час доби, швидкість переміщення, активність в інших додатках, навіть темп набору тексту або паузи при перегляді відео. Наприклад, пошукова система може інтерпретувати один і той самий запит «яблуко» по-рібному вранці в супермаркеті (пропонуючи інформацію про покупку фруктів) та ввечері біля магазину техніки (пропонуючи новинки про

продукцію Apple). Соціальні мережі, такі як TikTok чи Instagram, довели цей принцип до абсолюту: кожен скрол стрічки — це миттєва переоцінка сотень сигналів. Що ви дивилися секунду тому? Скільки часу зупинилися на попередньому відео? Чи поділилися ним? На основі цього наступний фрагмент контенту підбирається так, щоб максимізувати ймовірність утримання уваги, створюючи ефект безкінечного, ідеально відшліфованого потоку, що відчувається одночасно особистим і живим.

Ефекти реальної персоналізації найбільш помітні в сфері електронної комерції та цифрових сервісів. Сучасні маркетплейси та стрімінгові платформи працюють як віртуозні диригенти, оркеструючи інтерфейс під кожного відвідувача. Ціна готелю, яку ви бачите, може змінюватися в залежності від вашої локації, попиту в цей момент і навіть типу пристрою, з якого ви зайшли. Стрімінговий сервіс, аналізуючи, що ви переглянули до кінця, а що призупинили на півдорозі, динамічно змінює порядок рекомендованих категорій і навіть оформлення обкладинок фільмів. Музичні платформи, такі як Spotify, коректують ваші плейлисти в реальному часі, враховуючи, чи слухаєте ви музику під час пробіжки, роботи чи вечірнього відпочинку, адаптуючи темп та настрій композицій. Це вже не просто сервіси, а інтерактивні середовища, що навчаються і адаптуються зі швидкістю вашої власної поведінки.

Однак ця технологічна досконалість породжує суттєві філософські та соціальні виклики. Найбільш очевидним є феномен «фільтруючої бульбашки» або «ехокамери», яка посилюється в реальному часі. Алгоритми, оптимізовані на утримання уваги, природним шляхом ведуть користувача до контенту, що все більше радикалізує або підтверджує його існуючі погляди, обмежуючи зустріч з альтернативними думками. Інформаційний простір індивіда стає абсолютно унікальним і все менше збігається з досвідом інших, що може підривати формування спільної соціальної реальності. Крім того, динамічна персоналізація цін створює ризики дискримінації, коли різним користувачам пропонуються різні умови за однакові товари чи послуги на основі непрозорих критеріїв. Постійний моніторинг і аналіз поведінки також глибоко впливають на концепцію приватності, перетворюючи кожну онлайн-дію на дані, що живлять систему, яка прагне нас передбачити.

Таким чином, реальна персоналізація формує нову парадигму інтернету — інтернету як миттєвого, контекстно-залежного продовження нашої свідомості та поведінки. Вона пропонує безпрецедентний рівень зручності та релевантності, роблячи цифрові сервіси інтуїтивно зрозумілими та адаптивними. Але водночас вона ставить нас перед загрозою втрати автономії, зменшення інформаційної різноманітності та посилення маніпулятивних практик. Майбутнє цього «живого» інтернету залежатиме від розвитку етичних рамок, алгоритмічної прозорості та здатності самих користувачів усвідомлювати механізми, які формують їхній цифровий досвід кожну мить. Врешті-решт, реальна персоналізація — це не просто технологія, це дзеркало наших бажань і дій, що відображає їх назад до нас з неймовірною швидкістю і точністю, змушуючи задуматися про те, хто насправді керує цим діалогом: ми самі чи алгоритм, що вивчив нас до дрібниць.

1. КОНТРОЛЬНІ ПИТАННЯ

1. Визначення: У чому різниця між *кастомізацією* (користувач сам налаштовує інтерфейс, наприклад, темну тему) та *персоналізацією* (система сама підлаштовується під користувача)?

2. Рівні: Чому сегментація вважається застарілим методом порівняно з гіперперсоналізацією? Які обмеження має підхід "всі жінки люблять мелодрами"?

3. Дані: Що таке Zero-Party Data і чому медіа зараз полюють саме за ними? (Бо це дані, отримані за згодою, що знімає юридичні ризики).

4. Бізнес-ціль: Яка головна економічна мета персоналізації для медіа? (Збільшення LTV — Lifetime Value клієнта, зниження відтоку).

5. Етика: Наведіть приклад "моторошної персоналізації" (Creepy Personalization), яка може відлякати читача від ресурсу.

2. ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. Аудит "Я-клієнт". Об'єкт: Spotify, Netflix або YouTube. Дія. Відкрийте головну сторінку. Зробіть скріншот. Обведіть червоним блоки, які сформовані персонально для вас (наприклад, "Mix of the Week", "Recommended for you"). Обведіть зеленим блоки, які є загальними для всіх (наприклад, "Top 10 in Ukraine"). Висновок:

Який відсоток екрану займає персональний контент?

Завдання 2. Сценарій "Onboarding" (Збір даних). Умова: Ви запускаєте новий додаток про книги. Вам треба зібрати Zero-Party Data, щоб налаштувати стрічку. Завдання: Придумайте 3 питання для екрану вітання (Onboarding), які допоможуть зрозуміти смак користувача, не дратуючи його. *Погано*: "Введіть ваш річний дохід". *Добре*: "Які автори вам подобаються: Кінг чи Роулінг?"

Завдання 3. Дебати "Зручність проти Приватності". Тема: "Чи готові ви дозволити новинному сайту читати вашу геолокацію, щоб він надсилав новини про пробки у вашому районі?" Аргументація: Складіть список "За" (економія часу) і "Проти" (ризик сталкінгу).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. McKinsey & Company. *The value of getting personalization right.* (Звіт про те, як персоналізація збільшує доходи компаній на 40%).

2. Nielsen Norman Group. *Customization vs. Personalization in the User Experience.* (Класична стаття про UX-відмінності).

3. *Zuboff, S. The Age of Surveillance Capitalism.* (Критичний погляд на збір даних).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Personalization (Персоналізація): Використання технологій та даних для надання індивідуального досвіду користувачеві без його прямого запиту.

• Customization (Кастомізація): Зміна інтерфейсу чи контенту, яку ініціює *сам користувач* (наприклад, вибір рубрик у налаштуваннях профілю).

• Segmentation (Сегментація): Поділ аудиторії на групи зі схожими ознаками (вік, стать, географія).

• Hyper-personalization (Гіперперсоналізація): Найвищий рівень персоналізації, де контент адаптується під конкретну особу в реальному часі за допомогою ШІ та Big Data.

• Churn Rate (Показник відтоку): Відсоток клієнтів, які припинили користуватися сервісом. Персоналізація є головним інструментом боротьби з відтоком.

Розділ 3: Основи алгоритмічної журналістики
3.1. Визначення та сфера застосування. Що таке алгоритмічна журналістика

3.1.1. Що таке алгоритмічна журналістика.

Алгоритмічна журналістика — це практика використання комп'ютерних алгоритмів, автоматизованих систем та штучного інтелекту для збору, обробки, аналізу та створення журналістського контенту з мінімальним або без втручання людини на певних етапах виробництва новин. Це не просто використання комп'ютерів у редакції, що стало нормою ще в дев'яностих роках двадцятого століття, а фундаментальна зміна ролі машини у журналістському процесі: від інструменту в руках журналіста до активного учасника, а іноді й автора матеріалу.

Термін з'явився на початку двохтисячних років, але масове поширення отримав після 2010 року, коли технології машинного навчання та обробки природної мови досягли рівня, достатнього для практичного застосування в медіа. Сьогодні алгоритмічна журналістика охоплює широкий спектр практик: від автоматичної генерації спортивних звітів та фінансових оглядів до виявлення тенденцій у великих масивах даних, персоналізації новинних стрічок для кожного читача, автоматичної перевірки фактів та навіть створення повноцінних розслідувань на основі витоків документів. Ключова характеристика алгоритмічної журналістики полягає в тому, що алгоритм приймає редакційні рішення: що є новиною, як її представити, кому показати, які факти включити, який заголовок використати.

Важливо розрізняти кілька рівнів автоматизації в журналістиці. Перший рівень — це автоматизація процесів, коли комп'ютери допомагають журналістам працювати швидше та ефективніше: системи управління контентом, автоматичне транскрибування інтерв'ю, програми для аналізу даних, пошукові системи для баз даних. Журналіст залишається повністю контролює процес, комп'ютер лише прискорює виконання рутинних завдань. Другий рівень — це алгоритмічна асистенція, коли системи пропонують журналісту варіанти, але рішення приймає людина: алгоритм виявляє аномалії в даних та сигналізує журналісту про потенційну історію,

пропонує варіанти заголовків на основі аналізу попередніх успішних матеріалів, рекомендує джерела для цитування. Третій рівень — це напівавтоматична журналістика, коли алгоритм створює чернетку матеріалу, а журналіст редагує та доповнює: автоматично генеровані звіти про квартальні результати компаній, спортивні репортажі з попередньо створеного шаблону, погодні прогнози на основі метеорологічних даних. Четвертий, найвищий рівень — це повністю автоматична журналістика, коли алгоритм самостійно створює та публікує матеріал без втручання людини: автоматичні новини про землетруси одразу після реєстрації сейсмічної активності, фінансові оновлення про зміни котирувань акцій, персоналізовані дайджести новин для кожного користувача.

Алгоритмічна журналістика спирається на кілька ключових технологій. Natural Language Processing, обробка природної мови, дозволяє комп'ютерам розуміти та генерувати текст людською мовою: розпізнавання іменованих сутностей для виявлення людей, організацій, місць у тексті; sentiment analysis для визначення тональності висловлювань; автоматичне резюмування для створення коротких версій довгих текстів; машинний переклад для швидкої локалізації контенту. Machine Learning, машинне навчання, дозволяє системам навчатися на прикладах без явного програмування: класифікація новин за темами та важливістю, передбачення вірусного потенціалу матеріалу, виявлення патернів у даних для розслідувань, персоналізація контенту для кожного користувача. Data Mining, інтелектуальний аналіз даних, допомагає виявляти приховані закономірності та тенденції: аналіз великих масивів урядових документів для виявлення корупційних схем, моніторинг соціальних мереж для виявлення трендів та суспільних настроїв, обробка фінансових даних для виявлення інсайдерської торгівлі. Natural Language Generation, генерація природної мови, перетворює структуровані дані на читабельний текст: перетворення таблиць з результатами матчів на спортивні репортажі, перетворення фінансових звітів на новинні статті, створення погодних прогнозів з метеорологічних даних.

Економічна логіка алгоритмічної журналістики зрозуміла: медіакомпанії працюють у світі скорочуваних бюджетів та зростаючої потреби в контенті. Автоматизація дозволяє виробляти більше

контенту з меншими витратами. Один алгоритм може створювати тисячі локалізованих версій однієї новини для різних міст, що було б економічно неможливо з людськими журналістами.

3.1.2. Відмінності від традиційної журналістики.

Традиційна журналістика побудована на людській інтуїції, досвіді та судженні. Журналіст вирішує, що є новиною, спираючись на своє розуміння суспільної важливості, актуальності, впливу на аудиторію. Це суб'єктивний, творчий процес, який важко формалізувати. Журналіст шукає джерела, проводить інтерв'ю, перевіряє факти, вибудовує наратив, приймає етичні рішення про те, що публікувати, а що ні. Весь процес ґрунтується на людському розумінні контексту, нюансів, культурних кодів, етичних норм. Алгоритмічна журналістика працює інакше: вона оперує формальними правилами, статистичними закономірностями, математичними моделями. Алгоритм не "розуміє" новину в людському сенсі, він розпізнає патерни, які асоціюються з новинністю на основі навчальних даних.

Ключова відмінність полягає в природі рішень. Журналіст приймає рішення на основі професійного судження, яке формувалося роками через освіту, досвід, етичні принципи, редакційну політику. Це рішення часто інтуїтивне, важко пояснюване, але ґрунтується на глибокому розумінні контексту. Алгоритм приймає рішення на основі математичної оптимізації певної функції: максимізація кліків, часу читання, поширень у соцмережах, відповідності попереднім перевагам користувача. Рішення алгоритму детерміноване його програмуванням та навчальними даними. Якщо алгоритм навчений максимізувати залученість, він робитиме саме це, навіть якщо це суперечить журналістській етиці.

Швидкість та масштаб роботи радикально відрізняються. Досвідчений журналіст може написати дві-три якісні статті на день. Алгоритм може генерувати тисячі статей за годину. Журналіст може моніторити п'ять-десять джерел одночасно. Алгоритм може аналізувати мільйони документів, відстежувати всі соціальні мережі, всі урядові бази даних, всі корпоративні звіти одночасно. Ця різниця в масштабі означає, що алгоритми можуть виявляти історії, які людина фізично не здатна помітити: кореляції між тисячами змінних у великих даних, аномалії в роботі сотень урядових чиновників, патерни в мільйонах фінансових транзакцій.

Персоналізація — ще одна фундаментальна відмінність. Традиційна журналістика створює один продукт для всіх: одна версія газети, один випуск новин, одна стаття. Редактор вирішує, що важливо для аудиторії в цілому. Алгоритмічна журналістика може створювати унікальну версію новин для кожного користувача: різні статті, різний порядок, різні заголовки, різний рівень деталізації залежно від інтересів, локації, історії читання конкретної людини. Це переворот у концепції "єдиної порції новин для всіх", яка визначала журналістику з моменту винаходу друкарського верстата.

Об'єктивність та упередженість теж виглядають по-різному. Традиційна журналістика визнає, що повна об'єктивність неможлива, але прагне до неї через редакційні стандарти, перевірку фактів, баланс точок зору, відокремлення фактів від думок. Упередженість журналіста може бути ідентифікована та скоригована через редакторський процес. Алгоритмічна журналістика часто сприймається як об'єктивна, бо "комп'ютер не має упереджень". Це небезпечна ілюзія. Алгоритми несуть упередженість своїх створювачів, навчальних даних, цільових функцій. Але ця упередженість прихована в коді, математиці, даних — вона невидима для звичайного користувача і навіть для багатьох журналістів. Алгоритм, навчений на історичних новинних статтях, відтворюватиме всі упередження, присутні в цих статтях: расові, гендерні, класові, географічні.

Прозорість процесу кардинально відрізняється. У традиційній журналістиці читач може запитати журналіста, чому він написав статтю саме так, які джерела використав, чому вибрав цей ракурс. Журналіст може пояснити своє редакційне рішення. В алгоритмічній журналістиці часто неможливо пояснити, чому алгоритм прийняв конкретне рішення. Сучасні системи машинного навчання, особливо глибокі нейронні мережі, є "чорними скриньками": вони приймають рішення на основі мільйонів параметрів, і навіть їхні створювачі не можуть точно пояснити, чому для конкретного випадку система видала саме такий результат. Це створює проблему підзвітності: кого звинувачувати, якщо алгоритм опублікував неправдиву або шкідливу інформацію?

Творчість та інновації теж відрізняються. Журналіст може вийти за рамки звичного, знайти несподіваний ракурс, створити унікальний наратив, вигадати нову форму подачі матеріалу. Творчість — це

відхилення від норми, несподіваність, порушення правил. Алгоритм, навпаки, оптимізований відтворювати успішні патерни з минулого. Він аналізує, що спрацювало раніше, і робить більше того самого. Це створює ризик гомогенізації: всі алгоритмічні новини починають виглядати схоже, бо всі оптимізовані під ті самі метрики залученості. Справжні інновації, які спочатку можуть здаватися незвичними та ризикованими, відфільтровуються алгоритмами як неоптимальні.

Етична відповідальність розподіляється по-різному. Журналіст несе особисту відповідальність за свою роботу: його ім'я під статтею, його репутація на кону, він може бути притягнутий до відповідальності за наклеп, неправдиву інформацію, порушення приватності. Алгоритмічна журналістика розмиває відповідальність: хто винен, якщо автоматично згенерована стаття містить помилку — програміст, який написав код; редактор, який схвалив використання системи; компанія, яка володіє платформою; чи сам алгоритм? Це питання залишається юридично та етично нерозв'язаним у більшості юрисдикцій.

Взаємодія з аудиторією теж трансформується. Традиційний журналіст може спілкуватися з читачами, отримувати зворотний зв'язок, пояснювати свою роботу, коригувати помилки, вести діалог. Алгоритм не може відповісти на запитання, не може врахувати нюанси людської критики, не може пояснити свою логіку звичайною мовою. Це створює нову дистанцію між медіа та аудиторією, коли люди споживають контент, створений системами, які вони не розуміють і з якими не можуть вступити в діалог.

Економічна модель також змінюється. Традиційна журналістика вимагає великих інвестицій у людський капітал: навчання журналістів, зарплати, соціальні гарантії, робочі місця. Алгоритмічна журналістика вимагає великих початкових інвестицій у технології: розробка систем, навчання моделей, інфраструктура, але після створення має низькі граничні витрати на виробництво додаткового контенту. Один алгоритм може замінити десятки журналістів на рутинних завданнях, що створює тиск на ринок праці в медіа. Це не означає, що всі журналісти втратять роботу, але означає зміну структури професії: менше потрібно людей для виконання стандартизованих завдань, більше для складної аналітики, розслідувань, креативу, етичного нагляду за алгоритмами.

3.1.3. Спектр застосування: від аналітики до генерації контенту.

Алгоритмічна журналістика не є однорідним явищем, вона охоплює широкий спектр застосувань, які можна розмістити на континуумі від мінімального втручання машини до повної автоматизації журналістського процесу. На одному кінці спектру знаходиться використання алгоритмів для посилення можливостей журналістів, на іншому — повністю автоматичне створення та публікація контенту.

Автоматизований моніторинг та виявлення новин — це найпростіша форма алгоритмічної журналістики, але одна з найпотужніших. Системи безперервно сканують величезні обсяги даних, виявляючи аномалії, тренди, події, які можуть бути новинами. Earthquakebot від Los Angeles Times автоматично публікує повідомлення про землетруси протягом трьох хвилин після реєстрації сейсмічної активності, аналізуючи дані Геологічної служби США. Система не просто копіює дані, вона приймає редакційне рішення: чи є землетрус достатньо значущим для публікації, враховуючи магнітуду, глибину, відстань від населених пунктів. Quakebot працює з 2014 року і опублікував сотні попередніх новин про землетруси, часто випереджаючи традиційні медіа на годину і більше. Подібні системи існують для моніторингу фінансових ринків: алгоритми відстежують всі котирування акцій, валют, товарів і автоматично генерують новини про значні зміни. Bloomberg та Reuters використовують такі системи роками, публікуючи тисячі автоматичних фінансових оновлень щодня.

Моніторинг соціальних мереж — ще одна важлива сфера. Алгоритми аналізують мільйони постів у Twitter, Facebook, Instagram, Reddit, виявляючи трендові теми, вірусний контент, суспільні настрої, потенційні новини. Під час надзвичайних подій — терактів, стихійних лих, протестів — соцмережі часто є першим джерелом інформації. Алгоритми допомагають журналістам швидко ідентифікувати очевидців, знаходити фото та відео з місця подій, оцінювати масштаб події. Системи як Dataminr спеціалізуються на виявленні breaking news у соцмережах і продають доступ до своїх алертів медіакомпаніям. Однак тут виникає проблема верифікації: не все, що трендить у соцмережах, є правдою. Алгоритми можуть виявити потенційну історію, але журналіст повинен перевірити

факти перед публікацією.

Аналіз великих даних для розслідувань — це використання алгоритмів для виявлення патернів у величезних масивах документів, які людина не здатна проаналізувати вручну. Panama Papers та Paradise Papers, найбільші журналістські розслідування останнього десятиліття про офшорні фінанси, були б неможливі без алгоритмічного аналізу. Міжнародний консорціум розслідувальних журналістів отримав одинадцять з половиною мільйонів документів для Panama Papers. Жоден журналіст не може прочитати таку кількість. Алгоритми проіндексували всі документи, виявили зв'язки між людьми, компаніями, банківськими рахунками, створили візуалізації мереж офшорних схем. Журналісти з сімдесяти країн потім використовували ці аналітичні інструменти для пошуку історій, релевантних для їхніх країн. Результат: сотні розслідувань, відставки політиків, кримінальні справи проти корупціонерів.

Подібні техніки використовуються для аналізу урядових даних. ProPublica використовує алгоритми для моніторингу федеральних витрат, виявляючи незвичайні контракти, потенційну корупцію, неефективність. Journalists використовують скрейпінг даних для збору інформації з тисяч урядових вебсайтів, баз даних судових рішень, реєстрів власності, створюючи всеосяжні датасети для розслідувань. Алгоритми виявляють аномалії: чому це невелике місто витратило вдвічі більше на дорожній ремонт, ніж сусідні? Чому цей суддя виносить більш м'які вироки певній демографічній групі? Чому цей поліцейський департамент має втричі вище співвідношення скарг на надмірне застосування сили?

Автоматична генерація простого контенту — наступний рівень алгоритмічної журналістики, де машина не просто допомагає, а самостійно створює публікації. Це працює найкраще для стандартизованого, шаблонного контенту, де структура передбачувана, а дані структуровані. Спортивні репортажі — ідеальний кейс. Associated Press використовує систему Wordsmith від Automated Insights для генерації звітів про бейсбольні матчі низьких ліг. Система отримує structured data про гру: рахунок, статистика гравців, ключові моменти, і генерує читабельну статтю. Раніше AP взагалі не висвітлював матчі низьких ліг через обмежені ресурси. Тепер автоматизація дозволяє створювати тисячі таких репортажів, задовольняючи попит ло-

кальної аудиторії.

Фінансові звіти — ще одна сфера масової автоматизації. Компанії публікують квартальні результати в стандартизованому форматі. Алгоритми автоматично перетворюють ці цифри на новинні статті: "Apple повідомила про квартальний дохід у 90 мільярдів доларів, перевищивши очікування аналітиків на 5%. Продажі iPhone зросли на 10%, тоді як продажі iPad впали на 8%." Такі статті публікуються протягом секунд після оприлюднення фінансових даних, набагато швидше, ніж будь-який журналіст міг би написати. Bloomberg та Reuters генерують тисячі таких звітів щодня. Якість тексту не літературна, але фактична точність висока, а швидкість критична для фінансових ринків.

Погодні прогнози, результати виборів на рівні окремих дільниць, звіти про злочинність у районах, оновлення про дорожній трафік — все це приклади контенту, який масово автоматизується. Характерна риса: високий обсяг, низька складність, стандартизована структура, структуровані вхідні дані. Для медіакомпаній економіка очевидна: замість найму ста журналістів для написання ста локальних звітів про погоду, можна використати один алгоритм.

Персоналізація контенту та рекомендації — можливо, найпоширеніша форма алгоритмічної журналістики, хоча часто не розпізнається як така. Кожна новинна платформа від Google News до Facebook, від Apple News до Flipboard використовує алгоритми для визначення, які новини показувати кожному користувачу. Це редакційне рішення: що є важливим для цієї конкретної людини. Раніше це рішення приймав головний редактор для всієї аудиторії. Тепер його приймає алгоритм індивідуально для кожного з мільярдів користувачів. Алгоритм аналізує вашу історію читання, кліки, час на сторінці, соціальні зв'язки, демографію, локацію і на основі цього вирішує, що "Apple представила новий iPhone" для вас важливіше, ніж "Конфлікт на Близькому Сході", або навпаки.

Це має глибокі наслідки для формування суспільної думки. В епоху масмедіа всі отримували приблизно однакову порцію новин, що створювало спільну інформаційну базу для суспільного діалогу. В епоху алгоритмічної персоналізації кожен живе у своїй інформаційній реальності. Ви можете не знати про важливі події, бо алгоритм вирішив, що вони вас не зацікавлять. Це посилює filter

bubbles та echo chambers, про які багато говорили дослідники медіа. Алгоритми оптимізовані під engagement, а найбільшу залученість створює контент, який підтверджує наші існуючі переконання, а не викликає їх. Тому алгоритми мають тенденцію показувати нам більше того, з чим ми вже згодні, поглиблюючи поляризацію.

Автоматична перевірка фактів — порівняно нова, але швидко зростаюча сфера. Full Fact у Великобританії, ClaimBuster у США, інші організації розробляють алгоритми, які автоматично виявляють фактичні твердження в текстах політиків, перевіряють їх відповідність базам даних фактів і позначають потенційно хибні. Система не може повністю замінити людський факт-чекінг, особливо для складних нюансованих тверджень, але може значно прискорити процес, фільтруючи очевидні неправди та виявляючи твердження, які потребують ручної перевірки. Під час виборчих дебатів автоматичні системи можуть перевіряти сотні тверджень кандидатів у реальному часі, чого жодна людська команда не здатна зробити.

Автоматичний переклад та локалізація новин дозволяє медіакомпаніям швидко адаптувати контент для різних мов та регіонів. Reuters та AFP використовують машинний переклад для створення версій своїх новин десятками мов. Це не ідеально, люди-перекладачі все ще потрібні для редагування, але базовий переклад відбувається миттєво, дозволяючи публікувати breaking news багатьма мовами одночасно. Локалізація йде далі: алгоритм не просто перекладає, а адаптує контент, змінюючи приклади, посилання, контекст для конкретної аудиторії.

Генерація візуалізацій даних — алгоритми автоматично створюють графіки, карти, інфографіку з структурованих даних. Система аналізує датасет і вирішує, який тип візуалізації найкраще підходить: лінійний графік для трендів у часі, барчарт для порівнянь, карта для географічних даних. Потім автоматично генерує візуалізацію, яка вбудовується в статтю. The Guardian експериментує з автоматичною генерацією інтерактивних візуалізацій для статей на основі даних.

Автоматичне резюм Автоматичне резюмування — створення коротких версій довгих текстів. Алгоритми виявляють ключові речення, основні ідеї, важливі факти і створюють condensed версію. Це корисно для дайджестів новин, мобільних версій статей, voice

assistants, які зачитують новини. Yahoo News та інші платформи використовують автоматичне резюмування для створення "тизерів" статей у новинних стрічках.

Генеративний AI для повноцінних статей — найновіший та найконтроверсійніший рубіж. Системи як GPT-4 здатні генерувати оригінальні тексти, які часто неможливо відрізнити від написаних людиною. CNET та інші видання експериментували з використанням GPT для написання фінансових та технологічних статей. Результати mixed: деякі статті виглядали переконливо, інші містили фактичні помилки, галюцинації AI, плагіат. Це породжує питання: чи можемо ми довіряти журналістиці, створеній системами, які не "розуміють" того, про що пишуть, а лише імітують патерни людського письма? Чи повинні видання розкривати, коли контент створений AI? Хто несе відповідальність за помилки в AI-генерованих статтях?

3.1.4. Історія розвитку: від автоматизованих звітів до ШІ-журналістики.

Коріння алгоритмічної журналістики сягають глибше, ніж здається. Перші спроби автоматизації в медіа почалися ще в середині двадцятого століття, задовго до ери інтернету та штучного інтелекту. Розуміння цієї історії допомагає побачити, що сучасна алгоритмічна журналістика не раптовий розрив, а еволюція, що тривала десятиліттями.

П'ятдесяті-шістдесяті роки двадцятого століття: зародження комп'ютерної журналістики. Перші комп'ютери, величезні мейнфрейми, почали використовуватися для обробки виборчих даних телемережами. CBS у 1952 році використала комп'ютер UNIVAC для передбачення результатів президентських виборів на основі часткових даних. Це не була журналістика в сучасному розумінні, але демонструвала потенціал машин для швидкого аналізу даних. У шістдесятих деякі газети почали використовувати комп'ютери для набору тексту, замінюючи лінотипи. Це була автоматизація виробництва, а не контенту, але вона змінила робочі процеси редакцій.

Сімдесяті роки: Computer-Assisted Reporting. Філіп Мейєр, журналіст Detroit Free Press, використав комп'ютерний аналіз для розслідування расових заворушень 1967 року в Детройті. Він опитав жителів, завів дані в комп'ютер і провів статистичний аналіз, виявивши патерни, які суперечили поширеним припущенням. Це народження

Computer-Assisted Reporting, CAR, де комп'ютери допомагали журналістам аналізувати дані, але не створювали контент. Мейєр написав книгу "Precision Journalism" 1973, яка стала маніфестом використання соціально-наукових методів та обчислювальних інструментів у журналістиці. Проте доступ до комп'ютерів був обмежений, дані важко отримувати, техніки складні. CAR залишався нішевою практикою.

Вісімдесяті роки: розширення баз даних. Поява персональних комп'ютерів та електронних баз даних зробила комп'ютерний аналіз доступнішим. Журналісти почали використовувати spreadsheets як Lotus 1-2-3 та dBase для аналізу урядових даних. Деякі видання створили CAR-відділи. Проте це все ще була асистована журналістика, де комп'ютер як інструмент у руках журналіста.

Дев'яності роки: інтернет та перші автоматизовані системи. Поява вебу змінила все. Медіа почали публікувати онлайн, що вимагало постійного оновлення контенту. Перші автоматизовані системи з'явилися для фінансових новин. Bloomberg Terminal, хоча і не чиста журналістика, демонстрував можливості автоматичної агрегації та представлення фінансової інформації. Деякі фінансові сервіси почали експериментувати з автоматичною генерацією коротких звітів про котирування акцій, використовуючи прості template-based системи: "Акції [COMPANY] закрилися на рівні [PRICE], [UP/DOWN] [PERCENTAGE]% від попереднього дня."

Наприкінці дев'яностих Associated Press та Reuters почали інвестувати в системи для швидшого виробництва рутинного контенту. Алгоритми були примітивними за сучасними стандартами, в основному rule-based системи з заздалегідь визначеними шаблонами. Проте це були перші кроки до автоматичної генерації новин.

Двохтисячні: Data Journalism та раннє NLG. Термін "data journalism" набув популярності. Видання як The Guardian створили спеціалізовані data teams. Simon Rogers запустив Datablog у Guardian 2009, який став зразком для data-driven journalism. Журналісти почали створювати складні інтерактивні візуалізації, використовувати геопросторовий аналіз, працювати з великими датасетами. Це все ще була асистована журналістика, але складність аналізу та обсяги даних зростали.

Паралельно розвивалися технології Natural Language Generation.

Компанія Narrative Science, заснована 2010 року, створила систему Quill для автоматичної генерації спортивних звітів та бізнес-аналітики. Automated Insights запустила Wordsmith 2010. Ці системи використовували більш просунуті NLG-техніки, здатні створювати варіативний текст, а не просто заповнювати шаблони. Проте справжній прорив стався, коли великі медіакомпанії почали їх використовувати.

2014 рік: Associated Press автоматизує заробіткові звіти. AP оголосила, що використовуватиме Wordsmith для автоматичної генерації звітів про квартальні заробітки компаній. Раніше AP вручну писала близько трьохсот таких звітів щокварталу. Автоматизація дозволила збільшити це до чотирьох тисяч трьохсот, охопивши малі та середні компанії, які раніше ігнорувалися через обмежені ресурси. AP заявила, що це звільнить журналістів від рутини для складніших завдань. Це був переломний момент: якщо AP, золотий стандарт журналістики, автоматизує контент, інші послідують.

І вони послідували. Forbes почав використовувати "Bertie", AI-асистента, який пропонує журналістам шаблони для статей, рекомендує теми на основі трендів, навіть генерує чернетки заголовків та перші параграфи. Washington Post створив "Heliograf" 2016 для автоматичної генерації коротких новин про вибори та спортивні події. Під час виборів 2016 Heliograf створив понад п'ятсот коротких звітів про результати в різних округах. BBC експериментувала з "Juicer", системою для автоматичного резюмування та персоналізації новин.

2016-2018: Deep Learning революція. Прорив у глибокому навчанні, особливо в NLP, радикально розширив можливості. Transformer-архітектури, як BERT від Google 2018, показали безпрецедентне розуміння природної мови. Це дозволило створювати більш складні системи для аналізу тексту, автоматичного резюмування, виявлення fake news, sentiment analysis. Reuters створив News Tracer, систему на базі machine learning для виявлення breaking news у Twitter та автоматичної оцінки їх достовірності. Це критично під час надзвичайних подій, коли соцмережі заповнюються як справжньою інформацією, так і чутками.

Персоналізація новин стала повсюдною. Google News, Apple News, Facebook News Feed — всі використовували все більш складні ML-моделі для індивідуалізації контенту. Алгоритми почали враховувати не лише явні сигнали як клики, а й неявні як час читання,

контекст, соціальний граф, навіть емоційні реакції через emoji. Це створило безпрецедентний рівень персоналізації, але також підняло занепокоєння щодо filter bubbles.

2019-2020: GPT-2 та етичні дебати. OpenAI випустила GPT- 2, генеративну мовну модель, здатну створювати переконливі тексти на будь-яку тему. Спочатку OpenAI відмовилася публікувати повну модель через побоювання зловживань: генерація fake news, маніпуляції, спам. Це спровокувало дебати про етику автоматичної генерації контенту. Якщо AI може створювати нерозрізнені від людських тексти, як ми відрізнимо справжню журналістику від синтетичної пропаганди? Як захиститися від масового виробництва дезінформації?

Деякі медіа почали експериментувати з GPT-2 для генерації чернеток статей, але результати були нестабільними. Модель часто "галюцинувала" факти, створювала правдоподібно звучащі, але неправдиві твердження. Це підкреслило фундаментальну проблему: generative AI не "знає" фактів, воно лише передбачає правдоподібний текст на основі патернів у навчальних даних.

2020-2022: Пандемія та автоматизація локальних новин. COVID-19 прискорив автоматизацію. Потреба в оновленнях про кількість випадків, смертей, вакцинацій по тисячах локацій створила попит на автоматизоване виробництво контенту. Багато видань використовували алгоритми для генерації локалізованих звітів про пандемію на основі урядових даних. Одночасно пандемія погіршила економічну кризу в медіа, призвела до масових скорочень журналістів. Автоматизація стала не лише можливістю, а необхідністю для виживання багатьох видань.

Local newsrooms, особливо постраждалі від зникнення рекламних доходів, почали експериментувати з автоматизацією для продовження висвітлення своїх спільнот з меншою кількістю журналістів. Деякі стартапи як Arria NLG пропонували автоматичну генерацію локальних новин про погоду, спорт, бізнес для невеликих газет та радіостанцій.

2022-2024: Era of GPT-3, GPT-4 та генеративний AI mainstream. GPT-3 та особливо GPT-4 вивели генеративний AI на новий рівень. Модель могла писати складні статті, аналізувати дані, створювати різноманітний контент з мінімальними промптами. CNET, Insider та

інші видання почали експериментувати з використанням GPT для написання певних типів статей. Результати були контроверсійними: деякі AI-написані статті містили помилки, іноді плагіат з джерел у навчальних даних. CNET змушена була зробити корекції та більш чітко позначати AI-generated контент.

BuzzFeed оголосив про використання OpenAI для створення персоналізованого контенту та квізів, що призвело до зростання акцій компанії, але також до критики від журналістських спільнот. Дебати зосередилися на прозорості: чи повинні видання чітко позначати AI-generated контент? Чи можна довіряти фактам від AI? Хто несе відповідальність за помилки?

Регулятори почали реагувати. Європейський Союз розробляв AI Act, який вимагав прозорості для автоматизованих систем, особливо тих, що впливають на публічну інформацію. Деякі країни розглядали закони про обов'язкове позначення AI-generated контенту. Журналістські організації як Reuters Institute створювали гайдлайни для етичного використання AI в журналістиці.

2024 і далі: майбутнє гібридної журналістики. Сьогодні ми знаходимося на перетині, де алгоритмічна журналістика стає нормою, а не винятком. Практично кожна велика медіакомпанія використовує ту чи іншу форму автоматизації. Проте стає очевидним, що майбутнє не в повній заміні журналістів машинами, а в гібридних моделях: алгоритми роблять те, що роблять найкраще — швидкий аналіз великих даних, генерація рутинного контенту, персоналізація, моніторинг, а люди роблять те, що роблять найкраще — розслідування, інтерв'ю, етичні судження, креативність, критичне мислення, розуміння контексту.

Провідні редакції створюють нові ролі: AI editors, які курують алгоритмічні системи; data scientists у newsrooms; algorithmic transparency officers, які пояснюють аудиторії, як працюють їхні алгоритми. Журналістська освіта адаптується, включаючи навчання data science, machine learning, computational thinking поряд з традиційними навичками репортажу та письма.

Ключові тренди, що визначатимуть наступну фазу: більша прозорість алгоритмів, медіа під тиском будуть пояснювати, як їхні системи приймають рішення; розвиток explainable AI, систем які можуть обґрунтувати свої рішення; стандарти маркування AI-

generated контенту; інвестиції в перевірку фактів та модерацію AI-generated контенту; етичні frameworks для використання AI в журналістиці, розроблені спільнотами журналістів; технології для виявлення synthetic media, deepfakes, AI-generated дезінформації; зростання локальної автоматизованої журналістики для заповнення news deserts; експерименти з AI як співавтором, де журналіст та AI співпрацюють над матеріалом.

Історія алгоритмічної журналістики показує послідовну еволюцію від простої автоматизації рутини до систем, здатних приймати складні редакційні рішення. Кожен етап розширював можливості, але також створював нові етичні дилеми та виклики для професії. Розуміння цієї історії критично для навігації майбутнім, де межа між людською та машинною журналістикою стає все більш розмитою, а питання довіри, прозорості та підзвітності в медіа набувають нового значення.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допомагають сформувати чіткий понятійний апарат:

1. Дефініція: Що таке алгоритмічна журналістика (Automated Journalism)? Чому визначення *"процес автоматичної конвертації даних у розповідні тексти за допомогою НЛП"* є більш точним, ніж просто "роботи пишуть новини"?

2. Дані як паливо: Чому алгоритмічна журналістика можлива лише там, де є структуровані дані (спорт, фінанси, погода, нерухомість)? Чому роботу важко написати репортаж з мітингу або інтерв'ю?

3.Еволюція підходів: У чому різниця між простим методом підстановки у шаблони (Template-based), який нагадує гру Mad Libs, та сучасними методами машинного навчання, що генерують варіативний текст?

4. Сфера застосування: Які основні кейси використання у великих агенціях?

- *AP (Associated Press)*: Фінансові звіти компаній (збільшення обсягу в 12 разів).

- *Washington Post (Heliograf):* Висвітлення локальних виборів та Олімпіад.

- *LA Times (Quakebot):* Миттєві повідомлення про землетруси.

5. Гібридна модель (Cyborg Journalism): Як працює модель, де алгоритм робить чернетку з цифрами, а журналіст додає контекст та "людський голос" (наприклад, у Bloomberg)?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання спрямовані на розуміння логіки створення бота-журналіста:

Завдання 1. "Реверс-інжиніринг" новини

• Матеріал: Знайдіть новину про результати футбольного матчу або квартальний звіт компанії (бажано коротку замітку на фінансовому порталі).

• Завдання: Спробуйте відтворити алгоритм, за яким вона написана. Виділіть змінні.

- *Приклад:* "Команда [Змінна А] перемогла команду [Змінна Б] з рахунком [Змінна В]. Голи забили: [Список гравців]."

• Питання: Які умови (IF/THEN) використав бот? (Наприклад: *ЯКЩО різниця голів > 3, ТО писати "розгромила", ЯКЩО = 1, ТО писати "вирвала перемогу"*).

Завдання 2. SWOT-аналіз впровадження

• Ситуація: Ви головний редактор локального сайту новин. Ви розглядаєте можливість купити софт для автоматичного написання прогнозів погоди та зведень ДТП.

• Завдання: Складіть SWOT-аналіз:

- *Strengths:* (Швидкість, 24/7, звільнення людей від рутини).

- *Weaknesses:* (Шаблонність, відсутність іронії/стилю).

- *Opportunities:* (Гіперлокальність — прогноз для кожного села).

- *Threats:* (Втрата довіри читачів, якщо бот помилиться в цифрах; залежність від постачальника даних).

Завдання 3. Порівняння: NLG vs LLM

• Завдання: Попросіть ChatGPT написати новину про вигаданий футбольний матч. Потім порівняйте цей підхід з класичним NLG (Natural Language Generation).

• Висновок: Чому для фінансової звітності класичний алгоритм (правила) безпечніший за ChatGPT (ймовірності)? (Підказка: проблема галюцинацій).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. *Carlson, Matt. The Robotic Reporter.* (Фундаментальна праця про зміну ролі журналіста).

2. Diakopoulos, Nicholas. *Automating the News: How Algorithms Are Rewriting the Media.* (Обов'язкова книга для курсу; детальний розбір технологій та етики).

3. Guide by Associated Press. *The future of augmented journalism: A guide for newsrooms in the age of smart machines.* (Практичний посібник для індустрії).

4. Dörr, Konstantin. *Mapping the field of Algorithmic Journalism.* (Наукова стаття з класифікацією).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Алгоритмічна журналістика (Algorithmic / Automated Journalism): Автоматичне створення журналістського контенту за допомогою програмного забезпечення, що обробляє структуровані дані без прямого втручання людини в процес написання.

• NLG (Natural Language Generation): Розділ штучного інтелекту, що займається перетворенням комп'ютерних даних у читабельний текст природною мовою.

• Структуровані дані (Structured Data): Інформація, організована у чітко визначеному форматі (таблиці Excel, бази даних, JSON, XML), де кожне значення має свою категорію (наприклад, "Температура", "Рахунок", "Ціна акції"). Це "сировина" для робота.

• Шаблонний метод (Template-based approach): Найпростіший метод NLG, де програма підставляє дані у заздалегідь написані речення з пропусками.

• Гіперлокальність (Hyperlocality): Здатність алгоритмічної журналістики створювати тисячі унікальних новин для дуже малих спільнот (наприклад, новини про ціни на нерухомість для конкретної вулиці), що неможливо зробити силами людей-репортерів.

3.2. Типи алгоритмічної журналістики

3.2.1. Дата-журналістика та обчислювальна журналістика: філософія, інструменти та практика.

Поняття «дата-журналістика» та «обчислювальна журналістика» часто використовуються як взаємозамінні в повсякденному вжитку, однак між ними існує фундаментальна різниця у філософії, масштабі та інструментарії. Найпростіше їх розрізнити так: якщо дата-журналістика — це вміння оповідати історії за допомогою даних, то обчислювальна журналістика — це мистецтво створювати інструменти та процеси для виявлення цих історій. Разом вони формують основу сучасної алгоритмічної журналістики, яка застосовує потужність обчислень на всіх етапах новинного виробництва, дотримуючись при цьому базових цінностей журналістики, таких як точність і перевірка.

Дата-журналістика: історія з цифр. Філософія та підхід. Дата-журналістика ґрунтується на принципі evidence-based reporting, де цифри та масиви інформації стають першоджерелом. Мета — не просто проілюструвати статтю графіком, а відкрити, проаналізувати та наочно показати закономірності, які неможливо побачити неозброєним оком. Як визначають фахівці, це «переклад мови цифр на загальновживані мови». Вона дозволяє перейти від анекдотичних свідчень до системного аналізу, наприклад, показуючи реальну динаміку зростання статків депутатів, а не окремі випадки.

Технологічний процес та інструменти. Робочий процес дата-журналіста — це конвеєр перетворення «сирих» даних у зрозумілу історію. Він включає декілька етапів:

1. Пошук даних: Робота починається з гіпотези або теми, для якої потрібні підтвердження. Дані можуть надходити з відкритих реєстрів, публічних запитів до державних органів, аналітичних центрів або витоку масивів інформації. В українському контексті великим викликом є брак якісних агрегованих даних — часто інформацію доводиться збирати та структурувати вручну або напівавтоматично.

2. Очищення та аналіз: 80-90% часу дата-журналіста займає саме «чистка» — приведення даних до придатного для аналізу вигляду (форматування дат, усунення дублікатів, заповнення пропусків). Для цього використовують інструменти від Excel і Google Sheets

до складніших мов програмування Python або R, які дозволяють обробляти великі масиви інформації.

3. Візуалізація та розповідь: Виявлені інсайти треба подати аудиторії. Тут на допомогу приходять спеціалізовані інструменти для створення інфографіки, карт та інтерактивних графіків: Datawrapper, Flourish, Tableau та інші. Кінцевим продуктом може бути як інтерактивна карта забруднень, так і статична інфографіка про зростання цін.

Дата-журналістика — це зібрання навичок журналіста, статистика та дизайнера в одній особі. Вона доступна навіть тим, хто не є програмістом: почати можна з Excel та безкоштовних сервісів візуалізації, а потім нарощувати складність.

Обчислювальна журналістика: автоматизація та алгоритмічний нагляд. Філософія та підхід. Обчислювальна журналістика — ширше поняття, яке охоплює застосування комп'ютерних наук до всіх аспектів журналістики. Її суть не лише в аналізі наявних таблиць, а в створенні інструментів для їх отримання та обробки. Це перехід від ролі користувача софту до ролі розробника рішень. Фокус зсувається на автоматизацію рутинних завдань та розслідування систем, включаючи алгоритми, що впливають на життя людей (algorithmic accountability).

Ключові методи та продукти.

• Скрапінг (Scraping): Написання спеціальних програм (скрейперів) для автоматизованого збору даних з веб-сайтів, коли пряме завантаження неможливе. Наприклад, щоденний моніторинг цін на продукти в онлайн-магазинах для аналізу інфляції. Це дозволяє створити власні бази даних там, де їх офіційно немає.

• Алгоритмічне розслідування (Algorithmic Accountability): Вивчення та розслідування впливу алгоритмів — від рекомендаційних систем соціальних мереж до систем прийняття рішень в урядах та корпораціях. Прикладом є дослідження, яке показало потенційну дискримінацію в алгоритмах ціноутворення Uber.

• Аналіз неструктурованих даних: Використання методів штучного інтелекту (NLP — обробка природної мови) для аналізу текстів, аудіо, відео. Це дозволяє виявляти закономірності в великих обсягах документів, як це робить команда texty.org.ua, аналізуючи бази російської пропаганди.

3.2.2. Автоматизована генерація новин (Automated Journalism). Застосування мовних моделей для створення коротких, стандартизованих новинних повідомлень на основі даних, наприклад, спортивних результатів або фінансових звітів. Ще 2017 року The Washington Post за допомогою системи Heliograf публікував сотні таких новин.

Кінцевим продуктом обчислювальної журналістики часто є не готова стаття, а інструмент або платформа: база даних, чат-бот для комунікації з читачами або система моніторингу, що автоматично сповіщає про важливі події.

Еволюція: від CAR до AI Journalism. Розвиток цієї галузі можна простежити через три хвилі:

1. CAR (Computer-Assisted Reporting), 1970-1990-ті: Початок ери, коли журналісти почали використовувати мейнфрейми та перші бази даних (наприклад, для аналізу переписів населення). Засновником напрямку вважають журналіста Філіпа Меєра.

2. Дата-журналістика, 2000-2010-ті: Зростання потужності персональних комп'ютерів, поява доступу до відкритих даних та розвиток візуалізації. Акцент змістився на складні інтерактивні історії для інтернет-аудиторії.

3. AI Journalism (журналістика зі ШІ), 2020-ті: Інтеграція потужних мовних моделей (LLM) та інших інструментів штучного інтелекту. Це дозволяє працювати з раніше «сліпими зонами» — неструктурованими даними (тексти документів, розшифровки, скани), автоматизувати складні аналітичні завдання та генерувати контент.

Висновок: синергія для розслідувань та інновацій. Дата-журналістика та обчислювальна журналістика не конкурують, а органічно доповнюють одна одну в арсеналі сучасного медіапрофесіонала. Дата-журналістика надає фінальний продукт — переконливу, візуально привабливу історію. Обчислювальна журналістика надає інженерні методи та інструменти, щоб цю історію знайти, витягнути та обробити з навколишнього інформаційного океану. Разом вони роблять журналістику могутнішим інструментом для аналізу складних соціальних явищ, контролю за владою та розповіді про світ, керований даними.

Таблиця 3.1.

Порівняння дата-журналістики та обчислювальної журналістики

Аспект	Дата-журналістика (Data Journalism)	Обчислювальна журналістика (Computational Journalism)
Основна філософія	Розповісти історію на основі аналізу даних. Інтерв'ю з даними.	Застосування обчислювального мислення та комп'ютерних наук до всіх процесів журналістики.
Ключовий підхід	Evidence-based reporting: пошук доказів у структурованих даних.	Автоматизація рутини, створення інструментів, розслідування алгоритмів.
Процес	Збір, очищення, аналіз, візуалізація готових або отриманих масивів даних.	Програмування, скрейпінг, робота з API, аналіз неструктурованої інформації, створення алгоритмів.
Типовий інструментарій	Excel, Google Sheets, Tableau, Datawrapper, Flourish.	Мови програмування (Python, R, JavaScript), фреймворки для скрейпінгу, бібліотеки AI/ML (наприклад, для NLP).
Кінцевий продукт	Інфографіка, інтерактивна карта, стаття з графіками та візуалізаціями.	База даних, чат-бот, автоматизована система сповіщень, розслідувальний алгоритм, дослідження впливу алгоритмів.
Аналогія	Водій таксі: доставляє пасажира (читача) з точки А (питання) до точки Б (відповідь), використовуючи існуючі дороги (дані) та навігатор (інструменти).	Інженер-конструктор: проектує нові моделі автомобілів (інструменти), будує дороги (платформи) та аналізує рух транспорту (системи) для покращення перевезень в цілому.

Складено автором. Джерело. "Варіативність тлумачення data journalism" // Філологічні трактати (2020). https://www.philol. vernadskyjournals.in.ua /journals /2020/2_2020/part_4/23.pdf

Дата-журналістика фокусується на аналізі готових даних для створення історій та візуалізацій (Excel → Datawrapper → стаття). Обчислювальна журналістика створює алгоритми, які автоматизують весь процес — від збору до публікації (Python → автостатті → новина). Перша — це навичка роботи з даними, друга — програмування журналістських процесів. Для українських редакцій оптимальний

шлях: Datawrapper + Tableau (початок) → Python/R автоматизація (масштабування).

3.2.3. Алгоритмічне курування та дистрибуція: нові "ворітарі" цифрової інформації та їх виклики.

Ми живемо в епоху інформаційного потопу, де обсяг контенту, що створюється щохвилини, на порядки перевищує споживчі можливості будь-якої людини. У цьому хаосі алгоритмічне курування стало тим фільтром, який одночасно рятує нас від перевантаження та фундаментально змінює ландшафт медіа, передаючи контроль над вибором інформації від людей до машин. Цей процес породжує нову модель "ворітарства" (gatekeeping), трансформує механізми дистрибуції і ставить серйозні етичні та суспільні питання щодо упередженості та прозорості.

Новий ворітар: від суспільної значущості до персональної релевантності. Традиційно роль ворітаря — редактора, який вирішував, які новини варті уваги суспільства, — виконувала людина. Її рішення базувалися на професійних критеріях: суспільна значущість, актуальність, новизна, розкриття конфлікту інтересів. Алгоритмічний редактор працює за іншою логікою. Його першочергові метрики — це персональна релевантність та ймовірність залучення (engagement). Він аналізує вашу минулу поведінку (на що клікали, скільки часу витрачали, з ким спілкувалися), щоб передбачити, що вас найбільше зацікавить і змусить залишитись на платформі.

Це проявляється в роботі двох ключових типів платформ:

• Агрегатори новин (Google News, Apple News, Flipboard): Це системи без живого випускового редактора. ШІ постійно сканує тисячі джерел, групує схожі новини в кластери (щоб уникнути дублювання) і ранжує їх у вашій стрічці. Ранжування залежить не тільки від вашої історії, а й від авторитетності джерела, свіжості та географічної близькості події. Однак головною метою залишається утримати вашу увагу.

• Стрічки соціальних мереж (Facebook, X/Twitter, Instagram): Вони давно відмовилися від хронологічного порядку. Алгоритм показує пости, які, на його думку, мають максимальний потенціал отримати лайк, коментар, репост або тривалий перегляд. Це створює середовище, де контент, що викликає сильні емоції (часто негативні) або підтверджує існуючі переконання, має перевагу перед зваженим

або інформаційним.

Механізми дистрибуції: інтелектуальна доставка контенту. Алгоритмічне курування виходить далеко за межі головної стрічки. Воно пронизує всі канали дистрибуції, роблячи доставку контенту індивідуальною та прогностичною.

• Рекомендаційні віджети ("Читайте також", "Вас може зацікавити"): Раніше їх наповнювали редактори вручну. Сьогодні це роблять спеціалізовані ШІ-системи (як Taboola або Outbrain) або власні алгоритми видань. Вони аналізують не тільки вашу поведінку на сайті, але й контекст статті, яку ви читаєте, щоб запропонувати максимально "липкий" супутній контент, часто з метою монетизації.

• Розумні сповіщення (Smart Notifications): ШІ вирішує, кому і коли надіслати push-повідомлення про новину. Він враховує ваші інтереси (наприклад, не надсилати спортсменам-аматорам політичні новини о півночі), історичну активність та навіть годину доби. Мета — максимальна ймовірність кліку без спаму, що може призвести до інформаційної ізоляції.

• Персоналізовані розсилки (Email-дайджести): Замість єдиного листа для всієї аудиторії, ШІ формує унікальний набір заголовків і анонсів для кожного підписника. Це підвищує конверсію, але означає, що двоє людей, які підписалися на одну й ту саме розсилку, отримують різну картину дня.

Проблема "Чорної скриньки" та фундаментальні ризики. Найбільша загроза алгоритмічного курування полягає в його непрозорості. Логіка прийняття рішень часто є комерційною таємницею ("чорна скринька"), що унеможливлює зовнішній аудит і суспільний контроль. Це породжує два ключові ризики:

1. Вбудовані упередження (Bias): Алгоритми навчаються на історичних даних, які часто містять соціальні упередження. Якщо статистично чоловіки частіше клікають на новини про технології, а жінки — про культуру, система може почати систематично пропонувати їм відповідний контент, посилюючи гендерні стереотипи і обмежуючи інформаційний раціон користувачів. Це стосується не лише гендеру, але й расових, політичних та економічних аспектів.

2. Формування "ехо-камер" (Echo Chambers) та "фільтрувальних бульбашок" (Filter Bubbles): Персоналізуючи контент, алгоритм прагне показувати нам те, з чим ми, швидше за все, погодимося. По-

ступово людина опиняється в інформаційному середовищі, що лише підтверджує її існуючі переконання. Це обмежує стичність з альтернативними точками зору, посилює соціальну поляризацію, спрощує складні суспільні проблеми до бінарних опозицій і робить демократичний діалог все складнішим.

Таблиця 3.2.

Порівняння традиційного та алгоритмічного курування

Критерій	Традиційне (людське) Курування	Алгоритмічне (машинне) Курування
Критерій відбору	Суспільна значущість, новизна, професійні стандарти журналістики.	Персональна релевантність, ймовірність залучення (engagement), утримання уваги.
Основна мета	Інформувати громадян, виконувати роль "вартового демократії".	Максимізувати тривалість та активність перебування на платформі.
Прозорість	Відповідальність редакції, публічні стандарти, можливість звернення до редактора.	"Чорна скринька", комерційна таємниця, неможливість зрозуміти логіку вибору.
Різноманітність контенту	Намагання представити спектр поглядів у межах редакційної політики.	Прагнення до гомогенізації та персоналізації, що веде до "ехо-камер".
Упередження	Можуть бути суб'єктивними, але публічно видимими та підконтрольними критиці.	Вбудовані в дані та логіку, системні, масштабні і важкі для виявлення.

Складено автором. Джерело. "Автоматизоване ухвалення рішень в медіа" // ConstJournal (2025). https://www.constjournal.com/wp-content/uploads/ issues/ 2025-3/pdfs/7-andrii-hachkevych

Висновок. Алгоритмічне курування та дистрибуція стали невід'ємною інфраструктурою сучасного інформаційного світу. Вони пропонують беззаперечну ефективність у боротьбі з перевантаженням, але замінюють публічні критерії новинного відбору приватною логікою максимізації уваги. Це переносить ризики з площини суб'єктивних редакційних рішень у площину системних, масштабних та важковловимих упереджень. Майбутнє інформаційного суспільства залежатиме від здатності розробників, регуляторів, медіа та самих користувачів вимагати більшої прозорості алгоритмів, розвивати

критичну медіаграмотність та знаходити нові моделі, які поєднують ефективність технологій з цінностями громадянської журналістики. Боротьба йде не за повернення до минулого, а за те, щоб технології служили суспільним інтересам, а не лише комерційним.

3.2.4. Верифікація та фактчекінг з використанням ШІ

Мета та актуальність модуля. Метою цього навчального модуля є ознайомлення студентів з сучасними інструментами та методологіями верифікації інформації в епоху штучного інтелекту. Умовами цифрового середовища, де швидкість поширення інформації на порядки перевищує можливості її традиційної перевірки, а генеративний ШІ створює переконливі фейки (тексти, зображення, відео), класичний фактчекінг зазнає трансформації. Модуль розглядає концепцію Augmented Fact-Checking — посиленого фактчекінгу, де ШІ виконує рутинну, технічну роботу з аналізу великих масивів даних, пошуку артефактів і звірки тверджень, а журналіст зосереджується на критичному аналізі, інтерпретації та винесенні остаточного професійного вердикту. Це навичка життєво необхідна для сучасного медіафахівця, який працює в умовах інформаційної війни та гонки технологій.

Основні теми та зміст. Філософія Augmented Fact-Checking: ШІ як помічник, а не замінник журналіста. Етика використання автоматизованих інструментів. Детекція синтетичного тексту: Принципи роботи AI-детекторів, їх обмеження та ризики. Верифікація зображень та відео (Media Forensics): Методи виявлення deepfakes, аналіз метаданих, зворотний пошук. Автоматизована перевірка тверджень: Робота з базами перевірених фактів та стандартом ClaimReview. Інтеграція ШІ-інструментів у редакційний workflow: Від розшифровки аудіо до моніторингу дезінформаційних кампаній.

Детекція синтетичного тексту. Генеративні мовні моделі (ChatGPT, Gemini та ін.) здатні створювати когерентні, але часто неправдиві або маніпулятивні тексти. Інструменти-детектори (AI Text Classifiers), такі як ZeroGPT або Copyleaks, намагаються відрізнити людський текст від машинного. Вони аналізують дві ключові характеристики:

• Perplexity (Складність): Вимірює, наскільки текст є непередбачуваним для моделі. ШІ-генерований текст зазвичай має низьку складність, оскільки модель вибирає статистично найімовірніші наступні слова. Людська мова більш хаотична та творча.

• Burstiness (Вибуховість): Оцінює варіативність довжини та структури речень. Текст ШІ часто має однакову, «рівномірну» структуру.

Важливе застереження: Ці інструменти є ненадійними та часто дають хибно-позитивні результати, особливо для текстів, написаних формальною або науковою мовою. Їхній результат слід розглядати лише як перший сигнал для поглибленої перевірки, а не як доказ. Фінальне рішення про авторство завжди залишається за людиною, яка аналізує контекст, логічні невідповідності та можливі джерела тексту.

Детекція візуальних фейків (Deepfake Detection). Виявлення згенерованих фото та відео (deepfakes) — це окрема дисципліна цифрової криміналістики. Журналіст повинен володіти базовими методами аналізу:

• Візуальний аналіз артефактів: ШІ часто помиляється в деталях, які людина помічає інтуїтивно: нереальна кількість пальців на руці, нечитабельні або абсурдні написи на вивісках, асиметричні сережки, надто ідеальна або «пластикова» текстура шкіри, дивні локони волосся або зуби. Треба звертати увагу на фон — ШІ часто генерує його з частин інших зображень.

• Технічний аналіз: Професійні інструменти аналізують невидимі артефакти. Наприклад, Error Level Analysis (ELA) показує різницю в рівнях стиснення різних ділянок зображення, що може вказати на накладення одного фото на інше. Перевірка метаданих (EXIF) може виявити відсутність інформації про камеру або дивні дати.

• Зворотний пошук: Найважливіший перший крок. Інструменти Google Reverse Image Search або TinEye дозволяють знайти оригінальне зображення та перевірити, чи не було воно використане в іншому контексті чи маніпульоване. TinEye особливо корисний фільтрами «найстаріший» та «найбільш змінений». Для відео використовуються такі сервіси, як InVID/WeVerify, які дозволяють розкласти відео на кадри та перевірити кожен окремо.

Автоматизована перевірка тверджень (Automated Claim Checking)

Цей підхід працює за аналогією з «Shazam для фейків». Система порівнює підозріле твердження з базою вже перевірених фактів. Принцип роботи: Журналіст вводить цитату або тезу. ШІ розбиває її на окремі фактичні компоненти та автоматично звіряє їх з базами

авторитетних фактчекінгових організацій, таких як Snopes, PolitiFact, BBC Reality Check або українські StopFake та VoxCheck. Ключовий інструмент — Google Fact Check Explorer: Це безкоштовний пошуковик по мільйонам перевірок фактів з усього світу. Досить ввести ключове слово або завантажити зображення, щоб дізнатися, чи вже хтось перевіряв це твердження та який вердикт був винесений. Це значно прискорює початкову стадію розслідування. Стандарт ClaimReview: Це технічний стандарт, що дозволяє фактчекерам розмічати свої статті спеціальним кодом. Пошукові системи (Google) розпізнають цю розмітку і можуть надавати її в результатах пошуку, безпосередньо вказуючи на висновок фактчекерів. Це підвищує видимість та ефективність боротьби з дезінформацією.

Практичні інструменти та ресурси для роботи. Для ефективної роботи фактчекеру сьогодні необхідно вміти користуватись цілим арсеналом цифрових інструментів.

Таблиця 3.3.

Порівняння ключових категорій інструментів та їх призначення.

Категорія інструментів	Призначення	Приклади сервісів та ресурсів
Для роботи з текстом та даними	Розшифровка аудіо/відео, переклад, аналіз документів, пошук інформації.	Google's AI Transcription Tool, NoteGPT, Deepl, Perplexity AI (пошук з посиланнями на джерела).
Для верифікації зображень	Зворотний пошук, аналіз метаданих, виявлення маніпуляцій.	Google Reverse Image Search, TinEye (з фільтрами), Forensically, EXIF data перевірка.
Для верифікації відео	Розбиття на кадри, пошук оригіналу, аналіз метаданих.	InVID/WeVerify, YouTube Data Viewer, Watch Frame by Frame.
Бази перевірених фактів	Швидка перевірка тверджень на вже існуючі розслідування.	Google Fact Check Explorer, PolitiFact, Snopes, StopFake.
Моніторинг дезінформації	Відстеження наративів та трендів у соцмережах.	Osavul, Hamilton 2.0 Dashboard.
Перевірка джерел	Аналіз доменів, історія сайтів.	Whois, Wayback Machine, Domain Digger.

Складено автором. Джерело. Hootsuite OwlyGPT Docs https:// www. hootsuite. com/ owlygpt

Висновок. Модуль демонструє, що в сучасній «гонці озброєнь» між творцями дезінформації та журналістами останні мусять активно впроваджувати ШІ-інструменти у свою практику. Ключовим принципом залишається Augmented Intelligence — посилений інтелект, де технологія не замінює, а розширює можливості професійного журналіста. Навичка критично оцінювати результати роботи алгоритмів, сумніватися в них і завжди робити остаточний, обґрунтований висновок на основі множинних джерел і власного розсудку — це те, що формує незамінну цінність та етичну основу професії в цифрову епоху.

3.2.5. Персоналізовані новинні стрічки (The Feed)

Персоналізована новинна стрічка (англ. The Feed) є основним інтерфейсом сучасних цифрових платформ, що представляє собою безперервно оновлюваний потік контенту, автоматично сформований алгоритмами штучного інтелекту під індивідуальні вподобання та поведінку користувача. Цей механізм ознаменував кінець епохи універсальної «головної сторінки» та спричинив фундаментальний зсув у логіці отримання інформації — від активного пошуку до пасивного споживання, коли контент «знаходить» користувача самостійно.

Еволюція логіки формування стрічки: від Social Graph до Interest Graph

Історично формування контенту в соціальних мережах базувалося на концепції Соціального графа (Social Graph). Користувач бачив публікації тих людей та сторінок, на які він був підписаний. Ця модель створювала обмеження: якість стрічки повністю залежала від активності та інтересів знайомих, що могло призводити до формування інформаційних «бульбашок» та обмеження кругозору. У новій парадигмі, яку часто називають «TikTok-ізацією», домінує Граф інтересів (Interest Graph). Алгоритм аналізує не соціальні зв'язки, а мікро-вподобання користувача: час перегляду, вподобання, коментарі, навіть паузи та повторні перегляди. В результаті в стрічці з'являється контент від абсолютно незнайомих авторів, який система прогнозує як максимально релевантний. Це призвело до радикальних змін для медіа: лояльність бренду та велика кількість підписників втратили першочергове значення. Кожна окрема публікація мусить боротися за увагу в загальному алгоритмічному потоці, що породжує

гонку за клікабельністю та миттєвим залученням.

Архітектура та балансування вмісту: проблема Міх-фактору

Створення ідеальної стрічки є складною задачею балансування різних типів контенту, аналогічною до складання збалансованого раціону. У цьому контексті виділяють три ключові категорії:

1. «Цукерки» (Candy) — розважальний, сенсаційний або клікбейтний контент, що забезпечує миттєвий дофаміновий відгук і високу залученость.

2. «Овочі» (Vegetables) — важливий, але часто складний або стресовий контент (політика, економіка, соціальні проблеми), необхідний для інформованості громадянина.

3. «Вітаміни» або серендипіті (Serendipity) — контент, що відкриває нові, несподівані інтереси та розширює кругозір.

Роль алгоритму в прогресивних системах (як Google News) полягає в оптимальному змішуванні цих інгредієнтів. Мета — утримати користувача від «інформаційного переїдання» легким контентом, але й не спричинити втоми або бажання втекти через надмір «важкої» інформації. Ефективні системи використовують стратегії прогресивної експозиції та контекстуального змішування, адаптуючи стрічку до часу доби, тривалості сесії та емоційного відгуку користувача.

Психологія нескінченного скролу та етичні виклики. Техніка нескінченного скролу (Infinite Scroll) є фундаментальним механізмом утримання уваги, заснованим на принципі змінної винагороди (Variable Reward), запозиченому з дизайну азартних ігор. Користувач, прокручуючи стрічку, отримує контентну «винагороду» (цікаву новину) не регулярно, а в непередбачуваний момент. Це запускає в мозку потужну дофамінову реакцію очікування та формує поведінкову петлю, описану Моделлю залучення (Hook Model): тригер (нудьга) → дія (скрол) → змінна винагорода → інвестиція (лайк, коментар). Ця модель гарантує високу залежність і постійний повернення до додатку.

Однак для новинних медіа та суспільства в цілому це створює серйозні етичні дилеми. Чи можна використовувати механіки, що викликають залежність, для інформування про важливі події? Як балансувати між метою надання корисної інформації та бізнес-потребою максимізації часу на платформі? Персоналізація також

підриває спільний інформаційний простір, де кожен користувач отримує унікальну версію реальності, що ускладнює суспільний діалог і консенсус.

Таким чином, персоналізована стрічка є не просто технологічним інструментом, а соціально-технічним феноменом, що глибоко впливає на сприйняття інформації, формування суспільної думки та практики журналістики. Її розвиток вимагає не лише вдосконалення алгоритмів, але й широкої громадської дискусії про цінності, прозорість та відповідальність у цифрову епоху.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допоможуть класифікувати різні види автоматизованих медіасистем:

1. Класифікація за функцією: Назвіть три основні типи алгоритмів у ньюзрумі: *Створення контенту* (Reporting), *Курування/Дистрибуція* (Curation) та *Розслідування/Аналіз* (Investigative augmentation). У чому різниця їхніх цілей?

2. Alerting Systems (Системи сповіщення): Як працює *Quakebot* (LA Times) або алгоритми Bloomberg First Word? Чому в цьому типі журналістики швидкість (мілісекунди) важливіша за стиль написання?

3. Описова журналістика (Descriptive Journalism): Це найпоширеніший тип (спорт, фінанси). Як такі алгоритми перетворюють таблицю з даними на зв'язний текст? Чому вони ідеальні для "довгого хвоста" (наприклад, опису матчів нижчих ліг)?

4. Investigative Augmentation (Допомога у розслідуванні): Алгоритми не пишуть текст, а шукають аномалії. Як консорціум ICIJ використовував граф-бази даних та алгоритми пошуку сутностей (NER) для аналізу "Panama Papers"?

5. Генеративна/Креативна журналістика: Чим новітні спроби використовувати LLM (Large Language Models) для написання колонок або есе відрізняються від традиційних rule-based звітів? (Проблема фактажу vs. проблема стилю).

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання спрямовані на розуміння специфіки кожного типу:

Завдання 1. Проєктування Alert-бота (Сповіщувача)

• Сценарій: Ви працюєте в локальному медіа "Київ Екологічний".

• Завдання: Розробіть логіку бота, який пише новини про якість повітря.

- *Тригер:* (Дані з датчика AQI > 100).

- *Шаблон заголовка:* "Увага! У районі [Район] повітря небезпечне."

- *Шаблон тексту:* "Датчики зафіксували рівень [Значення]. Рекомендуємо зачинити вікна."

- *Умова публікації:* Публікувати автоматично, без премодерації (чому?).

Завдання 2. Порівняльний аналіз: Звіт vs Історія

• Матеріал: Знайдіть автоматично згенерований фінансовий звіт (наприклад, на Yahoo Finance) і аналітичну статтю про ту ж подію від живого журналіста.

• Аналіз: Складіть таблицю порівняння.

- *Факти:* (Чи однакові цифри?).

- *Контекст:* (Чи пояснює бот причини падіння акцій? Чи згадує чутки/інсайди?).

- *Емоційне забарвлення:* (Як відрізняється лексика?).

Завдання 3. Кейс "Panama Papers": Алгоритм як детектив

• Ситуація: У вас є 11 мільйонів документів.

• Завдання: Сформулюйте 3 запити (алгоритмічні правила), які допомогли б знайти корупцію:

- *Приклад:* "Знайти всіх осіб, які є політиками (зі списку PEP) І мають рахунок, відкритий на Британських Віргінських островах".

• Мета: Зрозуміти тип "Investigative Augmentation", де алгоритм не пише статтю, а підносить патрони журналісту.

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. ICIJ (International Consortium of Investigative Journalists). *How we analyzed the Panama Papers.* (Опис технологічного стеку розслідувачів).

2. Caswell, David. *Structured Journalism and the Robot Reporter.* (Класифікація типів автоматизації).

3. BBC News Labs. *Juicer project description.* (Приклад системи, що тегує та сортує контент — тип "Курування").

4. Schapals, A. K., & Porlezza, C. *Assistance or Replacement? Types of Automation in Newsrooms.*

4. ГЛОСАРІЙ ТЕРМІНІВ

• Alerting Systems (Системи сповіщення): Тип алгоритмічної журналістики, пріоритетом якого є миттєва публікація коротких повідомлень про критичні події (землетруси, вбивства, стрибки акцій) на основі даних моніторингу.

• Rule-based Systems (Системи на основі правил): Алгоритми, що працюють за чіткою логікою "Якщо X, то Y" (IF-THEN). Це найнадійніший тип для фінансової та спортивної журналістики, оскільки він виключає "галюцинації".

• Investigative Augmentation (Доповнена розслідувальна журналістика): Використання алгоритмів для пошуку зв'язків (link analysis), розпізнавання імен (NER) та виявлення патернів у великих масивах витоків даних, які людина не здатна прочитати фізично.

• Cyborg Journalism (Кіборг-журналістика): Підхід, де людина і машина співпрацюють над одним текстом: алгоритм створює чернетку з даними, а журналіст додає аналіз, цитати та перевіряє факти.

3.3. Технології, що забезпечують алгоритмічну журналістику

Алгоритмічна журналістика спирається на комплекс сучасних технологій штучного інтелекту та машинного навчання, які дозволяють автоматизувати процеси збору, аналізу та представлення інформації. Ці технології трансформують традиційні журналістські практики, забезпечуючи швидкість, масштабованість та точність обробки величезних обсягів даних. Розвиток обчислювальних потужностей, доступність великих наборів даних та прогрес у галузі штучного інтелекту створили унікальне технологічне середовище, яке фундаментально змінює спосіб виробництва та споживання новинного контенту.

Сучасна алгоритмічна журналістика не обмежується використанням окремих технологій, а інтегрує їх у складні системи, здатні виконувати повний цикл журналістської роботи – від моніторингу джерел інформації до публікації готових матеріалів. Це відкриває нові можливості для медіаорганізацій різного масштабу, дозволяючи невеликим редакціям конкурувати з великими медіахолдингами завдяки ефективному використанню автоматизації, а великим видавництвам – масштабувати виробництво контенту до рівня, недосяжного традиційними методами.

3.3.1. Обробка природної мови (NLP).

Обробка природної мови є фундаментальною технологією алгоритмічної журналістики, що дозволяє комп'ютерним системам розуміти, інтерпретувати та генерувати людську мову в усій її складності та багатогранності. NLP поєднує методи лінгвістики, інформатики та штучного інтелекту для створення систем, здатних працювати з текстом так само, як це робить людина, але з набагато вищою швидкістю та можливістю обробки величезних обсягів інформації.

Історично розвиток NLP у журналістиці пройшов кілька етапів. Ранні системи, що з'явилися у 1990-х роках, використовували прості шаблони та правила для генерації стандартизованих новинних повідомлень, таких як спортивні зведення чи фінансові звіти. Ці системи базувалися на жорстких граматичних правилах та обмежених словниках, що робило їх застосування досить вузьким. Проте вже на цьому етапі було продемонстровано потенціал автоматизації

рутинних журналістських завдань.

Сучасні NLP-системи використовують статистичні методи та машинне навчання, що дозволяє їм адаптуватися до різних стилів мовлення, контекстів та жанрів. Одним із ключових застосувань NLP у журналістиці є автоматичне створення новинних текстів на основі структурованих даних. Наприклад, агентство Associated Press використовує технологію Wordsmith для генерації тисяч фінансових звітів щокварталу, аналізуючи дані про прибутки компаній та автоматично створюючи читабельні статті, що містять ключові показники, порівняння з попередніми періодами та контекстну інформацію.

Резюмування великих документів є іншим важливим застосуванням NLP. Журналісти часто стикаються з необхідністю опрацювання об'ємних звітів, судових рішень, наукових публікацій чи урядових документів. Системи автоматичного резюмування можуть виділяти найважливіші тези, створювати короткі виклади основного змісту та навіть генерувати різні версії резюме залежно від цільової аудиторії. Існує два основних підходи до резюмування: екстрактивний, який виділяє найважливіші речення з оригінального тексту, та абстрактивний, який генерує нові формулювання, зберігаючи ключовий зміст.

Переклад матеріалів між мовами став значно точнішим завдяки нейронним мережам. Технології машинного перекладу, такі як Google Translate чи DeepL, використовують складні моделі, що враховують контекст, ідіоматичні вирази та культурні особливості. Це дозволяє редакціям швидко адаптувати контент для міжнародної аудиторії, перекладати іноземні джерела та розширювати географію висвітлення подій. Деякі видання, як Reuters та Bloomberg, інтегрували автоматичний переклад безпосередньо у свої робочі процеси, що дозволяє публікувати новини кількома мовами одночасно.

Аналіз тональності публікацій став незамінним інструментом для розуміння громадських настроїв. NLP-системи можуть визначати, чи є текст позитивним, негативним або нейтральним, а також ідентифікувати конкретні емоції – радість, гнів, страх, здивування. Це особливо корисно при аналізі реакцій на політичні події, корпоративні оголошення чи соціальні кампанії. Журналісти використовують аналіз тональності для вимірювання громадської

думки, виявлення змін у суспільних настроях та доповнення своїх матеріалів даними про емоційний контекст подій.

Екстракція ключових фактів із прес-релізів та інших джерел дозволяє автоматично ідентифікувати дати, місця, імена, організації, числові дані та інші важливі елементи інформації. Система розпізнавання іменованих сутностей (Named Entity Recognition, NER) може обробити сотні документів за лічені хвилини, виділяючи структуровану інформацію, яку журналіст може використати для створення матеріалу. Це значно прискорює процес дослідження та дозволяє редакціям швидко реагувати на події.

Автоматичне створення фінансових звітів стало однією з найуспішніших реалізацій NLP у журналістиці. Компанії отримують структуровані дані про квартальні результати, і система може автоматично створити базову новину, що містить зміни у доходах, прибутках, ключових показниках ефективності та порівняння з очікуваннями аналітиків. Журналіст може потім доповнити цей текст експертними коментарями, аналізом та контекстом, але рутинна робота зі збирання та форматування даних вже виконана.

Моніторинг згадок у медіа допомагає відстежувати, як конкретні теми, персони чи організації висвітлюються у різних джерелах. NLP-системи можуть аналізувати тисячі публікацій, виявляти тенденції у висвітленні, порівнювати підходи різних видань та ідентифікувати зміни в тональності з часом. Це корисно як для самих медіа, що відстежують власну репутацію, так і для журналістів-розслідувачів, які аналізують медіаландшафт навколо певної теми.

Виявлення трендів у суспільному дискурсі можливе завдяки аналізу великих обсягів текстів із різних джерел – новинних сайтів, соціальних мереж, форумів, блогів. Системи тематичного моделювання (topic modeling) можуть автоматично ідентифікувати основні теми обговорення, відстежувати їхню динаміку та виявляти зв'язки між різними дискусіями. Це дає журналістам уявлення про те, що турбує суспільство, які теми набирають популярності та як змінюється публічна порядок денна.

Розпізнавання іменованих сутностей та встановлення зв'язків між подіями та особами створює основу для побудови знаннєвих графів – структурованих представлень інформації про світ. Ці графи показують, як різні персони, організації та події пов'язані між

собою, що надзвичайно корисно для розслідувальної журналістики. Наприклад, система може автоматично відстежувати всі згадки певного політика, його зв'язки з бізнесовими структурами, участь у різних подіях та заяви у медіа, створюючи комплексну картину його діяльності.

Визначення емоційного забарвлення текстів виходить за межі простого позитивного чи негативного sentiment і може ідентифікувати складні емоційні стани, іронію, сарказм та багатошаровість значень. Сучасні системи використовують контекстуальні моделі, що враховують не лише окремі слова, а й їхнє розташування, синтаксичну структуру та культурні конотації.

Виявлення потенційної дезінформації через аналіз мовних патернів стало критично важливим у боротьбі з fake news. NLP-системи можуть ідентифікувати характерні ознаки маніпулятивних текстів – використання емоційно забарвленої лексики, відсутність конкретних фактів, суперечності у формулюваннях, характерні для певних джерел мовні штампи. Дослідники розробили моделі, що аналізують не лише зміст, а й стиль написання, що дозволяє виявляти координовані кампанії дезінформації та ботів, що імітують людську комунікацію.

3.3.2. Машинне навчання та глибоке навчання.

Машинне навчання становить основу для створення прогнозних моделей та виявлення прихованих патернів у журналістських даних, забезпечуючи здатність комп'ютерних систем навчатися на прикладах без явного програмування кожного правила. Ця технологія трансформувала алгоритмічну журналістику, дозволивши створювати адаптивні системи, що покращують свою роботу з часом та можуть справлятися з завданнями, які раніше вважалися виключно людськими.

Основні типи машинного навчання знаходять різне застосування у журналістиці. Навчання з вчителем (supervised learning) використовується, коли є набір прикладів з відомими правильними відповідями. Наприклад, для створення класифікатора новин за темами журналісти розмічають кілька тисяч статей, вказуючи їхню тематику (політика, економіка, спорт, культура тощо), а алгоритм навчається розпізнавати характерні ознаки кожної категорії. Після навчання система може автоматично класифікувати нові статті з

високою точністю.

Навчання без вчителя (unsupervised learning) застосовується для виявлення прихованих структур у даних без попередньо заданих категорій. Кластеризація статей може виявити несподівані тематичні групи, зв'язки між різними новинними подіями чи нові тренди, які ще не оформилися у чіткі категорії. Це особливо корисно для новинної аналітики та стратегічного планування контенту.

Навчання з підкріпленням (reinforcement learning) використовується для оптимізації послідовних рішень, наприклад, у системах рекомендацій контенту. Алгоритм отримує зворотний зв'язок від користувачів (клік, час читання, шерінг) і поступово навчається пропонувати матеріали, що максимізують залученість аудиторії, балансуючи між популярністю та різноманітністю контенту.

Класифікація новин за темами є однією з базових задач машинного навчання у журналістиці. Сучасні редакції отримують тисячі новинних повідомлень щодня з різних джерел – агентств, власних кореспондентів, соціальних мереж. Автоматична класифікація дозволяє швидко організувати цей потік, направляючи матеріали відповідним редакторам, формуючи тематичні дайджести та забезпечуючи ефективну навігацію на сайті. Багатокласова та багаторівнева класифікація дозволяє присвоювати статті кілька тегів одночасно та організовувати ієрархічну структуру тем.

Персоналізація контенту для різних аудиторій використовує складні моделі, що враховують історію читання користувача, демографічні характеристики, час доби, тип пристрою та безліч інших факторів. Системи рекомендацій, подібні до тих, що використовують Netflix чи Spotify, адаптовані для новинного контенту, допомагають кожному читачеві знаходити найцікавіші для нього матеріали. Проте це створює і ризик "інформаційних бульбашок", тому провідні медіа намагаються балансувати персоналізацію з забезпеченням різноманітності перспектив.

Прогнозування вірусності публікацій допомагає редакціям передбачити, які матеріали мають потенціал широкого поширення. Моделі аналізують характеристики контенту (тема, стиль, емоційний тон, наявність візуальних елементів), характеристики автора, час публікації та поточний інформаційний контекст. Це дозволяє оптимально розподіляти редакційні ресурси, промо-підтримку та

розміщення на головних сторінках.

Виявлення аномалій у даних є критично важливим для моніторингу різних сфер життя. Системи машинного навчання можуть автоматично детектувати незвичні патерни у фінансових операціях, відхилення від очікуваних результатів виборів, нетипову активність у соціальних мережах чи несподівані зміни у статистичних показниках. Такі аномалії часто свідчать про важливі події, що заслуговують журналістської уваги – від корупційних схем до епідемічних спалахів.

Глибоке навчання, як підгалузь машинного навчання, застосовує багатошарові нейронні мережі для розв'язання складних завдань, що вимагають виявлення ієрархічних структур у даних. Архітектури глибокого навчання, такі як згорткові нейронні мережі (CNN), рекурентні нейронні мережі (RNN), трансформери та їхні різновиди, досягли вражаючих результатів у різних аспектах обробки інформації.

Генерація природномовних текстів високої якості стала можливою завдяки архітектурам seq2seq (послідовність-до-послідовності) та механізмам уваги (attention mechanisms). Системи можуть створювати зв'язні, граматично правильні тексти, що адаптовані до певного стилю та аудиторії. The Washington Post використовує систему Heliograf для автоматичного створення локальних новин про спортивні події, вибори та інші структуровані події, звільняючи журналістів для більш складних аналітичних матеріалів.

Створення узагальнень документів із збереженням ключових смислів використовує моделі, здатні "розуміти" глибинну структуру тексту. Абстрактивне резюмування генерує нові формулювання, а не просто витягує речення з оригіналу, що робить резюме більш природними та адаптованими до контексту. Це особливо корисно при роботі з технічними документами, де потрібно пояснити складні концепції доступною мовою.

Автоматичне доповнення статей релевантною інформацією може включати додавання контексту, пояснень термінів, посилань на попередні публікації з теми, статистичних даних чи експертних цитат із баз даних. Системи можуть аналізувати чернетку статті, ідентифікувати теми та концепції, що потребують додаткового пояснення, та автоматично інтегрувати релевантну інформацію зі структурованих джерел.

Ідентифікація фейкових новин через аналіз множинних джерел використовує комплексний підхід, що поєднує аналіз змісту, джерел, поширення та контексту. Моделі глибокого навчання можуть виявляти суперечності між різними твердженнями, перевіряти факти через порівняння з надійними базами даних, аналізувати авторитетність джерел та відстежувати, як інформація змінюється при поширенні. Проекти типу Full Fact у Великій Британії чи Duke Reporters' Lab використовують машинне навчання для автоматизації процесу фактчекінгу.

Ці технології дозволяють редакціям обробляти величезні масиви інформації, які неможливо опрацювати вручну. Наприклад, при витоку великих баз даних (як Panama Papers чи Pandora Papers) журналісти використовували машинне навчання для аналізу мільйонів документів, виявлення зв'язків між особами та організаціями, ідентифікації підозрілих транзакцій та пріоритизації матеріалів для детального розслідування.

Персоналізований досвід споживання контенту виходить за межі простих рекомендацій та включає адаптацію формату, довжини, стилю презентації матеріалів. Деякі користувачі віддають перевагу коротким резюме, інші – детальному аналізу; когось цікавлять візуалізації даних, а когось – текстові пояснення. Машинне навчання дозволяє автоматично адаптувати контент до індивідуальних переваг.

3.3.3. Комп'ютерний зір для аналізу візуального контенту.

Комп'ютерний зір розширює можливості алгоритмічної журналістики на візуальний контент, дозволяючи автоматично аналізувати фотографії, відео та інфографіку з рівнем деталізації, недосяжним для ручної обробки величезних обсягів візуальних матеріалів. Ця технологія трансформує як процес виробництва візуального контенту, так і методи його верифікації та аналізу.

Історично комп'ютерний зір розвивався від простих алгоритмів виявлення країв та геометричних форм до складних нейромережевих архітектур, здатних розуміти семантичний зміст зображень. Прорив відбувся у 2012 році з появою глибоких згорткових мереж (CNN), що продемонстрували драматичне покращення точності розпізнавання об'єктів у конкурсі ImageNet. З того часу технологія швидко еволюціонувала, і сьогодні системи комп'ютерного зору часто перевершують людей у специфічних завданнях розпізнавання.

Розпізнавання об'єктів на зображеннях дозволяє автоматично ідентифікувати людей, місця, предмети, тварин, транспортні засоби та тисячі інших категорій. Сучасні моделі, такі як YOLO (You Only Look Once), Faster R-CNN чи EfficientDet, можуть обробляти зображення у реальному часі, визначаючи не лише що зображено, а й де саме об'єкт розташований на фото. Це дозволяє автоматично тегувати фотоархіви, полегшуючи пошук потрібних зображень, та прискорює процес підготовки ілюстрацій до публікацій.

Для журналістики особливо важливим є розпізнавання місць та архітектурних об'єктів. Система може ідентифікувати, в якому місті зроблено фото, розпізнати відомі будівлі, пам'ятники чи ландшафти. Це критично важливо для верифікації візуального контенту з соціальних мереж, коли потрібно підтвердити або спростувати заяви про місце зйомки. Геолокаційні інструменти використовують комп'ютерний зір для порівняння зображень із супутниковими знімками та фотографіями з баз даних.

Виявлення подій та дій на фотографіях дозволяє автоматично класифікувати зображення за змістом – протест, концерт, спортивний матч, природна катастрофа. Моделі можуть розпізнавати складні сцени, що включають взаємодію декількох осіб, та інтерпретувати контекст ситуації. Це допомагає журналістам швидко сортувати великі обсяги візуального матеріалу, що надходить з місць подій.

Системи виявлення облич використовуються для автоматичного тегування персон на світлинах, що особливо корисно при висвітленні масових заходів, політичних подій, прес-конференцій. Технологія розпізнавання облич може ідентифікувати людину навіть за частковим зображенням обличчя, в різних ракурсах, при різному освітленні. Проте це створює серйозні етичні питання щодо приватності та можливості стеження, тому багато медіаорганізацій розробили етичні керівництва для використання цієї технології.

Провідні фотоагентства, такі як Getty Images та Reuters, використовують розпізнавання облич для автоматичного додавання підписів до фотографій. Коли фотограф завантажує знімки з політичного саміту чи церемонії нагородження, система автоматично ідентифікує присутніх політиків, знаменитостей чи спортсменів, значно прискорюючи процес підготовки фото до публікації. Це особливо важливо в умовах жорсткої конкуренції, де різниця у кілька

хвилин може визначити, чиї фото будуть використані виданнями.

Аналіз відеоконтенту включає кілька різних завдань. Сегментація відео на сцени дозволяє автоматично розділити довге відео на логічні частини за змістом. Виявлення ключових моментів ідентифікує найважливіші або найцікавіші сегменти – наприклад, голи у футбольному матчі, критичні заяви на прес-конференції чи драматичні моменти в документальному фільмі.

Автоматичне створення резюме довгих відеороликів генерує короткі версії, що містять найважливіші моменти. Це корисно для створення тізерів, адаптації контенту для різних платформ (повна версія для YouTube, коротка для Instagram) та допомоги глядачам швидко оцінити зміст перед переглядом повної версії. BBC використовує такі системи для автоматичного створення коротких версій новинних випусків та документальних програм.

Виявлення ключових моментів у трансляціях у реальному часі дозволяє автоматично створювати highlights спортивних подій, політичних дебатів чи інших прямих ефірів. Система аналізує візуальні та аудіо сигнали – зміни в інтенсивності руху, реакції натовпу, тональність мовців – щоб ідентифікувати моменти, що заслуговують особливої уваги.

Генерація субтитрів поєднує комп'ютерний зір з обробкою мовлення. Системи розпізнають, хто говорить у кадрі, синхронізують текст із відповідними спікерами та форматують субтитри для оптимальної читабельності. Автоматична генерація субтитрів робить контент доступнішим для людей із порушеннями слуху та дозволяє переглядати відео без звуку, що важливо для споживання контенту у громадських місцях.

Детекція маніпуляцій із зображеннями стала критично важливою для боротьби з дезінформацією. Системи комп'ютерного зору можуть виявляти різні типи фальсифікацій – від простого фотошопу до складних дипфейків. Аналіз метаданих EXIF перевіряє інформацію про камеру, параметри зйомки та можливі редагування. Проте досвідчені маніпулятори можуть видаляти чи підробляти метадані, тому необхідні додаткові методи перевірки.

Error Level Analysis (ELA) виявляє області зображення з різним рівнем компресії, що може свідчити про редагування. Коли частина зображення вставлена з іншого файлу або відредагована, вона матиме

інший рівень компресії JPEG, що виявляється при спеціальному аналізі. Clone Detection ідентифікує скопійовані та вставлені ділянки зображення – техніку, що часто використовується для приховування небажаних елементів або додавання об'єктів.

Аналіз узгодженості освітлення перевіряє, чи відповідають тіні та відблиски фізичним законам. Якщо на композитному зображенні об'єкти освітлені з різних напрямків, це свідчить про маніпуляцію. Перевірка перспективи та пропорцій виявляє об'єкти, масштаб яких не відповідає оточенню.

Виявлення дипфейків – відео з підробленими обличчями чи голосами – є особливо складним завданням. Моделі глибокого навчання аналізують мікро-вирази, неприродні рухи очей, артефакти навколо країв обличчя, невідповідність руху губ звуку. Детектори дипфейків постійно еволюціонують у відповідь на покращення методів створення підробок, ведучи своєрідну гонку озброєнь між створювачами та детекторами синтетичних медіа.

Консорціум великих технологічних компаній та медіаорганізацій створив Deepfake Detection Challenge, що стимулює розробку все точніших методів виявлення синтетичних медіа. Проте експерти наголошують, що технологічні рішення мають доповнюватися медіаграмотністю аудиторії та журналістськими стандартами верифікації.

Комп'ютерний зір також застосовується для аналізу супутникових знімків у розслідуваннях. Журналісти використовують системи автоматичного виявлення змін для моніторингу будівництва, вирубки лісів, переміщень військ, наслідків природних катастроф. Організації типу Bellingcat піонерно використовують комп'ютерний зір разом із open-source розвідкою для розслідування воєнних злочинів, корупції та порушень прав людини.

Моніторинг змін ландшафту дозволяє відстежувати екологічні порушення, незаконне будівництво, зміни у використанні земель. Алгоритми порівнюють знімки одного й того ж місця, зроблені в різний час, автоматично виділяючи зони змін. Це дозволило розслідувачам документувати масштабну вирубку тропічних лісів, незаконний видобуток корисних копалин, розширення військових об'єктів у закритих зонах.

Документування подій у реальному часі через аналіз візуального

контенту з соціальних мереж став важливим інструментом під час конфліктів, природних катастроф та соціальних рухів. Коли професійні журналісти не мають доступу до місця події, комп'ютерний зір допомагає обробляти тисячі фото та відео від очевидців, верифікувати їхню автентичність, встановлювати хронологію подій та створювати просторову реконструкцію того, що сталося.

Розпізнавання тексту на зображеннях (OCR - Optical Character Recognition) дозволяє автоматично витягувати текстову інформацію з фотографій документів, вивісок, плакатів, написів. Це корисно при аналізі великих обсягів сканованих архівів, обробці витоків документів, моніторингу візуальної пропаганди. Сучасні системи OCR можуть працювати з різними мовами, рукописним текстом, спотвореними зображеннями.

Аналіз візуальної естетики та композиції використовується для автоматичного відбору найякісніших зображень з великих наборів фотографій. Системи оцінюють технічну якість (різкість, експозицію, баланс білого), композиційні принципи (правило третин, золотий переріз, ведучі лінії), емоційний вплив. Це допомагає фоторедакторам швидко знаходити найсильніші кадри серед сотень знімків з події.

Створення автоматичних описів зображень для доступності генерує текстові альтернативи візуального контенту для людей з порушеннями зору. Системи Image Captioning використовують комп'ютерний зір разом із генерацією природної мови для створення змістовних описів того, що зображено на фото. Це не лише покращує доступність, а й допомагає у індексації та пошуку візуального контенту.

3.3.4. Аналіз соціальних мереж.

Соціальні мережі стали невід'ємним джерелом інформації для сучасної журналістики, а технології їхнього аналізу дозволяють виявляти важливі тренди та події на ранніх стадіях, коли традиційні медіа ще не встигли відреагувати. Аналіз соціальних мереж поєднує методи обробки природної мови, машинного навчання, теорії графів та статистики для створення комплексного розуміння інформаційних потоків у цифровому суспільстві.

Еволюція аналізу соціальних мереж у журналістиці відображає зміни у самих платформах та поведінці користувачів. Ранні дослідження фокусувалися на простому підрахунку згадок та базовому

аналізі тональності. Сучасні системи використовують складні мережеві моделі, виявляють координовані кампанії, аналізують мультимодальний контент (текст, зображення, відео) та відстежують каскади поширення інформації через різні платформи.

Алгоритми відстеження трендів моніторять обговорення у Twitter, Facebook, Instagram, TikTok, Reddit та інших платформах, ідентифікуючи теми, що набирають популярності. Ці системи використовують різні метрики – частоту згадок, швидкість зростання обговорень, географічне поширення, залученість аудиторії. Складність полягає у відокремленні органічних трендів від штучно створених через координовані кампанії або ботів.

Виявлення emerging topics використовує алгоритми, що відстежують не лише абсолютну кількість згадок, а й аномальне зростання відносно базового рівня. Тема, що раптово набирає 1000% зростання обговорень, ймовірно сигналізує про важливу подію. Системи раннього попередження дозволяють журналістам реагувати на події швидше за конкурентів, часто ще до того, як новина з'явиться в традиційних медіа.

Просторово-часовий аналіз трендів показує, як теми поширюються географічно та еволюціонують з часом. Деякі події починаються локально та поступово набувають національного чи глобального значення. Інші вибухають одночасно у різних місцях. Розуміння цих патернів допомагає журналістам оцінити масштаб та важливість події, спрогнозувати її розвиток.

Тематичне моделювання соціальних дискусій виявляє приховані теми в масивах постів. На відміну від простого пошуку за ключовими словами, тематичне моделювання (наприклад, через алгоритм LDA - Latent Dirichlet Allocation) ідентифікує групи слів, що часто зустрічаються разом, навіть якщо немає очевидного об'єднуючого терміну. Це дозволяє виявити нові аспекти обговорення, альтернативні назви для тих самих явищ, зв'язки між різними темами.

Аналіз мережевих зв'язків допомагає розуміти, як інформація поширюється між користувачами, виявляти впливових акторів та картографувати спільноти за інтересами. Теорія графів застосовується для побудови соціальних мереж, де вузли представляють користувачів, а ребра – зв'язки між ними (підписки, згадки, ретвіти, коментарі). Аналіз цих мереж розкриває структуру інформаційних потоків.

Ідентифікація opinion leaders та influencers використовує різні метрики центральності. Degree centrality вимірює кількість прямих зв'язків користувача. Betweenness centrality показує, наскільки користувач є "мостом" між різними групами. Eigenvector centrality враховує не лише кількість зв'язків, а й важливість тих, з ким користувач пов'язаний. PageRank, алгоритм, що лежить в основі пошуку Google, адаптований для виявлення найвпливовіших голосів у соціальних мережах.

Виявлення спільнот та ехо-камер застосовує алгоритми кластеризації для ідентифікації груп користувачів, що інтенсивно взаємодіють між собою, але мало спілкуються з іншими групами. Це важливо для розуміння поляризації громадської думки, виявлення ізольованих інформаційних екосистем, де домінують специфічні наративи без зовнішньої критики.

Аналіз каскадів поширення інформації відстежує, як конкретний фрагмент контенту (новина, меми, чутки) поширюється через мережу. Деякі повідомлення мають вірусну структуру поширення з експоненційним зростанням, інші – лінійну. Розуміння механізмів поширення допомагає передбачити, які повідомлення досягнуть великої аудиторії, та виявити точки втручання для корекції дезінформації.

Технології sentiment analysis визначають емоційну реакцію аудиторії на події, бренди чи персони, що дає журналістам додатковий контекст для їхніх матеріалів. На відміну від традиційних опитувань громадської думки, аналіз соціальних мереж надає даані у реальному часі та з значно більшої вибірки, хоча й з упередженнями, пов'язаними з демографією користувачів соціальних мереж.

Багатокласова класифікація емоцій виходить за межі простого позитивного/негативного поділу та ідентифікує конкретні емоції – радість, гнів, страх, сум, здивування, огиду. Це особливо важливо при аналізі реакцій на кризові події, де розуміння емоційного стану аудиторії критично для адекватного висвітлення.

Аспектний аналіз тональності (aspect-based sentiment analysis) визначає ставлення до конкретних аспектів теми. Наприклад, при обговоренні політичної реформи користувачі можуть позитивно ставитися до мети, але негативно до запропонованих методів. Таке деталізоване розуміння дає набагато глибші інсайти, ніж загальна

тональність.

Темпоральна динаміка емоцій показує, як емоційна реакція змінюється з часом. Після травматичної події початковий шок може змінюватися гнівом, потім сумом, потім мобілізацією для дій. Відстеження цих змін допомагає журналістам адаптувати тон та фокус своїх матеріалів до поточного емоційного стану аудиторії.

Порівняльний аналіз емоцій між різними демографічними групами, регіонами чи політичними орієнтаціями виявляє розбіжності у сприйнятті подій. Те, що викликає ентузіазм в однієї групи, може спричинити обурення в іншої. Розуміння цих відмінностей критично для збалансованого висвітлення контроверсійних тем.

Виявлення координованих кампаній та ботів стало критично важливим інструментом для розслідувань маніпуляцій громадською думкою. Автоматизовані акаунти (боти) та координовані групи людей можуть штучно ампліфікувати певні повідомлення, створювати ілюзію масової підтримки, атакувати опонентів, поширювати дезінформацію.

Виявлення ботів використовує множинні сигнали. Аналіз активності виявляє нелюдські патерни – постинг у строго регулярні інтервали, надзвичайно висока частота публікацій, активність 24/7 без перерв. Аналіз контенту ідентифікує повторювані шаблони, копіювання повідомлень, відсутність оригінального контенту. Аналіз мережі виявляє підозрілі патерни взаємодії – групи акаунтів, що завжди ретвітять один одного, синхронне створення акаунтів, ідентичні списки підписок.

Виявлення координованої неавтентичної поведінки (Coordinated Inauthentic Behavior) фокусується на групах акаунтів, що діють узгоджено для маніпулювання дискусією, незалежно від того, чи є вони ботами чи керованими людьми. Платформи типу Facebook та Twitter публікують звіти про видалені мережі такої поведінки, а журналісти використовують подібні методи для власних розслідувань.

Аналіз часових патернів активності може виявити координацію. Якщо велика група акаунтів починає публікувати на певну тему одночасно, це може свідчити про організовану кампанію. Аналіз лінгвістичних патернів виявляє використання спільних говіркових пунктів (talking points), специфічних фраз, хештегів, що поширюються координовано.

Атрибуція кампаній намагається визначити, хто стоїть за координованими операціями впливу. Це може включати аналіз мовних особливостей (чи є ознаки перекладу, чи використовуються специфічні діалекти), часових зон активності, таргетованих тем та аудиторій, технічної інфраструктури (IP-адреси, використовувані додатки).

Аналіз соціальних мереж також використовується для пошуку очевидців подій. Під час breaking news журналісти моніторять геолокаційні пости з місця події, звертаються до користувачів, що публікували релевантний контент, збирають свідчення та візуальні матеріали. Це значно розширює можливості висвітлення подій у віддалених чи небезпечних локаціях.

Верифікація контенту, створеного користувачами (User-Generated Content, UGC), є критично важливою. Журналісти використовують комбінацію технологічних інструментів та ручної перевірки для підтвердження автентичності фото та відео з соціальних мереж. Зворотний пошук зображень виявляє попереднє використання фото. Аналіз метаданих перевіряє час та місце створення. Крос-референсування з іншими джерелами підтверджує деталі.

Розуміння демографічних характеристик аудиторії різних медіа через аналіз їхніх підписників у соціальних мережах допомагає у стратегічному плануванні контенту. Вік, стать, географічне розташування, інтереси, освітній рівень підписників можна оцінити через аналіз їхніх профілів та поведінки. Це дозволяє адаптувати контент, час публікацій, платформи для максимальної ефективності.

Моніторинг репутації та кризовий менеджмент використовує real-time аналіз соціальних мереж для виявлення потенційних проблем до їх ескалації. Раптове зростання негативних згадок, поява критичних хештегів, вірусне поширення критичного контенту можуть сигналізувати про кризу, що розгортається. Швидке виявлення дозволяє організаціям своєчасно реагувати.

Етичні виклики аналізу соціальних мереж включають питання приватності, згоди, репрезентативності. Користувачі часто не усвідомлюють, що їхні публічні пости можуть бути використані у журналістських матеріалах чи дослідженнях. Демографія соціальних мереж не репрезентує загальне населення, що може призводити до упереджених висновків. Журналістські організації розробляють

етичні керівництва для відповідального використання даних з соціальних мереж.

3.3.5. Великі мовні моделі (LLM).

Великі мовні моделі представляють найновіший та найбільш трансформаційний етап розвитку технологій обробки природної мови, ставши революційним інструментом для алгоритмічної журналістики та медіа-індустрії загалом. LLM – це нейронні мережі, навчені на величезних корпусах текстів (часто сотні мільярдів або навіть трильйони слів), що дозволяє їм розуміти та генерувати людську мову з безпрецедентною якістю та гнучкістю.

Архітектурна революція, що призвела до появи сучасних LLM, розпочалася у 2017 році з публікації статті "Attention is All You Need", що представила архітектуру трансформера. Механізм уваги (attention mechanism) дозволив моделям ефективно обробляти довгі послідовності тексту, фокусуючись на релевантних частинах контексту. Це стало основою для серії все потужніших моделей: GPT (Generative Pre-trained Transformer) від OpenAI, BERT від Google, T5, LLaMA від Meta, Claude від Anthropic та багатьох інших.

Принцип роботи LLM базується на навчанні в два етапи. Попереднє навчання (pre-training) на величезних корпусах текстів дозволяє моделі засвоїти загальні патерни мови, світові знання, логічні зв'язки. Моделі навчаються передбачувати наступне слово в послідовності або відновлювати замасковані слова, що вимагає глибокого розуміння контексту. Тонке налаштування (fine-tuning) адаптує загальну модель для специфічних завдань – генерації новин, резюмування, перекладу, відповідей на запитання.

У журналістиці LLM використовуються для автоматичного створення чернеток статей на основі фактичних даних. На відміну від ранніх систем генерації тексту, що працювали з жорсткими шаблонами, LLM можуть створювати природний, різноманітний текст, адаптований до контексту. Журналіст може надати модель ключовими фактами – результати виборів, фінансові показники, деталі події – і отримати базову статтю, що викладає ці факти зв'язним текстом.

Проте важливо розуміти обмеження. LLM можуть генерувати фактично неточну інформацію (явище, відоме як "галюцинації"), можуть відтворювати упередження з тренувальних даних, можуть

створювати контент, що звучить переконливо, але є неправдивим. Тому згенерований контент завжди вимагає ретельної перевірки журналістом. LLM найкраще працюють як інструменти допомоги, а не заміни журналістів.

Генерація різних варіантів заголовків та лідів дозволяє A/B тестування для оптимізації залученості аудиторії. LLM можуть створити десятки варіантів заголовка для однієї статті, кожен з різним акцентом, тоном, довжиною. Редактори можуть тестувати ці варіанти на невеликій частині аудиторії та визначити, який працює найкраще. Деякі видання автоматизували цей процес, використовуючи LLM для генерації варіантів та машинне навчання для вибору оптимального.

Адаптація контенту для різних аудиторій та платформ використовує можливість LLM змінювати стиль, тон, складність мови. Одна й та ж стаття може бути автоматично адаптована для академічної аудиторії (з технічною термінологією та деталями), масового читача (з простішою мовою та поясненнями), молодіжної аудиторії у соціальних мережах (з неформальним тоном та емоджі), професіоналів галузі (з фокусом на специфічних аспектах).

Створення персоналізованих розсилок та рекомендацій використовує LLM для генерації унікального контенту для кожного передплатника. На основі історії читання, інтересів, демографічних характеристик модель може створити персоналізоване резюме новин, рекомендації статей з поясненнями, чому вони можуть бути цікаві цьому конкретному читачеві.

Аналіз складних документів став значно ефективнішим з LLM. Моделі можуть обробляти довгі документи (сотні сторінок), виділяти ключові тези, відповідати на конкретні запитання про зміст, порівнювати кілька документів, виявляти суперечності. Це особливо корисно при роботі з урядовими звітами, судовими рішеннями, корпоративними документами, науковими публікаціями.

Питання-відповідь (Question Answering) дозволяє журналістам інтерактивно досліджувати документи. Замість читання сотень сторінок в пошуку конкретної інформації, журналіст може поставити пряме питання, і модель знайде релевантні фрагменти та сформулює відповідь. Це прискорює дослідження та дозволяє ефективніше аналізувати великі обсяги документів.

Попередня перевірка фактів через порівняння різних джерел

використовує здатність LLM розуміти та порівнювати твердження. Модель може взяти заяву політика та автоматично знайти релевантні факти у базах даних, порівняти з попередніми заявами цієї ж особи, ідентифікувати потенційні суперечності. Це не замінює ретельний фактчекінг журналістом, але значно прискорює початкову фазу перевірки.

LLM також застосовуються як асистенти для журналістів, допомагаючи у дослідженнях, пропонуючи кути подачі історій та генеруючи ідеї для матеріалів. Журналіст може обговорити з моделлю тему, отримати альтернативні перспективи, ідентифікувати прогалини у висвітленні, згенерувати питання для інтерв'ю. Це функціонує як форма "інтелектуального мозкового штурму", де LLM пропонує ідеї, а журналіст їх оцінює та розвиває.

Генерація питань для інтерв'ю на основі досьє інтерв'юйованого, контексту теми та цілей матеріалу може допомогти журналістам підготуватися ефективніше. Модель може запропонувати як базові, так і провокаційні питання, ідентифікувати потенційні суперечності у заявах особи, запропонувати follow-up питання.

Дослідження історичного контексту автоматизується через здатність LLM синтезувати інформацію з великих обсягів текстів. Журналіст може попросити модель створити резюме попередніх подій з теми, ідентифікувати ключові фігури, пояснити еволюцію ситуації. Це особливо корисно при висвітленні складних, довготривалих історій.

Мультимодальні можливості найновіших LLM дозволяють працювати не лише з текстом, а й з зображеннями. Моделі можуть описувати що зображено на фото, відповідати на питання про візуальний контент, порівнювати зображення, навіть генерувати зображення на основі текстових описів. Це розширює застосування у фотожурналістиці, верифікації візуального контенту, створенні мультимедійних матеріалів.

Інтерактивні новинні застосунки використовують LLM для створення персоналізованих, діалогових досвідів. Читач може "поговорити" з новинною статтею, ставити уточнюючі питання, просити пояснення складних термінів, дізнатися більше про певні аспекти історії. Це робить складні теми більш доступними та залучає аудиторію глибше.

Автоматизація рутинних комунікацій, таких як відповіді на часті запитання читачів, модерація коментарів, обробка підписок, дозволяє журналістам фокусуватися на творчій роботі. Chatbots на базі LLM можуть обробляти стандартні запити, направляти складніші до людей, навіть допомагати читачам навігувати сайтом та знаходити релевантний контент.

Етичні виклики використання LLM у журналістиці численні та серйозні. Прозорість вимагає розкриття, коли контент генерований або допоможений AI. Читачі мають право знати, чи стаття написана людиною, моделлю, чи є колаборацією. Багато видань розробляють політики розкриття використання AI.

Упередження в LLM відображають упередження у тренувальних даних, що часто походять з інтернету та існуючих текстів. Моделі можуть відтворювати гендерні, расові, політичні стереотипи. Журналісти мають бути обізнані про ці упередження та активно їх коригувати. Дослідники працюють над методами деб'ясингу (debiasing) моделей, але це залишається складною проблемою.

Галюцинації – генерація фактично неточної інформації, що звучить переконливо – є особливо небезпечними у журналістиці, де точність критична. LLM можуть вигадувати цитати, статистику, події, джерела. Багатошарова верифікація та фактчекінг обов'язкові для будь-якого контенту, згенерованого або допоможеного AI.

Питання авторства та інтелектуальної власності виникають, коли контент створений моделлю. Чи може AI бути автором? Хто володіє правами на контент, згенерований моделлю, що навчена на чужих текстах? Різні юрисдикції по-різному вирішують ці питання, і правові рамки все ще формуються.

Вплив на зайнятість у медіа є реальною занепокоєністю. Якщо AI може генерувати базові новини, що станеться з журналістами початкового рівня, для яких такі завдання були точкою входу у професію? Оптимісти стверджують, що AI звільнить журналістів від рутини для більш складної, творчої роботи. Песимісти побоюються масових скорочень. Імовірно, реальність буде десь посередині, з трансформацією, а не зникненням професії.

Залежність від технологічних компаній виникає, оскільки найпотужніші LLM контролюються кількома великими корпораціями. Це створює ризики для медіаплюралізму та незалежності. Розвиток

open-source моделей, таких як LLaMA, Mistral, намагається адресувати цю проблему, даючи організаціям можливість запускати власні моделі.

Дезінформація та пропаганда можуть отримати новий потужний інструмент у вигляді LLM. Моделі можуть масово генерувати правдоподібні, але хибні новини, персоналізовані під упередження конкретних груп. Поєднання LLM з технологіями deepfake для створення переконливих відео- та аудіо-фальсифікатів становить особливу небезпеку для довіри до медіа. Боротьба з цим вимагає розвитку не лише технологій детекту, але й медіаграмотності аудиторії.

Людський контроль та редакційна відповідальність залишаються фундаментальними принципами. LLM – це інструмент, а не автономний автор. Кінцеве рішення про формулювання, кут подачі та публікацію має завжди залишатися за журналістом. Нові ролі, як-от «редактор AI» або «промпт-інженер для медіа», можуть стати невід'ємною частиною редакційних команд, відповідаючи за ефективне та етичне використання технологій.

Таким чином, LLM в журналістиці – це синергетичний інструмент підсилення можливостей, а не заміщення. Найефективніше вони працюють у моделі «людина в петлі» (human-in-the-loop), де креативність, критичне мислення, етичний компас та редакційна відповідальність журналіста поєднуються зі швидкістю, масштабом і аналітичною потужністю моделей. Майбутнє успішної журналістики лежить не в конкуренції з AI, а в освоєнні навичок його грамотного застосування для посилення місії – пошуку істини, інформування суспільства та зміцнення демократичних інститутів.

Висновок. Перехід до епохи AI у медіа вимагає формування нової етичної, правової та професійної рамки. Журналістика має активно впливати на розвиток та впровадження цих технологій, захищаючи свої основоположні цінності – точність, справедливість, незалежність та служіння суспільству.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання перевіряють розуміння технологічного стеку:

1. Класичний пайплайн NLG: Опишіть стандартний процес генерації природної мови за Райтером і Дейлом (Reiter & Dale): *Ана-*

ліз контенту (Signal Analysis) → Планування документа (Document Planning) → Мікропланування (Microplanning) → Поверхнева реалізація (Surface Realization). Що відбувається на кожному етапі?

2. Шаблони проти Нейромереж: Чому технологія Template-based NLG (заповнення пропусків) гарантує 100% точність фактів, але звучить "роботизовано", тоді як Neural NLG (на базі трансформерів) пише красиво, але схильна до вигадок?

3. Архітектура Transformer: Як механізм "уваги" (Self-Attention), представлений Google у 2017 році, змінив здатність машин розуміти контекст і призвів до появи ChatGPT?

4. RAG (Retrieval-Augmented Generation): Це найважливіша технологія для сучасної журналістики. Як RAG дозволяє поєднати креативність ChatGPT з надійністю власної бази даних медіа, забороняючи боту вигадувати факти?

5. API (Application Programming Interface): Яку роль відіграють API у постачанні "палива" для алгоритмів? Чому закриття безкоштовного доступу до API Твіттера (X) стало ударом по дата-журналістиці?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання демонструють логіку роботи технологій:

Завдання 1. Логіка Template-based системи (IF/ELSE)

• Завдання: Складіть псевдокод для генерації заголовка про погоду.

• Вхідні дані: Temp (температура), Condition (стан неба).

• Логіка:

- *IF* Temp > 30 *AND* Condition == "Sunny" *THEN PRINT* "Спека повертається: синоптики обіцяють безхмарний день".

- *IF* Temp < 0 *AND* Condition == "Snow" *THEN PRINT* "Зима показує зуби: очікується мороз та снігопади".

- *ELSE* PRINT "Погода на сьогодні: [Temp] градусів, [Condition]".

• Висновок: У чому обмеженість такого підходу, якщо погода просто "нормальна"?

Завдання 2. Промпт-інжиніринг (Робота з LLM)

• Інструмент: ChatGPT або Claude.

• Контекст: Ви хочете, щоб ШІ написав новину, але не брехав.

• Завдання: Розробіть "System Prompt" (системну інструкцію),

який обмежує модель.

- *Приклад поганого промпту*: "Напиши цікаву статтю про вибори".

- *Приклад хорошого промпту*: "Ти — об'єктивний новинний репортер. Використовуй ТІЛЬКИ надані нижче факти. Не додавай власних суджень. Не вигадуй цитат. Якщо інформації не вистачає, напиши 'Дані відсутні'. Тон: формальний, інформаційний".

• Тест: Спробуйте "згодувати" боту набір розрізнених фактів і подивіться, як працює ваша інструкція.

Завдання 3. Порівняння архітектур

• Завдання: Зробіть таблицю порівняння Automated Insights (Wordsmith) та OpenAI (GPT-4).

• Критерії:

- *Передбачуваність результату* (Висока / Низька).

- *Потреба в структурованих даних* (Обов'язкова / Необов'язкова).

- *Ризик галюцинацій* (Нульовий / Високий).

- *Сфера застосування* (Звіти про прибутки / Творчі колонки).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Reiter, E., & Dale, R. *Building Natural Language Generation Systems.* (Класичний підручник з інженерії NLG).

2. Vaswani et al. *Attention Is All You Need.* (Оригінальна стаття, що описує архітектуру Transformer — для тих, хто хоче зрозуміти корінь революції ШІ).

3. Lewis, P. et al. *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.* (Про технологію RAG).

4. Francesco Marconi. *Newsmakers: Artificial Intelligence and the Future of Journalism.* (Огляд інструментів, які використовують WSJ та AP).

4. ГЛОСАРІЙ ТЕРМІНІВ

• NLG (Natural Language Generation): Процес перетворення даних на текст. "Мозок" алгоритмічного журналіста.

• Transformer (Трансформер): Архітектура нейронних мереж, яка дозволяє моделям обробляти цілі речення одночасно (а не по слову), краще розуміючи контекст і зв'язки між словами.

• LLM (Large Language Model): Велика мовна модель (наприклад,

GPT, Claude, Llama), навчена на величезних масивах тексту, здатна генерувати зв'язні, але імовірнісні (не завжди фактичні) відповіді.

• Галюцинація (AI Hallucination): Явище, коли генеративна модель впевнено видає неправдиву інформацію, тому що вона передбачає найбільш вірогідне слово, а не перевіряє факт.

• RAG (Retrieval-Augmented Generation): Гібридний підхід, коли ШІ перед генерацією відповіді спочатку шукає інформацію у надійній базі знань (наприклад, в архіві газети), і використовує знайдене як контекст для відповіді. Це "ліки" від галюцинацій у медіа.

3.4. Роль журналіста в алгоритмічну епоху

Цифрова трансформація медіа не просто змінила інструменти журналістської роботи — вона фундаментально переосмислила саму суть професії. У світі, де алгоритми визначають, яку інформацію побачать мільйони людей, а штучний інтелект може генерувати тексти за секунди, роль журналіста еволюціонує від традиційного «сторожового пса демократії» до багатовимірного фахівця, який поєднує функції автора, куратора, верифікатора та етичного арбітра технологій.

3.4.1.Від автора до куратора та верифікатора.

Традиційна журналістика XX століття будувалася навколо фігури автора — професіонала, який збирав, аналізував і представляв інформацію аудиторії. Журналіст був привілейованим посередником між подією та читачем, його експертиза полягала насамперед у вмінні розповідати історії та формулювати думки. В алгоритмічну епоху цей монополізм зруйновано.

Сьогодні журналіст працює в екосистемі інформаційного надлишку, де кожної секунди публікуються мільйони повідомлень, фото, відео та коментарів. Платформи на кшталт Twitter, TikTok чи Telegram стали конкурентами традиційних медіа у швидкості поширення новин, а користувачі — активними учасниками інформаційного процесу. У цьому контексті журналістська роль розширюється у трьох напрямках.

По-перше, журналіст стає куратором — професіоналом, який фільтрує, контекстуалізує та організовує потоки інформації. Якщо раніше проблемою було знайти інформацію, то тепер — відібрати достовірне з-поміж шуму. Кураторська функція передбачає не лише відбір фактів, а й створення наративів, які допомагають аудиторії зрозуміти складні явища.

По-друге, журналіст перетворюється на верифікатора — спеціаліста з перевірки фактів у добу дезінформації та deep-fake технологій. Фактчекінг, який раніше був імпліцитною частиною журналістської роботи, виокремився в окрему спеціалізацію. Організації як VoxCheck, StopFake, або міжнародні мережі на кшталт International Fact-Checking Network розробили методології та стандарти верифікації, які стали невід'ємною частиною сучасної

журналістики. Верифікація охоплює не лише перевірку текстових заяв, а й аналіз візуального контенту, виявлення маніпуляцій із зображеннями та відео, ідентифікацію ботів і координованих кампаній у соцмережах.

По-третє, журналіст залишається автором, але в оновленому розумінні. Авторство тепер означає не просто написання тексту, а створення унікальної аналітичної чи розслідувацької цінності, яку не може замінити алгоритм. Це глибокі розслідування, які вимагають місяців роботи, інсайдерської інформації та критичного мислення; це репортажі з місць подій, що передають людський вимір історії; це аналітичні есеї, які пропонують нові перспективи розуміння складних проблем.

3.4.2. Нові компетенції: дата-грамотність та промт-інженерія.

Трансформація ролі журналіста вимагає опанування принципово нових навичок, які ще десятиліття тому не входили до журналістських навчальних програм. Дві ключові компетенції визначають професійну спроможність сучасного медіапрацівника: дата-грамотність та промт-інженерія.

Дата-грамотність (data literacy) стала критичною навичкою для журналістів XXI століття. Вона охоплює здатність працювати з великими масивами даних, розуміти статистичні методи, візуалізувати інформацію та критично оцінювати дані як джерело доказів. Журналісти провідних видань регулярно працюють із урядовими базами даних, витоками документів, результатами наукових досліджень, фінансовими звітами компаній.

Дата-журналістика дозволила створити резонансні матеріали, які змінили суспільну дискусію. Розслідування Panama Papers, Pandora Papers ґрунтувалися на аналізі мільйонів документів за допомогою спеціалізованого програмного забезпечення. Проєкти з візуалізації наслідків пандемії COVID-19, моніторингу виборчих процесів, аналізу публічних закупівель — усі вони вимагають від журналістів розуміння методології роботи з даними.

Дата-грамотність не означає, що кожен журналіст має стати програмістом чи статистиком. Йдеться радше про базове розуміння принципів роботи з даними: як збирати та очищати інформацію, які статистичні помилки можуть спотворити висновки, як відрізнити кореляцію від причинно-наслідкового зв'язку, які інструменти

візуалізації найкраще підходять для конкретних типів даних. Ця компетенція також передбачає критичне ставлення до даних — розуміння, що цифри не є нейтральними, а завжди відображають чиїсь рішення щодо збору, класифікації та інтерпретації інформації.

Промт-інженерія — це принципово нова навичка, що виникла з появою великих мовних моделей та генеративного штучного інтелекту. Вона полягає у вмінні формулювати запити (промти) до AI-систем таким чином, щоб отримати максимально релевантні, точні та корисні результати. Це не просто технічна навичка, а форма комунікації з машиною, що вимагає розуміння можливостей і обмежень AI.

Для журналістів промт-інженерія відкриває нові можливості у дослідженні, написанні та редагуванні матеріалів. AI може допомогти в аналізі великих текстових корпусів, генерації ідей для матеріалів, створенні чернеток рутинних текстів (звіти про спортивні події, фінансові підсумки), перекладі інтерв'ю, транскрибуванні аудіозаписів. Проте ефективність цих інструментів залежить від того, наскільки вміло журналіст формулює завдання.

Хороший промт має бути конкретним, контекстуальним і структурованим. Замість загального «напиши про зміни клімату» професійний промт може виглядати як «проаналізуй останні три звіти IPCC та виділи п'ять ключових змін у прогнозах щодо підвищення рівня моря, з акцентом на регіон Південно-Східної Азії, у форматі для нетехнічної аудиторії». Така специфічність вимагає від журналіста чіткого розуміння того, що саме потрібно, і здатності декомпонувати складні завдання на етапи.

Важливо розуміти, що промт-інженерія не замінює журналістську експертизу — вона її підсилює. AI залишається інструментом, який не володіє критичним мисленням, не може самостійно перевіряти факти, не розуміє етичних нюансів і не несе відповідальності за результат. Журналіст, який використовує AI, зобов'язаний верифікувати згенерований контент, доповнювати його власним аналізом та приймати остаточні редакційні рішення.

3.4.3. Етична відповідальність за алгоритмічні рішення.

Впровадження алгоритмів у журналістську практику створює принципово нову конфігурацію етичної відповідальності, яка виходить далеко за межі традиційних журналістських етичних дилем. Якщо в традиційній журналістиці відповідальність за контент

чітко покладалася на конкретних людей — журналістів, редакторів, головних редакторів — то в алгоритмічну епоху ця відповідальність розподіляється між людьми, які створюють алгоритми, тими, хто їх налаштовує, системами, що приймають автоматизовані рішення, та журналістами, які використовують результати роботи цих систем. Така дифузія відповідальності створює небезпечну ситуацію, коли кожен учасник процесу може вважати, що відповідальність лежить на комусь іншому, а в результаті ніхто не несе повної відповідальності за кінцевий продукт.

Перша і найфундаментальніша етична проблема алгоритмічної журналістики полягає у визначенні природи самої відповідальності в умовах, коли рішення приймаються системами, що часто є занадто складними для повного розуміння навіть їхніми творцями. Сучасні системи машинного навчання, особливо ті, що базуються на глибоких нейронних мережах, функціонують як "чорні скриньки" — вони можуть давати точні прогнози та ефективні рекомендації, але логіка, що призводить до конкретних рішень, залишається непрозорою. Коли алгоритм обирає одну новину замість іншої для персоналізованої стрічки конкретного користувача, навіть інженери, що створили цей алгоритм, не завжди можуть точно пояснити, чому було прийняте саме таке рішення. Ця непрозорість ставить під питання саму можливість відповідальності в її традиційному розумінні, де відповідальність передбачає розуміння причин та наслідків своїх дій.

Етична відповідальність за алгоритмічні рішення починається ще на етапі проектування систем, коли приймаються фундаментальні рішення про те, які цілі оптимізації закладати в алгоритм. Здавалося б технічне рішення про те, що саме максимізувати — час перебування на сайті, кількість кліків, глибину залучення, різноманітність споживаного контенту, інформаційну цінність матеріалів — насправді є глибоко етичним вибором, що визначає, яким чином алгоритм впливатиме на інформаційне середовище та поведінку користувачів. Якщо алгоритм оптимізується виключно на короткострокову залученість, він схильний рекомендувати емоційно заряджений, сенсаційний, часто поляризуючий контент, що максимізує миттєві кліки, але може шкодити довгостроковим інтересам користувачів та суспільства. Якщо ж алгоритм намагається балансувати залученість із інформаційною цінністю, різноманітністю перспектив, фактичною точністю — він

може бути менш ефективним у короткостроковій конкуренції за увагу, але більш відповідальним з точки зору журналістської місії.

Розробники алгоритмів несуть відповідальність за вибір цілей оптимізації, але також і за те, які дані використовуються для тренування моделей. Алгоритми машинного навчання "вчаться" на історичних даних, і якщо ці дані містять упередження — системні перекоси в репрезентації певних груп, точок зору, типів контенту — алгоритм не просто відтворить ці упередження, але може їх посилити. Наприклад, якщо алгоритм тренується на історичних даних про те, який контент отримував найбільше уваги, і ці дані відображають історичну недопредставленість жінок-експертів у новинах, алгоритм "навчиться", що статті з цитатами чоловіків-експертів більш "цікаві" для аудиторії, і буде надавати їм вищий пріоритет у рекомендаціях, створюючи замкнене коло дискримінації. Подібні динаміки можуть стосуватися расових, етнічних, соціально-економічних груп, географічних регіонів — будь-яких категорій, де існують історичні нерівності в медіарепрезентації.

Хто приймає рішення про те, що є дезінформацією? Які стандарти доказовості використовуються? Наскільки прозорим є процес модерації? Які механізми апеляції існують для тих, чий контент був помилково видалений? Це не суто технічні питання — вони стосуються фундаментальних цінностей свободи преси, свободи вираження, права на інформацію. Етично відповідальний підхід вимагає не лише ефективних алгоритмів, але й належних процедур, людського нагляду для складних випадків, прозорості критеріїв та рішень, можливостей для оскарження.

Алгоритмічні системи також піднімають питання відповідальності за непередбачувані взаємодії та системні ефекти. Коли множина алгоритмів від різних платформ взаємодіють у складній медіаекосистемі, їхні комбіновані ефекти можуть бути непередбачуваними та потенційно шкідливими. Наприклад, алгоритми новинних платформ, соціальних мереж та пошукових систем можуть ненавмисно створювати ехо-камери, де помилкова інформація циркулює та ампліфікується, поки не стає домінуючим наративом для певних груп користувачів. Кожен окремий алгоритм може функціонувати "правильно" згідно своїх параметрів, але їхня взаємодія створює дисфункціональну інформаційну екосистему.

Відповідальність в такому контексті вимагає мислення на рівні екосистеми, розуміння того, що алгоритми не існують ізольовано, а є частиною складних соціотехнічних систем. Медіаорганізації мають враховувати не лише те, як їхні алгоритми працюють самі по собі, але й як вони взаємодіють з алгоритмами інших платформ, як вони впливають на ширшу інформаційну екосистему. Це може вимагати координації та співпраці між конкуруючими платформами — складне завдання, але необхідне для адресування системних проблем.

Важливим аспектом етичної відповідальності є також питання довгострокового впливу алгоритмічних систем на журналістську професію та якість демократичного дискурсу. Якщо алгоритми оптимізуються виключно на короткострокову залученість, вони можуть систематично недооцінювати складну, нюансовану журналістику на користь простих, емоційних, поляризуючих матеріалів. З часом це може створювати економічні стимули проти якісної журналістики, підриваючи фінансову життєздатність серйозних новинних організацій.

Алгоритми також можуть впливати на те, який тип журналістики виробляється. Якщо журналісти знають, що алгоритми віддають перевагу певним форматам, довжинам, стилям, тональностям, вони можуть, свідомо чи підсвідомо, адаптувати свою роботу до цих алгоритмічних переваг. Це може призвести до гомогенізації журналістичного контенту, втрати різноманітності стилів та підходів, ерозії журналістської автономії. Етична відповідальність вимагає усвідомлення цих динамік та активних зусиль для підтримки різноманітної, якісної журналістики, навіть якщо вона не завжди оптимальна з точки зору алгоритмічних метрик.

Концепція "людини в циклі" (human-in-the-loop) є спробою адресувати багато з цих етичних викликів, забезпечуючи, що критичні рішення завжди включають людське судження. Це може означати, що алгоритми надають рекомендації, але люди приймають остаточні рішення; що автоматизовані системи виявляють потенційні проблеми, але люди верифікують та приймають рішення про дії; що ШІ генерує чернетки, але журналісти редагують, перевіряють факти та приймають відповідальність за публікацію.

Проте навіть концепція людини в циклі має свої обмеження та складнощі. Коли системи обробляють величезні обсяги даних та

приймають мільйони рішень, людський нагляд за кожним рішенням стає неможливим. Люди можуть стати "rubber stamps", автоматично схвалюючи алгоритмічні рекомендації без справжньої критичної оцінки, особливо коли алгоритм має високу точність. Може виникнути надмірна довіра до технології, коли люди приписують алгоритмам більшу об'єктивність чи точність, ніж ті насправді мають. Ефективна людина в циклі вимагає не просто формальної присутності людини в процесі, але й справжнього критичного залучення, компетентності для оцінки алгоритмічних висновків, організаційної культури, що заохочує оспорювання автоматизованих рішень.

Етична відповідальність за алгоритмічні рішення також вимагає механізмів підзвітності — способів виявлення, коли щось йде не так, та виправлення проблем. Це може включати регулярні аудити алгоритмів на предмет упереджень та справедливості, системи для отримання та розгляду скарг користувачів, прозорість про те, як алгоритми приймають рішення, готовність публічно визнавати помилки та вносити корекції. Деякі організації створюють етичні ради або комітети з алгоритмічної відповідальності, що включають різноманітні перспективи та експертизи для оцінки етичних наслідків алгоритмічних систем.

Регуляторне середовище також еволюціонує, створюючи нові форми зовнішньої підзвітності. Законодавство як EU AI Act створює правові рамки для відповідальності за алгоритмічні системи, вимагаючи оцінки ризиків, прозорості, людського нагляду для високоризикових застосувань. Хоча регуляції можуть створювати додаткові зобов'язання для медіаорганізацій, вони також можуть допомогти встановити спільні стандарти та створити рівні умови конкуренції, де етична практика не є конкурентним недоліком.

Останньою, але, можливо, найважливішою складовою етичної відповідальності є визнання того, що алгоритми є не нейтральними інструментами, а втіленням цінностей та вибору їхніх творців. Кожне рішення про дизайн алгоритму — від цілей оптимізації до вибору даних, від архітектури моделі до інтерфейсу користувача — відображає певні цінності та пріоритети. Ці вибори мають реальні наслідки для того, яку інформацію отримують люди, як формується громадська думка, як функціонує демократія.

Етична відповідальність вимагає усвідомлення цих вбудованих

цінностей та свідомого, рефлексивного підходу до їхнього визначення. Які цінності мають керувати алгоритмічною журналістикою? Як балансувати комерційні інтереси з суспільною місією журналістики? Як забезпечити, що технологічний прогрес служить демократичним цінностям свободи преси, інформованого громадянства, плюралізму точок зору? Ці питання не мають простих технічних відповідей — вони вимагають постійного етичного діалогу за участю журналістів, технологів, етиків, представників громадськості.

Етична відповідальність за алгоритмічні рішення в журналістиці, зрештою, є колективною відповідальністю, що вимагає залучення всіх стейкхолдерів — від інженерів та журналістів до менеджерів та регуляторів, від самих медіаорганізацій до широкої громадськості. Це не одноразове завдання, а постійний процес навчання, адаптації, вдосконалення в міру еволюції технологій та нашого розуміння їхнього впливу. Успіх алгоритмічної журналістики залежить не лише від технічної досконалості, але й від здатності вбудувати в ці потужні інструменти фундаментальні етичні принципи журналістики — правдивість, справедливість, підзвітність, служіння суспільному благу.

3.4.4. Людина в циклі (human-in-the-loop).

Концепція «людини в циклі» (human-in-the-loop, HITL) стала ключовим принципом відповідальної автоматизації в журналістиці. Вона передбачає, що критичні рішення не повністю делегуються алгоритмам, а завжди включають людський нагляд, оцінку та втручання на критичних етапах процесу.

У журналістському контексті HITL може реалізовуватися на різних рівнях. На рівні виробництва контенту це означає, що навіть якщо AI генерує чернетку статті або підбирає цитати, журналіст завжди перевіряє факти, редагує текст, приймає остаточне рішення про публікацію. Агентство AP, яке використовує автоматизацію для створення звітів про корпоративні прибутки, встановило правило: кожен згенерований текст має пройти журналістську перевірку перед публікацією, навіть якщо система працює з високою точністю.

На рівні редакційних рішень HITL означає, що алгоритми можуть рекомендувати, які теми заслуговують уваги, які матеріали публікувати на головній сторінці, але остаточне слово залишається за редактором. The New York Times використовує алгоритми для

аналізу читацької поведінки та прогнозування популярності тем, але редакційна політика, баланс між різними рубриками, вибір головних матеріалів дня визначаються людьми, які враховують не лише метрики залученості, а й суспільну важливість, етичні міркування, редакційну місію.

На рівні модерації HITL особливо критична. Автоматизовані системи можуть попередньо фільтрувати коментарі, виявляючи ненависницькі висловлювання, спам чи дезінформацію, але складні чи неоднозначні випадки мають розглядатися модераторами-людьми. Facebook, YouTube, Twitter після численних скандалів з неправомірним видаленням контенту посилили участь людей у процесах модерації, хоча й досі отримують критику за недостатню прозорість.

Модель HITL також передбачає механізми зворотного зв'язку та навчання. Коли журналісти регулярно переглядають та коригують рішення алгоритмів, ці корекції можуть використовуватися для покращення систем. Це створює цикл безперервного вдосконалення, де людська експертиза не просто компенсує недоліки AI, а активно формує його розвиток.

Критично важливо розуміти різницю між людиною в циклі (human-in-the-loop), людиною над циклом (human-on-the-loop) та людиною поза циклом (human-out-of-the-loop). У першому випадку людина активно бере участь у кожному критичному рішенні; у другому — наглядає за процесом і втручається лише при виявленні проблем; у третьому — алгоритм діє повністю автономно. Для журналістики, де на кону стоять довіра суспільства, репутація, а часом і безпека людей, модель «людина поза циклом» неприйнятна для критичних рішень.

Водночас HITL не означає відмову від автоматизації. Навпаки, правильно реалізований підхід дозволяє максимізувати переваги AI (швидкість, масштабованість, аналіз великих даних) при мінімізації ризиків (помилки, упередженості, етичні прорахунки). Журналісти звільняються від рутинних завдань і можуть концентруватися на тому, що дійсно вимагає людської креативності, критичного мислення та етичного судження.

Еволюція ролі журналіста в алгоритмічну епоху — це не історія занепаду професії, а історія її трансформації. Від єдиноосібного автора

журналіст перетворюється на багатофункціонального професіонала: куратора в океані інформації, верифікатора в умовах епідемії дезінформації, критичного арбітра технологій, етичного навігатора алгоритмічних рішень. Нові компетенції — дата-грамотність та промт-інженерія — доповнюють традиційні журналістські навички, створюючи профіль фахівця, який однаково комфортно працює з текстами, даними та кодом.

Проте технологічна майстерність без етичної рефлексії може перетворити журналістику на технократичне підприємство, де ефективність переважає над цінностями. Саме тому принцип «людини в циклі» є не просто технічною практикою, а фундаментальним етичним вибором: у світі, де все більше рішень делегується алгоритмам, журналістика наполягає на збереженні людського судження, відповідальності та критичного мислення в серці інформаційного процесу. Це не страх перед технологіями, а усвідомлення того, що деякі функції — етична оцінка, контекстуальне розуміння, емпатія до наслідків — залишаються незамінно людськими.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допомагають переосмислити професію:

1. Від Gatekeeper до Sense-maker: Традиційно журналіст був "гейткіпером" (вирішував, що потрапить у газету). В епоху безкінечного контенту ця роль померла. Чому зараз ключовою стає роль "сенсмейкера" (того, хто пояснює контекст і значення подій)?

2. Human-in-the-loop (Людина в контурі): Чому навіть найдосконаліші системи (Bloomberg, AP) вимагають фінальної перевірки людиною? Які типи помилок ("зсув контексту", етичні порушення) алгоритм не може помітити сам?

3. Algorithmic Accountability Reporting: Це новий жанр розслідувань. Чому журналіст майбутнього повинен вміти не лише брати інтерв'ю у чиновників, а й "допитувати" алгоритми (наприклад, перевіряти, чи не дискримінує алгоритм мерії певні райони міста)?

4. Емпатія як конкурентна перевага: Алгоритм може написати звіт про кількість жертв ДТП. Але чому він не може написати репортаж про горе родини? У чому полягає незамінність "емоційного інтелекту" в історіях?

5. Нові компетенції: Як змінюється набір навичок (hard skills)?

Чому "обчислювальне мислення" (computational thinking) та базове розуміння статистики стають важливішими за швидкісний набір тексту?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання моделюють реальні ситуації співпраці людини та машини:

Завдання 1. Рольова гра "Редактор бота". Ситуація: ШІ згенерував чернетку новини про зростання злочинності в місті на 20%. Проблема: ШІ не розуміє контексту (насправді, просто поліція почала краще реєструвати заяви через онлайн-портал, реальна злочинність не зросла). Завдання: Виступіть у ролі редактора. Знайдіть "сліпу зону" алгоритму. Допишіть два абзаци, які змінюють тональність з "паніки" на "аналітику". Сформулюйте правило (prompt), щоб бот не робив цієї помилки в майбутньому.

Завдання 2. Пітчинг теми: "Що не зможе ШІ?" Завдання: Запропонуйте три теми для репортажів, які неможливо автоматизувати в найближчі 5 років. Критерії. Потреба у фізичній присутності (запахи, тактильні відчуття). Робота з джерелами, які бояться говорити (побудова довіри). Складні моральні дилеми. Обґрунтування: Чому ChatGPT провалив би це завдання?

Завдання 3. Аудит алгоритму (Algorithmic Auditing). Сценарій: Читачі скаржаться, що стрічка новин вашого сайту показує лише кримінал. Завдання: Розробіть план перевірки редакційного алгоритму. Які метрики ви перевірите? (Клікбейтність заголовків? Час публікації?). Яке "редакційне завдання" ви поставите програмістам, щоб виправити перекос? (Наприклад: "Впровадити коефіцієнт різноманіття тем").

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Diakopoulos, Nicholas. *Automating the News.* (Розділ про відповідальність журналістів за аудит алгоритмів).

2. ProPublica. *Machine Bias.* (Легендарне розслідування про те, як журналісти викрили расизм у судовому алгоритмі — приклад нової ролі журналіста).

3. Marconi, Francesco. *Newsmakers*. (Про гібридні ньюзруми, де люди керують машинами).

4. Lewis, Seth C. *The Rise of the "Algorithmically Infused" Journalist.*

4. ГЛОСАРІЙ ТЕРМІНІВ

• Human-in-the-loop (HITL): Модель взаємодії, де людина бере участь у циклі роботи алгоритму (навчає, перевіряє результат, коригує помилки). Гарантія якості та етичності в медіа.

• Algorithmic Accountability Reporting (Звітність про підзвітність алгоритмів): Форма розслідувальної журналістики, об'єктом якої є не люди-чиновники, а алгоритми (владні, корпоративні), що впливають на життя суспільства.

• Сенсмейкінг (Sense-making): Процес надання сенсу розрізненим даним. Якщо алгоритм дає "інформацію" (хто, де, коли), то журналіст дає "знання" (чому це важливо і що буде далі).

• Промпт-інжиніринг (Prompt Engineering): Професійна навичка формування точних інструкцій для генеративних моделей ШІ для отримання якісного результату. Нова форма "письменництва".

• Обчислювальне мислення (Computational Thinking): Вміння формулювати проблеми так, щоб їх міг вирішити комп'ютер (декомпозиція задачі, пошук патернів, алгоритмізація).

Висновок до розділу 3

Розділ присвячено комплексному аналізу трансформаційного впливу штучного інтелекту та алгоритмічних систем на сучасну журналістську практику. Головним висновком є констатація того, що алгоритмічна журналістика не являє собою заміщення професійного журналіста, а стає його потужним синергетичним інструментом, який фундаментально переформатовує всі етапи роботи — від первинної генерації ідей та дослідження до фінальної дистрибуції контенту та аналізу аудиторної взаємодії.

Алгоритмізація спричинила зміну парадигми інформаційного споживання — від редакційного монологу до персоналізованого діалогу. Механізми на кшталт персоналізованої стрічки (The Feed), побудованої на графі інтересів (Interest Graph), перенесли акцент з універсальної «головної сторінки» на індивідуальний потік контенту. Це змусило медіа-індустрію переорієнтуватися на боротьбу за увагу до кожної окремої публікації, підвищивши значення миттєвого залучення, формату та адаптивності. Однак паралельно виникли серйозні суспільні ризики, такі як формування герметизованих

інформаційних бульбашок, ерозія спільного суспільного простору для дискусії та поглиблення поляризації через надмірну персоналізацію.

У практичній площині великі мовні моделі (LLM) інтегруються в редакційні процеси як інтелектуальні ко-пілоти, виконуючи функції асистента для мозкового штурму, глибокого дослідження тем, підготовки структурованих інтерв'ю, попередньої верифікації фактів та автоматизації рутинних операцій. Це відкриває шлях до створення інноваційних форматів, серед яких інтерактивні статті у форматі діалогу, персоналізовані аудіодайджести та динамічні дата-історії, що трансформують роль читача з пасивного споживача в активного учасника інформаційного процесу.

Прогрес у цій галузі виокремив низку критичних етичних і професійних викликів. Центральними серед них є необхідність абсолютної прозорості щодо використання штучного інтелекту, боротьба з успадкованими алгоритмами упередженнями та небезпечними явищами «галюцинацій», коли модель генерує правдоподібну, але хибну інформацію. Незмінним залишається принцип «людина в петлі» (human-in-the-loop), де остаточне редакційне рішення, відповідальність за точність і етичність матеріалу залишаються за журналістом. Окрему загрозу становить потенціал AI для масштабної генерації дезінформації та пропагандистичного контенту, що потребує розвитку як складних технологій детекту, так і всезагальної медіаграмотності аудиторії.

У перспективі майбутнє журналістики формується як гібридна модель. Професія не зникає, але зазнає кардинальної трансформації, вимагаючи від фахівців нових компетенцій: промпт-інжинірингу, роботи з даними, алгоритмічної грамотності та критичного управління AI-інструментами. Успішними будуть ті журналісти та редакції, які навчаться використовувати силу алгоритмів для посилення власних унікальних сильних сторін — глибокої аналітики, творчої креативності, редакційної незалежності та людяного співчуття, які неможливо повністю алгоритмізувати.

Таким чином, алгоритмічна журналістика являє собою нову парадигму, в якій технологічна ефективність та швидкість повинні знаходитися в постійній та свідомій рівновазі з фундаментальними журналістськими цінностями: прагненням до істини, об'єктивністю, точності, справедливістю та служінням суспільним інтересам.

ЧАСТИНА ІІ: ТЕХНОЛОГІЇ ТА ІНСТРУМЕНТИ

Розділ 4: Рекомендаційні системи
4.1. Принципи роботи

Рекомендаційні системи стали невидимою, але всепроникною силою, що формує наш цифровий досвід. Кожного разу, коли Netflix пропонує фільм, Spotify — пісню, YouTube — відео, Amazon — товар, а Facebook — публікацію у стрічці, за цими рекомендаціями стоять складні алгоритми, що аналізують мільярди взаємодій для передбачення наших уподобань. У медіасфері рекомендаційні системи визначають, які новини побачать читачі, які теми стануть вірусними, які голоси будуть почуті, а які залишаться в тіні. Розуміння принципів їхньої роботи є критичним не лише для технічних спеціалістів, а й для журналістів, редакторів, дослідників медіа — усіх, хто прагне усвідомити механізми формування сучасного інформаційного простору.

На концептуальному рівні завдання рекомендаційної системи виглядає обманливо просто: передбачити, наскільки конкретному користувачу сподобається конкретний елемент контенту, який він ще не бачив. Проте за цією простотою ховається фундаментальна епістемологічна проблема — як вивести майбутні переваги з минулої поведінки, як балансувати між тим, що людині вже подобається (експлуатація), і тим, що може їй сподобатися (дослідження), як враховувати контекст, настрій, зміни смаків у часі. Протягом останніх трьох десятиліть було розроблено численні підходи до розв'язання цієї проблеми, від елегантних математичних методів до складних архітектур глибинного навчання. Розглянемо еволюцію цих підходів, їхні сильні сторони, обмеження та застосування в медіаконтексті.

4.1.1. Колаборативна фільтрація: мудрість натовпу.

Колаборативна фільтрація (collaborative filtering) базується на інтуїтивній ідеї: якщо два користувачі мали схожі смаки в минулому, вони ймовірно матимуть схожі уподобання і в майбутньому. Це принцип «мудрості натовпу», застосований до передбачення переваг. Замість того, щоб аналізувати властивості самого контенту, колаборативна фільтрація виявляє патерни у колективній поведінці

користувачів.

User-based колаборативна фільтрація (користувацько-орієнтована) працює наступним чином. Спочатку система будує матрицю взаємодій користувачів із контентом — наприклад, рейтинги фільмів, де рядки відповідають користувачам, стовпці — фільмам, а значення — оцінкам від 1 до 5. Ця матриця зазвичай дуже розріджена (sparse), адже жоден користувач не переглянув усіх фільмів. Для конкретного користувача А система шукає інших користувачів, які мають найбільш схожий профіль оцінок — так званих «сусідів». Схожість зазвичай обчислюється за допомогою метрик на кшталт косинусної відстані або кореляції Пірсона між векторами оцінок.

Припустімо, користувачі А і Б обоє високо оцінили фільми «Inception», «The Matrix», «Interstellar», а низько — «Transformers». Система визнає їх схожими. Якщо користувач Б також високо оцінив «Blade Runner 2049», який користувач А ще не бачив, система рекомендує цей фільм користувачу А. Формально, передбачена оцінка обчислюється як зважене середнє оцінок схожих користувачів, де вагами є коефіцієнти схожості.

User-based підхід інтуїтивно зрозумілий і працює добре для невеликих спільнот із стабільними смаками. Проте він стикається з серйозними проблемами масштабованості. На платформі з мільйонами користувачів обчислення схожості для кожної пари стає computationally prohibitive. Крім того, цей підхід вразливий до проблеми «холодного старту» для нових користувачів, які ще не надали достатньо оцінок для знаходження схожих «сусідів».

Item-based колаборативна фільтрація (елементно-орієнтована) розв'язує частину цих проблем, змінюючи фокус з схожості користувачів на схожість елементів контенту. Замість пошуку схожих користувачів система визначає схожі статті, відео чи товари на основі того, як їх оцінювали користувачі. Два фільми вважаються схожими, якщо їх оцінювали схоже одні й ті самі користувачі.

Логіка така: якщо користувач позитивно взаємодіяв із статтею про зміни клімату, система шукає інші статті, які позитивно оцінювали ті самі люди, що й цю статтю про клімат. Можливо, це виявляться матеріали про відновлювальну енергетику або екологічну політику. Ключова перевага item-based підходу — стабільність схожості між елементами. Якщо профілі користувачів можуть швидко змінюватися

(нові оцінки, зміна смаків), схожість між елементами контенту є більш постійною, що дозволяє попередньо обчислити її й зберегти, значно прискорюючи рекомендації в реальному часі.

Amazon популяризував item-based підхід у своїй системі «Customers who bought this item also bought...». У медіаконтексті це перетворюється на «Читачі цієї статті також читали...» — механіку, яку використовують практично всі новинні сайти. Item-based колаборативна фільтрація стала промисловим стандартом у 2000-х роках саме завдяки поєднанню ефективності, масштабованості та якості рекомендацій.

Проте обидва підходи колаборативної фільтрації мають спільні фундаментальні обмеження. По-перше, проблема розрідженості даних: якщо користувачі оцінили лише крихітну частку доступного контенту, знайти надійні схожості стає складно. По-друге, проблема холодного старту: для нових користувачів або нового контенту немає історії взаємодій, що робить рекомендації неможливими. По-третє, проблема популярності: колаборативні методи схильні рекомендувати вже популярний контент, створюючи ефект багатіють багаті (rich-get-richer), де нішеві або нові матеріали важко пробиваються до аудиторії.

4.1.2. Контент-базовані рекомендації: аналіз властивостей.

Якщо колаборативна фільтрація ігнорує властивості контенту, фокусуючись виключно на патернах користувацької поведінки, то контент-базовані (content-based) рекомендаційні системи діють протилежним чином — аналізують характеристики самого контенту для знаходження схожих елементів.

Логіка контент-базованого підходу будується на припущенні: якщо користувачу сподобалася стаття з певними характеристиками (тема, автор, стиль, ключові слова), йому ймовірно сподобаються інші матеріали з подібними властивостями. Система створює профіль уподобань користувача на основі характеристик контенту, з яким він взаємодіяв, і шукає новий контент, що відповідає цьому профілю.

Для текстового контенту, що домінує в новинних медіа, контент-базовані системи використовують методи обробки природної мови (Natural Language Processing, NLP). Найпростіший підхід — представлення документів через модель «мішок слів» (bag-of-words) з ваговою схемою TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF надає більшу вагу словам, що часто зустрічаються

в конкретному документі, але рідко в усьому корпусі, таким чином виділяючи характерні терміни.

Припустімо, користувач регулярно читає статті про квантові комп'ютери. Система аналізує ці статті, виділяє ключові терміни («квантова заплутаність», «кубіт», «суперпозиція», «квантова перевага») та їхні ваги TF-IDF, формуючи профіль інтересів. Коли з'являється нова стаття, система обчислює косинусну схожість між вектором TF-IDF цієї статті та профілем користувача. Статті з високою схожістю рекомендуються.

Для мультимедійного контенту використовуються інші методи виділення ознак. Для зображень це можуть бути колірні гістограми, текстурні дескриптори або, в сучасних системах, ембедінги від згорткових нейронних мереж. Для відео — аналіз ключових кадрів, аудіодоріжки, метаданих. Для подкастів — транскрипція й текстовий аналіз плюс акустичні особливості.

Контент-базований підхід має кілька важливих переваг. По-перше, він не потребує даних про інших користувачів, що розв'язує проблему холодного старту для нових користувачів — достатньо знати їхні початкові переваги. По-друге, він може рекомендувати нішевий контент, який ще не має багато взаємодій. По-третє, рекомендації пояснювані: систему можна спитати, чому вона рекомендує статтю, і отримати відповідь на кшталт «тому що ви часто читаєте матеріали про штучний інтелект, а ця стаття містить аналіз нових AI-моделей».

Проте контент-базовані системи мають і серйозні недоліки. Головний з них — схильність до надмірної спеціалізації (overspecialization) або створення «інформаційних бульбашок». Якщо система рекомендує виключно контент, схожий на те, що користувач уже споживав, вона обмежує exposure до різноманітних перспектив і несподіваних відкриттів. Читач, який цікавиться виключно прогресивною політикою, отримуватиме лише прогресивні погляди; той, хто читає про технології, ніколи не побачить культурні чи соціальні теми.

Крім того, контент-базовані системи вимагають значних зусиль для виділення якісних ознак контенту. Для тексту це може бути відносно автоматизовано, але для інших медіа, особливо мультимодальних, це складніше. Також ці системи можуть пропускати неочевидні зв'язки — дві статті можуть бути дуже різними за ключовими словами,

але звертатися до схожої аудиторії через тональність, складність викладу або культурні референції.

4.1.3. Матрична факторизація: латентні фактори уподобань.

Матрична факторизація (matrix factorization) представляє математично елегантний підхід до колаборативної фільтрації, який став домінуючим у 2000-х роках, особливо після успіху на конкурсі Netflix Prize. Ключова ідея полягає в тому, що матрицю взаємодій користувачів із контентом можна апроксимувати як добуток двох матриць меншої розмірності, що представляють латентні (приховані) фактори.

Уявімо матрицю R розміром m×n, де m — кількість користувачів, n — кількість елементів контенту, а R[i,j] — рейтинг користувача i для елемента j. Більшість значень R невідомі (користувачі не оцінили більшість контенту). Матрична факторизація розкладає R на добуток двох матриць: R ≈ P × Q^T, де P — матриця розміром m×k (користувачі × латентні фактори), Q — матриця n×k (елементи × латентні фактори), а k — кількість латентних факторів, зазвичай набагато менша за m і n (наприклад, 20-200).

Що означають ці латентні фактори? Для фільмів це можуть бути неявні жанрові характеристики, які не обов'язково відповідають традиційним жанрам. Один фактор може представляти спектр від «легкого розважального» до «серйозного арт-хаусного», інший — від «екшн-орієнтованого» до «діалог-орієнтованого», третій — міру наявності романтичної лінії. Для новинного контенту латентні фактори можуть відображати політичні нахили, рівень складності викладу, локальність/глобальність теми, емоційну тональність.

Кожен користувач представляється вектором у k-вимірному просторі латентних факторів, що відображає його переваги відносно цих характеристик. Кожен елемент контенту також представляється вектором у тому самому просторі, що відображає його властивості. Передбачений рейтинг обчислюється як скалярний добуток цих векторів — чим ближче користувач і елемент у латентному просторі, тим вищий очікуваний рейтинг.

Навчання моделі матричної факторизації зводиться до оптимізаційної задачі: знайти матриці P і Q, які мінімізують різницю між відомими рейтингами та їхніми апроксимаціями. Найпопулярніший метод — мінімізація функції втрат методом

градієнтного спуску або альтернованих найменших квадратів (Alternating Least Squares, ALS). Функція втрат зазвичай включає regularization для запобігання перенавчанню.

Одна з найуспішніших варіацій — Singular Value Decomposition (SVD) та її адаптації для розріджених матриць, як-от SVD++. Netflix Prize, конкурс на покращення рекомендаційної системи Netflix, який тривав з 2006 до 2009 року з призом у мільйон доларів, був виграний командою, що використала ансамблі методів, де ключовим компонентом була саме матрична факторизація. Переможна модель покращила точність передбачення рейтингів на 10% порівняно з існуючою системою Netflix.

Матрична факторизація має кілька важливих переваг. По-перше, вона ефективно працює з розрідженими даними, знаходячи латентну структуру навіть коли більшість оцінок відсутні. По-друге, вона масштабується до великих датасетів значно краще, ніж класична user-based або item-based фільтрація. По-третє, латентні фактори можуть виявити неочевидні зв'язки між користувачами та контентом, які не видимі при прямому порівнянні.

Однак і цей підхід не позбавлений обмежень. Матрична факторизація залишається методом колаборативної фільтрації, тому страждає від проблеми холодного старту для нових користувачів і нового контенту. Крім того, стандартна матрична факторизація не враховує часових динамік — вона припускає статичні переваги, тоді як насправді смаки людей еволюціонують. Для подолання цього було розроблено темпоральні варіації, що моделюють зміну латентних факторів у часі.

4.1.4. Deep learning підходи: нейронні мережі для рекомендацій.

З розвитком глибинного навчання (deep learning) у 2010-х роках рекомендаційні системи пережили чергову революцію. Нейронні мережі привнесли здатність автоматично навчатися складним, нелінійним взаємозв'язкам між користувачами, контентом і контекстом, значно перевершуючи можливості традиційних методів.

Одна з перших успішних архітектур — нейронна колаборативна фільтрація (Neural Collaborative Filtering, NCF). Замість обчислення скалярного добутку між векторами користувача та елемента, як у матричній факторизації, NCF використовує багатошарову нейронну мережу для вивчення складної функції взаємодії. Користувач і

елемент кодуються у векторні ембедінги (навчені представлення), які подаються на вхід нейронної мережі. Мережа через послідовність прихованих шарів із нелінійними активаціями вивчає, як комбінувати ці ембедінги для передбачення ймовірності позитивної взаємодії.

Рекурентні нейронні мережі (RNN) та їхні вдосконалені версії — LSTM (Long Short-Term Memory) і GRU (Gated Recurrent Unit) — дозволили моделювати послідовності взаємодій користувачів. Замість того, щоб розглядати всі взаємодії як незалежні, ці моделі враховують порядок і часовий контекст. Наприклад, якщо користувач протягом тижня читав статті про початок війни, потім про міжнародну реакцію, потім про гуманітарну кризу, RNN може виявити цю логічну послідовність та передбачити, що наступною логічною темою може бути аналіз воєнної стратегії або матеріали про біженців.

Attention механізми та трансформери, які революціонізували обробку природної мови (NLP), знайшли застосування й у рекомендаціях. Self-attention дозволяє моделі визначати, які попередні взаємодії користувача найбільш релевантні для передбачення наступної. Архітектури на кшталт BERT4Rec адаптували bidirectional encoder representations з BERT для моделювання історії користувача, досягаючи вражаючих результатів у передбаченні наступного елемента, який зацікавить користувача.

Автокодувальники (autoencoders) використовуються для вивчення compressed representations взаємодій користувачів. Collaborative Denoising Autoencoder (CDAE) навчається реконструювати вектор взаємодій користувача з його зашумленої версії, примушуючи мережу вивчати robust латентні представлення. Variational Autoencoders (VAE) додають імовірнісний компонент, дозволяючи генерувати різноманітні рекомендації та оцінювати невизначеність передбачень.

Графові нейронні мережі (Graph Neural Networks, GNN) розглядають рекомендації як задачу на графі, де користувачі та елементи є вузлами, а взаємодії — ребрами. GNN можуть ефективно поширювати інформацію через граф, виявляючи складні залежності високого порядку. Наприклад, Pinterest використовує PinSage — GNN-архітектуру для рекомендацій на графі з мільярдами вузлів. У медіаконтексті граф може включати не лише користувачів і статті, а й автора, теми, згадані сутності, джерела, створюючи багате семантичне представлення.

Мультимодальне глибинне навчання дозволяє інтегрувати різні типи даних. Для новинних медіа це означає одночасне використання тексту статті, супровідних зображень, відео, метаданих, соціального контексту. Згорткові мережі (CNN) обробляють зображення, рекурентні або трансформери — текст, а пізні шари об'єднують ці модальності для комплексного розуміння контенту. YouTube використовує схожий підхід, комбінуючи візуальні, аудіальні та текстові ознаки відео з поведінковими сигналами користувачів.

Deep learning підходи продемонстрували значні покращення в метриках точності рекомендацій, особливо на великих датасетах. Вони можуть автоматично виявляти складні патерни без hand-crafted feature engineering, адаптуватися до змінюваних контекстів, інтегрувати різнорідні джерела даних. Проте ці переваги мають свою ціну.

По-перше, нейронні мережі вимагають величезних обсягів даних для ефективного навчання — luxury, яким володіють лише великі платформи. По-друге, вони computationally expensive, потребують потужних GPU для тренування та inference. По-третє, вони є «чорними скриньками» — інтерпретувати, чому саме нейронна мережа рекомендувала конкретний контент, набагато складніше, ніж для традиційних методів. Ця непрозорість створює етичні проблеми, особливо в новинному контексті, де користувачі мають право розуміти, чому їм показують саме цю інформацію.

4.1.5. Контекстуальні бандити та reinforcement learning.

Всі попередні підходи в основному фокусувалися на передбаченні статичних рейтингів або ймовірностей взаємодії. Проте реальні рекомендаційні системи діють у динамічному середовищі, де кожна рекомендація впливає на поведінку користувача, а система має балансувати між експлуатацією (використання відомих уподобань) та дослідженням (тестування нових гіпотез). Саме ці виклики адресують підходи на основі контекстуальних бандитів (contextual bandits) та навчання з підкріпленням (reinforcement learning, RL).

Проблема багаторуких бандитів (multi-armed bandit problem) — класична задача теорії прийняття рішень. Уявіть казино з кількома ігровими автоматами (бандитами), кожен з яких має невідому ймовірність виграшу. Ваше завдання — максимізувати загальний виграш, вибираючи автомати для гри. Дилема: чи грати на автоматі,

який показав найкращі результати (exploitation), чи спробувати інші автомати, щоб виявити можливо кращі опції (exploration)?

У контексті рекомендацій кожен елемент контенту — це «рука бандита», користувач — це гравець, а «виграш» — позитивна взаємодія (клік, дочитування, лайк). Класичні алгоритми як epsilon-greedy (з ймовірністю epsilon вибирати випадковий варіант, інакше — найкращий відомий), Upper Confidence Bound (UCB, вибір варіантів з урахуванням невизначеності оцінок) або Thompson Sampling (байєсівський підхід із sampling з posterior розподілів) визначають стратегію балансу між експлуатацією та дослідженням.

Контекстуальні бандити розширюють цю концепцію, додаючи контекст. Ймовірність «виграшу» залежить не лише від вибраної «руки», а й від контексту — характеристик користувача, часу дня, пристрою, попередніх взаємодій. Формально, в кожному раунді t система отримує контекстний вектор x_t, вибирає дію a_t (яку рекомендацію показати) та отримує винагороду r_t. Мета — навчити політику, що максимізує кумулятивну винагороду.

LinUCB (Linear Upper Confidence Bound) — популярний алгоритм контекстуальних бандитів, що моделює очікувану винагороду як лінійну функцію контексту та підтримує exploration через confidence bounds. Його використовують для персоналізації новинних стрічок, де контекст включає профіль користувача, характеристики статті, часові фактори. Кожна показана рекомендація та наступна реакція користувача оновлюють модель у реальному часі, дозволяючи швидко адаптуватися до змін.

Reinforcement Learning (навчання з підкріпленням) іде далі, розглядаючи рекомендації як послідовний процес прийняття рішень. На відміну від контекстуальних бандитів, де кожна взаємодія незалежна, RL моделює довгострокові наслідки рекомендацій. Рекомендація, показана зараз, впливає не лише на поточну реакцію користувача, а й на його майбутній стан (настрій, рівень engagement, довгострокову лояльність до платформи).

Рекомендації моделюються як Марковський процес прийняття рішень (Markov Decision Process, MDP): стан s_t описує поточний контекст і історію користувача; дія a_t — рекомендований контент; винагорода r_t — immediate feedback (клік, час перегляду); наступний стан s_{t+1} — оновлений контекст після взаємодії. Мета — навчити

політику π(a|s), що максимізує очікувану кумулятивну винагороду (часто з дисконтуванням майбутніх винагород).

Алгоритми як Deep Q-Networks (DQN), Policy Gradient методи (REINFORCE, Actor-Critic), Proximal Policy Optimization (PPO) застосовуються для навчання таких політик. YouTube у своїй рекомендаційній системі використовує RL для оптимізації довгострокового engagement, а не лише immediate clicks. Модель навчається балансувати між показом вірусного clickbait (високий короткостроковий engagement) і якісного контенту (вищий довгостроковий retention та задоволення користувачів).

Критично важлива концепція в RL для рекомендацій — функція винагороди (reward function). Наївне визначення винагороди як кліку призводить до clickbait. Більш витончені підходи включають час взаємодії, завершеність споживання (чи дочитав статтю, доглянув відео), diverse signals (лайки, репости, коментарі), навіть довгострокові метрики як retention користувача через тиждень після рекомендації.

Офлайн RL (offline reinforcement learning або batch RL) — важлива адаптація для практичних застосувань. Класичні RL алгоритми потребують активної взаємодії з середовищем — показувати рекомендації, отримувати feedback, оновлювати модель. Проте на реальній платформі безпосереднє експериментування ризиковане — погані рекомендації відштовхують користувачів. Офлайн RL навчається на історичних даних взаємодій, намагаючись витягти оптимальну політику без активного втручання. Це вимагає техніки як importance sampling, щоб коригувати зміщення в даних, зібраних попередньою політикою.

Підходи на основі бандитів і RL особливо цінні в новинному контексті, де переваги користувачів швидко еволюціонують у відповідь на події. Під час breaking news ефективна система має швидко виявити зростаючий інтерес до теми та адаптувати рекомендації. Exploration дозволяє пропонувати різноманітні перспективи, не замикаючи користувачів у інформаційних бульбашках. Довгострокова оптимізація через RL може балансувати між immediate engagement і побудовою довіри та лояльності через якісну журналістику.

Проте ці підходи мають і серйозні виклики. По-перше, визначення правильної функції винагороди — нетривіальна задача, особливо

коли бажані метрики (інформованість громадян, плюралізм думок) важко квантифікувати. По-друге, RL моделі можуть виявляти unexpected exploitation стратегії — наприклад, показувати емоційно маніпулятивний контент, який максимізує short-term engagement на шкоду довгостроковому благополуччю користувачів. По-третє, ці системи вимагають великих обсягів даних і обчислювальних ресурсів, роблячи їх доступними переважно для великих платформ.

Еволюція рекомендаційних систем відображає загальний тренд у машинному навчанні — від простих евристик до складних, data-intensive, адаптивних алгоритмів. Кожен підхід має свої сильні сторони, обмеження та оптимальні сфери застосування. Колаборативна фільтрація ефективна, коли є багато даних про взаємодії користувачів. Контент-базовані методи працюють для нішевого контенту та нових користувачів. Матрична факторизація забезпечує елегантну математичну основу для виявлення латентних структур. Deep learning розкриває складні нелінійні залежності на великих датасетах. Контекстуальні бандити та RL дозволяють адаптивно балансувати exploration-exploitation та оптимізувати довгострокові метрики.

На практиці найефективніші рекомендаційні системи комбінують множину підходів. Netflix, Spotify, YouTube, Amazon використовують гібридні системи, що інтегрують колаборативну фільтрацію, контент-базовані методи, deep learning моделі та RL компоненти. Така композиція дозволяє компенсувати слабкості окремих методів та адаптуватися до різноманітних сценаріїв.

Для медіаіндустрії розуміння цих технічних підходів є фундаментальним. Рекомендаційні системи визначають, яка інформація досягає громадськості, які теми домінують у публічній сфері, які голоси будуть почуті. Вони можуть як сприяти інформаційному плюралізму та serendipitous discovery, так і створювати filter bubbles та поляризацію. Технічні рішення — вибір архітектури, функції винагороди, балансу exploration-exploitation — мають глибокі соціальні наслідки. Саме тому журналісти, редактори, дослідники медіа повинні не лише критикувати ці системи ззовні, а й розуміти їх внутрішню логіку, щоб вести інформовану дискусію про майбутнє інформаційного простору.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю)

Ці питання розмежовують головні підходи до прогнозування смаків:

1. Дві філософії: У чому фундаментальна різниця між Content-Based Filtering (фільтрація за змістом: "Ти любиш фантастику, ось тобі ще фантастика") та Collaborative Filtering (колаборативна фільтрація: "Ти схожий на Петра, а Петро купив це, тому тобі це теж сподобається")?

2. Проблема "Холодного старту" (Cold Start Problem): Чому рекомендаційні системи безсилі, коли на платформу приходить новий користувач або додається новий товар? Які існують стратегії вирішення цієї проблеми (наприклад, онбординг-опитування)?

3. Явне vs Неявне фідбек (Explicit vs Implicit Feedback): Чому Netflix відмовився від системи "5 зірок" (явний фідбек) на користь "палець вгору/вниз", але насправді найбільше покладається на час перегляду (неявний фідбек)?

4. Матриця "Користувач-Об'єкт" (User-Item Matrix): Як виглядає світ очима алгоритму? (Уявіть гігантську таблицю, де 99% клітинок порожні — *sparsity*, і задача алгоритму — заповнити ці пропуски прогнозованими оцінками).

5. Парадокс бульбашки: Як надто точна робота рекомендаційної системи призводить до перенавчання (Overfitting), коли користувач перестає бачити щось нове і нудьгує?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання допоможуть зрозуміти математику на пальцях:

Завдання 1. "Паперовий алгоритм" (Колаборативна фільтрація)

• Дано: Матриця оцінок (від 1 до 5) для трьох користувачів і трьох фільмів.

- *Анна:* "Титанік" (5), "Аватар" (4), "Термінатор" (?).
- *Богдан:* "Титанік" (5), "Аватар" (5), "Термінатор" (2).
- *Віктор:* "Титанік" (1), "Аватар" (?), "Термінатор" (5).

• Завдання: Спрогнозуйте оцінку Анни для "Термінатора".

• Логіка: Анна схожа на Богдана (їм обом подобаються перші два фільми). Богдан поставив "Термінатору" низьку оцінку (2). Отже, алгоритм має спрогнозувати Анні низьку оцінку.

Завдання 2. Тегування для Content-Based системи

• Об'єкт: Оберіть статтю або відео.

• Завдання: Розбийте цей контент на "вектор ознак" (Feature Vector), як це робить Pandora для музики (Music Genome Project).

- *Жанр:* (0 або 1).
- *Настрій:* (Сумний/Веселий - шкала 0..1).
- *Темп:* (Швидкий/Повільний).
- *Ключові слова:* (список).

• Мета: Зрозуміти, що для Content-Based алгоритму пісня — це просто набір цифр.

Завдання 3. Стратегія подолання Cold Start

• Ситуація: Ви запускаєте новий додаток для читання книг. У вас нуль даних про поведінку користувачів.

• Завдання: Запропонуйте 3 методи, як рекомендувати книги в перший день:

- *Метод 1:* (Популярне / Бестселери).
- *Метод 2:* (Контекстний — книги про осінь, якщо надворі жовтень).
- *Метод 3:* (Соціальний — підключити Facebook і показати, що читають друзі).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Кріс Андерсон. *Довгий хвіст (The Long Tail)*. (Фундаментальна книга про те, чому рекомендаційні системи економічно вигідні — вони продають нішеві товари).

2. Netflix Tech Blog. *Netflix Recommendations: Beyond the 5 stars*. (Опис реальних принципів роботи індустріального гіганта).

3. GroupLens Research. (Академічні статті від розробників MovieLens — піонерів колаборативної фільтрації).

4. Jannach, D. et al. *Recommender Systems: An Introduction.* (Базовий підручник).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• Колаборативна фільтрація (Collaborative Filtering): Метод прогнозування, що базується на зборі вподобань від багатьох користувачів. Основна ідея: "Людям, схожим на вас, сподобалось оце".

• Фільтрація на основі змісту (Content-Based Filtering): Метод, що рекомендує об'єкти, схожі на ті, що користувач обирав раніше, базу-

ючись на характеристиках самого об'єкта (жанр, автор, теги).

• Проблема холодного старту (Cold Start Problem): Труднощі з наданням точних рекомендацій для нових користувачів (немає історії дій) або нових товарів (ніхто ще не оцінив).

• Розрідженість даних (Data Sparsity): Ситуація, типова для великих платформ, коли більшість користувачів не взаємодіяли з більшістю товарів (матриця заповнена переважно нулями), що ускладнює пошук закономірностей.

• Гібридна система (Hybrid System): Поєднання різних методів (наприклад, колаборативного та контентного) для компенсації недоліків кожного з них. Використовується Netflix, Spotify, Amazon.

4.2. Метрики ефективності

Оцінка якості рекомендаційних систем — це набагато складніша задача, ніж може здатися на перший погляд. На відміну від класичних завдань машинного навчання, де існує чітка ground truth (правильна відповідь), у рекомендаціях межа між «хорошою» та «поганою» рекомендацією часто розмита. Більше того, різні стейкхолдери мають різні, іноді конфліктуючі критерії успіху. Користувачі хочуть релевантних, але і несподіваних рекомендацій. Платформи прагнують максимізувати engagement та монетизацію. Контент-творці потребують справедливого розподілу уваги. У медіаконтексті додаються суспільні цілі: інформаційна різноманітність, протидія дезінформації, демократичний плюралізм думок.

Ця багатовимірність якості рекомендацій відображається в розмаїтті метрик, які використовуються для їх оцінки. Жодна метрика не захоплює всю складність проблеми, тому практики комбінують множину вимірювань, часто балансуючи між конфліктуючими цілями. Розглянемо ключові метрики та компроміси між ними, звертаючи особливу увагу на їхнє значення для новинних медіа.

4.2.1. Точність та повнота: фундаментальні метрики релевантності.

Точність (precision) та повнота (recall) — це базові метрики з інформаційного пошуку, адаптовані для рекомендаційних систем. Вони оцінюють, наскільки точно система ідентифікує релевантний контент для користувача.

Точність відповідає на запитання: з усіх рекомендацій, які ми показали користувачу, яка частка виявилася релевантною? Формально, precision = (кількість релевантних рекомендацій) / (загальна кількість рекомендацій). Якщо система запропонувала 10 статей, і користувач знайшов 7 з них цікавими, точність становить 0.7 або 70%. Висока точність означає, що користувач не витрачає час на нерелевантні матеріали — критично важливо в умовах інформаційного перевантаження.

Повнота відповідає на іншу запитання: з усього релевантного контенту, який існує для користувача, яку частку ми йому показали? Формально, recall = (кількість релевантних рекомендацій) / (загальна кількість релевантного контенту). Якщо в каталозі є 20 статей, які користувач знайшов би цікавими, а система рекомендувала 7 з

них, повнота становить 0.35 або 35%. Висока повнота означає, що користувач не пропускає важливий для нього контент.

Ці дві метрики перебувають у trade-off відносинах. Легко досягти 100% повноти, рекомендуючи абсолютно весь контент — але точність буде жахливою. Так само легко досягти високої точності, рекомендуючи лише один елемент, в якому система найбільш впевнена — але повнота буде мізерною. Балансуючою метрикою є F-міра (F-measure або F1-score), що обчислюється як гармонійне середнє точності та повноти: $F1 = 2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$.

У практичних рекомендаційних системах часто використовують варіації цих метрик, адаптовані до формату рекомендацій:

Precision@K та Recall@K оцінюють точність і повноту серед топ-K рекомендацій. Наприклад, precision@5 питає: скільки з топ-5 рекомендацій виявилися релевантними? Це відображає реальність — користувачі зазвичай взаємодіють лише з кількома першими рекомендаціями, тому якість саме цих позицій критична.

Mean Average Precision (MAP) враховує не лише присутність релевантних елементів у топі, а й їхні позиції. Релевантний елемент на першій позиції цінніший, ніж на десятій. MAP обчислює average precision для кожного користувача (середня точність на кожній релевантній позиції) та усереднює по всіх користувачах.

Normalized Discounted Cumulative Gain (NDCG) іде ще далі, дозволяючи gradual relevance — релевантність не бінарна (релевантний/нерелевантний), а має градації (дуже релевантний, помірно релевантний, трохи релевантний). NDCG також дисконтує цінність релевантності залежно від позиції — елемент на позиції 10 дисконтується сильніше, ніж на позиції 2. Це відображає поведінку користувачів, які приділяють більше уваги вищим позиціям.

Проте всі ці метрики мають фундаментальну проблему: як визначити, що релевантно? У контрольованих експериментах можна використовувати експліцитні рейтинги (користувачі оцінювали контент від 1 до 5), але в реальних системах доводиться покладатися на імпліцитні сигнали — кліки, час взаємодії, дочитування статті, лайки. Кожен з цих сигналів має свої проблеми: клік не обов'язково означає задоволення (користувач міг розчаруватися після кліку); відсутність кліку не означає нерелевантність (користувач міг не

побачити рекомендацію або вже знати інформацію).

У новинному контексті проблема релевантності ще глибша. Чи має система рекомендувати лише те, що користувачу «сподобається», чи також те, що йому «потрібно знати»? Стаття про складну політичну реформу може бути менш «клікабельною», ніж розважальний контент, але більш важливою для інформованості громадянина. Традиційні метрики точності та повноти не захоплюють цю нормативну різницю.

4.2.2. Різноманітність рекомендацій: вихід за межі filter bubbles.

Різноманітність (diversity) рекомендацій стала критичною метрикою в епоху зростаючого занепокоєння інформаційними бульбашками та ехо-камерами. Навіть якщо всі рекомендації релевантні (висока точність), вони можуть бути надмірно схожими, обмежуючи exposure користувача до різних перспектив, тем, стилів.

Розрізняють кілька типів різноманітності:

Внутрішньосписочна різноманітність (intra-list diversity, ILD) вимірює, наскільки несхожі елементи в одному списку рекомендацій. Якщо система рекомендує користувачеві п'ять статей, і всі вони про одну й ту саму тему, написані в однаковому стилі, різноманітність низька. Навіть якщо користувач цікавиться цією темою, така гомогенність обмежує його перспективу.

ILD можна обчислювати різними способами залежно від того, що вважати «схожістю». Для текстового контенту це може бути косинусна схожість між TF-IDF векторами або ембедінгами від мовних моделей. Для новинних статей можна враховувати тематичну категорію, географічну прив'язку, політичний нахил джерела, формат (новина, аналітика, думка). Формула ILD часто виглядає як: ILD = (1 / |L|²) × Σ_{i,j∈L, i≠j} dist(i, j), де L — список рекомендацій, dist(i,j) — відстань між елементами i та j.

Охоплення тематик (topic coverage або categorical diversity) оцінює, скільки різних категорій представлено в рекомендаціях. Ідеальна новинна стрічка може включати політику, економіку, культуру, науку, спорт — навіть якщо користувач частіше читає лише політику. Це забезпечує serendipitous exposure до тем, про які користувач не знав, що вони йому цікаві.

Різноманітність джерел (source diversity) особливо важлива в медіаконтексті. Показувати статті з різних видань, що представляють різні редакційні лінії та політичні перспективи, допомагає

користувачам формувати збалансоване розуміння подій. Метрика може обчислюватися як кількість унікальних джерел у топ-N рекомендаціях або через індекси концентрації на кшталт Herfindahl-Hirschman Index, де нижчі значення вказують на більш рівномірний розподіл.

Темпоральна різноманітність враховує час публікації. Балансування між свіжим контентом (breaking news, trending topics) і evergreen матеріалами (поглиблені аналізи, historical context) забезпечує комплексний інформаційний досвід.

Проте максимізація різноманітності має свої компроміси. Надмірна різноманітність може зменшити релевантність — якщо користувач глибоко цікавиться квантовою фізикою, рекомендація статті про моду може бути різноманітною, але нерелевантною. Критичне питання: скільки різноманітності достатньо? Дослідження показують, що користувачі цінують певну міру різноманітності (вона робить рекомендації менш передбачуваними та цікавішими), але надмірна різноманітність сприймається як шум.

У новинному контексті різноманітність має також нормативний вимір. Демократичні суспільства потребують громадян, які exposure до різних точок зору та здатні розуміти позиції, відмінні від власних. Рекомендаційні системи, що створюють filter bubbles, потенційно підривають цю передумову демократії. Тому деякі дослідники та регулятори наполягають на включенні різноманітності не просто як опційної метрики, а як обов'язкового constraint у дизайні рекомендаційних систем новинних платформ.

4.2.3. Новизна та серендипність: несподівані відкриття.

Новизна (novelty) та серендипність (serendipity) — метрики, що оцінюють здатність системи пропонувати несподівані, але цінні рекомендації. Вони виходять за межі простої релевантності, питаючи: чи дізнається користувач щось нове? Чи приємно він здивований?

Новизна вимірює, наскільки рекомендований контент є новим для користувача. Найпростіший підхід — перевіряти, чи взаємодіяв користувач раніше з рекомендованим контентом або схожим. Більш витончені метрики враховують популярність: рекомендація малопопулярної, нішевої статті, про яку користувач навряд чи дізнався б іншим шляхом, має вищу новизну, ніж рекомендація вірусного trending топіку, який користувач побачив би в будь-якому

випадку.

Формально, новизна може обчислюватися через несхожість між рекомендаціями та попередньою історією користувача: novelty = (1 / |R|) × Σ_{i∈R} min_{j∈H} dist(i, j), де R — множина рекомендацій,H— історія споживання користувача. Альтернативний підхід — novelty через непопулярність: novelty(i) = -log₂(popularity(i)), де популярність вимірюється як частка користувачів, що взаємодіяла з елементом.

Серендипність іде далі, додаючи елемент несподіванки та цінності. Рекомендація є серендипною, якщо вона: (1) несподівана — користувач не очікував би її з огляду на свою історію; (2) релевантна — користувач все ж таки цінує її після споживання. Це складніше виміряти, оскільки потребує оцінки як несподіванки, так і ex-post задоволення.

Одна з формул серендипності: serendipity(i) = relevance(i) × unexpectedness(i), де relevance — фактична релевантність (чи сподобалося користувачеві), unexpectedness — міра несподіванки (наскільки рекомендація відрізняється від очікуваного профілю користувача). Альтернативно, серендипність можна обчислювати як релевантні елементи, яких не рекомендувала б проста baseline-модель (наприклад, popular items або typical collaborative filtering).

У медіаіндустрії серендипність має особливу цінність. Вона дозволяє користувачам виходити за межі своїх звичних інформаційних зон, відкривати нові теми, автора, перспективи. Читач, який зазвичай споживає лише політичні новини, може серендипно відкрити захоплюючу науково-популярну статтю. Це збагачує користувацький досвід та протидіє звуженню інформаційного горизонту.

Проте серендипність також має компроміси. Надмірна несподіванка може сприйматися як нерелевантність — користувач захоче релевантного контенту, а система пропонує щось зовсім інше. Балансування серендипності та релевантності — делікатна задача. Дослідження показують, що користувачі цінують periodic серендипні рекомендації (1-2 з 10), але не хочуть, щоб весь список був несподіваним.

Технічно, досягнення серендипності часто пов'язане з exploration у контекстуальних бандитах або RL — система свідомо інколи відхиляється від найбільш очікуваної рекомендації, щоб

дослідити альтернативи. Це не лише приносить користувачам цінні несподіванки, а й допомагає системі навчатися, виявляючи нові preference patterns.

4.2.4. Охоплення каталогу.

Охоплення каталогу (catalog coverage або aggregate diversity) вимірює, яка частка доступного контенту фактично рекомендується користувачам. Навіть якщо індивідуальні рекомендації різноманітні, система може глобально страждати від проблеми popularity bias — концентрації на невеликій частці популярного контенту, залишаючи більшість каталогу невидимою.

Формально, catalog coverage = (кількість унікальних елементів, рекомендованих хоча б одному користувачу) / (загальна кількість елементів у каталозі). Якщо платформа має 10,000 статей, але рекомендаційна система показує лише 1,000 з них, охоплення становить 10%. Це означає, що 90% контенту фактично невидимі для користувачів.

Альтернативна метрика — Gini coefficient розподілу рекомендацій. Коефіцієнт Джині вимірює нерівність розподілу — від 0 (ідеальна рівність, усі елементи рекомендуються однаково часто) до 1 (максимальна нерівність, лише один елемент отримує всі рекомендації). Низький Gini означає більш рівномірний розподіл уваги між контентом.

Чому охоплення каталогу важливо? По-перше, з точки зору контент-творців — журналістів, авторів, видань. Якщо рекомендаційна система концентрує всю увагу на кількох популярних матеріалах, інші не отримують аудиторії, незалежно від їхньої якості. Це створює winner-takes-all динаміку, де багато якісної журналістики залишається непоміченою.

По-друге, з точки зору користувачів. Низьке охоплення означає, що всі бачать приблизно той самий контент, що обмежує колективну різноманітність exposure. Навіть якщо індивідуальні списки рекомендацій здаються різноманітними, на агрегатному рівні платформа може створювати homogenized інформаційний простір.

По-третє, з точки зору платформи. Підтримка різноманітної екосистеми контент-творців забезпечує довгострокову життєздатність платформи. Якщо лише небагато авторів отримують увагу, інші втратять мотивацію створювати контент, збіднюючи каталог.

Проте максимізація охоплення каталогу також має trade-offs. Рекомендувати непопулярний контент заради рівномірності може зменшити релевантність — деякі матеріали непопулярні, бо вони справді низької якості або нішеві. Балансування між meritocracy (рекомендувати найкраще) та egalitarianism (давати шанс усьому контенту) — складне етичне та технічне питання.

Деякі платформи експериментують з quota-based підходами: гарантувати, що певний відсоток рекомендацій іде на менш популярний контент. Інші використовують calibrated popularity bias — коригують популярність як фактор ранжування, щоб дещо знизити перевагу вже популярних елементів. Exploration у RL також природно сприяє охопленню каталогу, оскільки система періодично тестує менш відомі опції.

4.2.5. Задоволеність користувачів: кінцева метрика успіху.

Задоволеність користувачів (user satisfaction) — це найбільш важлива, але і найскладніша для вимірювання метрика. Усі попередні метрики — точність, різноманітність, новизна, охоплення — є проксі (непрямими індикаторами) того, що насправді хочемо досягти: щоб користувачі були задоволені своїм досвідом з рекомендаційною системою.

Виклик полягає в тому, що задоволеність — суб'єктивний, багатовимірний конструкт, який важко безпосередньо виміряти. Різні підходи до її оцінки включають:

Експліцитні опитування — найпрямолінійний метод. Після взаємодії з рекомендаціями запитати користувача: «Наскільки ви задоволені цими рекомендаціями?» (шкала від 1 до 5). Або більш детальні опитувальники, що оцінюють різні аспекти: релевантність, різноманітність, корисність, довіру до системи. Проблема — низькі response rates та можлива упередженість (користувачі, що відповідають, можуть не бути репрезентативними).

Імпліцитні поведінкові сигнали — вимірювати задоволеність через спостережувану поведінку. Різні метрики включають:

• Click-through rate (CTR) — частка рекомендацій, на які користувачі клікнули. Проста, але обманлива метрика. Високий CTR не обов'язково означає задоволення — clickbait заголовки можуть мати високий CTR при низькому ex-post задоволенні.

• Dwell time або engagement time — скільки часу користувач про-

вів з рекомендованим контентом. Для статей це може бути час читання, для відео — час перегляду. Довша взаємодія часто (але не завжди) свідчить про більшу цінність контенту.

• Completion rate — частка користувачів, що дочитали статтю до кінця або доглянули відео повністю. Висока completion rate свідчить, що контент виправдав очікування, створені заголовком та початком.

• Post-click actions — що користувач робить після споживання контенту: лайкає, коментує, репостить, додає в закладки, шукає більше від того самого автора? Ці дії сигналізують глибше engagement і задоволення.

• Return rate або retention — чи повертається користувач на платформу регулярно? Довгострокова задоволеність проявляється в sustained engagement, а не лише в immediate реакціях.

A/B тестування — золотий стандарт для оцінки рекомендаційних систем на практиці. Частина користувачів отримує рекомендації від нової версії алгоритму (група B), інша частина — від базової версії (група A, контроль). Порівнюючи метрики між групами (CTR, dwell time, retention тощо), можна оцінити, чи нова система покращує користувацький досвід.

Проте A/B тестування має обмеження. По-перше, воно вимірює short-term метрики, тоді як довгострокові ефекти (наприклад, зростання довіри до платформи або, навпаки, інформаційна поляризація) виявляються лише з часом. По-друге, оптимізація під метрики A/B тестів може призвести до Goodhart's law — коли метрика стає ціллю, вона перестає бути хорошою метрикою. Системи, оптимізовані виключно під CTR або watch time, можуть максимізувати ці показники способами, що шкодять довгостроковому благополуччю користувачів (через adictive patterns, clickbait, emotional manipulation).

Довгострокова задоволеність вимагає longitudinal вимірювань. Це може включати:

• Lifetime value (LTV) — скільки цінності користувач приносить платформі протягом усього періоду використання. Для підписних моделей це пряма грошова метрика; для ad-supported — очікувані доходи від реклами.

• Churn rate — частка користувачів, що припиняють користуватися платформою. Низький churn свідчить про довгострокову задоволеність.

• Net Promoter Score (NPS) — «Наскільки ймовірно, що ви порекомендуєте цю платформу другу?» Показник лояльності та задоволеності.

• Trust metrics — особливо важливо для новинних медіа. Чи довіряють користувачі рекомендованому контенту? Чи вважають вони платформу надійним джерелом інформації?

У медіаконтексті виникає фундаментальне питання: чию задоволеність оптимізувати? Immediate gratification (миттєве задоволення від clickable, емоційно-заряджених матеріалів) може конфліктувати з informed citizenship (довгостроковим благом бути інформованим громадянином). Користувач може «задовольнятися» стрічкою розважальних новин, уникаючи складних, але важливих тем. Чи повинна рекомендаційна система потурати цим перевагам чи challenge їх?

Деякі дослідники розрізняють hedonic satisfaction (задоволення від приємного досвіду) та eudaimonic satisfaction (задоволення від значущого, ціннісно-орієнтованого досвіду). Стрічка котиків може приносити hedonic задоволення, але поглиблений репортаж про соціальну несправедливість може приносити eudaimonic задоволення навіть якщо він емоційно важкий. Збалансована рекомендаційна система, особливо в новинах, повинна враховувати обидва типи.

Метрики ефективності рекомендаційних систем утворюють складний, багатовимірний простір, де різні виміри часто конфліктують. Висока точність може зменшувати різноманітність. Максимізація серендипності може знижувати immediate релевантність. Оптимізація short-term engagement може шкодити long-term задоволеності. Фокус на популярному контенті покращує average задоволеність, але погіршує охоплення каталогу.

Не існує єдиної «правильної» комбінації метрик — вона залежить від цілей платформи, характеристик контенту, очікувань користувачів, соціального контексту. Для e-commerce точність може бути paramount; для новинних медіа різноманітність і довіра критичніші; для розважальних платформ серендипність і novelty додають цінність.

Ключовий висновок для медіаіндустрії: технічні метрики — CTR, dwell time, engagement — є лише частиною картини. Повна оцінка рекомендаційних систем повинна включати соціальні, епістемічні, демократичні виміри: чи сприяє система інформованості громадян?

Чи підтримує вона плюралізм думок? Чи забезпечує справедливий розподіл уваги між голосами? Ці питання не можуть бути повністю редуковані до цифрових метрик, але повинні направляти дизайн і оцінку алгоритмів, що формують наш інформаційний простір.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання розмежовують технічні та бізнес-показники:

1. Офлайн vs Онлайн метрики: У чому різниця між оцінкою алгоритму на історичних даних (Offline Evaluation, наприклад, точність передбачення рейтингу) та перевіркою на живих користувачах (A/B Testing)? Чому офлайн-успіх не гарантує зростання прибутку?

2. Проблема "Точності" (Accuracy Paradox): Чому алгоритм, який рекомендує лише найпопулярніші блокбастери, може мати високу математичну точність, але бути абсолютно марним для бізнесу та користувача? (Підказка: відсутність новизни).

3. Precision@k (Точність у топі): Чому в YouTube чи Google News нам байдуже на загальну точність, а важлива лише метрика Precision@Top5? (Ефект "першого екрана").

4. Serendipity (Несподіваність) та Novelty (Новизна): Чим відрізняються ці поняття? (Новизна — це те, чого я не бачив. Серендипність — це те, чого я не бачив, не шукав, але воно мені несподівано сподобалось). Чому це важливо для утримання інтересу?

5. Diversity (Різноманіття): Якщо користувач подивився "Гаррі Поттер 1", найточнішою рекомендацією буде "Гаррі Поттер 2". Але чому з точки зору метрики *Diversity* це погана рекомендація?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають оцінювати якість видачі:

Завдання 1. Розрахунок Precision@5

• Ситуація: Алгоритм видав користувачеві 5 статей.

1. Стаття А (Клікнув, лайкнув) — *Релевантна.*
2. Стаття Б (Пропустив) — *Нерелевантна.*
3. Стаття В (Клікнув, вийшов через 5 сек) — *Нерелевантна.*
4. Стаття Г (Пропустив) — *Нерелевантна.*
5. Стаття Д (Клікнув, дочитав) — *Релевантна.*

• Завдання: Розрахуйте Precision@5.

• Рішення: 2 релевантних / 5 загальних = 40% (або 0.4). Чи до-

статньо цього для утримання уваги?

Завдання 2. Вибір метрики під бізнес-ціль

• Завдання: Оберіть *одну* головну метрику (North Star Metric) для трьох різних сервісів і обґрунтуйте вибір:

- *Spotify (Музика):* (Варіанти: Кількість прослуховувань чи Час прослуховування? Відповідь: Час. Бо пісні короткі).

- *Amazon (Магазин)*: (Варіанти: Кліки чи Конверсія в покупку? Відповідь: Конверсія).

- *Tinder (Дейтинг):* (Варіанти: Кількість свайпів чи Кількість "Match"? Відповідь: Match).

Завдання 3. Аналіз "Довгого хвоста" (Coverage)

• Ситуація: У каталозі онлайн-кінотеатру 10 000 фільмів. Алгоритм рекомендує лише 500 топ-фільмів усім підряд.

• Аналіз:

- Якою буде метрика *Coverage* (Покриття)? (5%).

- Чому це погано для економіки платформи? (Ми платимо за ліцензії на 9500 фільмів, які ніхто не бачить. Алгоритм має "розкопувати" архів).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Shani, G., & Gunawardana, A. *Evaluating Recommender Systems.* (Академічний стандарт з методології оцінювання).

2. Netflix Tech Blog. *A/B Testing and Make Decisions*. (Як індустрія насправді перевіряє гіпотези).

3. Herlocker, J. L. et al. *Evaluating collaborative filtering recommender systems.* (Класична праця про метрики).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• A/B Тестування (A/B Testing): Метод маркетингового дослідження, при якому користувачів випадковим чином ділять на групи: Група А бачить старий алгоритм, Група Б — новий. Перемагає той, що дає кращі бізнес-показники (гроші, час).

• Precision (Точність): Частка релевантних (цікавих) об'єктів серед усіх рекомендованих. Відповідає на питання: "Наскільки ми влучили?".

• Recall (Повнота): Частка знайдених цікавих об'єктів серед усіх цікавих об'єктів, що існують у базі. Відповідає на питання: "Скільки

всього корисного ми пропустили?".

• Серендипність (Serendipity): Метрика, що оцінює здатність системи надавати рекомендації, які є одночасно привабливими та несподіваними (неочевидними).

• Покриття (Coverage): Відсоток товарів або контенту з каталогу, які система здатна порекомендувати хоча б комусь. Низьке покриття означає, що система ігнорує "довгий хвіст".

4.3. Кейси провідних платформ

4.3.1. Netflix: як визначається, що подивитись далі.

Кейс Netflix є фундаментальною ілюстрацією того, як сучасні цифрові платформи перевизначають взаємини між контентом і споживачем. Це не просто приклад використання алгоритмів рекомендацій, а повноцінна система управління увагою, побудована на глибокому машинному навчанні та психології поведінки. Їх головна бізнес-мета сформульована кристально чітко: мінімізувати час прийняття рішення (час, коли користувач пошукує контент) і максимізувати час активного перегляду. Дослідження Netflix показало, що користувач, який не знаходить щось цікаве протягом 60-90 секунд, з високою ймовірністю залишить платформу. Уся їхня алгоритмічна архітектура підпорядкована подоланню цього критичного часового бар'єру.

Першим ключовим інноваційним кроком став радикальний відхід від традиційної демографічної моделі сегментації аудиторії. Для Netflix не має первинного значення, чи є користувач чоловіком 35 років із Києва, якщо його реальна поведінка говорить про інші смаки. Замість цього платформа аналізує мільярди точок даних від понад 250 мільйонів підписників, групуючи їх у більш ніж 2000 так званих «спільнот смаків» (Taste Communities). Це динамічні, поведінкові кластери, що об'єднують людей на основі подібності їхніх переглядів і оцінок, незалежно від географії, віку чи статі. Якщо ваша історія переглядів включає темне наукове фентезі на кшталт «Чорного дзеркала» та документальні фільми про природу, алгоритм помістить вас у один кластер з користувачем з іншої частини світу, чиї вподобання збігаються на 80%. Отже, якщо той користувач вподобає новий анімаційний серіал, система з високою ймовірністю запропонує його вам. Ця гібридна фільтрація — поєднання колаборативної (на основі схожості користувачів) та контентної (на основі атрибутів фільмів) — дозволяє з неймовірною точністю передбачати смаки та відкривати контент, про існування якого користувач міг би не здогадуватися.

Другим рівнем персоналізації, що став візитівкою Netflix, є алгоритмічний підбір обкладинок (Artwork Personalization). Це тонкий інструмент мікроманіпуляції увагою. Кожен фільм або

серіал має десятки заздалегідь підготовлених варіантів постерів, і нейромережа в реальному часі обирає той один, який має максимальну ймовірність кліку саме для вас. Наприклад, для фільму «Кримінальне чтиво» система може показати постер з Умою Турман тому, хто часто дивиться фільми з нею або жіночі драми; фанатам Джона Траволти або класичних бойовиків буде запропоновано кадр з ним і зброєю; а тому, хто схильний до комедій, можуть показати образ Семюеля Л. Джексона в більш гумористичному контексті. Мета полягає не в тому, щоб відобразити суть фільму, а в тому, щоб створити миттєвий, підсвідомий зв'язок між незнайомим продуктом і вашими вже сформованими вподобаннями, тим самим усуваючи останні сумніви перед кліком.

Третім критичним стратегічним рішенням став перехід у 2017 році від звичної п'ятибальної системи оцінок до бінарної схеми «Вподобай»/«Не вподобай» (Thumbs Up/Down). Цей крок був заснований на глибокому розумінні психологічного розриву між публічною оцінкою та приватною поведінкою. Люди часто ставили високі оцінки «важливому» або «якісному» контенту (наприклад, документальним фільмам про війну), але для щоденного перегляду обирали легкий розважальний контент (комедії Адама Сендлера), якому ставили середні бали. Зірочки вимірювали абстрактне «схвалення якості», тоді як бінарна система дала відповідь на єдине важливе для алгоритму питання: «Чи хоче ця людина отримати більше такого контенту?». Це зробило зворотний зв'язок набагато чеснішим інструментом прогнозування майбутньої поведінки, що, за заявами Netflix, покращило точність рекомендацій на 200%. Разом із пасивним збором даних (процент перегляду, час зупинки, перегляд до кінця) це формує надзвичайно точний профіль залучення.

Крім впливу на дистрибуцію, дані Netflix активно формують і сам процес створення контенту. Платформа аналізує успіх певних акторів у конкретних жанрах, комбінації тем, що показують високу завершеність перегляду, та популярність окремих сюжетних арок. Це дозволяє приймати рішення про продюсування нових оригінальних серіалів, а також оптимізувати сценарії та монтаж вже існуючих (наприклад, скорочуючи сцени, які часто перемотують). Таким чином, алгоритм трансформується з посередника у співтворця, стираючи межу між аналітикою споживання і мистецьким процесом.

Однак ця ефективна система породжує серйозні етичні питання та соціальні ризики. Надмірна персоналізація загрожує створенням герметичних «фільтрувальних бульбашок», де користувач потрапляє в нескінченний потік контенту, що лише підтверджує його існуючі смаки, обмежуючи культурний кругозір. Оптимізація під максимальне залучення може використовувати психологічні механізми, близькі до формування залежності, змушуючи користувача пасивно скролити та переглядати. Крім того, data-driven підхід до продюсування несе ризик гомогенізації культурного продукту, коли креативні, ризиковані ідеї поступаються місцем безпечним, алгоритмічно перевіреним формулам.

Висновок для медіаіндустрії. Кейс Netflix демонструє, що майбутнє цифрових медіа полягає в переході від масового мовлення до створення адаптивних персоналізованих середовищ. Уроки Netflix можуть бути трансльовані в журналістику: відмова від єдиної головної сторінки на користь динамічних стрічок, сформованих на основі поведінкових кластерів; A/B тестування не лише заголовків, а й зображень та анонсів для різних груп аудиторії; зосередження на метриках глибокого залучення (час читання, прокрутка) замість поверхневих кліків. Алгоритм стає не просто інструментом сортування, а новим мовним конструктом, що формує і спосіб подачі інформації, і самі принципи її створення. Головний виклик для медіа полягає в тому, щоб адаптувати ці потужні механізми, не поступаючись журналістським цінностям — об'єктивності, різноманітності точок зору та соціальній відповідальності.

4.3.2. YouTube: алгоритм рекомендацій та "rabbit holes".

YouTube як платформа унікальна своєю подвійною природою: це друга за величиною пошукова система у світі, водночас більшість часу (близько 70%) користувачі проводять не в активному пошуку, а в пасивному споживанні контенту, рекомендованого алгоритмом. Цей алгоритмічний рушій трансформував платформу з архіву відео в потужну машину утримання уваги, логіка якої пройшла значну еволюцію та має серйозні суспільні наслідки.

Архітектура рекомендаційної системи: принцип «Двох веж». Для ефективної роботи з мільярдами відео алгоритм YouTube використовує складну дворівневу архітектуру, відому як Two-Tower Architecture.

На першому етапі, Candidate Generation (Генерація кандидатів), система здійснює «грубу фільтрацію». Вона аналізує вашу детальну історію переглядів, підписки, геолокацію та інші контекстні фактори, щоб з усієї бази відібрати кілька сотень (або тисяч) відео, які мають певну ймовірність вас зацікавити. Цей відбір багато в чому працює на принципах колаборативної фільтрації: якщо велика кількість користувачів, які дивилися відео А, також дивилися відео Б, система запропонує відео Б вам, навіть якщо ви ніколи безпосередньо не шукали подібного контенту.

На другому етапі, Ranking (Ранжування), складніша нейромережа бере цей короткий список кандидатів і присвоює кожному відео індивідуальний прогнозний бал. На цьому етапі враховуються сотні детальних ознак, як про користувача (час доби, тип пристрою), так і про відео (тривалість, залученість аудиторії, актуальність). Відео з найвищими балами потрапляють на головну сторінку (Home) та в панель «Наступне» (Up Next), формуючи тим самим індивідуальну стрічку контенту.

Еволюція цілей алгоритму: від кліків до тривалості перегляду та задоволеності. Принципова зміна в логіці YouTube відбулася приблизно в 2012 році, коли основною метрикою оптимізації з кількості кліків (Clicks) стала загальна тривалість перегляду (Total Watch Time). Це була реакція на побічний ефект першої метрики — розквіт клікбейту: провокаційних мініатюр та оманливих заголовків, які залучали кліки, але не могли утримати глядача. Нова метрика навчала алгоритм оцінювати не лише перше враження, але й реальну цінність контенту: якщо користувач клікає, але закриває відео через 10 секунд, це негативний сигнал; якщо ж він дивиться його 20 хвилин або навіть повністю — це сильний позитивний сигнал.

Згодом система ускладнилася ще більше, почавши прагнути до оптимізації задоволеності користувача (User Satisfaction). Для цього алгоритм почав враховувати «м'які» сигнали: співвідношення лайків та дизлайків, участь у опитуваннях («Чи сподобалося відео?»), частоту додавання до плейлистів, а також коментарі. Це дозволило намагатися відрізняти «якісний час перегляду» (коли користувач отримує цінність) від «марного часу» (наприклад, коли він пасивно споживає низькоякісний або маніпулятивний контент, хоча і проводить перед екраном багато часу). Така еволюція відображає

спробу балансувати між бізнес-інтересом (максимізація часу на платформі) і довгостроковою лояльністю користувача.

Побічний ефект: феномен «кролячої нори» та заходи боротьби з ним. Прагнення алгоритму максимізувати тривалість перегляду безпосередньо спричинило появу суспільно небезпечного феномену — «кролячої нори» (Rabbit Hole). Алгоритм емпірично «дізнався», що контент, який викликає сильні емоції (тривогу, обурення, здивування), конспірологічні теорії або поляризовані судження, зазвичай утримує увагу краще за збалансований, поміркований аналіз. Таким чином, починаючи з невинного відео про здорове харчування, система може поступово рекомендувати все більш радикальні матеріали: від критики окремих продуктів до теорій змови про фармацевтику та навіть до заперечення основ науки. Це створює ефект алгоритмічної радикалізації, коли користувача поступово зсувають у бік більш крайніх поглядів для підтримки його залучення новими «дозами» емоційного контенту.

Усвідомлюючи цю небезпеку, YouTube вживає заходів протидії. Платформа ввела концепцію «прикордонного контенту» (Borderline Content) — матеріалів, що не порушують безпосередньо правил, але можуть сприяти дезінформації або шкоді. Такі відео песимізуються алгоритмом: їх не рекомендує головна сторінка, вони демонетизуються або супроводжуються сповіщеннями від фактчекерів. Однак боротьба з цим феноменом залишається складним викликом, оскільки він є прямим наслідком внутрішньої логіки роботи самої рекомендаційної системи.

Висновок. Алгоритм YouTube є яскравим прикладом того, як технічні рішення щодо оптимізації певних метрик (від кліків до часу перегляду та задоволення) безпосередньо формують інформаційний ландшафт і поведінку мільярдів людей. Його архітектура «двох веж» демонструє потужність машинного навчання в управлінні увагою, а феномен «кролячої нори» — глибокі етичні та соціальні дилеми, що виникають, коли бізнес-цілі системи стикаються з її впливом на суспільство. Для медіафахівців розуміння цієї логіки критично важливе не тільки для просування контенту, але й для осмислення власної ролі у формуванні інформаційного середовища, вільної від алгоритмічних перекосів.

4.3.3. Аналіз кейсу Spotify. Алгоритмічна алхімія музичного смаку.

Успіх Spotify як домінантної аудіоплатформи ґрунтується не на ексклюзивності музичного каталогу, який є практично ідентичним у всіх основних стрімінгових сервісів, а на безпрецедентній здатності передбачити та сформувати музичні бажання користувача ще до того, як він усвідомлює ці бажання. Завданням Spotify є трансформувати пасивне споживання музики в безперервний, персоналізований потік звуку, мінімізуючи когнітивне навантаження на вибір та максимізуючи емоційний зв'язок. На відміну від платформ для відео чи товарів, Spotify використовує унікальну гібридну модель рекомендацій, яка поєднує аналіз соціальної поведінки, культурного контексту та фізичних властивостей самого звуку. Ця тристороння модель, що лежить в основі таких продуктів, як легендарний плейліст «Discover Weekly», включає три взаємодоповнюючі технології. Перша — це колаборативна фільтрація, яка працює за принципом «матриці смаків»: система знаходить користувачів зі схожою на вашу історією прослуховувань і пропонує вам треки з їхніх бібліотек, яких у вас ще немає. Друга — обробка природної мови (NLP), для якої Spotify сканує мільйони текстових джерел в інтернеті, від блогів до описів плейлістів, щоб виявити, якими словами, настроями та категоріями люди описують музику, надаючи їй багатий культурний контекст. Третя, найінноваційніша складова — це аналіз сирого аудіо за допомогою згорткових нейронних мереж, які вивчають спектрограму звуку, визначаючи темп, тональність, танцювальність та інші акустичні характеристики. Це слугує страховкою від проблеми «холодного старту», дозволяючи рекомендувати абсолютно нові пісні на основі їхньої звукової схожості з уже відомими виконавцями, навіть за відсутності будь-яких даних про прослуховування.

Щорічна кампанія Spotify Wrapped є геніальним прикладом того, як сухі дані перетворюються на потужну особисту історію, перетворюючи користувачів на безкоштовних амбасадорів бренду. Вона працює на глибокому психологічному рівні, задовольняючи потреба в саморефлексії та самопрезентації, адже кожен отримує яскраво оформлений звіт, що підкреслює унікальність його музичного смаку. Одночасно вона активізує FOMO (Fear Of Missing Out) — страх пропустити щось важливе, коли всі в соціальних

мережах починають ділитися своїми підсумками. Це перетворює приватний досвід споживання на публічну соціальну валюту, а бренд із постачальника послуги перетворюється на куратора особистої культурної ідентичності слухача. Логічним кроком у еволюції від рекомендації до повноцінного керування звуковим середовищем став запущений у 2023 році AI DJ. Ця функція покликана вирішити дві ключові проблеми: «самотність» стрімінгу та втому від прийняття рішень (Decision Fatigue). Вона поєднує три технології: глибоку персоналізацію для створення динамічного міксу з ностальгійних треків, новинок і відкриттів; генеративний штучний інтелект (на основі LLM від OpenAI), який аналізує профіль слухача та пише персональний скрипт для діджея з поясненнями підбору; та гіперреалістичний синтез голосу (від Sonantic), що відтворює не лише слова, а й дихання, паузи та інтонації справжнього радіоведучого. Користувач отримує ілюзію живого супроводу, натискаючи лише одну кнопку.

Таким чином, кейс Spotify демонструє еволюцію алгоритмічного підходу від пасивної фільтрації до активного контекстного супроводу. Якщо Netflix оптимізував кліки на обкладинку, а YouTube — час перегляду, то Spotify прагне оптимізувати емоційний стан та особистий зв'язок користувача з контентом, формуючи його культурну ідентичність. Для медіаіндустрії це вказує на майбутнє, де перемога залежить не від ексклюзивності контенту, а від здатності створювати інтелектуальні, контекстно-чутливі інтерфейси. Такі інтерфейси можуть виступати одночасно куратором, гідом і співрозмовником, перетворюючи інформаційний потік на особисту наративну історію, де алгоритм виступає не радником, а автором цілісного досвіду.

4.3.4. TikTok: For You Page та вірусність.

Кейс TikTok є кульмінаційним для розуміння сучасної алгоритмічної логіки цифрових платформ. На відміну від соціальних мереж, побудованих на графі зв'язків, TikTok є розважальним двигуном на основі графа інтересів (Interest Graph). Ця фундаментальна відмінність радикально змінила правила гри: у Instagram користувач бачить контент, тому що підписаний на людину; у TikTok він бачить контент, тому що алгоритм прогнозує, що він йому сподобається, незалежно від популярності автора. Це перетворює платформу на механізм безперервного відкриття, де кожен новий ролик, незалежно

від числа підписників творця, має шанс стати глобальним вірусним явищем.

Алгоритмічна механіка: від тестових басейнів до глобального потоку. Серце системи — це повністю автоматизований процес послідовного тестування, що нагадує піраміду виживання. Його мета — максимально точно виявити контент, здатний тривало утримувати увагу.

1. Етап «Холодного старту»: Після завантаження відео алгоритм показує його невеликій тестовій групі користувачів (зазвичай 200–500 осіб). Ця аудиторія формується не з підписників автора, а на основі схожих інтересів, контексту відео (хештеги, звук) та демографічних даних. Вже на цьому етапі система заміряє ключові метрики залучення.

2. Оцінка та валідація: Якщо відео демонструє високі показники залучення, воно отримує «буст» і переходить до наступного етапу, де його показують вже тисячам або десяткам тисяч користувачів. Цей процес повторюється: кожен успіх на ширшій аудиторії відкриває доступ до ще більшої.

3. Вірусний потік: Якщо контент зберігає високі показники на масштабній аудиторії, алгоритм запускає його в «глобальний потік», що може призвести до мільйонів переглядів. Таким чином, кількість підписників творця не є вирішальним фактором успіху окремого відео; кожен новий пост починає з чистого аркуша в рівних умовах.

Ієрархія сигналів: що алгоритм цінує найбільше. Не всі дії користувачів однаково цінні для алгоритму. Їхня важливість формує чітку ієрархію, що визначає майбутнє відео.

Ключові інструменти для творця: звук, пошук та контекст. Для успішної роботи з алгоритмом творці повинні оволодіти кількома ключовими інструментами:

• Аудіо як навігаційна система: На TikTok звук є таким же потужним тегом, як і хештег. 88% користувачів вважають звук критично важливою частиною досвіду. Використання трендового звуку автоматично вписає відео в існуючу хвилю популярності. Користувачі можуть натиснути на «платівку» і побачити всі ролики з цим звуком, що відкриває додатковий шлях для відкриття контенту.

• TikTok SEO: Платформа все частіше використовується як пошукова система, особливо молоддю. Оптимізація відео за ключови-

ми словами в описі, субтитрах та текстових накладеннях дозволяє контенту з'являтися в пошукових запитах як у самому додатку, так і в Google.

• Включення в тренди та ніші: Алгоритм цінує як актуальність, так і релевантність. Участь у трендах (челленджах, використанні специфічних форматів) дає початковий імпульс для розповсюдження. Одночасно створення контенту для конкретних спільнот («ніш») підвищує ймовірність глибокого залучення та лояльності цільової аудиторії.

Таблиця 4.1.

Сигнал для алгоритму - значення та пояснення

Сигнал для алгоритму	Значення та пояснення
Rewatch Rate (Повторний перегляд)	Найсильніший сигнал. Вказує, що контент настільки цінний або розважальний, що користувач готовий переглядати його повторно. Безпосередньо пов'язаний з метою максимізації часу перегляду.
Completion Rate (Додивляння до кінця)	Ключовий індикатор якості та здатності утримувати увагу. Доведення відео до кінця — прямий сигнал про релевантність.
Share (Поширення)	Сигнал про соціальну цінність та віральний потенціал. Поширення в інші додатки (Instagram, Telegram) експортує трафік та довіру.
Comment (Коментар)	Показник глибокого залучення та співчуття. Довгі обговорення, відповіді на запитання в коментарях сигналізують про сильну реакцію аудиторії.
Like (Лайк)	Найслабший сигнал. Це пасивна, найлегша для виконання дія, яка слабко відрізняє просто приємний контент від такого, що захоплює.

Складено автором. Джерело. "Рекомендаційні системи Netflix" // UDU (2025). https://enpuirb.udu.edu.ua/server/api/core/bitstreams/f8546bfb-4eef-4bde-b38a-4a138af12f28/content

4.3.5. LinkedIn: персоналізація професійного контенту.

Кейс LinkedIn демонструє принципово іншу, ніж у розважальних платформ, філософію алгоритмічної персоналізації. На відміну від TikTok чи YouTube, мета LinkedIn полягає не в тому, щоб користувач проводив якомога більше часу на платформі, а в тому, щоб кожна хвилина, проведена там, мала вимірювану професійну або економічну цінність. Ця логіка закладена в концепції Economic Graph

(«Економічний граф») — цифрової карти всієї світової економіки, що зв'язує користувачів, компанії, навички, вакансії та знання. Алгоритм платформи прагне зміцнювати ці зв'язки, сприяючи не розвазі, а кар'єрному росту, навчанню та бізнесу.

Фундаментальна відмінність: боротьба за якість замість залучення. Першою і найважливішою рисою алгоритму LinkedIn є активна боротьба з низькоякісним, але залучуючим контентом. На відміну від інших соцмереж, де будь-яка взаємодія сприймається позитивно, LinkedIn впровадив жорсткий пре-фільтр. Коли користувач публікує пост, він проходить кілька етапів верифікації. Спочатку штучний інтелект миттєво класифікує контент, позначаючи його як «спам», «низькоякісний» або «чистий». Потім пост потрапляє до тестової аудиторії — невеликої групи найближчих контактів. Критично важливо, що на етапі оцінки залученості алгоритм звертає увагу не на лайки, а на якісні коментарі та тривалість читання (Dwell Time). Навіть якщо публікація не отримала реакції, але користувачі уважно читали її протягом довгого часу, це розглядається як сильний позитивний сигнал. Унікальним інструментом є Human Review (людська перевірка): якщо контент стає надзвичайно вірусним, його можуть перевірити живі модератори, щоб запобігти поширенню дезінформації або токсичного контенту. Зміни 2023-2024 років спрямовані саме на цю боротьбу за якість: алгоритм почав системно песимізувати (знижувати в рекомендаціях) так звану «broetry» чи «bro-poetry» — мотиваційні цитати без конкретики, оформлені як набір коротких рядків. Натомість пріоритет віддається постам, що містять Knowledge & Insights — конкретні професійні знання, кейси, аналіз помилок чи детальні інструкції.

Соціальна релевантність: три кола довіри та нішевий авторитет. Алгоритм LinkedIn будує рекомендації не на абстрактній цікавості, а на соціальній близькості та професійному авторитеті. Це реалізується через концепцію трьох кіл довіри. Найвищий пріоритет мають пости ваших безпосередніх контактів (перше коло). Однак алгоритм також розширює ваш контекст за принципом «на кого реагують ваші зв'язки». Якщо ваша колишня колега прокоментувала допис відомого експерта або CEO, ви з високою ймовірністю побачите цей допис у своїй стрічці з позначкою «Анна прокоментувала це». Це перетворює стрічку з потоку окремих думок на карту професійних дискусій

вашої мережі. Ключовим фактором є також нішевий авторитет (Niche Authority). Система аналізує ваш профіль — посаду, навички, історію публікацій — і оцінює, чи є ви експертом у темі, про яку пишете. Якщо фахівець з маркетингу пише про маркетингові інструменти, алгоритм дає такій публікації «зелене світло» для широкого охоплення. Якщо ж той самий маркетолог раптово публікує детальний аналіз квантової фізики, система, швидше за все, обмежить показ цього контенту, оскільки не сприймає автора як авторитетне джерело в цій сфері. Це стимулює користувачів глибше розвиватися у своїй галузі, а не гнатися за віральністю поза нею.

Психологія і метрики: як алгоритм читає між рядків. Оскільки аудиторія LinkedIn — дорослі професіонали, їхня поведінка відрізняється від поведінки користувачів розважальних платформ. LinkedIn розуміє, що людина може бути залучена в контент, але не залишати явних слідів: соромитися ставити лайк під чутливим постом про звільнення, щоб не привертати увагу керівництва, або просто бути занадто зайнятою для коментаря. Тому алгоритм покладається на неявні сигнали. Найважливішим серед них є Dwell Time — час, протягом якого користувач утримує пост на екрані. Уважне прочитання довгого тексту, розгортання блоку «See more» або перегляд відео до кінця без будь-яких активних дій — все це розцінюється як сильний сигнал якості. Це підвищує значення глибини та структурованості контенту, адже саме такі матеріали «тримають» на собі увагу.

Висновок для медіа та контент-стратегій. Підхід LinkedIn формує новий стандарт для професійних медіа та бізнес-комунікацій. Він доводить, що алгоритм може бути налаштований не на максимізацію залежності, а на посилення експертності та створення спільної цінності. Для контент-мейкерів це означає необхідність переходу від клікбейту до глибини: детальних кейсів, аналізу помилок, інструкцій, що реально навчають. Для медіа-брендів — це можливість будувати лояльність не через скандали, а через авторитет, стаючи частиною Economic Graph у своїй галузі. Алгоритм LinkedIn показує, що в цифрову епоху найважливішим ресурсом є не увага сама по собі, а довіра, побудована на знаннях та релевантних соціальних зв’язках.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання висвітлюють унікальні стратегії кожної платформи:

1. TikTok та "Граф інтересів": Чому TikTok вважають революційним у світі алгоритмів? У чому різниця між *Social Graph* (Facebook: "дивись те, що лайкнули друзі") та *Interest Graph* (TikTok: "дивись те, що цікаво тобі, навіть якщо у тебе 0 друзів")?

2. Netflix та персоналізація обкладинок: Як Netflix використовує алгоритми не лише для вибору *фільму*, а й для вибору *картинки* до нього? (Чому одному користувачеві на постері "Дивних див" покажуть романтичну сцену, а іншому — монстра?).

3. Spotify "Discovery Weekly": У чому секрет успіху плейлиста "Відкриття тижня"? Як працює правило "Золотоволоски" (баланс між знайомими хітами та абсолютною новинкою), щоб користувач не злякався нового?

4. YouTube і зміна метрики: У 2012 році YouTube змінив головну метрику алгоритму з *Кількості переглядів (Views)* на *Час перегляду (Watch Time)*. Як це змінило контент на платформі? (Поява летсплеїв, довгих подкастів та зникнення коротких клікбейтних відео).

5. Amazon і "Item-to-Item": Чому Amazon не намагається вгадати ваш "психологічний портрет", а використовує просту логіку "Люди, які купили цей молоток, купили й ці цвяхи"? Чому для e-commerce це ефективніше?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають деконструювати логіку платформ:

Завдання 1. "TikTok-тест": Швидкість навчання

• Інструмент: Створіть новий (чистий) акаунт у TikTok.

• Дія: Протягом 15 хвилин лайкайте лише відео певної ніші (наприклад, "гончарство" або "ремонт авто"). Пропускайте все інше.

• Аналіз: Засічіть час, через який ваша стрічка "For You" (FYP) на 80% складатиметься з цієї теми.

• Порівняння: Спробуйте зробити те саме в Instagram Reels. Де алгоритм адаптується швидше і чому? (Підказка: TikTok аналізує навіть мікро-паузи при скролі).

Завдання 2. Аналіз "Бульбашки Spotify"

• Ситуація: Spotify створює для вас "Daily Mix".

• Завдання: Проаналізуйте один мікс.

- Скільки пісень ви вже лайкали раніше? (База безпеки).

- Скільки пісень нових, але від відомих вам артистів? (Розширення контексту).

- Скільки абсолютно нових імен? (Ризик).

• Висновок: Яка пропорція (наприклад, 70/20/10) дозволяє вам не перемикати треки?

Завдання 3. Стратегія YouTube: "Кроляча нора" (Rabbit Hole)

• Завдання: Відкрийте відео на політичну тему в режимі "Інкогніто".

• Спостереження: Подивіться на бічну панель "Up Next" (Наступне).

• Гіпотеза: Чи правда, що алгоритм YouTube схильний пропонувати все більш радикальний/сенсаційний контент, щоб утримати увагу? Запишіть ланцюжок із 5 рекомендованих відео.

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Eugene Wei. *TikTok and the Sorting Hat.* (Культова стаття, що пояснює, чому алгоритм TikTok такий потужний — концепція "сортувального капелюха").

2. Netflix Tech Blog. *Artwork Personalization at Netflix.* (Детальний опис технології підбору картинок).

3. Covington et al. (YouTube Engineering). *Deep Neural Networks for YouTube Recommendations.* (Класична технічна стаття про архітектуру рекомендацій YouTube).

4. Spotify R&D. *How we built Discover Weekly.* (Історія створення одного з найкращих алгоритмічних продуктів).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Граф інтересів (Interest Graph): Модель рекомендацій, що будує зв'язки між *користувачем та темами*, які йому цікаві, а не між користувачем та його друзями. Це основа TikTok.

• Watch Time (Час перегляду): Сумарна кількість часу, яку користувачі витратили на перегляд відео. Головна метрика успіху для YouTube, що замінила "перегляди" (views).

• Algotorial (Алготоріал): Термін, що описує підхід Spotify до плейлистів (наприклад, Pollen), які куруються редакторами-людьми, але персоналізуються алгоритмами під кожного слухача.

• Rabbit Hole (Кроляча нора): Ефект алгоритмічної рекомендації, коли користувач, переходячи від відео до відео, заглиблюється у все більш вузьку, дивну або радикальну тему.

• Bandit Algorithms (Алгоритми-бандити): Клас алгоритмів (використовує Netflix), які тестують новий контент (наприклад, нову обкладинку фільму) на невеликій групі людей, щоб визначити переможця ("Exploration vs Exploitation").

4.4. Проблеми та обмеження

4.4.1. Холодний старт (нові користувачі, новий контент).

Проблема «холодного старту» (Cold Start Problem) є однією з найбільш фундаментальних та складних технічних перешкод у побудові ефективних рекомендаційних систем і медіа-платформ. Її суть полягає в критичній нестачі даних для формування прогнозів, що створює замкнуте коло: щоб зібрати дані про вподобання користувача або якість контенту, систему потрібно їх показати, але щоб показати їх правильно (тобто релевантно), потрібні вже наявні дані. Ця ситуація виникає на початковому етапі функціонування будь-якої мережі та загрожує її розвитку, оскільки користувачі, не знайшовши цінного контенту, швидко залишають платформу, посилюючи проблему. Для рекомендаційних алгоритмів, що працюють з «розрідженою матрицею» (Sparse Matrix) взаємодій, це виклик номер один, який подолувають комбінацією технічних та психологічних підходів.

Холодний старт користувача: як порекомендувати щось тому, кого не знаєш? Коли новий користувач реєструється на платформі, він являє собою «чистий аркуш» — немає ні історії переглядів, ні лайків, ні соціальних зв'язків. Алгоритм змушений вдаватися до обхідних шляхів, щоб зробити перші, найважливіші рекомендації, які визначать, чи залишиться людина.

Найбільш перспективним напрямом вважається неінвазивне використання зовнішніх даних (наприклад, соціальних мереж), оскільки воно не вимагає від користувача додаткових зусиль і може бути реалізоване миттєво, значно пом’якшуючи ефект холодного старту.

2. Холодний старт контенту: як просунути статтю, яку ще ніхто не читав? Новий матеріал у каталозі також починає з нуля. Класична колаборативна фільтрація («користувачам, які сподобались X, також сподобалось Y») тут безсила, оскільки ще немає жодної взаємодії. На допомогу приходить контентно-орієнтований аналіз (Content-Based Filtering), коли алгоритм вивчає власні атрибути об’єкта:

• Текстовий аналіз: ШІ розбирає заголовок, підзаголовки, ключові слова та основний текст статті. Якщо нова публікація — про «iPhone 15», її запропонують усім, хто раніше читав про «iPhone 14» чи «смартфони».

• Аудіовізуальний аналіз: Для відео та аудіо алгоритми (зокрема, згорткові нейронні мережі) аналізують пікселі кадрів або спектрограми звуку. Кадри з автомобілями та вибухами класифікують як «бойовик», а звукова доріжка з характерним битом — як «дріл-реп».

• Метаданні: Жанр, режисер, акторський склад, тривалість — все це стає основою для первинних гіпотез про цільову аудиторію.

Цей підхід дозволяє «прилаштувати» новий елемент до вже існуючих категорій та інтересів, даючи йому шанс бути виявленим.

Таблиця 4.2.

Класифікація підходів до подолання проблеми холодного старту

Категорія рішення	Конкретні підходи та їх логіка
Безпечні, неперсоналізовані	Популярний контент (Global Popularity): Показ «топ-10 за день». Це гарантує базову якість та широкий інтерес, але не формує особистого зв'язку.
Активний збір даних	Онбордінг (Onboarding): Пряме опитування про інтереси під час реєстрації. Це Zero-Party Data — інформація, якою користувач ділиться свідомо та добровільно.
Використання зовнішніх даних	Перенесення знань (Transfer Learning): Імпорт вподобань із пов'язаних акаунтів (наприклад, соціальних мереж) за згодою користувача. Аналіз активності в соцмережах: Сучасні дослідження показують, що поведінковий профіль користувача можна побудувати на основі його публічної активності в соціальних мережах (наприклад, X/Twitter), що дозволяє персоніфікувати рекомендації без явних запитань.
Контекстуальні та технічні сигнали	Геолокація, пристрій, час доби: Користувачеві з iPhone у Києві можуть запропонувати технологічні новини та події столиці, формуючи гіпотези на основі контексту.

Складено автором. Джерело. Cold Start in Recommender Systems // freeCodeCamp (2025). https://www.freecodecamp.org/news/cold-start-problem-in-recommender-systems/

3. Дилема «Дослідження vs. Використання» (Exploration vs. Exploitation). Ефективна система не може лише експлуатувати вже перевірені дані — інакше вона закостеніє. Тому алгоритми постійно балансують між двома режимами:

• Використання (Exploitation): Показувати користувачеві те, що з високою ймовірністю йому сподобається, базуючись на відомій історії. Це безпечно та максимізує задоволеність у короткій перспективі.

• Дослідження (Exploration): Ризикнути та показати новий або незвичний контент невеликій, випадковій групі користувачів, щоб зібрати перші дані та розширити знання системи.

Для керування цим балансом використовуються бандитські алгоритми (Bandit Algorithms), які умовно виділяють частину трафіку (наприклад, 5-10%) на такі експерименти. Якщо реакція на новий контент у цій тестовій групі позитивна, його починають рекомендувати широко (переводять у режим «використання»). Це забезпечує постійне оновлення рекомендацій і запобігає «застарінню» системи.

Висновок та перспективи. Проблема холодного старту — це не просто технічний баг, а системний виклик для будь-якої платформи, що побудована на мережевих ефектах. Її успішне подолання вимагає комбінованої стратегії: від простих безпечних методів (популярний контент) до передових підходів, таких як глибокий контент-аналіз, перенесення знань із зовнішніх джерел та інтелектуальне балансування між ризиком і гарантією. Майбутнє полягає в ще більшій інтеграції відкритих даних та контексту (соціальні профілі, реальний світ) для миттєвого створення цифрового «двійника» користувача ще до того, як він зробить свою першу взаємодію всередині платформи. Розуміння цих механізмів критично не лише для інженерів, а й для сучасних медіафахівців, які повинні знати, як технології «бачать» та просувають їхній контент для нової, незнайомої аудиторії.

4.4.2. Популярність vs релевантність (popularity bias).

Проблема Popularity Bias (зсув у бік популярності, або упередженість популярності) є фундаментальним і часто неусвідомленим дефектом логіки рекомендаційних алгоритмів, який спотворило споживання контенту в цифрову епоху. Це явище описує системну тенденцію алгоритмів пропонувати користувачам вже популярний контент (відео, статті, книги), що має велику кількість попередніх взаємодій, водночас ігноруючи менш відомі, але потенційно більш релевантні об'єкти. Ця динаміка створює структурну несправедливість: віртуальне стає ще віртуальнішим не обов'язково через вищу якість, а через самопосилюючу логіку системи, тоді як якісний, нішевий контент залишається в тіні, не отримуючи шансу бути відкритим.

Феномен Матвія: як алгоритми посилюють нерівність. Соціо-

логічний Ефект Матвія, названий на честь біблійного вислову «бо кожному, хто має, дасться й примножиться», ідеально описує механік популярності в цифровому середовищі. Алгоритми навчаються на історичних даних про поведінку користувачів. Відео з мільйоном переглядів, лайків і коментарів надає системі мільйон чітких, статистично значущих сигналів для навчання. Навпаки, відео, яке переглянули лише сто разів, для складної нейромережі є статистичним шумом — слабким сигналом, який важко інтерпретувати. Таким чином, алгоритм, прагнучи максимізувати ймовірність позитивної взаємодії (кліку, перегляду), природно схиляється до «безпечної ставки» — рекомендації вже популярного. Це створює потужну петлю зворотного зв'язку (Feedback Loop): популярний контент частіше рекомендують → він отримує ще більше взаємодій → алгоритм інтерпретує це як підтвердження його якості та релевантності → починає рекомендувати його ще агресивніше. В результаті популярність зростає не лінійно, а експоненціально, часто незалежно від внутрішніх якостей контенту.

«Проблема Гаррі Поттера»: ілюзія успіху алгоритму. Класичний приклад з книги про Гаррі Поттера розкриває глибоку суперечність між статистичною точністю алгоритму та його реальною користю для користувача. Якщо система рекомендує бестселер усім без винятку, показники її точності (Accuracy) будуть високими — адже більшість людей справді знайомі з цим твором або позитивно до нього ставляться. Однак ця рекомендація марна (shallow). Вона не відкриває для користувача нічого нового, не розвиває його смак, не задовольняє специфічну потребу. Справжня цінність персоналізованої системи, особливо для досвідченого користувача, полягає у здатності досліджувати «довгий хвіст» (The Long Tail) культури — знаходити маловідомих авторів, нішеві жанри, альтернативні точки зору, про які людина не знала. Popularity Bias системно придушує цю можливість, обрізаючи «довгий хвіст» і звужуючи культурний горизонт користувача до мейнстримного канону, сформованого попередніми масовими взаємодіями.

Конфлікт бізнес-метрик: боротьба за клік замість лояльності. Ця технічна проблема має пряму проекцію на бізнес-стратегії медіа та контент-платформ, викликаючи постійний внутрішній конфлікт між короткостроковими та довгостроковими цілями.

• Оптимізація під CTR (Click-Through Rate): Це логіка миттєвого трафіку. Алгоритми, що максимізують клікабельність, природно схиляються до контенту з найширшим потенційним охопленням: сенсаційним заголовкам, вірусним трендам, контенту про зірок. Це дає швидкий приплив відвідувачів і рекламних показів «тут і зараз», але часто веде до ескалації сенсаційності та якості контенту.

• Оптимізація під LTV (Lifetime Value): Це логіка довгострокової лояльності. Мета полягає в тому, щоб користувач бачив у платформі незамінне джерело релевантної, глибокої цінності. Наприклад, фанат архітектури готовий платити за підписку на медіа, яке постійно відкриває йому якісні матеріали про урбаністику, навіть якщо окремі статті збирають менше кліків, ніж плітки про знаменитостей. Глибока персоналізація та боротьба з Popularity Bias є ключем до такого LTV.

Редакції та продуктові команди постійно балансують між цими двома полюсами. Перевага короткострокових метрик CTR часто призводить до домінування популярності над релевантністю, що в довгостроковій перспективі може підірвати довіру аудиторії та призвести до втоми від контенту.

Висновок: наслідки та шляхи подолання. Popularity Bias — це не просто технічний артефакт, а соціокультурне явище, що формує нашу інформаційну реальність. Воно зміцнює мейнстрим, маргіналізує альтернативи та обмежує вибір, створюючи ілюзію консенсусу, де насправді відбувається алгоритмічне посилення вже існуючих нерівностей. Для подолання цієї упередженості необхідні свідомі зусилля на рівні дизайну алгоритмів: впровадження де-біаїнгових технік (debiasing), навмисне «підштовхування» нішевого контенту у рекомендації (exploration), розробка метрик якості, що оцінюють не кількість, а глибину залучення та різноманіття рекомендацій. Для медіа це означає необхідність відстоювати редакційну цінність і глибину, іноді всупереч тиску миттєвих метрик, розуміючи, що справжня лояльність будуються на релевантності та відкриттях, а не на безперервному пережовуванні вже відомого.

4.4.3. Прозорість та пояснюваність рекомендацій.

Епоха «чорної скриньки» та виклик для суспільства. Сучасні цифрові платформи оперують складними системами штучного інтелекту, які постійно приймають рішення про те, яку інформацію ми бачимо. Однак механізми цих рішень все частіше стають

непроникними не лише для користувачів, а й для самих розробників. Це формує так зване «Суспільство чорної скриньки» (Black Box Society) — середовище, де алгоритми, що ґрунтуються на глибокому навчанні, формують світогляд мільйонів людей, залишаючись незрозумілими. Концепції прозорості (Transparency) та пояснюваності (Explainability) виникають як критичні відповіді на цей виклик, прагнучи зробити взаємодію з технологією справедливою, довіреною та відповідальною. Фундаментальна проблема полягає в переході від традиційного, правилового програмування до машинного навчання. Раніше логіка була прозорою: «якщо користувач із Києва — показувати київські новини». У нейромережах глибокого навчання рішення приймаються через аналіз мільярдів параметрів та тисяч мікро-сигналів: від часу доби та швидкості скролу до сукупної історії переглядів за роки. Ця надзвичайна складність робить модель «чорним ящиком» (black box) — системою, внутрішній механізм якої незбагненний навіть для її творців. Наслідки серйозні: неможливість точно визначити, чому відео стало вірусним, ускладнює боротьбу з дезінформацією, упередженістю та непередбачуваною поведінкою алгоритму. Довіра користувача до системи, яка не може пояснити свої дії, закономірно падає.

У відповідь на цю проблему виникло ціле наукове спрямування — Поясненний штучний інтелект (Explainable AI, XAI). Його мета — створити ШІ, результати роботи якого можуть бути зрозумілими для людини. XAI прагне перетворити «чорний ящик» на «скляну коробку» (glass box) або модель «білого ящика» (white box), механізми якої зрозумілі експертам. Існують два основні підходи до пояснення. По-перше, постфактум пояснення (Post-hoc explanations) — це методи, що пояснюють конкретне рішення вже навченої складної моделі. До них належать такі інструменти, як LIME (Local Interpretable Model-agnostic Explanations) та SHAP (Shapley Additive exPlanations), які допомагають зрозуміти внесок окремих факторів у кожну рекомендацію. По-друге, прозорість як публічність дизайну (Transparency as Design Publicity) — більш глибокий підхід, який полягає в публічному поясненні не окремих рішень, а цілей, принципів роботи та обмежень алгоритму в цілому. Це дозволяє суспільству оцінювати, чи виправдане існування та застосування такої системи.

На практиці в медіа це реалізується через інтерфейси для користувачів (User-facing explanations). Найпростішим прикладом є кнопка «Чому я це бачу?» (Why am I seeing this?), яка відкриває пряме пояснення на основі демографії, геолокації або минулої активності, наприклад: «бо ви лайкали сторінку про урбаністику». Іншим поширеним інструментом є контекстні мітки (маркування): підписи на зразок «Популярне серед ваших друзів» або «Рекомендовано, тому що ви дивилися...». Такі підписи не тільки пояснюють, але й знімають відчуття маніпуляції. Деякі платформи, як-от Instagram, тестують надання користувачам прямого контролю персоналізації через інструменти для явного додавання чи видалення інтересів, що впливають на рекомендації. Дослідження показують, що впровадження таких механізмів не лише підвищує довіру, але може покращити і технічну ефективність системи. Наприклад, інтеграція методів пояснюваності у рекомендаційні моделі дала приріст їхньої точності (precision) на 3%.

Сьогодні прозорість — це не просто етичний вибір, а юридична вимога. Європейське законодавство стало локомотивом змін. GDPR (Загальний регламент захисту даних) закріплює «право на пояснення» (Right to explanation). Користувач має право отримати інформацію про логіку, значення та наслідки автоматизованого прийняття рішень щодо нього. Новий DSA (Акт про цифрові послуги, Digital Services Act) піднімає планку ще вище, безпосередньо зобов'язуючи великі онлайн-платформи (TikTok, YouTube, тощо): по-перше, розкривати ключові параметри роботи своїх алгоритмів рекомендацій; по-друге, надавати альтернативу — можливість переглядати стрічку в хронологічному порядку без персоналізації; по-третє, проводити оцінку ризиків і вживати заходів проти системних загроз, таких як поширення дезінформації. Ці закони переводять дискусію про прозорість з теоретичної площини в практичну, змушуючи технологічні компанії будувати системи з можливістю звіту та обґрунтування.

Висновок: прозорість як основа довіри та відповідальності у цифрову епоху. Таким чином, прозорість та пояснюваність алгоритмів стають критичними елементами відповідального медіапростору. Це не технічне «додавання», а фундаментальний принцип, що пов'язує технічну можливість (методи XAI), інтерфейсний дизайн, юридичну

відповідальність та етичну норму. Для сучасного медіафахівця розуміння цих механізмів вже не є факультативним. Це — основа для критичного аналізу власних інструментів роботи, комунікації з аудиторією та відстоювання цінностей відкритого, довіреного суспільства в епоху алгоритмічно детермінованої уваги. Майбутнє інформаційної екології залежатиме від того, наскільки вдало ми інтегруємо принципи пояснюваності не як латку на проблему, а як основоположний елемент дизайну цифрових систем.

4.4.4. Маніпуляція та gaming the system.

Геймінг системи як системна вразливість. У цифровій епосі, де алгоритми визначають видимість і успіх, неминуче виникає феномен "геймінгу системи" (Gaming the System). Ця практика полягає у використанні формальних правил і механік алгоритмів для штучного досягнення результату, що суперечить їх початковому задуму — покращенню якості рекомендацій та відбору цінного контенту. Замість створення вартосного продукту, маніпулятори фокусуються на експлуатації вразливостей самої системи оцінки, що призводить до роздування популярності, спотворення інформаційного поля та ерозії довіри до платформи. Це не просто чесна конкуренція, а стратегічна атака на основи функціонування алгоритмічного суспільства.

Закон Гудхарта: фундаментальна причина маніпуляцій. У основі більшості спроб зламу алгоритмів лежить Закон Гудхарта (Goodhart's Law), сформульований економістом Чарльзом Гудхартом: "Коли міра стає ціллю, вона перестає бути хорошою мірою". Цей принцип ілюструє фатальну слабкість будь-якої системи, що покладається на кількісні показники для оцінки якісно складних явищ.

На практиці це працює так: платформа (наприклад, YouTube) визначає, що головним індикатором якості відео та залучення аудиторії є кількість коментарів. Алгоритм починає активніше просувати контент, що отримує багато коментарів. Незабаром автори виявляють цю кореляцію і, замість того щоб покращувати контент, переходять до створення провокаційних матеріалів ("Rage Bait") — відео з навмисними помилками в заголовках, поляризуючими твердженнями або образами, призначеними викликати хвилю обурення та суперечок у коментарях. В результаті метрика (коментарі) зростає, але реальна якість контенту, задум розробників (створення здорової дискусії) та користь для користувача — різко падають. Це змушує платформи

постійно ускладнювати та змінювати свої алгоритми, вводити нові, складніші для "хакання" метрики (наприклад, співвідношення лайків/дизлайків, тривалість перегляду, рівень утримання), що породжує нескінченну гонку озброєнь між розробниками та маніпуляторами.

Астротурфінг та ботоферми: штучне створення трендів. Одним з найнебезпечніших видів геймінгу системи є астротурфінг (Astroturfing) — штучна імітація масової, органічної (grassroots) суспільної підтримки певної ідеї, продукту чи політичної позиції. Механізм експлуатує фундаментальний принцип роботи алгоритмів трендів (у Twitter, Facebook, Instagram), які реагують на різкі сплески активності, інтерпретуючи їх як ознаку важливої та актуальної теми.

Атака відбувається за допомогою ботоферм — мереж із сотень або тисяч автоматизованих акаунтів (ботів). Ці акаунти в точно синхронізований момент часу починають масово публікувати пости, використовуючи однакові хештеги, ключові фрази або посилання. Алгоритм, виявляючи цю аномальну, скоординовану активність (Coordinated Inauthentic Behavior, CIB), помилково класифікує її як вияв справжнього суспільного інтересу. Внаслідок цього штучно створена тема потрапляє в "Тренди", починає рекомендуватися мільйонам справжніх користувачів і набуває видимості легітимного мейнстримного обговорення. Таким чином, маргінальна, фейкова або пропагандистська ідея може бути інжектована в суспільну дискусію, маніпулюючи суспільною думкою та політичною повесткою дня.

Атака на дані: отруєння джерела та SEO-спам. Маніпуляції можуть здійснюватися не тільки на рівні поведінки (лайки, коментарі), але й на більш фундаментальному рівні — рівні даних та самого контенту, що є паливом для алгоритмів.

SEO-спам — це класичний приклад, де контент створюється не для людей, а виключно для роботів пошукових систем. Це наповнення тексту ключовими словами, "невидимий" текст, створення штучних посилань. У своїй сучасній еволюції ця практика перетворилася на створення MFA-сайтів (Made for Advertising) — масштабних мереж сторінок, контент на яких часто повністю згенерований штучним інтелектом без перевірки якості чи точності. Єдина мета таких сайтів — займати високі позиції в пошуку за популярними запитами та монетизувати трафік через агресивну рекламу, витісняючи легітимні джерела.

Ще більш витонченою атакою є Data Poisoning (Отруєння даних). Якщо алгоритми навчаються на даних, то контроль над цими даними дає владу над їхніми майбутніми рішеннями. Зловмисники можуть свідомо "підгодовувати" алгоритм масивом неправильно позначених даних. Наприклад, конкурентна фірма може організувати кампанію, в ході якої тисячі акаунтів будуть масово позначати легальні листи певного бренду як спам у Gmail. Фільтр спаму, що навчається на цих даних, згодом почне автоматично блокувати цей бренд, наносячи йому серйозний репутаційний та фінансовий збиток. Подібні атаки можуть підривати кредитні системи, системи розпізнавання облич та рекомендаційні механізми.

Висновок: Вічна гонка та необхідність комплексного захисту. Феномен "геймінгу системи" демонструє, що алгоритмічне середовище є полем постійної боротьби між розробниками, які прагнуть до якості та релевантності, та маніпуляторами, які прагнуть до контролю та вигоди. Це не технічна дрібниць, а серйозний виклик для інформаційної безпеки, чесності публічного дискурсу та довіри до цифрових інститутів. Ефективна боротьба з цим вимагає не лише технічних патчів (покращення алгоритмів виявлення CIB, боротьби з MFA), а й розвитку медіаграмотності користувачів, прозорості з боку платформ та, часто, регуляторних втручань. Для медіапрофесіонала розуміння цих механізмів — це ключ до критичного аналізу інформаційного ландшафту, захисту власного контенту від нечесної конкуренції та усвідомлення своєї ролі у підтриманні здорового цифрового середовища.

1. КОНТРОЛЬНІ ПИТАННЯ

1. Функція: Чому Герберт Саймон (нобелівський лауреат) сказав, що "багатство інформації створює бідність уваги"? Як RecSys вирішують цю проблему?

2. Еволюція: Чим рекомендаційна система відрізняється від звичайного редакційного відбору (Curated list)? (Масштабом: редактор робить список для всіх, RecSys — для кожного окремо).

3. Дані: Чому Netflix відмовився від зірочок (рейтингу) на користь системи "палець вгору/вниз"? (Бо зірочки відображали *оцінку якості* фільму, а не *бажання* його дивитися).

4. Ризики: Що таке "Feedback Loop" (Петля зворотного зв'язку)

і як вона може звузити кругозір користувача? (Користувач клікає -> Алгоритм пропонує схоже -> Користувач клікає ще більше схожого -> Інші теми зникають).

5. Метрика: Що таке "Serendipity" (Серендипність) у контексті рекомендацій? (Здатність системи приємно здивувати неочікуваною пропозицією).

2. ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. Аудит "Явного vs Неявного"

• Платформа: YouTube.

• Дія:

1. Знайдіть у своїй історії відео, яке ви подивилися повністю, але не лайкнули. (Неявний сигнал).

2. Знайдіть відео, яке ви лайкнули, але не додивилися. (Явний сигнал).

3. Подивіться на рекомендації збоку. Яких відео більше схожих: на перше чи на друге?

• Висновок: Який сигнал для YouTube важливіший?

Завдання 2. Аналіз "Парадоксу вибору"

• Ситуація: Ви зайшли на стрімінговий сервіс, де зламалися рекомендації і показується просто алфавітний каталог усіх фільмів.

• Рефлексія: Опишіть свої емоції та дії. Скільки часу ви витратите на вибір? Чи є ризик, що ви підете ні з чим?

Завдання 3. Проєктування даних

• Роль: Ви розробляєте новинний додаток.

• Завдання: Напишіть список з 5 неявних сигналів (окрім кліку), які допоможуть зрозуміти, що стаття сподобалася читачеві. (Наприклад: шерування, копіювання тексту, час на сторінці > 2 хв, скрол до коментарів, повернення на статтю наступного дня).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Schwartz, B. *The Paradox of Choice: Why More Is Less.* (Фундаментальна книга про психологію вибору).

2. Netflix Tech Blog. *The Netflix Recommender System: Algorithms, Business Value, and Innovation.*

3. Anderson, C. *The Long Tail.* (Про те, як рекомендації дозволяють заробляти на нішевому контенті).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Recommender System (RecSys): Програмний комплекс, який прогнозує, наскільки користувачеві сподобається певний об'єкт, і формує індивідуальну видачу.

• Information Overload (Інформаційне перевантаження): Стан, коли обсяг вхідної інформації перевищує можливості людини її обробити, що призводить до стресу та неможливості прийняти рішення.

• Churn Rate (Показник відтоку): Відсоток клієнтів, які припинили користуватися сервісом за певний період. Головне завдання RecSys — зниження цього показника.

• Explicit Feedback (Явний зворотний зв'язок): Свідома оцінка користувачем контенту (лайк, рейтинг, відгук).

• Implicit Feedback (Неявний зворотний зв'язок): Дані про поведінку користувача, які збираються автоматично (перегляди, кліки, історія пошуку).

Висновок до розділу 4

Розділ присвячено комплексному аналізу рекомендаційних систем як основоположного механізму сучасного цифрового медіапростору. Розглянуто їх еволюцію від простих фільтрів до складних архітектур штучного інтелекту, які визначають не лише що ми читаємо, дивимось і слухаємо, але й як формується наш світогляд, культурні смаки та професійні зв'язки. Ключовим висновком є те, що рекомендаційні алгоритми сьогодні — це не технічний інструмент, а новий видавничий інститут, який замінив собою традиційних редакторів та кураторів, перетворивши увагу на основний ресурс, а персоналізацію — на основний принцип дистрибуції контенту.

Механізми та стратегії платформ: від "графа зв'язків" до "графа інтересів". Аналіз кейсів провідних платформ демонструє різницю в їх філософії та цілях. Netflix та TikTok стали хрестоматійними прикладами повного переходу від Social Graph до Interest Graph, де пріоритет надається не знайомству з автором, а прогнозованій цінності окремої одиниці контенту для конкретної людини. Їх алгоритми (від кластеризації смаків до тестових басейнів) побудовані на мікроменеджменті уваги з метою мінімізації часу пошуку. YouTube пройшов еволюцію від оптимізації кліків до максимізації

часу перегляду, що неминуче породило феномен «кролячих нор». Spotify довів, що успіх лежить у гібридному підході, де аналізуються одночасно соціальні зв'язки, культурний контекст та фізичні властивості звуку, а функції типу AI DJ ознаменували перехід до генеративного супроводу. LinkedIn, на противагу, показав, що алгоритм може бути оптимізований не під максимальне залучення, а під економічну цінність та якість професійної комунікації, активно боронячись від низькоякісного контенту. Це доводить, що немає єдиної ідеальної моделі — є різні алгоритмічні логіки, підпорядковані різним бізнес-цілям.

Фундаментальні проблеми: технічні обмеження та етичні виклики.

Незважаючи на неймовірну складність, рекомендаційні системи стикаються з низкою вбудованих проблем, які обмежують їхню об'єктивність і справедливість. Проблема холодного старту розкриває фундаментальну залежність алгоритмів від даних і постійну потребу балансувати між дослідженням нового та експлуатацією вже відомого. Popularity Bias (зсув популярності) та Закон Гудхарта демонструють, як метрики ефективності можуть спотворити сам контент, створюючи петлі зворотного зв'язку, що посилюють мейнстрим і маргіналізують «довгий хвіст». Непрозорість («чорна скринька») та практики геймінгу системи (від астротурфінгу до отруєння даних) показують, як технологічний інструмент може стати об'єктом маніпуляцій, загрожуючи цілісності інформаційного середовища.

Магістральний висновок: нова парадигма медіавиробництва та споживання. Підсумовуючи, рекомендаційні системи сформували абсолютно нову парадигму, де:

1. Контент атомізується. Цінність окремої статті, відео або треку тепер незалежна від автора чи бренду; кожна одиниця бореться за увагу в глобальному потоці.

2. Алгоритм стає автором та редактором. Він не лише фільтрує, а й формує контент (через data-driven продюсування, як у Netflix) та визначає контекст його споживання (через персоналізацію обкладинок, звукових хвиль).

3. Увага стає архітектурованою. Психологічні механізми (змінна винагорода, FOMO) і технічні механізми (нескінченний скрол, тестові басейни) інтегровані для створення максимально залучую-

чих досвідів.

4. Довіра стає ключовою метрикою. Успіх тепер залежить від здатності системи не просто «вгадати» контент, а стати достовірним куратором, що потребує прозорості, пояснюваності та боротьби зі шкідливими явищами.

Таким чином, розуміння логіки рекомендаційних систем перестає бути прерогативою інженерів та data scientist'ів. Для сучасного журналіста, медіаменеджера чи контент-мейкера це базова медіаграмотність. Майбутнє успішного медіа лежить у синергії між творчою, людяною експертизою та здатністю ефективно взаємодіяти з алгоритмічною логікою, використовуючи її для посилення релевантності, відкриття нових тем та глибокого залучення аудиторії, але ніколи не поступаючись фундаментальним цінностям — точності, якості та служінню суспільному інтересу.

Розділ 5: Збір та аналіз даних про аудиторію
5.1. Джерела даних

5.1.1. Поведінкові дані: кліки, перегляди, час перебування.

Від демографії до дії. У цифровій епосі справжнє розуміння аудиторії перестало обмежуватися знанням демографічних профілів (хто користувач: вік, стать, локація). На перший план виходять поведінкові дані (Behavioral Data) — об'єктивні цифрові сліди, які користувач залишає під час кожної взаємодії з продуктом. Ці дані не опитують наміри, а фіксують реальні дії, формуючи «мову тіла» читача або глядача. Для сучасних медіа саме ці сигнали стають найважливішим джерелом інсайтів для персоналізації, вдосконалення контенту та підвищення залучення.

Ієрархія сигналів: від поверхневої цікавості до глибокого задоволення. Не всі поведінкові сигнали мають однакову цінність. Їх можна вибудувати в ієрархію за рівнем довіри та інформативної щільності. Найнижчий, але наймасовіший рівень — це клік (Click або CTR — Click-Through Rate). Це сигнал початкової цікавості, що показує ефективність заголовка, зображення чи прев'ю. Однак його головна слабкість — відсутність гарантії задоволення. Яскравий приклад — клікбейт: заголовок може спровокувати високий CTR, але швидке розчарування від низькоякісного контенту не лише не розвиває лояльність, але й руйнує її.

Наступний, значно важливіший рівень — час перебування на сторінці (Dwell Time / Time on Page). Ця метрика вже наближається до вимірювання реального інтересу. Якщо користувач провів на сторінці статті дві-три хвилини, висока ймовірність, що він уважно її читав. Для якісних медіа, що роблять акцент на глибині, це своєрідна «валюта». Короткий час перебування (менше 10-15 секунд) розцінюється як негативний сигнал «відмови» (Bounce) і вказує на те, що контент не відповідає очікуванням, сформованим заголовком.

Ще точнішим індикатором є глибина прокручування (Scroll Depth), яка вимірює, яку частину контенту фактично спожив користувач (25%, 50%, 75% чи 100%). Цей сигнал прямо вказує на залученість. Він може виявити структурні проблеми матеріалу: якщо 80% читачів залишають сторінку на другому абзаці, проблема, швидше за все, не в темі, а в манері викладу, занадто довгому вступі

чи невдалому форматуванні.

Патерни навігації: аналіз шляху користувача (User Journey). Справжню силу поведінковим даним надає аналіз не окремих кліків, а їх послідовності — цілісного шляху користувача (User Journey). Алгоритми та аналітики виділяють ключові патерни, що свідчать про якість досвіду.

Класичним негативним патерном є «стрибки на батуті» (Pogo-sticking). Це ситуація, коли користувач клікає на результат пошуку (наприклад, в Google), потрапляє на статтю, але через кілька секунд повертається назад до пошукової видачі. Для пошукових систем та алгоритмів рекомендацій це прямий сигнал: «Контент не відповів на запит користувача або виявився низькоякісним». Регулярне повторення таких сценаріїв різко знижує рейтинг сторінки в видачі.

Протилежним, позитивним патерном є рециркуляція (Recirculation). Це модель поведінки, при якій користувач, переглянувши один матеріал (статтю А), переходить за внутрішнім посиланням до іншого матеріалу (статті Б) на тому ж сайті. Це яскравий сигнал для системи: «Високий рівень задоволеності та лояльності, хороший користувацький досвід (UX) і ефективна внутрішня перелинковка». Рециркуляція максимізує час на сайті та глибину занурення в контент.

Мікро-конверсії: індикатори найглибшої взаємодії. На вершині ієрархії знаходяться мікро-конверсії (Micro-conversions) — невеликі, часто неусвідомлені дії, що свідчать про активне, зацікавлене залучення. Вони коштують значно більше за простий клік. До них належать:

• Виділення тексту мишкою: Часто свідчить про бажання скопіювати цитату або про уважне, вдумливе читання.

• Розгортання зображення на весь екран: Індикатор інтересу до візуального контенту та деталей.

• Натискання кнопки відтворення на відео: Активний вибір споживання контенту у форматі, що вимагає більше часу.

• Збереження статті в закладки або «читати пізніше»: Пряме свідчення високої суб'єктивної цінності матеріалу для користувача.

• Активна участь в опитуванні або голосуванні в кінці статті.

Висновок: Від теорії до практики. Для сучасного журналіста чи редактора розуміння поведінкових даних перестає бути сферою ви-

нятково технічних фахівців. Це навичка медіаграмотності, що дозволяє:

1. Оцінювати реальний вплив контенту за глибокими метриками (Dwell Time, Scroll Depth), а не лише за поверхневими показниками переглядів.

2. Діагностувати проблеми матеріалів: чи дійсно статтю дочитали до кінця, де саме втрачається увага.

3. Будувати ефективні шляхи залучення (User Journey) через грамотну внутрішню перелинковку, що стимулює рециркуляцію.

4. Створювати контент, орієнтований на залученість, де мікроконверсії стають маркерами успіху.

Таким чином, поведінкові дані — це об'єктивне дзеркало, у яке може (і має) дивитися кожен медіапрофесіонал, щоб розмовляти з аудиторією тією ж мовою, якою вона діє.

5.1.2. Демографічні та психографічні дані.

Ефективна персоналізація в сучасних медіа вимагає глибокого розуміння аудиторії, що виходить далеко за межі простих статистичних ярликів. Для цього необхідно поєднати два фундаментальні типи даних: демографію, яка формує базовий «скелет» або портрет користувача, та психографію, що розкриває його внутрішній «характер», мотивацію та цінності. Лише синтез цих двох підходів дозволяє створювати по-справжньому релевантний та цінний контент.

Демографія: статичний портрет «хто вони?». Демографічні дані — це об'єктивні, статистично перевірені характеристики населення. Вони описують соціальну та економічну позицію користувача і залишаються відносно статичними. До ключових параметрів належать вік, стать, географія, рівень доходу, освіта, професія та сімейний стан. У медіаіндустрії демографія традиційно використовується для двох основних цілей: прецизійного таргетингу реклами (наприклад, показу оголошень про елітні авто чоловікам 40+ із столиці) та локалізації контенту (автоматичне налаштування новинної стрічки користувача згідно з його містом або регіоном). Однак основний недолік демографічного підходу в тому, що він більше не є визначальним фактором культурного споживання. Реальність суперечить узагальненим ярликам: 60-річна жінка може бути активним глядачем кіберспортивних турнірів, тоді як 20-річний хлопець — захопленим читачем блогів про в'язання. Демографія

задає рамки, але не відповідає на найважливіше питання — про мотивацію та інтереси.

Психографія: динамічна душа «чому вони?». Саме для відповіді на це питання використовується психографія — вивчення особистості, цінностей, ставлень, інтересів та стилю життя (Lifestyle). Вона розкриває не «хто» користувач, а «чому» він робить той чи інший вибір. Класичною моделлю для аналізу є модель AIO (Activities, Interests, Opinions), що охоплює:

• Діяльність (Activities): Професійна активність, хобі, спортивні та соціальні практики.

• Інтереси (Interests): Сфери стійкої уваги, такі як кулінарія, мода, технології, політика.

• Думки (Opinions): Переконання та ставлення до соціальних явищ, таких як сталий розвиток, гендерна рівність, економічна політика.

Практичне застосування психографії в медіа полягає в створенні тонких, змістовних рекомендацій. Знаючи, що користувач глибоко цінує ідеї сталого розвитку, система може рекомендувати йому репортаж про інновації в переробці сміття, навіть якщо за демографічним профілем (наприклад, бухгалтер 50 років) він не відповідає стереотипному образу молодого еко-активіста. Таким чином, психографія долає обмеження демографії, дозволяючи будувати зв'язок на основі спільних цінностей, а не зовнішніх ознак.

Яскравий приклад обмежень демографії: Оззі Осборн та король Чарльз III. Ілюстрацією непрацездатності суто демографічного підходу є класичний приклад двох чоловіків, які за всіма формальними параметрами виглядають ідентично: обидва народилися в 1948 році у Великій Британії, одружувались двічі, мають двох дітей, надзвичайно заможні, відомі та люблять собак. Демографічний алгоритм віднесе їх до одного сегменту та запропонує їм ідентичний контент, наприклад, квитки на оперу чи анонси класичних вистав. Однак реальність полягає в тому, що один із них — король Чарльз III, а інший — рок-зірка Оззі Осборн. Рекомендація опери, імовірно, буде успішною для першого та абсолютно провальною для другого. Цей приклад наочно демонструє, що справжня персоналізація стає можливою лише тоді, коли ми доповнюємо демографічний «скелет» психографічною «душею», яка розрізнить монарха та рок-музиканта.

Методи збору даних: від прямої розмови до аналізу поведінки. Для формування такого комплексного портрету використовуються два основні методи. Перший — декларативний, який передбачає пряме звернення до користувача через опитування, анкети під час реєстрації (онбординг) або форми зворотного зв'язку («Які теми вас цікавлять?»). Другий, значно потужніший та масштабований — інференційний метод (inferential). Він ґрунтується на виведенні психографічних характеристик із цифрової поведінки. Наприклад, якщо система аналізує, що читач постійно відкриває статті про венчурні інвестиції, стартапи та інтерв'ю з IT-підприємцями, вона може зробити висновок про наявність психотипу «підприємливий», «орієнтований на кар'єрний успіх» або «новатор», і на основі цього будувати подальші рекомендації.

Висновок: синергія для глибокого розуміння. Таким чином, демографія та психографія не є конкуруючими підходами, а взаємодоповнюючими рівнями аналізу. Демографія дає відповідь на питання «Хто?», задаючи контекст, тоді як психографія відповідає на питання «Чому?», розкриваючи мотивацію. Для сучасного медіапрофесіонала вміння працювати з цими даними означає можливість рухатися від масового охоплення до осмисленого залучення, створюючи контент, який резонує не зі статистичною категорією, а з живою людиною, її інтересами та цінностями. Це основа для побудови тривалої лояльності в епоху надлишку інформації.

5.1.3. Дані з соціальних мереж.

Соціальні мережі як ключ до цифрового профілю. Соціальні платформи (Facebook, X/Twitter, LinkedIn, TikTok) сьогодні є найбагатшими архівами цифрових досьє на мільярди користувачів. Вони зберігають не лише базову демографію, а й детальні карти соціальних зв'язків, інтересів, переконань та моделей поведінки. Для медіа організацій легальний та етичний доступ до частини цих даних (за згодою користувача) відкриває потужні можливості, зокрема для миттєвого вирішення проблеми «холодного старту» — ситуації, коли про нового користувача або контент ще немає жодних внутрішніх даних. Використання цих зовнішніх сигналів дозволяє з першого кліку створювати персоналізований досвід.

Соціальна автентифікація (Social Login / SSO): шлюз для даних

і довіри. Кнопка «Увійти через Facebook/Google» (Single Sign-On, SSO) — це не просто зручність для користувача, якому не потрібно запам'ятовувати нові паролі. Це фундаментальний механізм обміну даними між платформами. Для медіа це означає миттєве отримання верифікованого профілю (ім'я, email, аватар), що різко знижує ризик бот-активності та фальшивих акаунтів. За явної згоди користувача додаток може отримати доступ до розширених даних: списку друзів (соціальний граф), сторінок і публікацій, які він вподобав (граф інтересів), демографічної інформації. Це дає можливість з першої ж сесії побудувати базовий, але вже осмислений, портрет читача, пропонуючи йому контент, релевантний його заявленим у соцмережах інтересам, і тим самим утримати його увагу.

Соціальний граф та граф інтересів: дві картини однієї людини. Ключовим розрізненням у роботі з соціальними даними є розділення двох типів зв'язків: Social Graph (Соціальний граф) та Interest Graph (Граф інтересів).

• Соціальний граф відповідає на питання «Хто твої друзі?». Він будується на гіпотезі гомофілії («схожий тягне до схожого»): якщо п'ятеро ваших найближчих друзів у Facebook регулярно читають або обговорюють матеріали про криптовалюти, з високою ймовірністю ця тема може бути цікавою і вам, навіть якщо ви ще жодного разу не шукали її на медіа-сайті. Це дозволяє робити контекстні рекомендації на основі соціального оточення.

• Граф інтересів відповідає на питання «Що тобі подобається?». Він формується на основі явних дій користувача в відкритих платформах, таких як X (колишній Twitter) або TikTok: підписки на конкретних осіб (Ілон Маск, NASA), використання хештегів, лайки та репости. Якщо ваш профіль у Twitter показує стійкий інтерес до космонавтики та нових технологій, медіа-сайт може автоматично налаштувати вашу домашню стрічку на відповідні теми, пропонуючи статті про запуск ракет або інновації в галузі акумуляторів, ще до того, як ви прочитаєте на цьому сайті хоча б один матеріал.

Lookalike-аудиторії: від маркетингу до редакційних стратегій. Найпотужнішим інструментом, що перейшов із сфери таргетованої реклами в редакційну та продуктову роботу, є технологія Lookalike Audiences (аудиторії, схожі на цільові). Її логіка геніально проста та ефективна. Припустимо, у медіа є ядро найбільш лояльних

користувачів — тисяча людей, які регулярно донатять на підтримку видання, купують преміум-підписку або проводять на сайті надзвичайно багато часу. Завантаживши анонімізовані (хешовані) контакти цих користувачів (наприклад, email-адреси) у рекламний кабінет Facebook або іншої платформи, медіа запускає процес машинного навчання. Алгоритм платформи аналізує профілі цих «ідеальних читачів», знаходячи сотні спільних, часто неочевидних рис: не лише явні інтереси (джаз, італійська кухня), але й поведінкові патерни, демографічні збіги, географічну близькість. Після цього система сканує мільйони інших профілів і знаходить 10, 50 або 100 тисяч користувачів, які максимально схожі на вихідне ядро, але ще не знайомі з виданням. Це дозволяє будувати стратегію розширення аудиторії не навмання, а на основі даних про вже існуючу лояльність.

Обмеження та нова ера: після Cambridge Analytica. До 2018 року доступ до даних соціальних мереж, зокрема через відкриті API Facebook, був значно вільнішим, що дозволяло збирати навіть дані друзів користувача без їхньої прямої згоди. Скандал навколо Cambridge Analytica, яка використала ці дані для мікротаргетингу політичної реклами, став водо розділу. Після нього регулятори (наприклад, через GDPR) та самі платформи різко обмежили прямий доступ до детальних соціальних графів. Це змусило медіа та маркетологів змістити фокус з прямого парсингу профілів на Social Listening (соціальне прослуховування) та роботу з агрегованими, анонімізованими даними. Замість того щоб отримувати персональні списки друзів, медіа тепер аналізують публічні розмови, тренди, згадки брендів та реакції на контент у відкритих джерелах, що, з одного боку, зберігає приватність, а з іншого — вимагає більш витончених інструментів аналітики.

Висновок. Дані соціальних мереж залишаються критично важливим ресурсом для розуміння аудиторії та подолання проблеми холодного старту, проте способи роботи з ними кардинально змінилися. Від моделей, що спиралися на детальний соціальний граф, індустрія перейшла до комбінації: легального отримання даних через згоду (SSO), використання графа інтересів для миттєвої персоналізації, потужних, але непрямих методів розширення аудиторії (Lookalike) та пасивного Social Listening для вивчення публічного дискурсу. Для медіа це означає необхідність будувати

довіру з користувачем (щоб отримати його згоду на обмін даними) та інвестувати в аналітичні системи, здатні трансформувати агреговані соціальні сигнали в практичні редакційні та продуктові рішення.

5.1.4. Контекстуальні дані: геолокація, пристрій, час.

У сучасній медіаіндустрії персоналізація без розуміння контексту є сліпою. Ключовим принципом є те, що одна й та сама людина — наприклад, бізнесмен — має кардинально різні інформаційні потреби та моделі поведінки залежно від ситуації. О 8:00 ранку, перебуваючи в метро зі смартфоном у руці, його увага розсіяна, а час обмежений; контент має бути стислим і яскравим. О 21:00 вечора, розслабившись на домашньому дивані з планшетом, він готовий до глибшого, довшого занурення в довгі статті або документальні фільми. Контекстуальні дані — геолокація, тип пристрою та час доби — це те, що перетворює абстрактний профіль користувача на живого людину в конкретній ситуації, дозволяючи медіа адаптувати контент, його формат і тон не лише під людину, а й під момент її життя.

Знаючи, де фізично знаходиться користувач, глобальне медіа може стати його особистим гіперлокальним помічником. Точність визначення залежить від технології. IP-адреса дає зрозуміле уявлення про місто та країну, що достатньо для автоматичного підключення релевантних регіональних новин чи мовного інтерфейсу. GPS забезпечує точність до кількох метрів, але вимагає явного дозволу користувача; це основа для сервісів, де місцезнаходження критичне, як-от карти чи прогноз погоди. Окремим інструментом є технологія Geofencing (Геопаркан) — віртуальний периметр, який активує події: коли користувач заходить у зону аеропорту, додаток може автоматично запропонувати завантажити статті для офлайн-читання. Найкращим прикладом використання є гіперлокальні новини: користувачеві на Хрещатику буде набагато цікавіше отримати сповіщення про ярмарок саме на цій вулиці, а не загальну новину про подію в Києві. Це перетворює мас-медіа на персонального супутника.

Тип пристрою, з якого користувач заходить на сайт чи в додаток, безпосередньо вказує на його психофізіологічний стан і готовність до споживання інформації. Мобільний пристрій (смартфон) — це режим «перекусу» (Snacking): сесії короткі, користувач часто в русі. Контент має бути оптимізований під вертикальне споживання — короткі абзаци, булети, великі кнопки, перевага коротким відео. Десктоп

(ноутбук або ПК) — це режим «нахилу вперед» (Lean-forward), що передбачає робочу концентрацію; тут користувач готовий до довгого читання (лонгрідів), детальної інфографіки та складних інтерфейсів, з піком активності в робочий день. Планшет або Smart TV, навпаки, — це режим «відкидання назад» (Lean-back) для вечірнього відпочинку; контент має бути спрямований на пасивне, але тривале споживання: документальні фільми, фоторепортажі з мінімальною кількістю кліків.

Стратегія дейпартингу (Dayparting) полягає в автоматичній адаптації контентної пропозиції під типові потреби користувача в різний час доби. Ранок (7:00–10:00) — час поспіху; ідеальний формат це «бриф»: стислі дайджести головних подій, прогноз погоди, мінімум аналітики. Обід (12:00–14:00) — пауза в роботі; доречний легкий розважальний контент або короткі відео. Вечір (19:00–23:00) — основний час для особистого споживання; час для глибокої аналітики, довгих інтерв'ю та документальних проектів. На вихідних ритм сповільнюється, тому пріоритет отримують матеріали в стилі «повільного читання» (Slow reading) про подорожі, кулінарію та культуру.

Таким чином, контекстуальна персоналізація на основі геолокації, пристрою та часу — це не технологічна надмірність, а необхідність для виживання в конкурентному медіа-просторі. Вона дозволяє трансформувати універсальний контентний продукт в інтуїтивно зрозумілу послугу, яка передбачає потреби користувача в конкретний момент. Синтез цих трьох типів даних є основою для побудови справжнього, глибокого залучення, коли медіа стає не безособовим джерелом, а розумним і чуйним супутником у повсякденному житті.

5.1.5. Явний фідбек: лайки, дизлайки, коментарі, шерінг.

Явний фідбек (Explicit Feedback) представляє собою категорію даних, що виникає в результаті свідомої та цілеспрямованої дії користувача, спрямованої на безпосереднє вираження його ставлення до контенту. На відміну від пасивно зібраних поведінкових метрик, такі дії — лайк, дизлайк, коментар, поширення — є інтенційним сигналом, який, однак, потребує глибокої та критичної інтерпретації. Його цінність для алгоритмічної системи полягає не лише у факті взаємодії, але й у розумінні контексту, мотивації та потенційних спотворень, які такі сигнали можуть нести.

Фундаментальним обмеженням, що ставить під сумнів надійність

побудови системи виключно на явному фідбеку, є правило «1-9-90», яке описує нерівність участі у цифрових спільнотах. Згідно з цим правилом, лише приблизно 1% користувачів активно створює оригінальний контент або залишає розгорнуті коментарі; 9% беруть участь у легших формах взаємодії, таких як лайки та репости; а переважна 90% аудиторії є «луалерами» (lurkers) — мовчазними спостерігачами, які споживають інформацію, не залишаючи явних слідів. Отже, алгоритм, який оптимізує рекомендації виключно під голосну активну меншину (10%), ризикує створити інформаційне середовище, що системно ігнорує інтереси та уподобання мовчазної більшості. Це може призвести до посилення поляризації та формування стрічки, яка видається «цікавою» лише вузькому колу найбільш голосних учасників.

Через це критично важливим для розробників є розуміння ієрархії цінності дій (Signal Weight), де кожен тип взаємодії має різну інформаційну вагу та ступінь довіри. Найнижчу вагу традиційно несе лайк (Like). Ця дія є найлегшою для виконання і часто носить ритуальний або перформативний характер («performative liking»), коли користувач відзначає контент, щоб проєктувати певний образ (наприклад, інтелектуала), навіть не занурюючись у матеріал. На противагу цьому, дизлайк або дія «Приховати» (Dislike / Hide) має найвищу негативну вагу. Це сильний сигнал негативної реакції, адже користувач доклав свідомі зусилля, щоб видалити контент зі свого поля зору. Для алгоритму це має бути «сигналом до знищення» (Kill signal), який вимагає миттєвої корекції рекомендацій, щоб уникнути втрати користувача.

Значно вищу цінність мають більш складні форми фідбеку. Коментар (Comment) свідчить про глибоке емоційне або інтелектуальне залучення, проте його аналіз вимагає від системи проведення аналізу тональності (Sentiment Analysis). Сто коментарів під публікацією можуть вказувати як на її виняткову якість та здатність породжувати дискусію, так і на скандальний, образливий характер, що викликав хвилю обурення. Найвищу позитивну вагу та рівень довіри несе дія шерингу або репосту (Share / Repost). Поширюючи контент у своєму профілі, користувач робить репутаційну ставку, імпліцитно рекомендуючи його своєму колу як щось, що відображає його власні цінності або погляди. Окремим, найціннішим, але й

найскладнішим для відстеження явищем є «темний соціал» (Dark Social) — поширення посилань через приватні канали комунікації, такі як месенджери (Telegram, WhatsApp) або email. Для вебаналітики такий трафік часто виглядає як «прямий» (Direct Traffic), що маскує справжнє соціальне підґрунтя його поширення. Відстежити його можна лише за допомогою спеціальних технологічних рішень, наприклад, унікальних посилань з параметрами або кнопок «Скопіювати посилання», що дозволяє фіксувати факт поділу.

Розуміння мотивації, що стоїть за явним фідбеком, є завершальним ланцюжком у його інтерпретації. Користувачі взаємодіють з контентом не лише через його зміст, але й для реалізації власних психологічних та соціальних потреб. Побудова ідентичності (Identity building) проявляється, коли людина поширює новини про екологію, щоб підкреслити свою приналежність до певної спільноти та світогляд. Зміцнення зв'язків (Relationship bonding) — це мотив, коли мем або стаття ділиться з конкретною людиною для підтримання спільного контексту та емоційного резонансу. Альтруїзм (Altruism) спонукає поширювати корисні інструкції або важливі новини з бажання принести суспільну користь. Таким чином, кожен явний сигнал — це не просто точкове вподобання, а складний соціально-психологічний акт, розкодування якого є ключем до створення по-справжньому релевантних, глибоко персоналізованих рекомендаційних систем, здатних чути не лише голосних, але й розуміти потреби мовчазної більшості.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання розкривають механіку збору інформації платформами:

1. Явні vs Неявні дані: Ми знаємо, що соцмережа бачить наші лайки (явні дані). Але чому для алгоритму *неявні дані* (швидкість скролінгу, затримка пальця на фото, час доби, рівень заряду батареї) часто є важливішими для прогнозування поведінки?

2. Позаплатформні дані (Off-Facebook Activity): Як працюють трекінгові пікселі (наприклад, Meta Pixel), розміщені на сайтах новин чи інтернет-магазинів? Як Facebook дізнається, що ви купили кросівки на зовнішньому сайті, щоб показати рекламу шкарпеток?

3. Соціальний граф (Social Graph): Що таке "граф" у контексті баз даних? Як аналіз зв'язків (хто ваші друзі, друзі друзів) дозволяє

передбачити ваші політичні погляди чи рівень доходу, навіть якщо ви про це не пишете?

4. API як шлюз: Що таке API (Application Programming Interface) соціальних мереж? Чому закриття публічного доступу до API Твіттера (X) та Reddit у 2023 році стало катастрофою для академічних дослідників та дата-журналістів?

5. Тіньові профілі (Shadow Profiles): Чи збирає Facebook дані про людей, які *не* зареєстровані в соцмережі? Як це відбувається через список контактів їхніх друзів?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання спрямовані на аудит власного цифрового сліду:

Завдання 1. "Полювання на Піксель" (Pixel Hunting)

• Інструмент: Встановіть розширення для браузера "Meta Pixel Helper" (або аналог).

• Дія: Зайдіть на 5 популярних українських новинних сайтів та 5 інтернет-магазинів.

• Аналіз:

- Чи фіксує розширення наявність трекера Meta?

- Які саме події він передає? (PageView — перегляд сторінки, AddToCart — додавання в кошик).

• Висновок: Поясніть, як медіа використовують ці дані для ретаргетингу (показу новин тим, хто вже був на сайті).

Завдання 2. Аудит архіву даних (Download Your Data)

• Дія: Замовте вивантаження архіву своїх даних з Instagram або Facebook (це робиться в налаштуваннях приватності). *Примітка: архів може готуватися кілька годин, тому це домашнє завдання.*

• Аналіз файлів (JSON/HTML):

- Знайдіть папку "Ads Interests" (Рекламні інтереси). Наскільки точний цей список?

- Знайдіть історію геолокацій.

• Рефлексія: Чи здивував вас обсяг збереженої інформації (наприклад, видалені повідомлення або історія пошуку)?

Завдання 3. Симуляція роботи Graph API

• Ситуація: Ви — журналіст, який хоче проаналізувати реакцію на скандальний пост.

• Завдання: Складіть структуру даних, яку ви б хотіли отримати

від API (у форматі таблиці):

- *Поля:* ID користувача, Текст коментаря, Час (Timestamp), Кількість лайків на коментарі, Тип реакції (Сміх/Злість).

• Питання: Які з цих даних соцмережі віддають охоче, а які приховують (наприклад, список усіх, хто поширив пост) і чому?

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Бріттані Кайзер. *Targeted. (Таргетовані).* (Сповідь колишньої працівниці Cambridge Analytica про те, як дані використовуються для політичних маніпуляцій).

2. Documentation for Meta for Developers (Graph API). (Технічна документація, що показує, які саме поля даних доступні розробникам).

3. Shoshana Zuboff. *The Age of Surveillance Capitalism.* (Розділи про вилучення поведінкового надлишку).

4. Matthew Hindman. *The Internet Trap.* (Про економіку даних та увагу).

4. ГЛОСАРІЙ ТЕРМІНІВ

• API (Application Programming Interface): Програмний інтерфейс, який дозволяє двом додаткам (наприклад, сайту новин і Facebook) обмінюватися даними. Для журналістів це був основний спосіб автоматизованого збору даних для аналізу.

• Трекінговий піксель (Tracking Pixel): Прозоре зображення розміром 1x1 піксель (або шматок коду JavaScript), розміщене на вебсайті, яке дозволяє соцмережам відстежувати дії користувачів поза межами самої платформи.

• Соціальний граф (Social Graph): Структура даних, що моделює зв'язки між об'єктами в соцмережі (людьми, сторінками, групами). "Вузли" — це люди, "ребра" — це дружба або підписка.

• Парсинг / Скрейпінг (Scraping): Технологія автоматизованого збору даних з веб-сторінок (імітуючи дії людини), яка використовується, коли офіційний API недоступний або платний. Часто знаходиться в "сірій" правовій зоні.

• Поведінковий надлишок (Behavioral Surplus): Термін Ш. Зубофф; дані про поведінку користувача (голос, міміка, рухи мишкою), які не потрібні для покращення сервісу, але збираються для продажу рекламодавцям.

5.2. Методи збору даних у digital-маркетингу

5.2.1. Web Analytics (Google Analytics, Adobe Analytics).

Web-аналітика є фундаментальним методом збору даних про поведінку користувачів на веб-сайтах і в додатках. Провідні платформи, такі як Google Analytics та Adobe Analytics, надають комплексні інструменти для відстеження та аналізу цифрової активності.

Google Analytics залишається найпоширенішою безкоштовною платформою для веб-аналітики, яка дозволяє відстежувати відвідування сайту, джерела трафіку, поведінку користувачів, конверсії та багато іншого. Сучасна версія Google Analytics 4 (GA4) використовує подієву модель даних, що дозволяє гнучкіше налаштовувати відстеження та краще розуміти шлях користувача через різні платформи та пристрої. Система автоматично збирає базові метрики, такі як перегляди сторінок, тривалість сесій, показники відмов, географічні дані та технічні характеристики пристроїв відвідувачів.

Adobe Analytics представляє професійне enterprise-рішення з розширеними можливостями сегментації, атрибуції та прогнозної аналітики. Ця платформа особливо популярна серед великих корпорацій завдяки потужним інструментам для роботи з великими обсягами даних, можливостям реального часу та глибокої інтеграції з іншими продуктами Adobe Experience Cloud. Adobe Analytics пропонує складніші моделі атрибуції, що допомагають визначити справжній внесок кожного маркетингового каналу в конверсії.

Обидві платформи використовують код відстеження (тег), який інтегрується на всі сторінки веб-сайту і збирає дані про взаємодію користувачів, передаючи їх на сервери для обробки та аналізу. Важливим аспектом сучасної веб-аналітики є дотримання вимог конфіденційності, включно з GDPR та іншими регуляціями, що вимагає отримання згоди користувачів на збір даних.

5.2.2. Event Tracking та користувацькі події.

Event tracking або відстеження подій дозволяє збирати дані про конкретні дії користувачів, які не відстежуються автоматично стандартними налаштуваннями аналітики. Це включає кліки по кнопках, завантаження файлів, перегляди відео, взаємодії з формами,

скролінг сторінки та будь-які інші важливі користувацькі взаємодії.

Налаштування користувацьких подій дозволяє маркетологам та аналітикам отримувати точні дані про те, як користувачі насправді взаємодіють з контентом і функціоналом сайту. Наприклад, можна відстежувати, скільки користувачів натискають на певну кнопку Call-to-Action, на якому етапі форми реєстрації вони зупиняються, який відсоток відео переглядається до кінця, або які елементи інтерактивного контенту викликають найбільше зацікавлення.

Сучасні системи тег-менеджменту, такі як Google Tag Manager або Adobe Launch, спрощують процес налаштування подій без необхідності постійного втручання розробників. Це дозволяє маркетинговим командам швидко адаптувати відстеження до змін у стратегії чи дизайні сайту. Користувацькі події можуть включати параметри та значення, що надають додатковий контекст – наприклад, не просто "клік на продукт", а "клік на продукт категорії X ціною Y".

Правильно налаштоване відстеження подій створює детальну картину користувацького шляху, виявляє точки тертя в процесі конверсії та допомагає приймати обґрунтовані рішення щодо оптимізації. Важливо планувати структуру подій заздалегідь, створюючи логічну та послідовну систему найменувань і категорій, що полегшує подальший аналіз і звітність.

5.2.3. Опитування та дослідження аудиторії.

Опитування та дослідження аудиторії надають якісні дані, які доповнюють кількісні метрики з аналітичних систем. Цей метод дозволяє зрозуміти мотивації, потреби, очікування та задоволеність користувачів безпосередньо від них самих, розкриваючи "чому" за цифрами.

Онлайн-опитування можуть проводитися через різні канали: на самому веб-сайті (pop-up опитування, embedded форми), електронною поштою, через соціальні мережі або спеціалізовані платформи дослідження. Інструменти, такі як SurveyMonkey, Typeform, Google Forms або Qualtrics, спрощують створення, розповсюдження та аналіз опитувань. Важливо ретельно розробляти питання, уникаючи упередженості та забезпечуючи чіткість формулювань, щоб отримати надійні результати.

Різні типи опитувань служать різним цілям: опитування задоволеності (CSAT) вимірюють задоволення конкретною взаємодією або

продуктом; Net Promoter Score (NPS) оцінює загальну лояльність і ймовірність рекомендації; опитування про зручність використання виявляють проблеми з інтерфейсом; дослідження потреб допомагають розробляти нові продукти або функції.

Глибинні інтерв'ю та фокус-групи надають ще більше контексту, дозволяючи досліджувати складні теми та отримувати розгорнуті відповіді. Хоча вони вимагають більше ресурсів, ці методи розкривають нюанси, які неможливо зафіксувати стандартними опитуваннями. Користувацьке тестування, коли учасники виконують конкретні завдання на сайті під спостереженням дослідників, виявляє проблеми зручності використання в реальному часі.

Важливим аспектом є вибіркове дослідження – забезпечення того, щоб респонденти представляли цільову аудиторію. Також необхідно враховувати потенційну упередженість відповідей: люди, які беруть участь в опитуваннях, можуть відрізнятися від тих, хто їх ігнорує. Комбінування різних методів дослідження та перехресна перевірка результатів з іншими джерелами даних допомагають створити більш повну і точну картину.

5.2.4. Соціальне слухання (Social Listening).

Соціальне слухання являє собою процес моніторингу та аналізу розмов і згадок про бренд, продукти, конкурентів та галузеві теми в соціальних мережах та інших онлайн-каналах. Цей метод надає безцінну інформацію про сприйняття бренду, настрої аудиторії та актуальні теми в реальному часі.

Спеціалізовані платформи для social listening, такі як Brandwatch, Sprout Social, Hootsuite Insights, Mention або Talkwalker, автоматизують збір та аналіз величезних обсягів даних з платформ, таких як Twitter, Facebook, Instagram, LinkedIn, Reddit, форумів, блогів, новинних сайтів та коментарів. Ці інструменти використовують алгоритми обробки природної мови (NLP) для визначення тональності (позитивна, негативна, нейтральна) та виявлення ключових тем і трендів.

Соціальне слухання дозволяє відстежувати кілька важливих аспектів: згадки бренду та їх тональність, що показує загальне сприйняття компанії; обговорення конкурентів, що виявляє їхні сильні та слабкі сторони очима споживачів; скарги та проблеми клієнтів, які потребують уваги служби підтримки; виникаючі тренди

та теми, що цікавлять цільову аудиторію; потенційних амбасадорів бренду та інфлюенсерів; кризові ситуації на ранніх етапах, що дозволяє швидко реагувати.

Ефективне соціальне слухання виходить за межі простого підрахунку згадок – воно передбачає глибокий контекстуальний аналіз розмов, розуміння емоцій і мотивацій аудиторії. Результати можуть впливати на різні аспекти маркетингової стратегії: розробку продуктів, створення контенту, управління репутацією, обслуговування клієнтів, стратегію конкурентного позиціонування.

Важливо налаштовувати правильні ключові слова, хештеги та параметри пошуку, щоб відфільтрувати нерелевантний шум і зосередитися на значущих сигналах. Регулярний моніторинг та створення аналітичних звітів допомагає виявляти закономірності та своєчасно реагувати на зміни в суспільних настроях.

5.2.5. Інтеграції та API третіх сторін.

Інтеграції та API (Application Programming Interface) третіх сторін дозволяють об'єднувати дані з різних джерел і систем, створюючи цілісну картину маркетингової ефективності та поведінки клієнтів. Це критично важливо в сучасному маркетинговому ландшафті, де дані розосереджені між численними платформами та інструментами.

API забезпечують програмний доступ до даних і функціональності різних сервісів. Наприклад, API соціальних мереж (Facebook, Instagram, LinkedIn, Twitter) дозволяють витягувати дані про ефективність публікацій, зростання аудиторії, залученість і демографію підписників. API рекламних платформ (Google Ads, Facebook Ads, TikTok Ads) надають детальну інформацію про витрати, покази, кліки, конверсії та ROI кампаній. CRM-системи через API можуть обмінюватися даними про клієнтів з маркетинговими платформами, забезпечуючи персоналізацію на основі історії взаємодій.

Платформи інтеграції, такі як Zapier, Make (раніше Integromat), або Segment, спрощують підключення різних інструментів без необхідності писати код. Вони дозволяють автоматизувати передачу даних між системами: наприклад, автоматично додавати лідів з лендингу в CRM, синхронізувати дані про email-кампанії з аналітичною платформою, або оновлювати рекламні аудиторії на основі поведінки користувачів на сайті.

Для більш складних потреб створюються custom інтеграції,

які розробляються програмістами спеціально під вимоги бізнесу. Це дозволяє створювати унікальні потоки даних, об'єднувати інформацію з внутрішніх систем компанії (ERP, складських систем, систем обслуговування) з маркетинговими даними, забезпечуючи глибокий аналіз і автоматизацію.

Data warehouses або сховища даних, такі як Google BigQuery, Amazon Redshift або Snowflake, часто використовуються як централізовані репозиторії, куди через API збираються дані з усіх джерел. Це дозволяє проводити комплексний аналіз, створювати детальні звіти та будувати моделі машинного навчання для прогнозування та оптимізації.

Ключові переваги інтеграцій включають усунення ручного введення даних і пов'язаних помилок, економію часу на збір і консолідацію інформації, можливість створення єдиного джерела правди для всіх команд, швидший доступ до інсайтів і можливість автоматизувати маркетингові процеси на основі даних з різних систем. При цьому важливо забезпечувати безпеку даних, дотримуватися вимог конфіденційності та регулярно перевіряти працездатність інтеграцій.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допомагають структурувати знання про джерела інформації:

1. Ієрархія даних (Zero, First, Second, Third):

- *Zero-party:* Чому дані, які користувач вводить добровільно (наприклад, пройшовши квіз "Яка ти піца"), вважаються "золотим стандартом" маркетингу?

- *First-party:* Як медіа збирають дані про поведінку на власному сайті (історія переглядів)?

- *Third-party:* Чому купівля баз даних у брокерів (третіх сторін) стає незаконною та неефективною (Cookiepocalypse)?

2. Технологія Cookies: У чому різниця між Session Cookies (зникають після закриття вкладки) та Persistent Cookies (живуть місяцями, відстежуючи повернення користувача)? Як працюють "суперкуки" (Fingerprinting), які неможливо видалити?

3. Тегування (Tagging): Що таке Google Tag Manager (GTM)? Як контейнери тегів дозволяють маркетологам збирати дані про кліки

по конкретних кнопках без допомоги програмістів?

4. Соціальний логін (Social Login): Коли ви натискаєте "Увійти через Facebook" на сайті новин, які саме дані передаються? Чому це вигідно і сайту (верифікація), і соцмережі (збагачення профілю)?

5. Веб-скрейпінг (Web Scraping): Коли маркетологи використовують ботів (наприклад, Python-скрипти) для збору цін конкурентів або коментарів під постами? Де проходить межа між легальним парсингом і порушенням правил платформи?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання спрямовані на опанування інструментів збору:

Завдання 1. Проєктування стратегії Zero-party Data

• Ситуація: Google скасовує сторонні куки. Вашому медіа потрібно дізнатися інтереси читачів напряму.

• Завдання: Розробіть механіку збору даних (Lead Magnet).

- *Формат:* (Квіз / Опитування / Доступ до закритого контенту).

- *Питання:* Які 3 питання ви поставите користувачу, щоб краще персоналізувати йому контент? (Наприклад: "Як часто ви читаєте новини: зранку чи ввечері?").

- *Винагорода:* Що користувач отримає взамін? (PDF-звіт, знижку, бейдж).

Завдання 2. Робота з UTM-мітками

• Інструмент: Campaign URL Builder (Google).

• Завдання: Ви запускаєте статтю у Facebook, Telegram та Email-розсилці. Створіть 3 унікальні посилання з UTM-мітками, щоб відстежити джерело трафіку.

- utm_source (facebook / telegram / newsletter)

- utm_medium (social / email)

- utm_campaign (spring_promo)

• Аналіз: Поясніть, як ці "хвости" в посиланнях дозволяють аналітиці (Google Analytics) зрозуміти, звідки прийшла людина, без використання складних шпигунських скриптів.

Завдання 3. Аудит Cookie Banner (Згода на збір)

• Дія: Зайдіть на сайт великого європейського медіа (наприклад, BBC або DW) в режимі "Інкогніто".

• Аналіз: Прочитайте спливаюче вікно про куки.

- Які категорії можна вимкнути? (Marketing, Analytics, Functional).

- Які категорії є "Strictly Necessary" (Обов'язкові) і чому сайт не працюватиме без них?

• Висновок: Як GDPR змінив дизайн цих вікон? (Раніше була просто кнопка "ОК", тепер — складне меню вибору).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Google Analytics Academy. *Data Collection and Processing.* (Базовий курс про те, як працюють лічильники).

2. Brad Geddes. *Advanced Google AdWords.* (Розділи про ремаркетинг та збір аудиторій).

3. GDPR Official Guidelines. (Розділ про згоду користувача на обробку даних).

4. Interactive Advertising Bureau (IAB). *The Data Primer.* (Керівництво індустрії щодо типів даних).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Zero-party Data: Дані, які клієнт навмисно та проактивно надає бренду. Відрізняються від First-party даних тим, що це не "підглядання" за поведінкою, а прямий діалог (наприклад, налаштування вподобань у профілі).

• Cookies (Куки): Невеликі текстові файли, що зберігаються браузером. Дозволяють сайтам "запам'ятовувати" відвідувача. Сторонні куки (Third-party) дозволяють рекламним мережам стежити за користувачем між різними сайтами.

• UTM-мітка (Urchin Tracking Module): Спеціальний параметр, що додається до URL-адреси. Дозволяє системам аналітики точно ідентифікувати джерело трафіку (з якої саме рекламної кампанії прийшов клієнт).

• GTM (Google Tag Manager): Система керування тегами, що дозволяє маркетологам швидко оновлювати коди відстеження на вебсайті (пікселі, лічильники) без необхідності втручання в вихідний код сторінки програмістами.

• Fingerprinting (Зняття відбитка): Метод ідентифікації користувача за унікальною комбінацією налаштувань його браузера (роздільна здатність екрану, встановлені шрифти, версія ОС), навіть якщо він видалив cookies.

5.3. Аналітичні техніки у digital-маркетингу

5.3.1. Сегментація аудиторії.

Сегментація аудиторії є не просто технікою, а базовим принципом сучасного маркетингу, який визначає весь підхід до комунікації з ринком. Вона ґрунтується на простій, але глибокій ідеї: різні групи людей сприймають один і той самий продукт або послугу по-різному через відмінності в своїх життєвих обставинах, психології та поведінці. Тому суть процесу полягає у поділі широкого, неоднорідного ринку на чіткі, керовані сегменти — групи споживачів, об'єднаних спільними ознаками, що дозволяють передбачити їхню реакцію на маркетингові стимули. Мета такого поділу — прагматична: замість розпилювання сил і бюджету на всіх, компанія може сконцентруватися на найбільш перспективних або доступних групах, пропонуючи їм максимально точне рішення їхніх проблем чи задоволення потреб. Це дає змогу трансформувати абстрактний "ринок" у конкретні портрети реальних людей, яких можна зрозуміти, до яких можна знайти емоційний або раціональний підхід і яким можна ефективно продавати.

Розуміння механіки сегментації вимагає погляду на основні вектори, за якими вона проводиться. Найчастіше починають з демографічних та географічних критеріїв — це об'єктивні, легко вимірювані дані, такі як вік, стать, дохід, місце проживання. Вони дають перший, поверхневий, але важливий зріз аудиторії. Проте справжню глибину та життєздатність сегменту надає психографіка — аналіз цінностей, інтересів, способу життя та особистості людей. Саме тут бренд може знайти ті спільні "нотки", на яких будується довгострокова лояльність. Водночас, найбільш практичним індикатором часто є поведінка: як саме людина взаємодіє з брендом (чи вона постійний клієнт чи випадковий покупець), що шукає в продукті (вигоду, статус, надійність), і на якій стадії прийняття рішення вона знаходиться. Ідеальна модель сегментації зазвичай поєднує кілька цих критеріїв, створюючи багатовимірний і точний профіль.

Перехід від теорії до практики реалізується через чіткий процес, який починається зі збору даних — через дослідження, аналітику, опитування. Потім ці дані групуються за обраними критеріями, формуючи попередні сегменти. Кожен такий сегмент потім ретельно

описується, часто у вигляді "персони" — яскравого персонажа з ім'ям, історією та конкретними больовими точками, що робить його зрозумілим для маркетологів і креаторів. Наступний крок — оцінка привабливості кожного сегменту з точки зору його розміру, потенціалу зростання, прибутковості та відповідності можливостям компанії. Після цього відбувається стратегічний вибір цільових сегментів — рішення, на які саме групи буде направлена вся потужність бізнесу. Фінальним акордом є розробка цільового маркетингового міксу: адаптація самого продукту, його ціни, каналів розповсюдження та комунікаційної стратегії під вимоги та сприйняття обраних сегментів. Таким чином, сегментація — це не абстрактна класифікація, а потужний стратегічний інструмент, що безпосередньо впливає на продукт, комунікацію та, в кінцевому підсумку, на успіх компанії на ринку. Вона перетворює безликє "потреби ринку" на конкретні завдання для кожного підрозділу компанії.

5.3.2. Когортний аналіз.

Когортний аналіз – це потужна аналітична техніка, яка групує користувачів на основі спільного досвіду чи характеристики в певний період часу і відстежує їхню поведінку протягом наступних періодів. Найчастіше когорти формуються на основі дати першої взаємодії або реєстрації, що дозволяє порівнювати, як різні групи користувачів поводяться з плином часу.

Класичний приклад когортного аналізу – відстеження утримання користувачів (retention). Користувачі, які зареєструвалися в січні, формують одну когорту, ті, хто приєднався в лютому – іншу, і так далі. Потім аналізується, який відсоток кожної когорти повертається через тиждень, місяць, три місяці. Це виявляє патерни: чи покращується утримання з часом (що може свідчити про покращення продукту), чи певні когорти виявляються особливо лояльними (що може вказувати на успішні маркетингові кампанії того періоду).

Когортний аналіз допомагає відповісти на критичні бізнес-питання: як змінюється якість користувачів з різних джерел трафіку з часом, чи нові функції продукту покращують довгострокову залученість, наскільки швидко користувачі досягають моменту "ага" (момент усвідомлення цінності продукту), як швидко окупаються інвестиції в залучення користувачів з різних каналів, які періоди року приносять найбільш цінних клієнтів.

Існують різні типи когорт. Часові когорти базуються на даті першої події (реєстрація, перша покупка). Поведінкові когорти об'єднують користувачів за конкретними діями (наприклад, всі, хто скористався певною функцією). Розмірні когорти групують за атрибутами, такими як джерело трафіку, тип пристрою або демографічні характеристики.

Метрики, які зазвичай аналізуються в когортах, включають retention rate (відсоток користувачів, які повертаються), revenue per cohort (дохід, згенерований кожною когортою), lifetime value (загальна цінність користувачів когорти), engagement metrics (частота та глибина взаємодії) та conversion rate (швидкість конверсії в платних клієнтів).

Візуалізація когортного аналізу часто представляється у вигляді таблиць або теплових карт, де кожен рядок представляє когорту, кожен стовпець – період часу від початкової події, а кольори або цифри показують метрику. Це дозволяє швидко ідентифікувати тренди та аномалії. Наприклад, якщо утримання раптово погіршилося для когорт, які приєдналися після певного оновлення продукту, це сигналізує про потенційну проблему, яку необхідно досліджувати.

Когортний аналіз особливо цінний для SaaS-продуктів, мобільних додатків, підписних сервісів та e-commerce, де розуміння довгострокової цінності та поведінки користувачів критично важливе для прийняття стратегічних рішень.

5.3.3. Аналіз воронки (Funnel Analysis).

Funnel analysis або аналіз воронки – це методика відстеження користувачів через послідовність кроків, які ведуть до бажаної цілі чи конверсії. Воронка візуалізує шлях користувача від початкової взаємодії до фінальної дії, виявляючи на яких етапах відбувається найбільше відсіювання і де існують найбільші можливості для оптимізації.

Типова воронка складається з кількох послідовних етапів. Наприклад, воронка e-commerce може включати: відвідування головної сторінки, перегляд категорії продуктів, перегляд сторінки конкретного продукту, додавання до кошика, початок оформлення замовлення, введення платіжної інформації, підтвердження покупки. Для SaaS-продукту воронка може виглядати так: відвідування лендингу, реєстрація на trial, перше входження в продукт, активація ключової функції, перехід на платний план.

Ключові метрики воронки включають conversion rate (відсоток користувачів, які переходять від одного етапу до наступного), drop-off rate (відсоток користувачів, які залишають процес на кожному етапі), time to convert (середній час, необхідний для проходження всієї воронки або окремих сегментів). Аналізуючи ці метрики, маркетологи виявляють "вузькі місця" – етапи з найвищим відсіюванням, які потребують першочергової оптимізації.

Існують різні типи воронок. Макро-воронки охоплюють весь шлях клієнта від першого знайомства з брендом до покупки та retention. Мікро-воронки фокусуються на конкретних процесах, таких як реєстрація, оформлення замовлення або активація функції. Багатоканальні воронки враховують, що користувачі можуть взаємодіяти з брендом через різні touchpoints і пристрої перед конверсією.

Аналіз воронки не обмежується простим підрахунком конверсій на кожному етапі. Сегментований аналіз воронки порівнює ефективність для різних груп користувачів – наприклад, як воронка працює для мобільних versus desktop користувачів, для різних джерел трафіку, демографічних сегментів або географічних регіонів. Це виявляє специфічні проблеми та можливості для різних аудиторій.

Часовий аналіз воронки показує, скільки часу користувачі проводять на кожному етапі і як швидко вони проходять весь процес. Затримки на певних етапах можуть вказувати на плутанину, технічні проблеми або недостатню мотивацію. Когортний аналіз воронки показує, чи покращується конверсія з часом, що може вказувати на ефективність оптимізацій.

Інструменти для funnel analysis включають Google Analytics, Mixpanel, Amplitude, Heap Analytics та інші платформи продуктової аналітики. Ці інструменти дозволяють не лише візуалізувати воронки, але й проводити A/B тестування змін, застосовувати статистичний аналіз значущості різниць між сегментами та автоматично відстежувати зміни в ефективності воронки.

Оптимізація воронки – це ітеративний процес. Після виявлення проблемного етапу формулюються гіпотези про причини відсіювання, розробляються та тестуються зміни (покращення UX, зміна копірайтингу, спрощення форм, додавання соціальних доказів), вимірюється вплив на конверсію. Навіть невеликі покращення

конверсії на кожному етапі можуть призводити до значного зростання загальної конверсії завдяки кумулятивному ефекту.

5.3.4. Картування шляху користувача (User Journey Mapping).

User journey mapping або картування шляху користувача – це візуальне представлення повного досвіду взаємодії користувача з брендом, продуктом або сервісом від початкової обізнаності до післяпродажної підтримки та адвокації. На відміну від funnel analysis, який фокусується на конкретних конверсійних шляхах, journey mapping охоплює всі touchpoints, емоції, болючі точки та можливості покращення досвіду на кожному етапі.

Типова карта користувацького шляху включає кілька ключових елементів. Етапи journey зазвичай відповідають класичній моделі: awareness (обізнаність), consideration (розгляд), decision (рішення), retention (утримання), advocacy (адвокація). Touchpoints – це всі точки контакту користувача з брендом: реклама, соціальні мережі, веб-сайт, мобільний додаток, email, служба підтримки, фізичні магазини. Дії користувача описують конкретні кроки, які робить людина на кожному етапі: пошук інформації, читання відгуків, порівняння варіантів, спілкування з підтримкою.

Емоційна крива показує настрій та почуття користувача на різних етапах – від ентузіазму при виявленні продукту до потенційної фрустрації при складному процесі оформлення замовлення і задоволення від успішного використання. Болючі точки (pain points) ідентифікують моменти тертя, незручності, плутанини або розчарування. Можливості (opportunities) виділяють місця, де бренд може покращити досвід, усунути проблеми або створити додаткову цінність.

Процес створення user journey map починається зі збору даних з різних джерел: аналітика поведінки на сайті, дані CRM, записи розмов зі службою підтримки, результати опитувань та інтерв'ю з реальними користувачами, соціальне слухання, дані про відкриття email та взаємодію з контентом. Важливо базувати карту на реальних даних та інсайтах, а не на припущеннях про те, як користувачі "повинні" поводитися.

Створюються персони – деталізовані профілі типових користувачів з різними потребами, цілями та поведінковими патернами. Для кожної персони може бути створена окрема journey

map, оскільки різні сегменти аудиторії можуть мати суттєво відмінні шляхи та досвід. Наприклад, шлях молодого технічно грамотного користувача може кардинально відрізнятися від шляху старшого покоління, яке потребує більше підтримки.

Journey mapping виявляє критичні інсайти: момент істини (moments of truth) – критичні touchpoints, які найбільше впливають на рішення користувача; розриви в досвіді (experience gaps) – місця, де очікування користувача не відповідають реальності; можливості для автоматизації або персоналізації; потенційні точки відпливу користувачів; канали та контент, які найбільш ефективні на різних етапах.

Візуальне представлення journey map може приймати різні форми: лінійні діаграми, які показують хронологічну послідовність; циклічні моделі для продуктів з повторюваними покупками; багатоканальні карти, які показують паралельні взаємодії через різні платформи. Ключове – зробити карту зрозумілою та практичною для всіх зацікавлених сторін: маркетингу, продукту, дизайну, підтримки клієнтів.

User journey mapping не є одноразовою активністю. Карти повинні регулярно оновлюватися на основі нових даних, змін у поведінці користувачів, впровадження нових функцій або каналів. Вони служать інструментом для міжфункціональної співпраці, допомагаючи різним командам усвідомити свій вплив на загальний досвід користувача та узгодити зусилля навколо покращення критичних моментів у journey.

5.3.5. Предиктивна аналітика: прогнозування відтоку клієнтів, прогнозування LTV (довічної цінності клієнта).

Предиктивна аналітика використовує історичні дані, статистичні алгоритми та техніки машинного навчання для прогнозування майбутніх подій та поведінки користувачів. У digital-маркетингу дві критично важливі області застосування – це прогнозування відпливу користувачів (churn prediction) та передбачення життєвої цінності клієнта (LTV forecasting).

Churn prediction або прогнозування відпливу ідентифікує користувачів або клієнтів, які мають високу ймовірність припинити використання продукту або скасувати підписку. Моделі машинного навчання аналізують поведінкові патерни, які історично корелюють з відпливом: зниження частоти використання, відсутність взаємодії

з ключовими функціями, зменшення часу сесій, зміни в патернах споживання контенту, негативні взаємодії з службою підтримки, затримка платежів або зміна платіжного методу.

Створення моделі churn prediction включає кілька етапів. Спочатку визначається, що саме означає "churn" для конкретного бізнесу – це може бути відсутність активності протягом 30 днів, скасування підписки, відсутність повторної покупки протягом певного періоду. Потім збираються та підготовлюються дані: демографічні характеристики, історія взаємодій, транзакційні дані, метрики залученості, інформація про канали залучення.

Feature engineering – процес створення значущих змінних для моделі – критично важливий. Наприклад, замість просто "кількості входів" можна створити змінні типу "зміна кількості входів порівняно з попереднім місяцем" або "дні з останнього входу". Використовуються різні алгоритми машинного навчання: логістична регресія, дерева рішень, випадковий ліс (random forest), градієнтний бустинг (XGBoost, LightGBM), нейронні мережі. Кожен має свої переваги в залежності від характеристик даних та вимог до інтерпретованості моделі.

Модель оцінюється за метриками точності, recall (здатність виявляти реальні випадки churn), precision (точність передбачень) та AUC-ROC (загальна якість класифікації). Важливо знайти баланс між false positives (помилково ідентифіковані як ризик відпливу) та false negatives (пропущені реальні випадки відпливу), базуючись на бізнес-вартості помилок кожного типу.

Після побудови моделі створюються стратегії retention: користувачі з високим ризиком відпливу отримують персоналізовані пропозиції, спеціальні знижки, проактивну підтримку або освітній контент для збільшення залученості. Ефективність цих інтервенцій вимірюється через контрольовані експерименти, що дозволяє постійно вдосконалювати як модель, так і тактики утримання.

LTV (Lifetime Value) прогнозування передбачає загальний дохід або прибуток, який принесе клієнт протягом всього періоду взаємодії з брендом. Точне передбачення LTV дозволяє приймати обґрунтовані рішення про те, скільки можна інвестувати в залучення різних сегментів клієнтів, які канали найбільш прибуткові в довгостроковій перспективі, на яких клієнтів варто фокусувати retention зусилля.

Існують різні підходи до розрахунку та прогнозування LTV. Історичний LTV розраховується на основі фактичних даних про минулі транзакції. Предиктивний LTV використовує моделі для прогнозування майбутніх покупок та доходів на основі ранньої поведінки клієнта. Базова формула враховує середній дохід на клієнта, частоту покупок, період утримання клієнта та маржу прибутку.

Складніші моделі LTV використовують ймовірнісні підходи, такі як BG/NBD (Beta-Geometric/Negative Binomial Distribution) модель для прогнозування частоти транзакцій та Gamma-Gamma модель для прогнозування монетарної цінності. Моделі машинного навчання можуть інтегрувати численні змінні: демографічні характеристики, поведінку на сайті, історію взаємодій з маркетинговими кампаніями, соціально-економічні фактори, сезонність.

Когортний аналіз LTV показує, як змінюється цінність користувачів з різних періодів залучення, що допомагає оцінити ефективність змін у продукті або маркетингових стратегіях. Сегментований LTV розкриває, які канали залучення, демографічні групи або типи користувачів найбільш цінні, дозволяючи оптимізувати розподіл маркетингового бюджету.

Практичне застосування LTV прогнозування включає оптимізацію ставок у рекламних кампаніях (можна дозволити вищий CPA для сегментів з високим прогнозованим LTV), персоналізацію досвіду та пропозицій на основі потенційної цінності клієнта, приоритизацію зусиль служби підтримки (VIP обслуговування для high-LTV клієнтів), стратегічне планування та прогнозування доходів.

Обидві техніки – churn prediction та LTV forecasting – вимагають якісних даних, регулярного оновлення моделей, інтеграції з операційними системами для автоматизації дій на основі прогнозів та постійного моніторингу ефективності. При правильному впровадженні вони трансформують маркетинг від реактивного до проактивного, дозволяючи компаніям діяти на випередження та максимізувати цінність кожного клієнта.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання перевіряють здатність аналізувати причино-наслідкові зв'язки:

1. Моделі атрибуції (Attribution Models): Це "вічна проблема"

маркетингу. Якщо користувач побачив рекламу в Instagram, потім загуглив бренд, а купив підписку через тиждень з email-розсилки — який канал привів клієнта? У чому недолік моделі Last Click (останній клік) і перевага моделі Data-Driven (на основі даних)?

2. Когортний аналіз (Cohort Analysis): Чому зростання загальної відвідуваності сайту може приховувати катастрофу? Як аналіз поведінки груп людей, що зареєструвалися в один і той самий місяць (когорти), допомагає виявити проблеми з утриманням (Retention)?

3. RFM-аналіз: Як класичний метод сегментації за трьома параметрами (Recency — давність покупки, Frequency — частота, Monetary — сума) дозволяє автоматично виділити "Китів" (VIP-клієнтів) та тих, хто "спить"?

4. Unit-економіка (LTV vs CAC): Яке "золоте правило" співвідношення *Customer Lifetime Value* (скільки грошей принесе клієнт за все життя) до *Customer Acquisition Cost* (скільки коштувало його залучити)? Чому, якщо LTV/CAC < 3, бізнес приречений, навіть за високих продажів?

5. Аналіз шляху клієнта (Customer Journey Map): Як аналітики використовують "теплові карти" (Heatmaps) та записи сесій, щоб зрозуміти, на якому етапі воронки користувачі "відвалюються"?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають "читати" аналітичні звіти:

Завдання 1. Інтерпретація Когортного звіту (Retention)

• Матеріал: Уявіть таблицю утримання (Retention rate) для мобільного додатка новин.

- *Когорта "Січень":* Місяць 0 (100%) -> Місяць 1 (40%) -> Місяць 2 (35%).

- *Когорта "Лютий":* Місяць 0 (100%) -> Місяць 1 (20%) -> Місяць 2 (5%).

• Завдання: Ви — головний аналітик. Що сталося у лютому?

• Гіпотези: (Невдале оновлення дизайну? Залучення неякісного трафіку через клікбейт? Технічний баг на Android?). Опишіть ваші дії для перевірки.

Завдання 2. Битва за бюджет (Атрибуція)

• Ситуація:

- *Facebook-менеджер каже:* "Моя реклама принесла 100

продажів!" (Дивиться по показах).

- *SEO-спеціаліст каже:* "Ні, це органічний пошук приніс 100 продажів!" (Дивиться по Last Click).

- Реальна кількість продажів у CRM — всього 120.

• Завдання: Поясніть, чому сума звітів різних каналів більша за реальність (перетин аудиторій). Розрахуйте ефективність каналів за моделлю Linear Attribution (рівномірний розподіл цінності між усіма точками контакту).

Завдання 3. RFM-сегментація для Email-маркетингу

• Дано: База підписників.

• Завдання: Розробіть стратегію комунікації для трьох сегментів:

- *R=1, F=5, M=5 (Недавні, Часті, Платять багато):* "Чемпіони". (Дія: Дати ранній доступ до ексклюзиву, не давати знижок — вони і так куплять).

- *R=5, F=5, M=5 (Давні, Часто купували, Платили багато):* "У зоні ризику". (Дія: Терміново повернути! Персональна пропозиція, опитування "Що не так?").

- *R=1, F=1, M=1 (Недавні, Один раз, Мало):* "Новачки". (Дія: Welcome-серія листів, навчання користуванню продуктом).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. *Avinash Kaushik. Web Analytics 2.0.* (Біблія веб-аналітики. Хоча інструменти змінилися, філософія "про що говорять дані" актуальна).

2. *Alistair Croll, Benjamin Yoskovitz. Lean Analytics.* (Як обрати "одну метрику, що має значення" для стартапу).

3. Google Analytics 4 Certification Study Guide. (Офіційні матеріали Google про перехід на подійно-орієнтовану аналітику).

4. Mark Jeffery. *Data-Driven Marketing.* (15 ключових метрик, які повинен знати кожен маркетолог).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• LTV (Lifetime Value): Прогнозований чистий прибуток, який компанія отримає від усіх майбутніх відносин з клієнтом.

• CAC (Customer Acquisition Cost): Сукупна вартість залучення одного нового клієнта (витрати на маркетинг / кількість нових клієнтів).

• Churn Rate (Відтік): Відсоток клієнтів, які припинили користу-

ватися послугою за певний період. "Діряве відро" будь-якого бізнесу.

• Conversion Rate (CR, Конверсія): Відсоток відвідувачів, які виконали цільову дію (купили, підписалися).

• Funnel (Воронка): Модель шляху клієнта від першого знайомства до покупки. Аналіз воронки дозволяє знайти "вузьке місце" (bottleneck), де втрачається найбільше людей.

5.4. Інструменти та платформи для data-driven маркетингу

5.4.1. Google Analytics 4 та його можливості.

Google Analytics 4 (GA4) являє собою нове покоління веб-аналітики від Google, яке фундаментально відрізняється від попередньої версії Universal Analytics. Запущена у 2020 році і ставши обов'язковою з липня 2023 року, GA4 побудована на подієвій моделі даних, що забезпечує більш гнучке та всебічне відстеження користувацької поведінки.

Ключова відмінність GA4 – це event-based архітектура. На відміну від Universal Analytics, яка базувалася на сесіях та переглядах сторінок, GA4 трактує всі взаємодії як події. Це включає автоматично відстежувані події (перегляди сторінок, прокручування, клік по зовнішньому посиланню, перегляд відео, завантаження файлів), рекомендовані події для e-commerce (додавання до кошика, покупка, перегляд товару) та користувацькі події, які можна налаштувати під специфічні потреби бізнесу. Така гнучкість дозволяє відстежувати будь-яку взаємодію, яка має значення для конкретного бізнесу.

Міжплатформне відстеження – одна з найпотужніших можливостей GA4. Система може об'єднувати дані з веб-сайтів, мобільних додатків iOS та Android у єдиному звіті, створюючи цілісне уявлення про користувацький шлях незалежно від пристроїв та платформ. Використовуючи Google signals та User ID, GA4 може ідентифікувати одного користувача на різних пристроях, якщо він авторизований у своєму Google-акаунті або в системі сайту.

Інтеграція машинного навчання вбудована безпосередньо в GA4. Предиктивні метрики автоматично прогнозують ймовірність покупки (purchase probability), ймовірність відпливу (churn probability) та потенційний дохід від користувача (predicted revenue) на основі їхньої поведінки. Ці прогнози можна використовувати для створення аудиторій та їх активації в Google Ads або інших рекламних платформах. Інсайти та аномалії автоматично виявляються системою, яка сповіщає про незвичайні зміни в трафіку або поведінці користувачів.

Аналіз воронок та шляхів користувачів у GA4 значно вдосконалений. Exploration reports дозволяють створювати гнучкі аналітичні звіти: funnel exploration для аналізу конверсійних воронок

з можливістю додавання необмеженої кількості кроків і сегментації; path exploration для візуалізації реальних шляхів користувачів через сайт або додаток; segment overlap для розуміння перетину між різними сегментами аудиторії; cohort exploration для аналізу утримання та поведінки когорт; user lifetime для аналізу LTV та довгострокової цінності користувачів.

Можливості налаштування подій та конверсій у GA4 надзвичайно гнучкі. Можна модифікувати події через інтерфейс без зміни коду, створювати нові події на основі умов (наприклад, подія "engaged_session" коли користувач провів більше 10 секунд на сайті), встановлювати події як конверсії одним кліком, призначати монетарну цінність подіям для точнішого відстеження ROI.

Інтеграція з Google Ads та іншими продуктами Google надзвичайно глибока. Аудиторії, створені в GA4, автоматично синхронізуються з Google Ads для ремаркетингу та look-alike targeting. Дані про конверсії імпортуються для оптимізації кампаній. Інтеграція з Google BigQuery доступна навіть на безкоштовному рівні (раніше лише для платної версії 360), що дозволяє експортувати сирі дані для просунутого аналізу, створення custom звітів та інтеграції з іншими системами.

Privacy-first підхід GA4 відповідає сучасним вимогам конфіденційності. Система побудована з урахуванням майбутнього без cookies третіх сторін, використовує моделювання даних для заповнення прогалин там, де згода на відстеження не була надана, надає гнучкі налаштування збору даних відповідно до GDPR, CCPA та інших регуляцій, автоматично видаляє IP-адреси та не зберігає ідентифікуючу інформацію на рівні користувача.

Attribution modeling у GA4 включає різні моделі атрибуції: data-driven attribution (використовує машинне навчання для розподілу кредиту між touchpoints), last click, first click, linear, time decay, position-based. Порівняльні звіти показують, як різні моделі впливають на оцінку ефективності каналів, допомагаючи зрозуміти справжній внесок кожного каналу в конверсії.

Обмеження GA4 також варто враховувати: інтерфейс та логіка суттєво відрізняються від Universal Analytics, що вимагає періоду навчання; деякі звичні звіти відсутні або реалізовані по-іншому; сирі дані не завжди ідеально точні через вибіркову обробку (sampling) у великих датасетах; налаштування вимагає технічних знань для

повного використання можливостей.

5.4.2. Mixpanel, Amplitude для Product Analytics.

Mixpanel та Amplitude представляють категорію платформ product analytics, які спеціально розроблені для глибокого аналізу поведінки користувачів у цифрових продуктах, особливо в SaaS-додатках, мобільних застосунках та веб-платформах. На відміну від традиційної веб-аналітики, ці інструменти фокусуються на відстеженні конкретних дій користувачів та їхньому впливі на ключові продуктові метрики.

Mixpanel пропонує детальне event-based відстеження з акцентом на зручність використання для продуктових команд. Платформа дозволяє відстежувати будь-яку взаємодію користувача як окрему подію з необмеженою кількістю властивостей (properties). Наприклад, подія "завершив онбординг" може містити властивості: час завершення, кількість пропущених кроків, використаний пристрій, джерело реєстрації. Це дозволяє проводити надзвичайно гранульований аналіз.

Ключові можливості Mixpanel включають Insights – інтерактивні запити для аналізу трендів, сегментації та порівняння поведінки різних груп користувачів. Можна швидко відповісти на питання типу "скільки користувачів з iOS виконали подію X протягом останніх 30 днів" або "як змінилася частота використання функції Y після останнього релізу". Funnels дозволяють створювати воронки конверсії з необмеженою кількістю кроків, аналізувати час проходження кожного етапу, порівнювати різні сегменти користувачів і виявляти, де саме відбувається найбільше відпадання.

Retention analysis у Mixpanel показує, як часто користувачі повертаються для виконання певних дій. Можна аналізувати retention за будь-якою подією (не лише входом в додаток), сегментувати по когортах, порівнювати різні періоди та джерела залучення. Flows візуалізують реальні шляхи користувачів, показуючи найпопулярніші послідовності дій, виявляючи неочікувані патерни поведінки.

Користувацькі профілі в Mixpanel створюють повну картину кожного індивідуального користувача: всі виконані події, демографічні та поведінкові властивості, життєвий цикл, історія сегментів. Це дозволяє глибоко розуміти типових представників різних сегментів. Messaging та engagement tools дозволяють надсилати персоналізовані

повідомлення (email, push, in-app) безпосередньо з Mixpanel на основі поведінки користувачів, створюючи автоматизовані кампанії retention та активації.

Amplitude позиціонується як більш просунута платформа для enterprise-сегменту з особливим акцентом на предиктивну аналітику та поведінкові когорти. Архітектура Amplitude базується на концепції "північної зірки" (north star metric) – ключової метрики, яка найкраще відображає цінність продукту для користувачів та бізнесу. Платформа допомагає ідентифікувати цю метрику та зрозуміти, які дії користувачів найбільше корелюють з її зростанням.

Behavioral Cohorts – одна з найпотужніших можливостей Amplitude. Система автоматично групує користувачів на основі подібних патернів поведінки, використовуючи машинне навчання. Наприклад, може виявити когорту "power users", які регулярно використовують певну комбінацію функцій, або користувачів з ризиком відпливу на основі зменшення активності. Ці когорти можна використовувати для таргетованих кампаній або експериментів.

Pathfinder та патерни використання показують найпопулярніші та найефективніші шляхи до конверсії. Pathfinder Amplitude візуалізує, як користувачі, які досягли певної мети, рухалися через продукт, виявляючи оптимальні послідовності дій. Це допомагає оптимізувати онбординг та продуктовий досвід.

Compass – функція, яка використовує статистичний аналіз для виявлення поведінкових паттернів, що найбільше корелюють з утриманням, конверсією або іншими ключовими метриками. Система автоматично знаходить "aha moments" – дії, які, якщо користувач їх виконує, значно підвищують ймовірність довгострокового утримання. Наприклад, для соціальної мережі це може бути "додавання 7 друзів протягом перших 10 днів".

Experimentation platform в Amplitude дозволяє проводити A/B тести безпосередньо в інтерфейсі, автоматично розраховуючи статистичну значущість результатів, аналізуючи вплив на множинні метрики одночасно, виявляючи непередбачені наслідки експериментів на різні сегменти користувачів. Інтеграція експериментів з аналітикою дозволяє швидко ітерувати та приймати data-driven рішення.

Data taxonomy та governance в Amplitude допомагають великим командам підтримувати якість даних: tracking plan визначає, які

події повинні відстежуватися та з якими властивостями; валідація даних перевіряє, що події надсилаються правильно; data quality score показує надійність даних для прийняття рішень.

Обидві платформи пропонують потужні API для інтеграції з іншими системами, warehouses sync для синхронізації з data warehouses (Snowflake, BigQuery, Redshift), SDK для простої інтеграції в мобільні та веб-додатки, можливості експорту сирих даних для custom аналізу.

Вибір між Mixpanel та Amplitude часто залежить від специфічних потреб: Mixpanel зазвичай вважається більш доступним для малих та середніх команд, з простішим інтерфейсом та швидшим time-to-value. Amplitude пропонує більш просунуті можливості для великих команд з складними продуктами, особливо сильна в автоматизованих інсайтах та предиктивній аналітиці. Ціноутворення обох платформ базується на обсязі відстежуваних подій, роблячи їх доступними для стартапів на початкових етапах, але потенційно дорогими при масштабуванні.

5.4.3. Tableau, Power BI для візуалізації.

Tableau та Microsoft Power BI є провідними платформами для бізнес-аналітики та візуалізації даних, які трансформують складні датасети у зрозумілі, інтерактивні дашборди та звіти. Ці інструменти дозволяють маркетологам та аналітикам створювати compelling візуальні історії з даних, виявляти тренди та комунікувати інсайти зацікавленим сторонам.

Tableau славиться своїм інтуїтивним підходом до візуалізації через drag-and-drop інтерфейс. Користувачі можуть швидко створювати різноманітні типи візуалізацій: від базових стовпчастих та лінійних графіків до складних heat maps, treemaps, scatter plots, географічних карт та custom візуалізацій. Tableau автоматично пропонує найбільш відповідні типи візуалізацій на основі вибраних даних, що прискорює процес аналізу.

Можливості підключення даних у Tableau надзвичайно широкі. Платформа може з'єднуватися з десятками джерел: реляційні бази даних (MySQL, PostgreSQL, SQL Server), хмарні сховища даних (Snowflake, BigQuery, Redshift), файли (Excel, CSV, JSON), cloud applications (Salesforce, Google Analytics, Adobe Analytics), веб-дані через API та web data connectors. Live connections забезпечують роботу

з реальними даними в реальному часі, тоді як extracts створюють оптимізовані локальні копії для швидшої роботи.

Tableau Prep – окремий інструмент для підготовки даних, який дозволяє очищати, трансформувати та комбінувати дані з різних джерел візуальним способом. Це включає фільтрацію, агрегацію, об'єднання таблиць, pivoting, створення обчислюваних полів. Візуальний підхід робить ETL (Extract, Transform, Load) процеси доступними для нетехнічних користувачів.

Calculated fields та table calculations у Tableau дозволяють створювати складні метрики та KPI: від простих математичних операцій до статистичних функцій, forecasting, cohort analysis, running totals, window functions. Мова обчислень Tableau поєднує простоту з потужністю, підтримуючи складні аналітичні сценарії.

Interactivity дашбордів Tableau включає filters для динамічної фільтрації даних, parameters для створення what-if сценаріїв, actions для зв'язку між різними візуалізаціями (клік на одній діаграмі фільтрує інші), drill-down можливості для деталізації від високого рівня до гранулярних даних. Це перетворює статичні звіти на інтерактивні аналітичні інструменти.

Tableau Server та Tableau Online дозволяють публікувати дашборди для спільного використання в організації, налаштовувати права доступу, створювати автоматизовані розсилки звітів, вбудовувати візуалізації в інші додатки через embedding. Mobile apps забезпечують доступ до дашбордів на планшетах та смартфонах з адаптивним дизайном.

Microsoft Power BI пропонує схожі можливості з міцнішою інтеграцією в екосистему Microsoft та часто більш привабливим ціноутворенням для організацій, які вже використовують Microsoft 365. Power BI складається з кількох компонентів: Power BI Desktop – безкоштовний додаток для створення звітів, Power BI Service – хмарна платформа для публікації та спільного використання, Power BI Mobile – мобільні додатки, Power BI Embedded – для вбудовування в custom додатки.

Power Query у Power BI забезпечує потужну функціональність ETL, дозволяючи трансформувати та очищати дані з понад 100 різних джерел. M language (формульна мова Power Query) та DAX (Data Analysis Expressions) надають розширені можливості для маніпуляції

даними та створення складних обчислень. DAX особливо потужний для створення measures, calculated columns та time intelligence функцій.

Data modeling у Power BI дозволяє створювати складні моделі даних з множинними таблицями, relationships, hierarchies. Це забезпечує структуровану та ефективну роботу з великими обсягами даних. Automatic aggregations та incremental refresh оптимізують продуктивність при роботі з мільйонами рядків даних.

AI-powered features в Power BI включають Quick Insights – автоматичне виявлення цікавих паттернів у даних, Q&A – можливість запитувати дані природною мовою ("покажи продажі за регіонами за минулий рік"), Key Influencers visual – автоматичне виявлення факторів, що впливають на певну метрику, Decomposition Tree – інтерактивне дослідження того, як метрика розподіляється по різних dimensions.

Collaboration features дозволяють створювати workspaces для командної роботи, налаштовувати row-level security для контролю доступу до даних, створювати apps для дистрибуції наборів дашбордів кінцевим користувачам, інтегруватися з Microsoft Teams для обговорення інсайтів безпосередньо в контексті даних.

Integration з Azure ecosystem робить Power BI природним вибором для організацій, що використовують Azure: Azure Synapse Analytics, Azure Data Lake, Azure Machine Learning можуть безшовно інтегруватися з Power BI. Power Automate дозволяє створювати автоматизовані workflows на основі подій у даних (наприклад, сповіщення коли метрика перевищує поріг).

Вибір між Tableau та Power BI часто залежить від кількох факторів: Tableau традиційно сильніше в складних та естетично привабливих візуалізаціях, має більш інтуїтивний інтерфейс для ad-hoc аналізу. Power BI пропонує кращу інтеграцію з Microsoft екосистемою, більш привабливе ціноутворення (особливо якщо вже є Microsoft 365), сильніші можливості в data modeling та DAX для складних обчислень. Обидві платформи постійно розвиваються, додаючи нові можливості та зменшуючи різницю між собою.

5.4.4. Python/R для просунутої аналітики.

Python та R представляють потужні мови програмування для просунутої аналітики даних, статистичного моделювання та

машинного навчання в маркетингу. На відміну від готових платформ з графічним інтерфейсом, ці інструменти надають необмежену гнучкість та контроль над аналітичними процесами, дозволяючи вирішувати унікальні та складні задачі.

Python стала de facto стандартом для data science завдяки своїй універсальності, читабельності коду та величезній екосистемі бібліотек. Для маркетингової аналітики ключові бібліотеки включають Pandas – для маніпуляції та аналізу структурованих даних (dataframes), що дозволяє легко фільтрувати, групувати, об'єднувати та трансформувати датасети будь-якого розміру. NumPy забезпечує ефективні операції з числовими масивами та математичні обчислення.

Візуалізація даних у Python реалізована через Matplotlib – базову бібліотеку для створення статичних графіків, Seaborn – надбудову над Matplotlib з красивими статистичними візуалізаціями, Plotly – для інтерактивних візуалізацій та дашбордів. Ці інструменти дозволяють створювати publication-ready графіки повністю програмно з можливістю автоматизації.

Машинне навчання у Python доступне через Scikit-learn – найпопулярнішу бібліотеку для класичних ML алгоритмів (регресія, класифікація, кластеризація, dimensionality reduction), що включає інструменти для feature engineering, model selection, cross-validation, evaluation metrics. TensorFlow та PyTorch використовуються для глибокого навчання, neural networks та складних предиктивних моделей.

Специфічні маркетингові задачі вирішуються через спеціалізовані бібліотеки: Lifetimes для CLV (Customer Lifetime Value) моделювання та RFM аналізу, використовуючи probabilistic моделі типу BG/NBD; Marketing Attribution для multi-touch attribution modeling; PyMC3 для Bayesian статистики та A/B тестування з більш гнучкими підходами ніж frequentist методи; NLTK та spaCy для обробки природної мови, sentiment analysis соціальних медіа та аналізу customer feedback.

Web scraping та data collection через Beautiful Soup, Scrapy дозволяють збирати дані з веб-сайтів конкурентів, форумів, соціальних мереж. Requests та API-специфічні wrapper бібліотеки забезпечують програмний доступ до даних з Google Analytics, Facebook Ads, Twitter та інших платформ через їхні API.

Jupyter Notebooks створюють інтерактивне середовище для

аналізу, де код, візуалізації та текстові пояснення поєднуються в одному документі. Це ідеально для exploratory analysis, створення reproducible research та комунікації інсайтів. Notebooks можна легко конвертувати в презентації, звіти HTML або PDF.

R історично була мовою статистиків і до сих пір залишається надзвичайно потужною для статистичного аналізу та візуалізації. Tidyverse – колекція пакетів, яка включає dplyr для data manipulation (filter, select, mutate, group_by, summarize), ggplot2 для створення складних багатошарових візуалізацій на основі grammar of graphics, tidyr для reshaping даних, readr для швидкого читання файлів.

Статистичне моделювання у R неперевершене: базові статистичні функції вбудовані в мову, glm() та lm() для лінійної та узагальненої лінійної регресії, survival analysis для retention та churn modeling, time series analysis через forecast пакет для прогнозування трендів, seasonality detection. Caret та tidymodels забезпечують unified інтерфейс для сотень ML алгоритмів.

Специфічні маркетингові пакети включають BTYDplus для моделювання customer lifetime value, conjoint analysis пакети для аналізу переваг споживачів, MarketMatching для синтетичний контроль та quasi-experimental дизайни, multivariate testing frameworks. RMarkdown аналогічний до Jupyter, дозволяючи створювати dynamic documents з вбудованим R кодом.

Shiny framework дозволяє створювати інтерактивні веб-додатки та дашборди безпосередньо з R коду без необхідності знання web development. Це потужний інструмент для створення custom аналітичних інструментів для нетехнічних стейкхолдерів.

Практичні застосування Python/R у маркетингу включають A/B тестування з правильним статистичним аналізом (Bayesian A/B testing, multi-armed bandits, sequential testing), marketing mix modeling для визначення впливу різних маркетингових каналів на продажі, customer segmentation через clustering (K-means, hierarchical, DBSCAN), RFM аналіз, churn prediction models з feature engineering та model tuning, sentiment analysis соціальних медіа та відгуків клієнтів, recommendation systems, automated reporting з регулярною генерацією звітів та їх розсилкою.

Обидві мови мають активні спільноти, величезну кількість навчальних ресурсів та регулярно оновлюються. Вибір між ними

часто залежить від background команди (академічне середовище схиляється до R, tech компанії до Python) та специфічних задач. Багато data scientists використовують обидві мови, вибираючи найкращий інструмент для конкретної задачі. Для максимальної гнучкості обидві мови можуть інтегруватися одна з одною через reticulate (R пакет для виклику Python) або rpy2 (Python бібліотека для виклику R).

5.4.5. ШІ-платформи: Twilio Segment, Heap.

Нове покоління аналітичних платформ використовує штучний інтелект та автоматизацію для спрощення збору даних, створення інсайтів та активації аудиторій. Twilio Segment та Heap представляють різні підходи до AI-powered аналітики, автоматизуючи традиційно трудомісткі процеси та надаючи democratized доступ до складної аналітики.

Twilio Segment позиціонується як Customer Data Platform (CDP), яка централізує збір, очищення та маршрутизацію даних про клієнтів з усіх touchpoints. Основна ідея – зібрати дані один раз і надіслати їх до всіх необхідних інструментів, замість інтегрування кожного джерела з кожним destination окремо. Це створює single source of truth для клієнтських даних.

Архітектура Segment базується на Sources (джерела даних) → Segment → Destinations (пункти призначення). Sources включають веб-сайти, мобільні додатки, сервери, cloud apps (Salesforce, Zendesk, Stripe). Дані збираються через Segment SDK або API у стандартизованому форматі. Destinations включають понад 300 інтеграцій: аналітичні платформи (Google Analytics, Mixpanel, Amplitude), рекламні платформи (Facebook Ads, Google Ads), email tools (Mailchimp, SendGrid), data warehouses (Snowflake, BigQuery), CRM системи.

Protocols – ключова функція Segment для data governance. Tracking plan визначає, які події та properties повинні збиратися, їхні типи даних та валідаційні правила. Schema validation автоматично перевіряє, що дані відповідають специфікації, блокуючи або маркуючи невалідні події. Data Quality Score показує надійність даних для прийняття рішень. Це критично для великих організацій, де множина команд збирають дані, і легко виникає inconsistency.

Personas – вбудована функція для створення unified customer profiles, які об'єднують дані про одного користувача з різних джерел

та пристроїв. Identity resolution використовує різні ідентифікатори (email, user ID, cookie ID, mobile advertising ID) для створення single customer view. Computed traits автоматично обчислюють характеристики користувачів (total revenue, days since last purchase, favorite category) на основі їхньої поведінки.

Audiences в Segment дозволяють створювати динамічні сегменти користувачів на основі будь-яких комбінацій traits, events, та computed values. Ці аудиторії автоматично синхронізуються з advertising platforms для персоналізованих кампаній, email tools для таргетованих розсилок, websites для on-site персоналізації. Predictive traits використовують ML для прогнозування ймовірності покупки, churn risk, predicted LTV.

Reverse ETL функціональність дозволяє брати дані з data warehouse і надсилати їх назад до операційних інструментів. Наприклад, results ML моделі, побудованої в warehouse, можуть автоматично створювати аудиторії в рекламних платформах. Це замикає цикл між аналітикою та активацією.

Privacy Portal допомагає виконувати вимоги регуляцій (GDPR, CCPA): автоматизує data subject requests (доступ до даних, видалення), забезпечує consent management, blocked data forwarding до destinations коли згода не надана. Це критично в сучасному privacy-first середовищі.

Heap використовує радикально інший підхід до збору даних через автоматичне відстеження (automatic capture). Замість необхідності визначати кожну подію, яку потрібно відстежувати заздалегідь, Heap автоматично збирає всі взаємодії користувачів: кліки, submits, page views, field changes. Це вирішує проблему "ми не можемо аналізувати це, бо не відстежували" – якщо користувач це робив, дані є.

Retroactive analysis – ключова перевага Heap. Оскільки всі дані збираються автоматично, можна створювати events та аналізувати їх ретроспективно. Наприклад, якщо через місяць після релізу нової кнопки вирішили аналізувати клік по ній, дані за весь минулий період вже доступні. Не потрібно чекати накопичення нових даних після налаштування відстеження.

Event visualizer в Heap дозволяє визначати події візуально, клікаючи на елементи на сайті замість написання коду. Клік на кнопку в інтерфейсі Heap автоматично створює event definition. Auto-

captured properties включають інформацію про елемент (text, class, id), сторінку, session, користувача. Custom properties можна додавати через API для business-specific контексту.

Session replay – потужна функція для якісного аналізу. Heap записує фактичні сесії користувачів (рухи миші, кліки, прокручування, input), дозволяючи переглядати їх як відео. Це надзвичайно цінно для розуміння UX проблем, debugging, empathy з користувацьким досвідом. Privacy controls дозволяють маскувати чутливу інформацію (паролі, номери карт).

Illuminations – AI-powered insights від Heap, які автоматично виявляють аномалії, тренди та можливості. Наприклад, система може автоматично виявити, що конверсія на певній сторінці раптово впала, або що нова когорта користувачів має значно вищий retention. Це democratizes аналітику, допомагаючи нетехнічним користувачам виявляти інсайти без створення custom запитів.

Effort analysis показує, скільки "зусиль" користувачі вкладають у досягнення мети – кількість кліків, час, кількість відвідуваних сторінок. High-effort flows вказують на проблеми UX, де процес можна спростити. Порівняння effort між користувачами, які конвертували, та тими, хто ні, виявляє критичні точки тертя.

Data science capabilities включають SQL доступ до сирих даних через Heap Connect, що дозволяє data scientists використовувати знайомі інструменти для просунутого аналізу. Integrations з warehouses дозволяють об'єднувати Heap дані з іншими джерелами для holistic аналізу.

Обидві платформи представляють майбутнє аналітики, де AI автоматизує рутинні задачі, democratizes доступ до інсайтів та дозволяє фокусуватися на стратегічних рішеннях замість технічного implementation. Segment ідеальна для організацій, що потребують централізованої інфраструктури даних та складних активацій аудиторій. Heap особливо цінна для продуктових команд, що хочуть швидко ітерувати без залежності від engineering ресурсів для налаштування відстеження. Вибір залежить від специфічних потреб, технічних можливостей команди та архітектури існуючого tech stack.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допомагають розібратися в архітектурі інструментів:

1. Еволюція баз даних: У чому принципова різниця між CRM (Salesforce, HubSpot), яка працює з *відомими* клієнтами (ім'я, телефон), та DMP (Data Management Platform), яка працює з *анонімними* аудиторіями (cookie ID) для реклами?

2. Схід зірки CDP: Чому Customer Data Platform (CDP) (наприклад, Segment або Tealium) зараз вважається "золотим стандартом"? Як вона об'єднує дані з сайту, мобільного додатка та офлайн-каси в "Єдиний профіль клієнта" (Single Customer View)?

3. Візуалізація даних: Чому Excel більше не підходить для моніторингу медіа? У чому перевага BI-систем (Business Intelligence), *таких як Tableau, Power BI або Looker Studio*, які дозволяють будувати дашборди в реальному часі?

4. SEO та конкурентна розвідка: Які інструменти (Ahrefs, Semrush, SimilarWeb) дозволяють зазирнути "під капот" конкуренту: побачити, звідки він бере трафік та за якими ключовими словами рекламується?

5. Автоматизація (Marketing Automation): Як платформи типу Mailchimp або Klaviyo використовують тригери (подія "Користувач покинув кошик") для автоматичної відправки серії листів без участі людини?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання спрямовані на навички вибору та інтеграції софту:

Завдання 1. Побудова MarTech Stack

• Легенда: Ви — технічний директор нового онлайн-медіа про технології.

• Бюджет: Обмежений (стартап).

• Завдання: Оберіть по одному інструменту для кожної категорії та обґрунтуйте вибір (ціна/функціонал):

1. *CMS (Система управління контентом):* (WordPress? Ghost?).

2. *Analytics (Аналітика):* (Google Analytics 4? Matomo — якщо важлива приватність?).

3. *Email Marketing:* (Substack? MailerLite?).

4. *Project Management:* (Trello? Notion?).

• Результат: Намалюйте стрілочками, як дані про нового підписника потрапляють з сайту в систему розсилки.

Завдання 2. Створення дашборда в Looker Studio

• Інструмент: Google Looker Studio (безкоштовно).

• Джерело даних: Підключіть тестовий акаунт Google Analytics або просту Google-таблицю.

• Завдання: Створіть візуалізацію з трьома елементами:

- *Графік динаміки* (Line Chart) відвідувачів за місяць.

- *Кругова діаграма* (Pie Chart) джерел трафіку (Mobile vs Desktop).

- *Таблиця* топ-5 найпопулярніших статей.

• Мета: Зрозуміти різницю між "сирими даними" та "інсайтами".

Завдання 3. Шпигунство через SimilarWeb

• Інструмент: SimilarWeb (безкоштовна версія).

• Дія: Введіть адресу двох конкуруючих українських сайтів (наприклад, "pravda.com.ua" та "nv.ua").

• Аналіз: Порівняйте їхні джерела трафіку.

- Хто більше залежить від соцмереж?

- У кого краща глибина перегляду (Pages per Visit)?

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Scott *Brinker. Marketing Technology Landscape.* (Щорічна карта, що показує 10 000+ інструментів на ринку. Допомагає зрозуміти масштаб індустрії).

2. Avinash Kaushik. *Web Analytics 2.0.* (Розділи про вибір інструментарію).

3. Блог HubSpot / Salesforce. (Найкращі джерела актуальної інформації про те, як працюють CRM та автоматизація).

4. Google Data Analytics Professional Certificate (Coursera). (Для тих, хто хоче опанувати технічні навички роботи з даними).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• MarTech (Marketing Technology): Сукупність програмного забезпечення та технологічних інструментів, які маркетологи використовують для планування, виконання та вимірювання маркетингових кампаній.

• CRM (Customer Relationship Management): Система управління відносинами з клієнтами. База даних, де зберігається історія спілкування з конкретними, ідентифікованими людьми (ПІБ, телефон, історія покупок).

• CDP (Customer Data Platform): Програмне забезпечення, що зби-

рає дані про користувача з усіх джерел (сайт, додаток, email, офлайн) в єдиний профіль, доступний для інших маркетингових систем. На відміну від DMP, працює з персональними даними (PII).

• SaaS (Software as a Service): Модель продажу софту, коли ви не купуєте програму на диску, а платите за підписку (оренду) у хмарі (наприклад, Slack, Google Workspace).

• Lead Scoring (Оцінка лідів): Функція інструментів автоматизації, яка автоматично нараховує бали потенційному клієнту за його дії (відкрив лист +5 балів, відвідав сторінку цін +20 балів). Коли сума досягає ліміту, система сигналізує відділу продажів.

5.5. Приватність та етика збору даних

5.5.1. GDPR, CCPA та інші регуляції.

Персоналізація контенту та рекомендаційних систем неможлива без збору та аналізу даних, проте в останнє десятиліття ця діяльність перестала бути правовим вакуумом. Почалася нова ера, коли парадигма володіння даними змістилася від платформи до користувача. Цей перехід закріплено низкою міжнародних та національних регуляцій, найвпливовішими з яких стали європейський GDPR та каліфорнійський CCPA (згодом розширений CPRA). Для медіа-індустрії, особливо в контексті глобалізації та цифрової трансформації, розуміння та дотримання цих норм перетворилося з доброї волі на обов'язкову вимогу, що несе як юридичні ризики, так і репутаційні переваги.

GDPR: золотий стандарт та принцип екстериторіальності. Загальний регламент захисту даних (GDPR), що набрав чинності в ЄС у 2018 році, дійсно встановив найсуворіший у світі стандарт. Його фундаментальна відмінність полягає в застосуванні за принципом екстериторіальності: закон поширюється на будь-яку компанію в світі, яка обробляє персональні дані громадян ЄС, незалежно від її місцезнаходження. Таким чином, українське медіа, що має версію сайту польською мовою, аналізує аудиторію з ЄС або таргетує на неї рекламу, повністю потрапляє під його дію. Серце GDPR — принцип активної, інформованої згоди (Opt-in). Згода має бути добровільною, конкретною, недвусмисленою та даною на підставі чіткої інформації. Мовчання, заздалегідь відмічені галочки або бездіяльність не можуть слугувати її підставою.

Регламент надає користувачам низку потужних прав, які медіа-організації зобов'язані поважати: право на доступ, виправлення, обмеження обробки, перенесення даних (отримання своїх даних у структурованому форматі) та найвідоміше — право на забуття. Останнє дозволяє користувачеві вимагати видалення своїх персональних даних, що ставить перед редакціями складні технічні та архітектурні завдання. Відповідальність компаній включає не лише отримання згоди та забезпечення прав, але й прозорість, повідомлення про порушення безпеки протягом 72 годин та призначення уповноваженого із захисту даних (DPO) у певних

випадках. Порушення можуть коштувати до 20 млн євро або 4% від світового обороту компанії.

CCPA/CPRA: американська модель та право на відмову. Каліфорнійський закон про захист конфіденційності споживачів (CCPA), чинний з 2020 року, та його розширення — CPRA (California Privacy Rights Act), що набрав чинності у 2023 році, сформували власний впливовий стандарт, особливо для технологічного сектора. На відміну від проактивного Opt-in GDPR, американський підхід орієнтований на Opt-out — право на відмову. Компанії можуть збирати дані за замовчуванням, але зобов'язані чітко повідомляти про це та надавати механізми для відмови, зокрема кнопку «Do Not Sell or Share My Personal Information».

CPRA значно посилив CCPA, наблизивши його до європейських стандартів. Зокрема, було введено концепцію «чутливих персональних даних» (геолокація, біометрика, раса тощо), для обробки яких потрібне явне згода або надається право на їх обмеження. Також було розширено права споживачів, додавши право на виправлення даних та захист від автоматизованого прийняття рішень. Закон застосовується до компаній, які ведуть бізнес у Каліфорнії та відповідають певним критеріям обороту або обсягу даних.

Privacy by Design та Privacy by Default: етика, вбудована в архітектуру. Одним з найважливіших вимог GDPR, що безпосередньо впливає на дизайн медіа-продуктів, є принципи Privacy by Design (Захист приватності за задумом) та Privacy by Default (Приватність за замовчуванням). Це не про додавання заходів безпеки після факту, а про проактивне вбудовування приватності в архітектуру системи та бізнес-процеси ще на етапі проектування.

• Data Minimization (Мінімізація даних): Збирати лише ті дані, які абсолютно необхідні для конкретної заявленої мети. Наприклад, для email-розсилки не потрібен номер телефону користувача.

• Purpose Limitation (Обмеження мети): Дані, зібрані для однієї цілі (наприклад, підписки на новини), не можуть використовуватися для іншої (таргетингу реклами партнерів) без нової явної згоди.

• Privacy by Default: Налаштування конфіденційності мають бути максимально суворими за замовчуванням. Користувачеві не слід вручну відключати відстеження або обміну даними — за замовчуванням система має його захищати. Ці принципи, вперше сформу-

льовані Анною Кавукян, вимагають від медіа-компаній переосмислення всього життєвого циклу даних — від моменту їхнього збору до фінального видалення.

Український контекст: гармонізація з GDPR як стратегічний вектор. Україна дійсно активно гармонізує своє законодавство з європейським правом, включно з GDPR. Основним законом залишається Закон України «Про захист персональних даних» від 2010 року, заснований на старій євро-директиві. Однак у парламенті вже знаходиться новий проєкт Закону № 8153, який був прийнятий в першому читанні та прямує до повної адаптації принципів і термінології GDPR. Цей законопроєкт має запровадити зобов'язання щодо повідомлення про порушення, оцінки впливу на захист даних (DPIA) та суттєво підвищити штрафи.

Таблиця 5.1.

Ключові відмінності між українським правом та GDPR

Аспект	GDPR	Закон України «Про захист персональних даних» (нині)	Проєкт Закону № 8153 (майбутнє)
Територіальна дія	Екстериторіальна (на всіх, хто обробляє дані громадян ЄС)	Застосовується до обробки даних в Україні	Має містити механізми для нерезидентів
Згода	Явна, інформована, активна (Opt-in)	Може бути отримана шляхом дій (напр., заповнення форми)	Очікується наближення до стандартів GDPR
Права користувачів	Широкий спектр (доступ, виправлення, забуття, перенесення тощо)	Схожі, але немає права на перенесення даних	Очікується повна гармонізація з GDPR
Штрафи	До 20 млн євро або 4% від глобального обороту	Значно нижчі, в кілька тисяч гривень	Очікується суттєве підвищення

Аспект	GDPR	Закон України «Про захист персональних даних» (нині)	Проєкт Закону № 8153 (майбутнє)
DPO (Уповноважений з захисту даних)	Обов'язковий у певних випадках	Не обов'язковий	Очікується запровадження вимог

Складено автором. Джерело. "Відмінності GDPR та українського закону" // Юридична газета (2025). https://yur-gazeta.com/publications/ practice/ zahist-intelektualnoyi-vlasnosti-avtorske-pravo/vidminnosti-mizh-novoyu-ukrayinskoy

Для українських медіа дотримання цих норм — це не лише «квиток на глобальний ринок», але й стратегічна інвестиція в довіру аудиторії. В умовах цифрової трансформації та підвищеної обізнаності користувачів про свої права, прозорість і повага до приватності стають конкурентною перевагою та основою для побудови довгострокових відносин з читачами. Гармонізація законодавства робить вивчення принципів GDPR не просто вимогою для роботи з європейською аудиторією, а невід'ємною частиною професійної культути сучасного медіа-фахівця в Україні.

5.5.2. Consent management та прозорість.

Епоха безкарного масштабного збору даних без згоди користувача дійсно закінчилася, і на зміну їй прийшли суворі правові рамки, зокрема GDPR, які повністю переосмислили процес управління згодою. CMP (Consent Management Platform) виступає не просто «брамником», а центральним техніко-юридичним механізмом, що перетворює абстрактні правові вимоги на практичну дію. Його роль полягає в автоматизованому скануванні сайту, ідентифікації всіх трекерів (від Google Analytics до Facebook Pixel) та їх безумовному блокуванні до моменту отримання дійсної згоди. Крім того, CMP слугує цифровим нотаріусом, що фіксує факт, час та обсяг наданої згоди, надаючи беззаперечний доказ відповідності вимогам регулятора. Однак наявність CMP сама по собі не гарантує законності; її конфігурація та дизайн інтерфейсу повинні неухильно дотримуватися принципів GDPR, зокрема щодо добровільності, інформованості та простоти відмови.

Фундаментальною основою будь-якої системи управління згодою є принцип гранулярності, що безпосередньо випливає з вимоги GDPR про специфічну та інформовану згоду. Категорізація

куків, як-от «Строго необхідні», «Статистичні», «Функціональні» та «Маркетингові», не є формальністю, а технічною реалізацією права користувача контролювати, для яких саме цілей обробляються його дані. «Строго необхідні» куки, без яких сайт не функціонує (наприклад, для кошика покупок або автентифікації), є єдиними, для яких згода не потрібна. Для всіх інших категорій, особливо для маркетингу та створення профілю, вимога до згоди є найсуворішою. Будь-яке порушення цього принципу, наприклад, активація маркетингових трекерів при згоді лише на аналітику, робить всю згоду недійсною та веде до порушення закону.

Шаруватий підхід до прозорості інтерфісу є відповіддю на вимогу GDPR про надання інформації у зрозумілій та легкодоступній формі, використовуючи зрозумілу мову. Перший шар у вигляді простого банера з основним повідомленням слугує точкою входу. Другий шар («Налаштування») має забезпечувати реальний вибір, надаючи перемикачі для кожної категорії куків. Третій шар («Політика конфіденційності») має містити повну юридичну інформацію, включаючи перелік усіх партнерів-обробників даних. Судова практика в ЄС показує, що приховування цієї інформації (наприклад, за межами області прокрутки) або використання маніпулятивних формулювань (як «оптимальний досвід») робить згоду недійсною.

Найкритичнішим етичним та юридичним викликом сьогодні є боротьба з Cookie Fatigue та маніпулятивним дизайном, відомим як dark patterns. Законодавство та регулятори чітко посилюють позицію: відмова має бути такою ж легкою, як і згода. Це означає, що кнопка «Reject All» (Відхилити все) має бути не менш видимою, доступною та потребувати такої ж кількості кліків, як і кнопка «Accept All» (Прийняти все). Як зазначає регулятор Великої Британії, це є мінімальною вимогою для забезпечення вільного вибору. Адміністративний суд Ганновера в Німеччині у 2025 році підтвердив, що відсутність легкої можливості відмови робить отриману згоду недійсною та порушує GDPR. Це підтверджує прецедент зі штрафом для Facebook у Франції у розмірі 60 мільйонів євро саме за те, що відмова від куків була прихована в налаштуваннях.

Важливо розуміти, що управління згодою не закінчується одним кліком. Відповідно до GDPR, право на відкликання згоди є таким же фундаментальним, як і право на її надання, і реалізувати

його має бути настільки ж просто. Це вимагає від медіа-компаній впровадження постійно доступних та інтуїтивно зрозумілих інструментів (наприклад, в особистому кабінеті або через постійну посилання в підвалі сайту), що дозволяють користувачеві змінити свої налаштування або повністю видалити згоду в будь-який момент.

Для українських медіа, що прагнуть до глобальної аудиторії, дотримання цих стандартів — це не лише «квиток на європейський ринок». Через екстериторіальний принцип дії GDPR, будь-яка українська платформа, яка цілеспрямовано обслуговує користувачів з ЄС (навіть маючи лише локалізовану версію контенту), повністю підпадає під його юрисдикцію та ризикує величезними штрафами. Впровадження прозорого та етичного управління згодою стає, таким чином, необхідним компонентом не лише юридичного захисту, а й стратегії побудови довіри з міжнародною та вітчизняною аудиторією, яка все більше цінує контроль над своїми даними. Це перетворює CMP з обов'язкового технічного елемента на інструмент конкурентної переваги та демонстрації професійної медіа-культури.

5.5.3.Анонімізація та агрегація даних.

Виклик безпеки в епоху великих даних. Збереження «сирих» даних (Raw Data), що містять особисту інформацію, пов'язане з серйозними ризиками порушення конфіденційності. Завдання сучасних медіа-технологій полягає у знаходженні балансу між потребою в глибокій аналітиці для персоналізації та бізнес-аналітики та обов'язком захисту приватності користувачів. Відповіддю на цей виклик стають технічні методи агрегації, анонімізації та передові математичні моделі приватності, які дозволяють перетворити індивідуальну інформацію на анонімну статистику, зберігаючи її аналітичну цінність.

Агрегація: безпека через об'єднання. Агрегація — це фундаментальний процес об'єднання індивідуальних даних у групи (когорти). Замість спостереження за окремими «деревами» (наприклад, «Марія, 25 років, читала статтю про розлучення о 23:00») система аналізує цілий «ліс» («Вчора статтю про розлучення прочитали 150 жінок віком 20-30 років»). Цей підхід є основою таких сервісів, як Google Analytics, які надають зведені відомості про аудиторію, але не розкривають особисту інформацію окремих користувачів. Це ідеально підходить для стратегічного планування, маркетингу та вдосконалення продукту, однак повністю виключає

можливість індивідуальної персоналізації на основі таких даних.

Псевдонімізація проти анонімізації: принципова відмінність. Нерозуміння різниці між цими двома поняттями може призвести до критичних помилок у захисті даних. Псевдонімізація — це заміна прямих ідентифікаторів (ім'я, email) на унікальний код (User_ID_998877). Ця операція є оборотною: десь зберігається «ключ» (таблиця відповідності), що дозволяє пов'язати код з конкретною особою. Саме тому згідно з GDPR псевдонімізовані дані все ще вважаються персональними даними, оскільки за наявності додаткової інформації особу можна ідентифікувати. Анонімізація — це повне та незворотне видалення або зміна всіх ідентифікаторів таким чином, щоб відновити особу стало неможливо. Після успішної анонімізації дані виходять з-під дії GDPR, оскільки перестають бути даними про фізичну особу.

Передові техніки гарантованого захисту: k-анонімність та диференційна приватність. Для досягнення справжньої анонімізації недостатньо просто видалити імена. Необхідні математично обґрунтовані моделі.

• k-анонімність: Індивідуальний запис у наборі даних вважається захищеним, якщо існує щонайменше k-1 інших записів із точно такими самими квазі-ідентифікуючими атрибутами (наприклад, поштовий індекс, вік, стать). Якщо в базі даних є 50 записів «Жінка, 30 років, лікар, Київ», то виділити одну з них неможливо. Проте метод має слабкі місця: якщо значення атрибутів унікальні (наприклад, «Чоловік, 105 років, космонавт»), запис легко ідентифікувати.

• Диференційна приватність: Цей «золотий стандарт», який використовують такі компанії, як Apple і Google, вирішує проблеми попередніх методів. Його суть полягає в додаванні контрольованого статистичного «шуму» (випадкових похибок) до результатів запитів або самих даних. Це гарантує, що присутність або відсутність будь-якого окремого індивіда в наборі даних практично не впливає на кінцевий результат аналізу. Таким чином, тренди для групи залишаються вірними (наприклад, «60% читачів-чоловіків віком 30+ зацікавились статтею»), але визначити, чи міститься в цих 60% конкретний користувач Іван, стає математично неможливим.

Таблиця 5.2.

Порівняльна характеристика методів анонімізації та де-ідентифікації даних

Критерій	Агрегація	Псевдонімізація	k-Анонімність	Диференційна приватність
Основний принцип	Об'єднання даних у групи.	Заміна ідентифікатора на код.	Приховування особи в групі з k однакових записів.	Додавання математичного «шуму» для захисту будь-якого окремого запису.
Зворотність	Незворотна (індивідуальні дані втрачаються).	Зворотна (за наявності ключа).	Зазвичай незворотна, але вразлива до атак.	Незворотна та захищена від атак.
Правовий статус за GDPR	Не персональні дані.	Персональні дані.	Можуть бути персональними, якщо існує ризик деанонімізації.	Найвищий рівень захисту; правильно реалізовані дані не є персональними.
Основне використання	Бізнес-аналітика, звітність.	Внутрішня обробка даних із зниженням ризиків.	Публікація наборів даних для досліджень.	Захищені аналітичні запити, машинне навчання на чутливих даних.

Складено автором. Джерело. Сачик Т.В. "Порівняння алгоритмів k-анонімізації" // TNTU (2019). https://elartu.tntu.edu.ua/ bitstream/lib/30587 /12/ Dyp _%20 Sachyk_2019.pdf

Ризик деанонімізації: історія Netflix Prize як застереження. Навіть ретельно анонімізовані дані не застраховані від реідентифікації (de-anonymization) при поєднанні з іншими джерелами. Класичним прикладом є історія Netflix Prize. У 2006 році Netflix опублікував анонімний набір даних, що містив 100 мільйонів оцінок фільмів від 500 000 користувачів, з метою покращення рекомендаційної системи. Імена були замінені випадковими числами. Однак дослідникам вдалося деанонімізувати частину записів, зіставивши шаблони оцінок і часові мітки з публічними профілями користувачів на сайті IMDb (Internet Movie Database), де люди часто залишають оцінки під своїми іменами.

Цей кейс демонструє ключові ризики:

1. Унікальність поведінкових відбитків: Наші культурні вподобання (які фільми ми дивимося, які статті читаємо) часто є унікальними і виступають потужним ідентифікатором. Дослідження показало, що іноді достатньо всього 8 оцінок фільмів, щоб унікально ідентифікувати людину в великому наборі даних.

2. Легкість поєднання джерел: Сьогодні існує безліч публічних джерел даних (соціальні мережі, форуми, відкриті реєстри), за допомогою яких можна здійснити подібну атаку.

3. Підвищення загрози завдяки ШІ: Сучасні великі мовні моделі (LLM), такі як GPT, значно спрощують та автоматизують процес деанонімізації, роблячи його доступним навіть для тих, хто не володіє глибокими технічними знаннями.

Висновок для медіа-галузі. Для українських медіа, які прагнуть до глобальних стандартів і довіри аудиторії, ретельне застосування методів анонімізації є критично важливим. Порушення принципів захисту даних може призвести не лише до втрати довіри користувачів, але й до значних юридичних наслідків, таких як величезні штрафи за GDPR. Диференційна приватність сьогодні є найбільш перспективним та надійним шляхом для аналітики, що відповідає вимогам етики та законодавства. Кейс Netflix Prize назавжди залишиться нагадуванням про те, що справжня анонімізація — це складне мистецтво, що вимагає глибокого розуміння як технологій, так і потенційних загроз. Учасники медіа-ринку повинні будувати свої системи з урахуванням принципу «захист приватності за задумом» (Privacy by Design), інтегруючи найкращі практики анонімізації в саму архітектуру своїх продуктів.

5.5.4. Етичне використання персональних даних.

Етика в сфері даних — це принципи поведінки, які виходять за рамки формальної відповідності закону. Ключове правило можна сформулювати так: «Те, що технічно можна зробити, не обов'язково слід робити». Це питання морального вибору, особливо коли ніхто не контролює.

Одним із центральних понять є «моторошна межа» (The Creepy Line) — невидимий рубіж між корисною персоналізацією та діями, що сприймаються як переслідування або вторгнення. Корисна персоналізація ґрунтується на явних, обмежених даних: наприклад,

рекомендація статті про смартфон після перегляду схожої теми. Моторошною ж вона стає, коли використовуються детальні, виведені або таємно зібрані дані про реальну поведінку людини, наприклад, пропозиція знижки на основі точного відстеження перебування біля певної вітрини. Класичним прикладом порушення цієї межі є історія з мережею Target. Її алгоритм, аналізуючи зміни в покупках (вітаміни, лосьйон без запаху), ідентифікував вагітність школярки та почав надсилати їй відповідні рекламні пропозиції. Це стало відомо її батькові, що викликало скандал. Етичний урок полягає в тому, що виведені дані (Inferred Data) про надзвичайно чутливі аспекти життя (здоров'я, вагітність, фінанси) не можна використовувати для маркетингу без явної, інформованої та добровільної згоди людини.

Інший ключовий принцип — контекстуальна цілісність (Contextual Integrity), запропонована Хелен Ніссенбаум. Він стверджує, що дані не є абсолютно приватним або публічним благом; їхня допустимість визначається контекстом отримання та використання. Порушення виникає, коли інформація, передана в одному контексті довіри (наприклад, медичні відомості лікарю для лікування), використовується в зовсім іншому, часто комерційному контексті (продаж цих даних страховій компанії для таргетингу), навіть якщо це технічно дозволено дрібним шрифтом угоди. Аналогічно, email-адреса, надана для підписки на новини, дана в контексті медіаспоживання. Її використання для пошуку профілю користувача в соціальних мережах з метою аналізу його політичних поглядів є явним порушенням вихідного контексту взаємин.

Третій критичний аспект — алгоритмічна справедливість (Fairness). Етика даних стосується не лише збору, але й наслідків автоматизованої обробки інформації. Упередженість (Bias), вбудована в дані або логіку алгоритму, може призводити до системної дискримінації. Наприклад, якщо система рекомендації вакансій показує високооплачувані керівні посади переважно чоловікам лише на тій підставі, що історично їх переглядали чоловіки, вона лише посилює існуючу соціальну нерівність. Окремо стоїть етична проблема формування «ехо-камер» (фільтрованих бульбашок) в медіа. Питання в тому, чи є прийнятним з етичної точки зору показувати користувачеві лише контент, що підтверджує його існуючі переконання, заради максимізації його уваги та прибутків платформи.

Така практика, хоч і ефективна бізнесово, завдає шкоди суспільству, підриваючи плюралізм думок і здорову демократичну дискусію.

Отже, етичне поводження з даними вимагає не просто технічного захисту, а глибокого розуміння контексту, поваги до автономії людини та усвідомлення соціальних наслідків алгоритмічних рішень.

5.5.5. Захист приватності шляхом проектування (Privacy by design).

Концепція Privacy by Design (Захист приватності шляхом проектування) є фундаментальною філософією в сучасному світі обробки даних. Розроблена доктором Енн Кавукян, вона передбачає радикальну зміну підходу: замість того, щоб розглядати приватність як додаткову функцію або юридичне зобов'язання, яке можна «додати» після факту, PbD вимагає вбудовувати її принципи в саму архітектуру системи, продукту або послуги ще на етапі їх створення. Цей проактивний підхід став настільки важливим, що був інкорпорований як обов'язкова вимога в Статті 25 Загального регламенту захисту даних (GDPR), що формалізує його правовий статус для великої кількості організацій.

Серцем PbD є сім фундаментальних принципів, які взаємодоповнюють один одного. Перший принцип — проактивність, а не реактивність — закладає основу усієї концепції. Він зміщує фокус з реагування на порушення та сплати штрафів на попередження ризиків. Це означає проведення оцінки впливу на захист даних на самому початку проекту, щоб передбачити потенційні загрози приватності та вирішити їх до того, як вони реалізуються. Найважливішим, часто центральним принципом є Privacy as the Default (Приватність за замовчуванням). Він постулює, що максимальний рівень захисту персональних даних має автоматично застосовуватися до будь-якого користувача системи. Користувач не повинен виконувати складні дії в налаштуваннях, щоб захистити себе. Навпаки, якщо він бажає послабити захист для отримання додаткових функцій (наприклад, зробити свій профіль публічним), він повинен зробити це свідомо, ввімкнувши відповідну опцію. Це перевертає традиційну парадигму, де за умовчанням дані часто збираються максимально широко.

Третій принцип — вбудованість у дизайн — підкреслює, що приватність має бути невід'ємною частиною технічної реалізації, а не забезпечуватися лише юридичними документами, такими як

політика конфіденційності. Вона має бути закодована у функціях, алгоритмах та інфраструктурі. Принцип повної функціональності (Positive-Sum) спростовує поширений міф про те, що між приватністю та іншими цілями (як безпека, функціональність або прибуток) існує невблаганний компроміс. PbD стверджує, що за допомогою креативного проектування можна досягти всіх цілей одночасно, не жертвуючи однією заради іншої. Наскрізна безпека (End-to-End Security) вимагає реалізації належних заходів захисту (шифрування, контроль доступу, цілісність) на кожному етапі життєвого циклу даних — від моменту їх збору, через обробку та зберігання, до фінального і надійного знищення, коли в них більше немає потреби.

Останні два принципи стосуються відповідальності та контролю. Видимість і прозорість означають, що механізми захисту приватності повинні бути перевіряємими як користувачами (у доступній формі), так і незалежними аудиторами. Система повинна працювати так, як обіцяно, а не приховувати свою справжню діяльність. Нарешті, повага до приватності користувача робить інтереси та права людини абсолютним пріоритетом, вищим за зручність розробника чи бізнес-модель. Це користоцентричний підхід у чистому вигляді.

На практиці реалізація PbD в технологічних продуктах, зокрема в медіа, виражається в дотриманні правила мінімізації даних. Інженер, який керується PbD, керується мантрою: «Дані, які ви не зібрали, неможливо вкрасти, втратити чи зловжити ними». Це різко контрастує з традиційним «агресивним» підходом до збору даних, коли збирається все можливе «про запас» для потенційного майбутнього аналізу або навчання штучного інтелекту. Замість цього, PbD вимагає для кожної точки збору даних поставити питання: «Чи є ці дані абсолютно необхідними для надання конкретної заявленої послуги?». Наприклад, для відправки email-дайджесту новин номер телефону користувача не потрібен, отже, поле для нього слід взагалі прибрати з форми реєстрації. Це не лише знижує ризики, але й спрощує інтерфейс та зменшує юридичні зобов'язання компанії.

Ключовим інструментом для впровадження PbD є управління життєвим циклом даних (Data Lifecycle Management). Цей принцип вимагає визначення «терміну придатності» для кожної категорії даних. Дані не повинні зберігатися вічно «просто так». Система має бути спроектована таким чином, щоб автоматично видаляти

або анонімізувати інформацію, коли вона досягає заздалегідь встановленого часу життя (TTL — Time to Live). Наприклад, сесійні куки для аналітики можуть автоматично видалятися через 30 хвилин після неактивності користувача, а історія переглядів для системи рекомендацій — автоматично анонімізуватися через певний період (наприклад, 14 місяців). Цей процес має бути автоматизованим і не вимагати втручання адміністратора, що усуває людський фактор і забезпечує дотримання політик.

Таким чином, Privacy by Design — це не просто набір технічних вимог, а цілісна система мислення, яка інтегрує правові, етичні та інженерні аспекти. Вона перетворює приватність з пасивного об'єкта захисту на активний компонент якості та безпеки продукту, будуючи довіру з користувачами та створюючи стійкіші та більш відповідальні технології для цифрової епохи.

1. КОНТРОЛЬНІ ПИТАННЯ

1.Термінологія: У чому різниця між *Data-Driven* (керований даними) та *Data-Informed* (поінформований даними)? (Data-Driven сліпо виконує те, що кажуть цифри. Data-Informed використовує цифри як пораду, але залишає фінальне рішення за людиною/редактором).

2. Валюта: Чому email-адреса вважається "цифровим паспортом" в інтернеті? (Це унікальний ідентифікатор, який дозволяє об'єднати дані з різних пристроїв і платформ).

3.Walled Gardens: Що це таке і чому це проблема для медіа? (Закриті екосистеми, як Facebook чи Google, які не віддають медіа дані про їхніх читачів. Медіа бачить лише "трафік", але не знає, хто ці люди).

4 Data Rot (Гниття даних): Як швидко застарівають дані? (Поведінкові дані ("шукає піцу") живуть 1 годину. Демографічні ("є діти") — роками. Важливо не використовувати старі інтенційні дані).

2. ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. Аудит "Сліпих зон"

• Ситуація: Ваше медіа має сайт, Telegram-канал та YouTube.

• Питання: Чи знаєте ви, що користувач Іван, який коментує в Telegram, це той самий Іван, який дивиться YouTube?

• Дія: Якщо ні — у вас немає Single Customer View. Запропонуйте механізм об'єднання (наприклад, посилання в Telegram на ексклю-

зивну статтю на сайті, яка вимагає реєстрації).

Завдання 2. Класифікація даних

• Завдання: Розподіліть дані по категоріях (Zero, First, Third):

- Історія переглядів на вашому сайті. (First)
- Куплена база адрес "Власники авто Києва". (Third)
- Відповідь на опитування "Яку музику ви любите?". (Zero)
- Дані з Facebook Pixel про інтереси. (Third/Second).

Завдання 3. Вибір метрики полярної зірки (North Star Metric)

• Умова: Ви будуєте медіа за підпискою.

• Завдання: Оберіть одну головну метрику успіху.

- *Варіант А:* Кількість унікальних відвідувачів (Users).
- *Варіант Б:* Кількість переглядів сторінок (Pageviews).
- *Варіант В:* Кількість днів активності на місяць (DAU/MAU).
- *Відповідь:* Варіант В. Для підписки важлива звичка (Retention), а не разовий захід.

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Croll, A. & Yoskovitz, B. *Lean Analytics: Use Data to Build a Better Startup Faster.*

- *Про що:* Хоча книга про стартапи, це "біблія" для медіа-менеджерів. Вона вчить обирати "Єдину Метрику, Яка Має Значення" (OMTM) і не тонути в морі цифр.

2. The Reuters Institute. *Digital News Report*. (Щорічні звіти).

- *Про що:* Найавторитетніше джерело глобальних трендів. Розділи про те, як змінюється довіра до медіа і як аудиторія ставиться до збору даних.

3. Financial Times. *Defining a North Star Metric.* (Статті та кейс-стаді).

- *Про що:* FT — піонери використання метрики RFV (Recency, Frequency, Volume). Варто почитати, як вони перейшли від підрахунку кліків до підрахунку звички читання.

4. IAB (Interactive Advertising Bureau). *The Demise of Third-Party Cookies.*

- *Про що:* Гайди про те, як вижити медіа-бізнесу в світі без сторонніх куків (Cookiepocalypse).

4. ГЛОСАРІЙ ТЕРМІНІВ

• DIKW (Data, Information, Knowledge, Wisdom): Ієрархічна модель, що описує процес перетворення сирих даних (цифри) на мудрість (рішення). Головна мета аналітики — піднятися на вершину цієї піраміди.

• Zero-Party Data (Дані нульової сторони): Дані, які користувач надає бренду свідомо та добровільно (відповіді на опитування, налаштування вподобань у профілі). Найбільш цінні та точні дані.

• First-Party Data (Дані першої сторони): Дані про поведінку користувача, які компанія збирає безпосередньо на своїх платформах (сайт, додаток, email). Це актив, який належить редакції.

• Third-Party Data (Дані третьої сторони): Дані, придбані у сторонніх брокерів або зібрані через трекери на інших сайтах. Втрачають актуальність через блокування куків браузерами.

• Single Customer View (SCV / Єдиний профіль клієнта): Консолідований запис, який об'єднує всі дані про одного користувача з різних каналів (мобільний додаток, сайт, соцмережі) в одну картку.

• CDP (Customer Data Platform): Програмне забезпечення, яке збирає дані з усіх джерел, уніфікує їх у профілі користувачів і дозволяє іншим системам (наприклад, сайту) використовувати ці дані для персоналізації в реальному часі.

• North Star Metric (Метрика полярної зірки): Головний показник, який найкраще відображає основну цінність продукту для клієнта. Для медіа це часто "Час, проведений за читанням" або "Частота візитів", а не просто "Кількість переглядів".

• Walled Garden (Огороджений сад): Закрита екосистема платформи (наприклад, Facebook, Apple, Google), яка контролює всі операції всередині себе і не ділиться даними користувачів зі сторонніми видавцями.

Висновок до Розділу 5

Сучасна діяльність у медіа, маркетингу та цифровій продуктивності ґрунтується на зборі та аналізі даних про аудиторію. Однак цей процес є не лише технологічним завданням, а й комплексом етичних, правових і проектно-архітектурних викликів. Ключовим висновком розділу є те, що ефективний та відповідальний аналіз даних неможливий без вбудовування захисту приватності на всіх етапах — від ідеї до реалізації.

Основні підсумки можна сформулювати так:

1. Технічні методи захисту даних є базою, але не панацеєю. Агрегація, псевдонімізація, k-анонімність та диференційна приватність надають різні рівні захисту та зворотності. Їх вибір залежить від цілей обробки даних (наприклад, внутрішня аналітика чи публікація наборів для досліджень). Жоден метод не є ідеальним, а правовий статус результатів їх застосування (чи є дані персональними) може відрізнятися, що вимагає уважного аналізу.

2. Етика визначає межі прийнятного. Технічна можливість зібрати та проаналізувати глибоко особисту інформацію не означає морального права це робити. Концепції «моторошної межі» та контекстуальної цілісності чітко показують, що порушення довіри користувача та контексту, в якому він надав дані, призводить до втрати лояльності та репутаційних ризиків. Випадок Target є класичним попередженням про небезпеку використання виведених даних про чутливі аспекти життя.

3. Алгоритмічна справедливість — це обов'язок. Аналіз даних та алгоритми, побудовані на їх основі, можуть ненавмисно посилювати історичні упередження (bias) та соціальну нерівність, створюючи «ехо-камери». Відповідальний підхід вимагає активних зусиль з виявлення та усунення таких упереджень, щоб системи були справедливими та недискримінаційними.

4. Privacy by Design (PbD) — це обов'язковий стандарт. Найважливішим висновком є те, що захист приватності має бути проактивним і вбудованим у дизайн системи за замовчуванням. Принцип мінімізації даних («не збирай те, що тобі не потрібно») та управління життєвим циклом даних (автоматичне видалення за закінченням терміну) є практичними інструментами реалізації PbD. Це не перешкода для інновацій, а умова створення стійких, надійних і довірених продуктів, що відповідають вимогам GDPR та очікуванням суспільства.

У підсумку, успішний і відповідальний збір та аналіз даних про аудиторію вимагає поєднання технічної компетентності у використанні інструментів анонімізації, глибокого етичного мислення для оцінки наслідків та архітектурного підходу Privacy by Design на етапі проектування. Лише такий комплексний підхід дозволяє отримувати цінну аналітичну інформацію, не порушуючи фундаментальних прав і довіри людей, забезпечуючи тим самим сталий розвиток будь-якого цифрового продукту чи сервісу.

Розділ 6: Автоматизація контент-виробництва
6.1. Шаблонна генерація контенту

6.1.1. Natural Language Generation (NLG).

Natural Language Generation (NLG) є критично важливою підгалуззю штучного інтелекту, яка займається автоматичним створенням людямзрозумілих текстів зі структурованих даних. Фундаментально її роль можна окреслити як навчання машини писати, тоді як Natural Language Processing (NLP) навчає її читати. Ця технологія лягає в основу автоматизації контенту, перетворюючи цифри, фактори та метрики з таблиць і баз даних на зв'язні, інформативні текстові звіти, новини чи резюме. У контексті серйозних індустрій, таких як фінанси, спортивна аналітика чи бізнес-звітність, NLG виступає "двигуном", що трансформує об'єктивні, але німі дані в читабельні історії, зрозумілі для кінцевого користувача.

Еволюція та парадигми: Детермінована точність проти ймовірнісної творчості. Історично та в рамках шаблонної генерації NLG розвивався від Rule-Based (систем на основі правил) підходів до складніших моделей на базі Large Language Models (LLM), таких як ChatGPT. Кожна з цих парадигм має суттєво різні принципи роботи та сфери застосування.

Rule-Based NLG (детермінований підхід): Це класична, передбачувана система. Вона працює за жорстко заданими шаблонами та логічними правилами, запрограмованими людьми. Її процес можна уявити як заповнення пропусків у заздалегідь підготовлених фразах конкретними значеннями з бази даних («[Команда А] [дія] [Команду Б] з рахунком [рахунок]»). Її головна перевага — абсолютна точність і контроль над виводом. Система не "вигадує" факти, а лише відтворює їх у заданій формі. Це робить її незамінною там, де помилка в цифрі чи факті має критичні наслідки: у автоматизованих фінансових звітах, медичних заключеннях, спортивних дайджестах чи прогнозах погоди.

NLG на основі LLM, як-от ChatGPT (ймовірнісний підхід): Сучасні генеративні моделі працюють на іншому принципі. Вони передбачають наступне слово в послідовності, ґрунтуючись на статистичних паттернах, вивчених з колосальних масивів тексту. Це дозволяє їм створювати тексти, що виглядають більш творчими,

гнучкими та "людськими". Однак, вони схильні до "галюцинацій" — генерації правдоподібної, але вигаданої інформації. Висновок очевидний: для завдань, що вимагають креативності та стилістичного розмаїття (генерація ідей, креативний копірайтинг, неформальні діалоги), краще підходять LLM. Для завдань, що вимагають гарантованої, відтворюваної точності — незамінними залишаються rule-based системи. У серйозній журналістиці даних часто саме старі добрі правила забезпечують ту надійність, якої не вистачає сучасним "розумним" моделям.

Анатомія процесу: Конвеєр генерації тексту. Створення тексту машиною — це не одномоментний акт, а структурований конвеєр, що складається з декількох послідовних етапів (за класичною моделлю Рейтера і Дейла):

1. Паливо Визначення змісту (Content Determination): На цьому етапі система вирішує, що саме сказати з наявного масиву даних. Це процес, аналогічний роботі редактора: з сотні статистичних показників футбольного матчу для короткої новини можуть бути обрані лише ключові: рахунок, автори голів, результат. Інші дані (володіння м'ячем, кутові) будуть відфільтровані.

2. Структурування документа (Document Structructuring): Після вибору фактів система визначає, в якому порядку їх подати. Вона будучить логічну структуру, наприклад, для новини: головний результат -> деталі матчу -> цитата тренера -> позиції команд у таблиці -> майбутній суперник.

3. Агрегація (Aggregation): Цей етап відповідає за стилістичну економію та читабельність. Система об'єднує просту послідовність фактів у складніші, природніші речення. Наприклад, замість двох речень «Іван забив гол на 15-й хвилині. Петро забив гол на 77-й хвилині», система може сгенерувати: «Іван та Петро відзначилися голами на 15-й та 77-й хвилинах відповідно».

4. Лексикалізація (Lexicalization): Це вибір конкретних слів та фраз для передачі значення. Тут система може враховувати контекст та емоційне забарвлення. Наприклад, для перемоги з рахунком 1:0 вона обере нейтральне «перемогла», а для рахунку 5:0 — більш експресивне «розгромила».

Реалізація (Realization): Фінальний технічний етап, на якому застосовуються правила граматики, морфології та синтаксису

обраної мови для створення коректного тексту. Система узгоджує відмінки, роди, числа та часи, перетворюючи набір слів і фраз у граматично правильне речення для NLG: Критична важливість структурованих даних. NLG, особливо його детермінована, rule-based форма, принципово не може працювати з неструктурованою інформацією. Його «паливом» є чітко організовані, машиночитані дані. Формати, такі як XML, JSON, CSV чи реляційні бази даних SQL, надають інформацію у вигляді колонок, рядків, тегів та ключ-значень, що дозволяє системі однозначно ідентифікувати, яке значення в яке місце шаблону підставити. Звичайний текстовий документ (наприклад, файл Word з нотатками) для такої системи є непрозорим «повітрям» — в ньому відсутня явна структура, яку можна було б алгоритмічно обробити без попереднього складного аналізу засобами NLP. Таким чином, якість та структурованість вхідних даних є першою та обов'язковою умовою успішної генерації тексту.

У підсумку, Natural Language Generation, зокрема його rule-орієнтована гілка, представляє собою потужний інструмент для масштабованої, точної та надійної комунікації даних. Він не намагається імітувати людську творчість, а виконує іншу, не менш важливу функцію: стає високопродуктивним "перекладачем" з мови чисел на мову людей, забезпечуючи основу для автоматизації в тих сферах, де правда факту важливіша за красу слів.

6.1.2. Structured data to text: фінансові звіти, спортивні результати, погода.

На відміну від творчих есе або абстрактних роздумів, реальна сила сучасної шаблонної генерації контенту (Structured Data-to-Text) розкривається там, де мова йде про об'єктивні, вимірювані факти, представлені у формі чітких цифр і показників. Близько 90% всього автоматично створеного контенту в світі припадає на три ключові жанри: фінансові звіти, спортивні дайджести та прогнози погоди. Ці сфери є ідеальним паливом для rule-based систем Natural Language Generation (NLG), оскільки вони поєднують в собі необхідність масштабу, швидкості та абсолютної точності з доступністю високоякісних структурованих даних.

1. Фінансові звіти (Earnings Reports): Швидкість як конкурентна перевага. Фінансова звітність стала першим масштабним і успішним

кейсом впровадження NLG у практику великих медіа. Проблема полягала в обмежених ресурсах: щокварталу тисячі публічних компаній по всьому світу публікують звіти про прибутки та збитки, але редакції, на кшталт Associated Press (AP), силами журналістів встигали висвітлити лише топ-300–500 найбільших корпорацій. Це призводило до інформаційного вакууму навколо середніх та малих компаній, що негативно позначалося на їх інвестиційній привабливості та ліквідності акцій.

Рішенням стало партнерство AP з платформою Automated Insights. Алгоритм, отримуючи структуровані дані у вигляді машинних файлів (наприклад, XML або JSON) з показниками виручки, прибутку та прогнозами аналітиків, почав автоматично генерувати короткі, але інформативні новинні заметки. Логіка проста та детермінована: IF (Revenue > Expectation) THEN "Компанія [Назва] перевершила очікування аналітиків". Результат був революційним: обсяг висвітлення зріс з кількох сотень до понад 4000 звітів за квартал. Цінність полягає не лише в масштабі, але й у неймовірній швидкості — алгоритм публікує новину впродовж 0.1–0.3 секунди після офіційного релізу. Для біржових трейдерів, де кожна мілісекунда має значення, ця миттєвість є критичною для прийняття рішень.

2. Спортивні результати (Sports Recaps): Масштаб, глибина та персоналізація. Спорт, як царство статистики (голи, передачі, кидки, час), є родним середовищем для NLG. Автоматизація дозволяє вирішити проблему «довгого хвоста»: якщо про матчі англійської Прем'єр-ліги пишуть тисячі журналістів, то про гри третьої ліги, університетських чи дитячих чемпіонатів інформації практично немає.

Класичним експериментом був проєкт, де батьки через мобільний додаток заповнювали протокол матчу дитячої бейсбольної ліги (Little League). Система на основі цих структурованих даних генерувала короткий звіт про кожну гру. Ключовою була глибока персоналізація: бабуся юного гравця отримувала не загальну статтю, а текст з заголовком на кшталт «Ваш онук зробив вирішальний удар у перемозі команди». Інший масштабний приклад — Fantasy Sports (віртуальні спортивні ліги). Платформи, такі як Yahoo Sports, автоматично генерують мільйони унікальних персональних звітів для кожного гравця фентезі-ліги, аналізуючи його статистику, успіхи та невдачі

в порівнянні з іншими. Фізично ні одна команда авторів не змогла б написати 10 мільйонів індивідуальних листів за ніч.

3. Погода (Weather): Від загального прогнозу до гіперлокальних порад. Прогноз погоди для великого міста — це лише усереднене значення. Справжню цінність надає гіперлокальне висвітлення, можливе завдяки NLG. Роботи аналізують дані з тисяч мікросенсорів, метеостанцій та радарів у реальному часі.

Завдання системи — перетворити цей масив чисел не просто на фразу «По Києву дощ», а на конкретну текстову пораду, що несе практичну користь: «Дощ почнеться на вулиці Сагайдачного через 15 хвилин, буде помірним і закінчиться через 40 хвилин. Радимо захопити парасольку». Такий підхід перетворює метеодані з пасивної інформації на активний інструмент планування дня, що має значно більшу цінність для користувача.

4. Обмеження методу: Кордони можливостей шаблонної генерації.

Однак сила rule-based NLG — його детермінованість і залежність від структури — є одночасно його головною слабкістю. Система безсила перед неструктурованими, непередбачуваними подіями, яких немає в шаблоні вхідних даних.

Яскравий приклад — футбольний матч, який перервався через те, що на поле вибігла собака, або через відключення електроенергії. У стандартному статистичному протоколі (box score) немає граф для таких подій. Тому автоматично згенерована новина констатуватиме лише голі факт: «Матч перервано на 85-й хвилині», — без жодного пояснення причини, що зробить текст незрозумілим або неповним. У таких ситуаціях незамінною залишається людина-журналіст, здатна інтерпретувати контекст, розпізнати унікальну подію та розповісти повну історію. Таким чином, шаблонна генерація не замінює творчу журналістику, а вивільняє людські ресурси для роботи над складнішими, аналітичними та розслідувальними матеріалами, доповнюючи її на рівні масштабної, рутинної роботи з даними.

6.1.3. Інструменти: Automated Insights, Narrative Science.

Сучасна цифрова екосистема формує складний, але логічно пов'язаний ланцюг перетворення інформації, де кожна ланка є критично важливою. Цей процес починається з фундаментальних питань етики та захисту приватності при зборі даних про аудиторію,

проходить через технології їхньої анонімізації та безпечної обробки і завершується потужними інструментами автоматичного перетворення структурованих даних на зрозумілий текст. Розуміння цього ланцюга є ключовим для створення відповідальних, ефективних та масштабованих медійних і бізнес-рішень.

Основою всього процесу є етичний збір та захист даних. У цифрову епоху інформація про аудиторію стала найцінніш активом, але її збір не може відбуватися поза рамками моральних принципів та правових норм. Концепції, як-от «моторошна межа», яка відділяє корисну персоналізацію від переслідування, та контекстуальної цілісності Хелен Ніссенбаум, яка вимагає використовувати дані лише в тому контексті, в якому вони були надані, стають не просто етичними орієнтирами, а практичними правилами. Не менш важливою є боротьба з алгоритмічною упередженістю, яка може посилювати соціальну нерівність, та усвідомлення шкоди від створення інформаційних «ехо-камер». Технічним вираженням цієї етики є Privacy by Design (PbD) — проактивне вбудовування захисту приватності в архітектуру системи з самого початку. Принципи PbD, такі як приватність за замовчуванням, мінімізація даних (збирати лише найнеобхідніше) та управління життєвим циклом даних (автоматичне видалення після закінчення строку придатності), зобов'язують розробників ставити права користувача вище за зручність. Ці принципи підкріплюються конкретними технологіями захисту: агрегацією, псевдонімізацією, k-анонімністю та найбільш потужною — диференційною приватністю, яка захищає окремі записи шляхом додавання математичного «шуму». Висновок очевидний: будь-яка подальша робота з даними, особливо їх автоматизована обробка та публікація, можлива лише на основі безпечного, законного та етичного фундаменту.

На цьому фундаменті будується потужна надбудова — автоматизована генерація контенту на основі Natural Language Generation (NLG). NLG — це галузь ШІ, що навчає машини «писати», перетворюючи структуровані дані на природномовний текст. У контексті промислового застосування домінує шаблонна (rule-based) генерація, яка є детермінованою, передбачуваною та гарантує 100% точність фактів. Це різко відрізняє її від генеративних моделей типу ChatGPT, які, будучи творчими, схильні до галюцинацій. Процес у шаблонній

системі — це чіткий конвеєр: від визначення ключового змісту та структурування документа до агрегації речень, вибору лексики та фінального граматичного оформлення. «Паливом» для такого NLG слугують виключно чітко організовані дані з XML, JSON, CSV або SQL-баз.

Найбільшого успіху шаблонна генерація досягла в трьох жанрах, що становлять 90% всього автоматизованого контенту: фінансові звіти, спортивні дайджести та гіперлокальні прогнози погоди. У фінансах вона дозволяє миттєво, протягом мілісекунд, висвітлювати тисячі квартальних звітів компаній, що критично для ринків. У спорті вона охоплює «довгий хвіст» матчів, про які ніхто не пише, та створює мільйони персоналізованих звітів для уболівальників. У погоді — трансформує дані з тисяч сенсорів у конкретні текстові поради для вулиці чи району.

Для реалізації цих завдань існують спеціалізовані промислові платформи класу Enterprise, такі як Automated Insights (Wordsmith), яка стала стандартом індустрії завдяки інтерфейсу, схожому на «Excel для письменників», та потужному API для повної автоматизації. Narrative Science (Quill) зробила півот від медіа до бізнес-аналітики, генеруючи текстові пояснення до корпоративних дашбордів, і була інтегрована в екосистему Salesforce Tableau. Ринець також включає нішевих гравців: Arria NLG для складних галузей (медицина, нафтогаз), AX Semantics для масштабного e-commerce та United Robots для автоматизації локальних новин.

Життєздатність цих «простих» rule-систем у епоху GPT-4 пояснюється їх ключовою перевагою — аудитуємістю (Auditability) або статусом «білої скриньки». На відміну від «чорної скриньки» нейромереж, де неможливо зрозуміти логіку виведення, у шаблонних системах кожна помилка може бути відстежена до конкретного правила та виправлена. Ця прозорість, контроль та гарантія точності є абсолютно необхідними у фінансах, регульованих галузях і серйозній журналістиці.

Таким чином, сучасний цикл обробки інформації утворює замкнутий контур: етика та Privacy by Design створюють безпечну правову й технічну основу для збору якісних даних → ці структуровані дані стають сировиною для детермінованих, надійних систем NLG → які, у свою чергу, масштабно та миттєво трансформують їх у цінний,

точний контент для кінцевого користувача. Цей ланцюг демонструє, що технологічний прогрес у галузі ШІ та автоматизації нерозривно пов'язаний із поглибленням відповідальності, прозорості та поваги до прав людини.

6.1.4. Переваги: швидкість, масштабованість, консистентність.

Впровадження технологій автоматизованої генерації мови (NLG) у редакційні процеси керується не модними трендами, а суворою економічною та операційною доцільністю. У прямому порівнянні з людською працею алгоритми демонструють неперевершену перевагу у трьох фундаментальних параметрах, які кардинально трансформують можливості медіа та контент-виробників: швидкості, масштабованості та консистентності. Ці переваги роблять шаблонну генерацію не просто зручним інструментом, а стратегічним активом для охоплення інформаційних потоків у цифрову епоху.

1. Швидкість (Speed): Ера нульової затримки в публікації. Людська праця зі створення контенту має природні фізіологічні та когнітивні обмеження. Типовий цикл роботи журналіста включає отримання звіту, його аналіз (5-10 хвилин), безпосереднє написання тексту (10-30 хвилин), редагування та вичитування (5-10 хвилин) і фінальну публікацію. Загальний час часто перевищує 20-30 хвилин на один матеріал. На противагу цьому, роботизована система NLG працює на межі миттєвості. Її операційний цикл радикально спрощений: отримання машиночитаних даних та їхнє автоматичне перетворення у готовий текст займає менше секунди (близько 0.1-0.3 секунди).

Бізнес-цінність такої швидкості є абсолютно вирішальною в галузях, де час еквівалентний грошам. У фінансовій журналістиці (Bloomberg, Reuters, AP) новина про квартальні звіти компаній або зміни ринкових індексів, опублікована навіть на одну секунду раніше за конкурентів, може принести клієнтам мільйони доларів прибутку або запобігти аналогічним збиткам. Трейдери та алгоритмічні торгові системи, що отримують інформацію першими, отримують критичну перевагу для прийняття рішень. Таким чином, швидкість NLG трансформується з технічного параметра у прямий джерело фінансової та конкурентної переваги.

2. Масштабованість (Scalability): Економіка "нульових граничних витрат" та охоплення довгого хвоста. Масштабованість визначає здатність системи ефективно обробляти експоненційне зростання

обсягів роботи без зниження якості. Ключова відмінність між людиною та алгоритмом лежить у структурі граничних витрат. Для журналіста витрати майже лінійно залежать від обсягу: створення однієї статті коштує певну кількість годин праці, а тисяча статей потребуватиме приблизно тисячі разів більше ресурсів. Система NLG вимагає значних початкових інвестицій у розробку та налаштування шаблонів, логічних гілок та інтеграцію з джерелами даних. Однак після запуску гранична вартість кожної наступної згенерованої статті стрімко наближається до нуля. Генерація тисячі, десяти тисяч чи мільйона текстів не потребує додаткових людських зусиль.

Ця економічна логіка відкриває доступ до ефективного висвітлення явищ "довгого хвоста" — масових, але економічно невигідних для людини подій. Редакції тепер можуть автоматично охоплювати кожен матч регіональних спортивних ліг, публікувати щоденні звіти про ціни на нерухомість у кожному мікрорайоні, генерувати персоналізовані фінансові огляди для десятків тисяч малих компаній або створювати мільйони унікальних описів товарів для e-commerce. NLG перетворює раніше "німі" масиви даних у доступний і монетизований контент.

3. Консистентність (Consistency) та гарантована точність: Ідеальний виконавець редакційних стандартів. Людський фактор неминуче вносить ризик помилок: втома, неуважність, суб'єктивність, технічні описки. Роботизовані системи, побудовані на детермінованих правилах, позбавлені цих недоліків. Вони забезпечують безпрецедентну консистентність — незмінність якості, стилю, термінології та форматування навіть при генерації мільйонів текстів. Якщо в шаблоні чітко прописано використовувати певне написання ("в Україні"), певний формат чисел або послідовність викладу, система не відступить від цього правила жодного разу.

Більше того, шаблонний NLG стає еталоном об'єктивної, нефальсифікованої журналістики фактів. Алгоритм не має власних упереджень, емоцій або політичних поглядів. Він не напише "На жаль, акції впали" або "На щастя, конкурент збанкрутував", якщо такі оціночні судження не були явно закладені в логіку. Він просто констатує факт: "Акції компанії X впали на Y%". Ця свобода від суб'єктивізму та гарантія відтворюваності роблять його незамінним інструментом у сферах, де точність та нейтральність є обов'язковими

вимогами: у фінансовій звітності, науковій комунікації, юридичних документах та базовій новинній журналістиці.

4. Висновок: Концепція доповненої журналістики (Augmented Journalism) — звільнення для творчості та аналізу. Ключове розуміння переваг NLG полягає в тому, що його головна мета — не заміна людини, а її доповнення та звільнення. Ця концепція, відома як "доповнена журналістика", переосмислює роль фахівця. Коли роботи беруть на себе рутинну, ресурсомістку роботу з обробки великих масивів структурованих даних (наприклад, генерацію 4000 квартальних фінансових звітів), вони звільняють найцінніший ресурс — час та інтелектуальну енергію журналістів. Фахівці отримують можливість зосередитися на завданнях, де людські якості незамінні: на глибокому аналізі, розслідувальній журналістиці, інтерв'ю, створенні авторських репортажів, осмисленні складних соціальних явищ та роботі з неструктурованими подіями, які не вкладаються в шаблони.

Таким чином, переваги шаблонної генерації — швидкість, масштабованість і консистентність — створюють синергетичний ефект. Вони не лише підвищують операційну ефективність і економічну життєздатність медіа, але й сприяють якісній трансформації професії, спрямовуючи людський креативний та аналітичний потенціал на вирішення найскладніших завдань у інформаційному суспільстві.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання допомагають зрозуміти, чому прості технології іноді кращі за складні:

1. Принцип "Mad Libs": Як працює метод заповнення пропусків (Slot Filling)? Чому дитяча гра, де треба вставити прикметник у речення, є основою алгоритмів *Associated Press* для фінансових звітів?

2. Детермінізм vs Ймовірність: Чому шаблонна система є детермінованою (на одні й ті ж дані вона завжди видасть однаковий текст), а ChatGPT — імовірнісним (кожного разу текст різний)? Чому для банківської сфери підходить тільки перше?

3. Вимоги до даних: Чому шаблонна генерація неможлива без структурованих даних? Що станеться з шаблоном "Команда А забила [Кількість] голів", якщо у комірці [Кількість] буде пусто або написа-

но "багато"?

4. Проблема репетитивності: Як розробники ботів борються з "роботизованим" звучанням текстів? Що таке "бібліотека синонімів" та варіативність шаблонів (наприклад, мати 5 варіантів вступу для перемоги і 5 для поразки)?

5. Економічна доцільність: У яких випадках налаштування шаблонної системи (яке займає місяці) є виправданим? (Підказка: Висока частота публікацій + низька варіативність подій, як-от землетруси чи курси валют).

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають логіці програмування текстів:

Завдання 1. Створення "Скелету новини" (Псевдокод)

• Ситуація: Ви налаштовуєте бота для сайту нерухомості.

• Вхідні дані: Price (ціна), District (район), Change (зміна ціни у % за місяць).

• Завдання: Напишіть логічне дерево (Branching):

- *Гілка 1 (Ціна впала > 5%):* "Гарна новина для покупців! У районі [District] ціни обвалилися на [Change]%."

- *Гілка 2 (Ціна зросла > 5%):* "Нерухомість дорожчає: у [District] зафіксовано стрибок цін на [Change]%."

- *Гілка 3 (Зміни < 5%):* "Стабільність на ринку: у [District] ціни майже не змінилися."

Завдання 2. Порівняння: Шаблон vs LLM

• Завдання: Візьміть сухий фінансовий звіт (наприклад, "Прибуток компанії X зріс на 10%").

1. Напишіть шаблонний текст (сухо, тільки факти).

2. Попросіть ChatGPT написати про це новину "в стилі поеми" або "з гумором".

• Аналіз: У чому ризик другого варіанту для інвестора? (Наприклад, ШІ може пожартувати про банкрутство, що неприпустимо).

Завдання 3. Аудит даних для шаблону

• Дано: Таблиця результатів шкільних олімпіад, де в колонці "Місце" написано: "1", "Перше", "І місце", "Переможець".

• Проблема: Алгоритм зламається, бо дані не уніфіковані.

• Завдання: Опишіть процес нормалізації даних (Data Cleaning), який має передувати генерації. (Всі значення привести до формату

цифр: 1, 2, 3).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Reiter, E., *& Dale, R. Building Natural Language Generation Systems.* (Класичний підручник, де детально описано архітектуру шаблонних систем).

2. Automated Insights (Wordsmith). *White papers and Case Studies*. (Приклади того, як Yahoo використовує шаблони для фентезі-футболу).

3. Arria NLG. *Difference between NLG and LLMs.* (Технічне порівняння підходів для бізнесу).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Шаблонна генерація (Template-based NLG): Метод створення тексту, що використовує заздалегідь написані фрагменти (шаблони) з "дірками" (слотами), у які підставляються дані з бази.

• Змінна (Variable): Елемент даних, який змінюється в кожному новому тексті (наприклад, [Рахунок], [Назва команди], [Температура]).

• Розгалуження (Branching): Логічна структура "Якщо... То...", яка дозволяє вибирати різні шаблони залежно від значень даних (наприклад, якщо температура < 0, вибрати шаблон про мороз).

• Детермінованість (Determinism): Властивість системи завжди видавати однаковий результат при однакових вхідних даних. Гарантує відсутність "галюцинацій".

• Hard-coding: Процес ручного прописування правил і текстів у коді програми (на відміну від машинного навчання, де система вчиться сама).

6.2. ШІ-асистована журналістика

Штучний інтелект стає потужним інструментом у руках сучасного журналіста, допомагаючи оптимізувати рутинні процеси та зосередитися на творчих і аналітичних аспектах роботи. Водночас важливо розуміти, що ШІ — це саме асистент, а не заміна журналістської майстерності, критичного мислення та етичної відповідальності. У цьому розділі ми детально розглянемо п'ять ключових напрямків використання штучного інтелекту в журналістиці, які вже сьогодні змінюють професійні практики медіаіндустрії.

6.2.1. Дослідження та пошук джерел.

Одним із найбільш трудомістких етапів журналістської роботи є збір інформації та пошук релевантних джерел. Традиційно журналісти витрачали години, а іноді й дні на пошук необхідних даних, перегортання архівів, вивчення документів та виявлення експертів у потрібній галузі. ШІ-асистенти можуть суттєво прискорити цей процес, хоча вони не замінюють ретельної перевірки фактів та журналістської інтуїції.

Можливості ШІ у дослідницькій роботі. Сучасні ШІ-системи здатні швидко обробляти величезні обсяги текстової інформації, знаходити приховані зв'язки між різними темами, виявляти закономірності та тренди, а також пропонувати потенційні джерела інформації. Це особливо цінно в епоху інформаційного перевантаження, коли обсяг доступних даних перевищує можливості людини їх опрацювати вручну.

Наприклад, при підготовці розслідувального матеріалу про вплив певної корпорації на екологію регіону, журналіст може попросити ШІ-асистента знайти всі публічно доступні звіти компанії за останні п'ять років, проаналізувати їхню динаміку, виявити розбіжності між офіційними заявами та реальними показниками, знайти схожі кейси в інших країнах. Система може запропонувати релевантні наукові дослідження, урядові документи, судові рішення, публікації екологічних організацій та інші матеріали, які варто розглянути.

ШІ також може допомогти у пошуку експертів. Замість того, щоб годинами шукати фахівця з вузької теми, журналіст може попросити систему знайти дослідників, які публікувалися з цього питання, проаналізувати їхні роботи та навіть підказати, хто з них найбільш

авторитетний у даній галузі та хто може дати збалансовану оцінку ситуації.

Робота з різнорідними джерелами. Особливу цінність ШІ має при роботі з різнорідними типами джерел. Сучасні системи можуть одночасно аналізувати текстові документи, статистичні дані, відкриті бази даних, соціальні мережі, архіви ЗМІ. Наприклад, при підготовці матеріалу про корупційну схему, ШІ може допомогти зіставити дані з реєстрів власності, декларацій чиновників, тендерних закупівель, судових рішень та публікацій у ЗМІ, виявивши підозрілі збіги та нестиковки.

ШІ-інструменти також ефективні у роботі з іноземними джерелами. Вони можуть не лише перекласти тексти, а й знайти релевантні публікації на різних мовах, що розширює горизонти журналістського дослідження. Це особливо важливо для міжнародної журналістики або при висвітленні глобальних тем.

Системи машинного навчання можуть виявляти закономірності у великих масивах даних, які людина може просто не помітити. Наприклад, аналізуючи тисячі урядових контрактів, ШІ може виявити, що певні компанії систематично отримують замовлення за завищеними цінами, або що існує патерн у призначеннях на певні посади. Це дає журналісту зацепки для подальшого розслідування.

Обмеження та ризики. Однак критично важливо пам'ятати про обмеження ШІ-асистентів у дослідницькій роботі. По-перше, ШІ може помилятися або надавати застарілу інформацію. Системи тренуються на певних даних до певного моменту часу, і їхні знання можуть бути неповними або неактуальними. По-друге, ШІ може не розуміти контекст так, як його розуміє людина. Система може запропонувати джерело, яке формально відповідає запиту, але насправді не є релевантним через нюанси, які ШІ не вловила.

По-третє, існує ризик упередженості даних. Якщо ШІ навчалася на текстах, що містять певні стереотипи або однобічні погляди, вона може відтворювати їх у своїх рекомендаціях. Наприклад, при пошуку експертів з певної теми система може непропорційно часто пропонувати чоловіків, якщо у навчальних даних переважали саме чоловічі голоси.

По-четверте, ШІ не завжди може оцінити достовірність джерела. Система може запропонувати публікацію з непевного сайту або

дослідження сумнівної якості, якщо вони формально відповідають критеріям пошуку. Професійний журналіст знає, як оцінити репутацію джерела, перевірити методологію дослідження, зважити потенційні конфлікти інтересів.

Професійні стандарти роботи з ШІ-дослідженням. Саме тому кожен факт, кожне джерело, знайдене за допомогою ШІ, потребує незалежної перевірки журналістом. Професійний підхід передбачає використання ШІ як інструменту попереднього пошуку, а не остаточного арбітра достовірності. ШІ-асистент — це відправна точка для дослідження, а не його завершення.

Журналіст має перевіряти знайдені ШІ дані за принципом триангуляції — знаходити підтвердження з незалежних джерел. Якщо ШІ знайшла цікаву статистику, варто переконатися, що вона походить з надійного джерела, актуальна та правильно інтерпретована. Якщо система запропонувала експерта, варто перевірити його кваліфікацію, попередні публікації, можливі конфлікти інтересів.

Важливо також документувати процес дослідження. Якщо ШІ допомогла знайти важливе джерело, журналіст має зафіксувати не лише саме джерело, а й як воно було знайдене, які ще джерела розглядалися, чому саме це джерело було обране. Це питання прозорості та професійної етики.

Корисною практикою є комбінування ШІ-пошуку з традиційними методами журналістського дослідження. Наприклад, використовувати ШІ для швидкого огляду теми та виявлення основних напрямків, а потім доповнювати це інтерв'ю з експертами, аналізом первинних документів, польовою роботою. ШІ допомагає швидше орієнтуватися в темі, але не замінює глибокого занурення.

Практичні сценарії використання. Розглянемо кілька конкретних сценаріїв, як журналісти можуть використовувати ШІ для дослідження:

Швидкий брифінг на незнайому тему. Коли журналіст отримує термінове завдання про тему, в якій він не є експертом, ШІ може за лічені хвилини надати огляд основних фактів, ключових гравців, актуальних дискусій. Це дозволяє швидко ввійти в контекст і підготувати адекватні питання для експертів.

Пошук трендів у даних. При роботі з великими масивами статистичних даних ШІ може допомогти виявити цікаві закономірності, які

стануть основою для журналістської історії. Наприклад, аналізуючи дані про закриття підприємств у регіоні, система може помітити, що це відбувається хвилями, збігається з певними політичними подіями або концентрується в окремих галузях.

Контекстуалізація подій. Коли відбувається певна подія, ШІ може швидко знайти схожі випадки з минулого, показати історичний контекст, пояснити складні взаємозв'язки. Це допомагає журналісту надати читачам не лише опис поточної ситуації, а й розуміння її місця в ширшій картині.

Моніторинг змін. ШІ може відстежувати зміни в документах, публікаціях, заявах офіційних осіб, що особливо важливо для розслідувальної журналістики. Наприклад, система може помітити, що компанія тихо змінила формулювання у своєму звіті про екологічний вплив, що може бути зацепкою для розслідування.

Отже, ШІ-асистенти стають потужним інструментом для журналістського дослідження, але лише в руках професіонала, який розуміє їхні можливості та обмеження, критично оцінює результати та дотримується високих стандартів перевірки фактів.

6.2.2. Транскрипція інтерв'ю.

Транскрибування аудіо- та відеозаписів інтерв'ю традиційно було однією з найбільш рутинних та трудомістких частин журналістської роботи. Досвідчені журналісти пам'ятають, як годинна розмова перетворювалася на три-чотири години роботи з перенесення кожного слова на папір або в текстовий файл. Сучасні ШІ-технології розпізнавання мови кардинально змінили цю ситуацію, перетворюючи годинні записи на текст за лічені хвилини.

Технологічний прогрес у розпізнаванні мови. Сучасні системи автоматичного розпізнавання мови (ASR, Automatic Speech Recognition) базуються на глибоких нейронних мережах, які навчаються на величезних масивах аудіоданих. За останні роки точність цих систем драматично зросла, наближаючись до рівня людської точності для якісних записів стандартною мовою.

Системи транскрипції не лише розпізнають слова, а й можуть виконувати складніші завдання. Вони ідентифікують різних співрозмовників (speaker diarization), що критично важливо для інтерв'ю з кількома учасниками. Система може автоматично розставляти розділові знаки на основі інтонації та пауз, форматувати

текст, розбиваючи його на абзаци. Деякі просунуті системи навіть позначають емоційні інтонації, сміх, паузи, що допомагає журналісту краще зрозуміти контекст розмови при подальшій роботі з текстом.

Особливо корисні функції автоматичного створення тайм-кодів — позначок часу для кожної фрази або абзацу. Це дозволяє журналісту швидко знайти потрібний момент в аудіозаписі, якщо потрібно уточнити формулювання або перевірити інтонацію. Деякі системи також створюють автоматичні саммарі інтерв'ю, виділяючи ключові теми та найважливіші цитати.

Практичні переваги для журналістів. Економія часу — найочевидніша, але далеко не єдина перевага автоматичної транскрипції. Журналіст може провести кілька інтерв'ю підряд і отримати всі транскрипти ввечері того ж дня, замість того, щоб витрачати наступний тиждень на ручне транскрибування. Це особливо важливо для термінових новинних матеріалів або коли потрібно швидко опублікувати реакцію експерта на поточну подію.

Можливість швидко переглянути текст інтерв'ю також покращує якість журналістської роботи. Маючи повний текст перед очима, журналіст може легше виявити найцікавіші цитати, побачити загальну структуру розмови, помітити теми, які варто розвинути у додаткових запитаннях. Часто найцінніші інсайти народжуються саме при повторному аналізі інтерв'ю, коли журналіст бачить зв'язки між різними частинами розмови.

Автоматична транскрипція також сприяє кращій організації матеріалів. Текстові файли легше індексувати, шукати по ключовим словам, порівнювати між собою. Журналіст, який працює над великим проєктом з десятками інтерв'ю, може швидко знайти всі згадки певної теми або порівняти, як різні експерти коментують одну й ту саму ситуацію.

Для подкастів та відеоконтенту транскрипти відкривають додаткові можливості. Вони роблять контент доступним для людей з вадами слуху, покращують SEO (пошукові системи індексують текст), дозволяють аудиторії швидко переглянути зміст перед прослуховуванням. Транскрипти також можна адаптувати в окремі текстові матеріали — статті, есеї, соціальні пости.

Технічні обмеження та виклики. При цьому варто чесно визнавати обмеження сучасних технологій транскрипції. Точність

розпізнавання суттєво залежить від якості аудіозапису. Фоновий шум, погана акустика приміщення, віддалений мікрофон, перекриття голосів — все це знижує якість транскрипції. Професійний журналіст має розуміти важливість якісного запису не лише для власного прослуховування, а й для ефективної роботи ШІ-систем.

Акценти та діалекти можуть створювати серйозні проблеми. Більшість систем транскрипції навчаються на стандартній літературній мові, і регіональні особливості вимови можуть призводити до помилок. Для української мови це особливо актуально через різноманіття діалектів та суржику в повсякденній мові. Інтерв'ю з людьми похилого віку, які використовують архаїчну лексику, або з представниками певних професій, які оперують специфічною термінологією, можуть транскрибуватися з численними помилками.

Технічна та спеціалізована термінологія — ще один виклик. Медичні терміни, юридичні поняття, назви хімічних сполук, іншомовні слова часто розпізнаються неправильно. Власні назви — імена людей, назви компаній, географічні локації — також можуть викликати труднощі, особливо якщо вони незвичні або записуються не так, як вимовляються.

Швидкість мови та чіткість артикуляції також впливають на якість. Люди, які говорять дуже швидко, "проковтують" закінчення слів або мають нечітку дикцію, створюють додаткові виклики для системи розпізнавання. Емоційно забарвлена мова — крики, шепіт, плач — теж може розпізнаватися гірше.

Критична важливість перевірки. Саме через ці обмеження професійний журналіст завжди звіряє автоматичну транскрипцію з оригінальним аудіозаписом, особливо для прямих цитат, які увійдуть у фінальний матеріал. Навіть одна неправильно розпізнана літера може кардинально змінити зміст фрази. Класичний приклад: "не може бути" vs "може бути" — одна пропущена частка повністю інвертує значення.

Для критично важливих цитат, які стануть основою матеріалу або можуть мати юридичні наслідки, варто не лише перевірити текст, а й кілька разів прослухати оригінал. Контекст, інтонація, паузи, емоційне забарвлення — все це може мати значення для правильного розуміння сказаного та етичного використання цитати.

Паузи та незавершені фрази в усному мовленні теж потребують

уваги. Автоматична система може просто пропустити паузу або невпевненість у голосі, які насправді були важливою частиною відповіді. Наприклад, коли політик затримується з відповіддю на незручне запитання або експерт довго шукає правильне формулювання — ця пауза може бути красномовнішою за саму відповідь.

Професійні стандарти використання. Журналістська етика вимагає дотримання певних стандартів при використанні автоматичних транскриптів. По-перше, будь-яка пряма цитата має бути перевірена на відповідність оригіналу. По-друге, при редагуванні усного мовлення для друкованої форми (прибирання повторів, "еканнь", незавершених фраз) треба зберігати оригінальний зміст та інтенцію висловлювання.

По-третє, якщо транскрипт містить помилку, яка потрапила до публікації, це залишається відповідальністю журналіста, а не системи транскрипції. Не можна виправдовувати фактичну помилку тим, що "ШІ неправильно розпізнала". Журналіст несе повну відповідальність за все, що публікується під його ім'ям.

По-четверте, варто бути чесним щодо процесу редагування цитат. Якщо усне мовлення значно відредаговане для читабельності, можна позначити це словами "у літературній обробці" або зберегти більш природне звучання, якщо контекст дозволяє.

Вибір інструментів транскрипції. На ринку існує безліч інструментів для автоматичної транскрипції — від вбудованих функцій у програмах для відеоконференцій до спеціалізованих сервісів для журналістів. При виборі варто враховувати кілька факторів:

Підтримка української мови. Не всі системи однаково добре працюють з українською. Деякі взагалі її не підтримують, інші дають погану точність через недостатній обсяг навчальних даних.

Конфіденційність даних. Інтерв'ю можуть містити чутливу інформацію. Важливо розуміти, чи зберігаються записи на серверах сервісу, хто має до них доступ, чи використовуються вони для навчання моделей. Для розслідувальних проєктів може бути критично важливим використовувати системи з гарантіями конфіденційності або навіть локальні рішення, які обробляють аудіо на комп'ютері журналіста без відправки в хмару.

Додаткові функції. Автоматичне розпізнавання співрозмовників,

тайм-коди, можливість редагування транскрипту разом з аудіо, інтеграція з іншими інструментами журналіста — все це може суттєво полегшити роботу.

Вартість. Деякі сервіси безкоштовні, але обмежені за обсягом або функціональністю. Професійні рішення можуть коштувати від кількох доларів на годину до місячної підписки. Редакція має оцінити, скільки інтерв'ю проводиться регулярно і чи виправдані ці витрати економією робочого часу.

Отже, автоматична транскрипція — це вже не футуристична технологія, а щоденний інструмент сучасного журналіста. Вона кардинально економить час, але вимагає професійного підходу до перевірки та використання результатів.

6.2.3. Допомога у написанні: розширення, резюме, рефрейминг.

ШІ-асистенти можуть стати корисними помічниками на етапі написання та редагування текстів, пропонуючи різні підходи до подання матеріалу та допомагаючи журналісту знайти оптимальну форму для своїх ідей. Важливо розуміти, що йдеться не про повну автоматизацію написання, а про асистування творчому процесу, надання інструментів для експериментування з формою та змістом.

Розширення тексту та ідей. Розширення (expansion) допомагає журналісту розгорнути стислу ідею або тезу в повноцінний абзац чи розділ. Це особливо корисно на етапі чернеток, коли журналіст фіксує ключові думки у вигляді коротких нотаток і потім потребує їх розгорнути в читабельний текст.

Наприклад, маючи стислу тезу "Зміна клімату впливає на сільське господарство України", журналіст може попросити ШІ запропонувати кілька варіантів її розкриття з різними акцентами. Система може запропонувати економічний кут зору (втрата врожаїв, зростання цін, вплив на експорт), соціальний аспект (міграція сільського населення, зміна традиційних практик господарювання), технологічний підхід (впровадження нових сортів, систем зрошення, агротехнологій) або екологічну перспективу (деградація ґрунтів, зміна біорізноманіття, ризики опустелювання).

Журналіст може обрати найбільш релевантний для свого матеріалу підхід, скомбінувати елементи різних варіантів або використати пропозиції ШІ як джерело натхнення для власного унікального кута зору. Ключовий момент: ШІ генерує варіанти, журналіст приймає

творче рішення.

Розширення також корисне для подолання письменницького блоку. Коли журналіст знає, що хоче сказати, але не може знайти правильні слова, ШІ може запропонувати кілька формулювань, які послужать відправною точкою. Навіть якщо жоден з запропонованих варіантів не підходить ідеально, вони можуть розблокувати власний творчий процес журналіста.

Важливо розуміти різницю між використанням ШІ для розширення власних ідей та копіюванням згенерованого тексту. Професійна етика вимагає, щоб фінальний текст був справжньою журналістською роботою, а не просто відредагованим варіантом згенерованого ШІ. Автор має переосмислити пропозиції системи, додати власні інсайти, факти, приклади, цитати експертів — все те, що робить журналістський матеріал цінним та унікальним.

Створення резюме та узагальнень. Резюме (summarization) особливо корисне при роботі з об'ємними документами, звітами, науковими публікаціями чи довгими інтерв'ю. ШІ може виділити ключові тези, головні факти та основні висновки, допомагаючи журналісту швидше орієнтуватися в матеріалі та виокремлювати найважливіше для свого матеріалу.

Існує два основні типи резюме: екстрактивне та абстрактивне. Екстрактивне резюме просто виділяє найважливіші речення з оригінального тексту та компонує їх разом. Абстрактивне резюме переформулює зміст своїми словами, що більше нагадує те, як працює людина при створенні саммарі. Сучасні ШІ-системи здатні створювати обидва типи, причому абстрактивні резюме стають дедалі якіснішими.

Для журналіста це означає можливість швидко опрацювати великі обсяги матеріалу. Наприклад, маючи 200-сторінковий урядовий звіт про економічну ситуацію, журналіст може попросити ШІ створити:

- Загальне резюме на одну сторінку з головними висновками
- Детальне резюме кожного розділу
- Виділення всіх конкретних цифр та статистичних даних
- Список рекомендацій та пропозицій
- Порівняння з попередніми звітами (якщо вони теж доступні)

Це дозволяє журналісту швидко зрозуміти структуру документа, виділити найцікавіші частини для детального вивчення, сформулю-

вати питання для експертів. Замість того, щоб читати весь звіт від початку до кінця, журналіст може стратегічно зосередитися на найрелевантніших частинах.

Водночас важливо пам'ятати, що автоматичне резюме може упустити нюанси або важливі деталі, які людина з професійним досвідом помітить одразу. ШІ може сконцентруватися на статистиці, пропустивши важливе застереження в примітці. Система може виділити формальні висновки, не звернувши уваги на приховану критику в одному з абзаців. Вона може пропустити контекст, який робить певний факт особливо значущим.

Тому резюме від ШІ варто розглядати як попередній огляд, а не як заміну власного читання. Для критично важливих частин документа, які стануть основою публікації, журналіст має обов'язково звернутися до оригінального тексту. Це питання не лише професіоналізму, а й етики — читач довіряє журналісту точну інтерпретацію джерел.

Рефрейминг та переосмислення. Рефрейминг (reframing) — це перефразування або переосмислення тексту з іншого кута зору, для іншої аудиторії або з іншим акцентом. Той самий факт або подію можна подати через призму різних наративів: людської історії, економічного аналізу, політичного контексту, історичної перспективи, технологічного виміру.

Наприклад, новина про закриття заводу може бути представлена як:

• Економічна історія про структурні зміни в промисловості

• Соціальна драма про долі сотень сімей, які залишилися без роботи

• Політична історія про провал регіональної влади захистити робочі місця

• Екологічна історія про вплив виробництва на довкілля та можливості для відновлення природи

• Урбаністична історія про трансформацію промислової зони міста

ШІ може запропонувати кожен із цих фреймів, допомагаючи журналісту побачити різні можливості для подання матеріалу. Це особливо цінно при роботі над складними темами, де не існує одного "правильного" способу розповісти історію.

Рефрейминг також корисний для адаптації контенту під різні

аудиторії. Складний технічний матеріал можна переформулювати для загальної аудиторії, використовуючи прості аналогії та уникаючи жаргону. І навпаки, загальну новину можна розширити технічними деталями для спеціалізованого видання. ШІ може допомогти знайти правильний баланс між доступністю та глибиною.

Перефразування також допомагає уникнути монотонності тексту. Якщо журналіст помічає, що використовує одні й ті самі звороти або починає кожен абзац однаково, ШІ може запропонувати альтернативні формулювання, зберігаючи зміст, але урізноманітнюючи форму.

Етичні та професійні стандарти. При використанні ШІ для асистування у написанні критично важливо дотримуватися професійних стандартів. По-перше, фінальний текст має бути результатом журналістської роботи, а не просто відредагованою версією згенерованого ШІ контенту. Журналіст додає факти, перевірені цитати, власні спостереження, аналітичні інсайти — все те, що робить матеріал журналістським продуктом.

По-друге, будь-які фактичні твердження у тексті мають бути перевірені, навіть якщо вони з'явилися в результаті роботи з ШІ. Система може згенерувати правдоподібно звучаще, але фактично неточне твердження. Відповідальність за точність завжди лежить на журналісті.

По-третє, важливо зберігати власний голос та стиль. ШІ може генерувати грамотний, але часто безликий текст. Журналістська майстерність проявляється саме в унікальному способі розповідати історії, в особливій чутливості до деталей, в умінні знайти несподіваний кут зору. Надмірне покладання на ШІ ризикує зробити всі тексти схожими один на одного.

По-четверте, деякі редакції та видання можуть мати специфічні політики щодо використання ШІ у створенні контенту. Журналіст має бути знайомим із цими політиками та дотримуватися їх. У деяких випадках може вимагатися розкриття факту використання ШІ-асистування.

Практичні поради щодо використання. Ефективне використання ШІ для допомоги у написанні вимагає певних навичок та підходів:

Чіткі інструкції. Чим точніше журналіст формулює запит до ШІ, тим корисніший буде результат. Замість загального "розширь цю думку" краще сказати "розширь цю думку з акцентом на соціальних

наслідках, додай приклад та збережи нейтральний тон".

Ітеративний підхід. Рідко коли перший варіант від ШІ буде ідеальним. Корисно працювати ітераціями: отримати перший варіант, оцінити його, дати додаткові інструкції для покращення, повторити процес кілька разів.

Критичне мислення. Кожна пропозиція ШІ має оцінюватися критично: чи відповідає вона фактам, чи підходить за тоном, чи не містить стереотипів або упереджень, чи зберігає етичні стандарти журналістики.

Комбінування підходів. Найкращі результати часто досягаються при комбінуванні різних можливостей ШІ. Наприклад, спочатку створити резюме документа, потім використати рефрейминг для знаходження цікавого кута, після чого розширити ключові тези власними доповненнями.

Збереження контролю. ШІ — це інструмент, який має залишатися під контролем журналіста. Фінальне рішення про кожне речення, кожне формулювання, кожен акцент приймає людина, а не машина.

Отже, ШІ-асистування у написанні відкриває нові можливості для журналістської творчості, але вимагає професійного та етичного підходу. Інструмент лише настільки хороший, наскільки добре його використовують.

6.2.4. Генерація альтернативних заголовків та лідів.

Заголовок та лід — найважливіші елементи будь-якого журналістського матеріалу, від яких залежить, чи зацікавиться читач публікацією, чи відкриє її в переповненій стрічці новин, чи дочитає до кінця. Створення ефективних заголовків та лідів — це мистецтво, яке поєднує точність, креативність, розуміння аудиторії та журналістську інтуїцію. ШІ-асистенти можуть генерувати десятки варіантів заголовків і лідів з різними акцентами та стилістиками, надаючи журналісту ширше поле для вибору та натхнення.

Різноманітність підходів до заголовків. Існує безліч типів заголовків, кожен із яких має свої переваги та контексти використання. ШІ може генерувати всі ці типи, допомагаючи журналісту експериментувати з різними варіантами:

Інформативні заголовки чітко повідомляють основний факт матеріалу. Наприклад, для статті про освітню реформу: "Міністерство освіти виділило 5 мільярдів гривень на оновлення шкільних

програм". Такі заголовки працюють для серйозних новин, де читач шукає конкретної інформації та очікує прямолінійності.

Заголовки-питання інтригують та залучають читача до роздумів: "Чи готові українські школи до цифрової трансформації?". Вони добре працюють для аналітичних матеріалів, де немає однозначної відповіді, та для тем, що викликають суспільні дискусії.

Заголовки з людською історією фокусуються на конкретному герої: "Як реформа змінила життя вчительки з Тернополя". Такий підхід добре працює для репортажів та feature-матеріалів, де важлива емоційна складова.

Цитатні заголовки використовують яскраву фразу з тексту: "'Це катастрофа для малих шкіл' — директори про освітню реформу". Вони привертають увагу та одразу дають голос головним героям матеріалу.

Цифрові заголовки акцентують на конкретних даних: "7 ключових змін в українській освіті у 2025 році". Особливо популярні в онлайн-медіа, де конкретика та структурованість привертають увагу.

Провокативні заголовки використовують несподівані формулювання або контрасти: "Реформа, якої всі боялися, виявилася порятунком для села". Ризиковані, але можуть бути дуже ефективними.

ШІ може згенерувати всі ці типи для одного матеріалу, дозволяючи журналісту та редактору порівняти різні підходи. Журналіст може побачити, як по-різному можна позиціонувати той самий матеріал, та обрати варіант, який найкраще відповідає редакційній політиці, цільовій аудиторії та контексту публікації.

SEO-оптимізація та онлайн-специфіка. У цифрову епоху заголовки виконують подвійну функцію: привертають увагу людей та оптимізовані для пошукових систем. ШІ може допомогти знайти баланс між цими двома вимогами, генеруючи заголовки, що містять ключові слова для пошуку, але залишаються природними та привабливими для читачів.

Наприклад, для матеріалу про зміни в податковому законодавстві ШІ може запропонувати:

• SEO-орієнтований: "Податки 2025: нові ставки ПДВ та зміни для ФОП в Україні"

• Більш креативний: "Що зміниться у вашому гаманці: гід по новому податковому кодексу"

• Збалансований: "Нові податки 2025: як зміни вплинуть на підприємців та найманих працівників"

Крім того, ШІ може генерувати різні версії заголовка для різних платформ. Заголовок для головної сторінки сайту може бути довшим та описовим, для Facebook — більш емоційним та особистим, для Twitter — стислим та таким, що миттєво привертає увагу, для Google Discover — оптимізованим під пошукові запити.

Створення лідів різних типів. Лід — це вступний абзац, який має утримати увагу читача, що вже клікнув на заголовок, та переконати його прочитати весь матеріал. ШІ може генерувати ліди різних типів відповідно до жанру та мети публікації:

Класичний новинний лід відповідає на ключові запитання: хто, що, де, коли, чому, як. "У середу, 15 січня, Міністерство освіти України оголосило про виділення 5 мільярдів гривень на оновлення шкільних програм у зв'язку з цифровою трансформацією освіти, яка має охопити всі регіони країни до кінця 2026 року."

Анекдотичний лід починає з конкретної історії або сцени: "Марія Петренко, вчителька математики зі Львова, пам'ятає, як три роки тому її учні скаржилися, що підручники застарілі та нецікаві. Сьогодні, завдяки новій реформі, вона викладає за допомогою інтерактивних програм, які роблять алгебру захопливою пригодою."

Описовий лід створює атмосферу: "Ранкова тиша сільської школи на Полтавщині зникає, коли діти заходять до класу, оснащеного новітніми планшетами та інтерактивною дошкою — результат освітньої реформи, яка докорінно змінює українську школу."

Лід-питання залучає читача до роздумів: "Чи може цифрова освіта замінити традиційного вчителя? Українські школи перевіряють цю гіпотезу на практиці, впроваджуючи найамбітнішу реформу за останні 30 років."

Лід з цитатою одразу дає голос головному герою: "'Я 25 років працюю вчителькою і вперше відчуваю, що держава справді інвестує в освіту,' — каже Оксана Іваненко, директорка сільської школи, яка отримала грант на технологічну модернізацію."

Статистичний лід починає з вражаючих цифр: "85% українських шкіл досі не мають достатньої кількості комп'ютерів для всіх учнів. Нова реформа обіцяє змінити цю статистику за два роки, виділивши 5 мільярдів гривень на цифрове оснащення."

ШІ може згенерувати всі ці варіанти, дозволяючи журналісту обрати той, що найкраще відповідає тону матеріалу, очікуванням аудиторії та редакційному стилю видання.

Тестування та A/B-експерименти. Одна з найцінніших можливостей, яку надає ШІ, — це швидке генерування великої кількості варіантів для тестування. Онлайн-медіа часто використовують A/B-тестування, коли різним сегментам аудиторії показують різні заголовки для того самого матеріалу, а потім аналізують, який варіант збирає більше кліків та генерує вищу читабельність.

ШІ може згенерувати десяток варіантів заголовків за секунди, що робить можливим систематичне тестування різних підходів. Наприклад, можна протестувати:

- Емоційні проти фактичних заголовків
- Короткі проти довгих формулювань
- Заголовки з цифрами проти описових
- Прямі проти інтригуючих формулювань

Дані з таких тестів допомагають редакціям краще розуміти свою аудиторію та поступово вдосконалювати стратегію заголовків. ШІ може навіть аналізувати результати попередніх тестів та на їх основі генерувати нові варіанти, які з більшою ймовірністю будуть успішними.

Небезпеки та обмеження. Однак використання ШІ для створення заголовків несе певні ризики, про які журналісти мають пам'ятати:

Кликбейт та сенсаціоналізм. ШІ, навчена на великих масивах онлайн-контенту, може мати схильність до генерування провокативних або сенсаційних заголовків, які максимізують кліки, але можуть спотворювати зміст матеріалу. Професійний редактор має критично оцінювати, чи не перетинає запропонований заголовок етичну межу між привабливістю та маніпуляцією.

Відсутність контексту. ШІ не завжди повністю розуміє нюанси матеріалу, культурний контекст або потенційні чутливі аспекти теми. Заголовок, що виглядає привабливо алгоритмічно, може бути неприйнятним через культурні, політичні або етичні причини.

Стереотипність. Системи, навчені на існуючому контенті, можуть відтворювати стереотипи та кліше в заголовках. Наприклад, матеріал про жінку-лідерку може отримати заголовок, що непотрібно акцентує на гендері, або матеріал про мігрантів може використовувати

проблемні формулювання.

Втрата унікальності. Якщо всі видання використовують ШІ для генерації заголовків, існує ризик, що заголовки стануть більш однорідними та передбачуваними. Журналістська креативність та унікальний голос видання можуть загубитися в морі алгоритмічно оптимізованих формулювань.

Професійна відповідальність та етика. Фінальне рішення щодо заголовка та ліду завжди має залишатися за журналістом чи редактором. ШІ генерує варіанти, людина приймає редакційні рішення на основі професійних стандартів та етичних принципів.

Етична відповідальність включає:

• Точність. Заголовок має точно відображати зміст матеріалу, не перебільшуючи та не спотворюючи факти.

• Чесність. Уникати маніпулятивних формулювань, які можуть ввести читача в оману.

• Повага. Особливо при висвітленні трагедій, конфліктів, чутливих тем — заголовок має поважати гідність людей, про яких іде мова.

• Відповідальність. Усвідомлювати вплив заголовка на формування громадської думки та уникати тих, що можуть поглибити поляризацію або підживлювати ворожнечу.

Деякі редакції встановлюють внутрішні правила використання ШІ для заголовків: наприклад, вимога обов'язкової людської перевірки перед публікацією, заборона на певні типи провокативних формулювань, або обмеження на використання ШІ для особливо чутливих тем.

Практичні поради. Для ефективного використання ШІ у створенні заголовків та лідів. Надавайте контекст. Чим більше інформації про матеріал, аудиторію, тон ви дасте ШІ, тим релевантніші будуть варіанти. Генеруйте багато варіантів. Не зупиняйтеся на першому варіанті. Попросіть 10-15 різних підходів, щоб мати ширше поле для вибору. Комбінуйте елементи. Часто найкращий заголовок виходить із комбінації елементів кількох згенерованих варіантів. Тестуйте на колегах. Перш ніж публікувати, покажіть кілька варіантів колегам та отримайте зворотний зв'язок. Аналізуйте результати. Відстежуйте, які типи заголовків працюють краще для вашої аудиторії, та використовуйте ці знання для вдосконалення. Зберігайте людський фінальний контроль. Завжди редагуйте та адаптуйте згенеровані

варіанти відповідно до редакційних стандартів.

Отже, ШІ може бути потужним помічником у створенні заголовків та лідів, розширюючи креативні можливості журналіста та економлячи час. Але фінальне рішення, що балансує між привабливістю, точністю та етикою, завжди має приймати професіонал.

6.2.5. Переклад та адаптація контенту.

Переклад та адаптація контенту. Глобалізація медійного простору вимагає від журналістів здатності працювати з контентом різними мовами. Важлива міжнародна новина може з'явитися китайською, важливе дослідження — німецькою, цікава людська історія — іспанською. Сучасні ШІ-системи машинного перекладу досягли значного прогресу, що відкриває нові можливості для журналістської роботи, хоча й не замінюють професійних перекладачів у всіх випадках.

Еволюція машинного перекладу. Машинний переклад пройшов довгий шлях від примітивних пословних перекладів 1950-х років до сучасних нейронних систем. Ранні системи працювали на основі словників та граматичних правил, часто видаючи комічно неправильні результати. Статистичні системи 1990-2000-х покращили ситуацію, але все ще мали проблеми з контекстом та ідіомами.

Революцію здійснили нейронні мережі, особливо архітектури на основі трансформерів, які з'явилися у 2017 році. Сучасні системи, такі як Google Translate, DeepL, а також спеціалізовані ШІ-асистенти, не просто замінюють слова одне на одне, а розуміють контекст цілих речень та абзаців. Вони враховують ідіоматичні вирази, культурні особливості, навіть тон та стиль оригіналу.

Для української мови це особливо важливо, адже довгий час ми були на периферії уваги розробників перекладацьких систем. Але останніми роками ситуація значно покращилася: з'явилися якісні моделі для перекладу з/на українську, що відкриває нові можливості для українських журналістів у роботі з міжнародними джерелами.

Журналістські сценарії використання. Для журналістів ШІ-переклад відкриває кілька критично важливих можливостей:

Моніторинг іноземних медіа. Журналіст може стежити за публікаціями в іноземних ЗМІ, швидко перекладаючи заголовки та ліди, щоб виявити цікаві історії для адаптації або теми, що набирають актуальності в інших країнах. Наприклад, український журналіст

може відстежувати європейські видання для виявлення трендів, що невдовзі можуть прийти і до України.

Робота з первинними джерелами. Замість покладатися на чужі переклади офіційних документів, заяв, звітів, журналіст може отримати доступ до оригіналу. Наприклад, ознайомитися з рішенням Європейського суду французькою мовою, з доповіддю ООН англійською, з заявою німецького уряду німецькою. Це знижує ризик неточностей при перекладі через посередників.

Інтерв'ю з іноземцями. Хоча для фінальної публікації краще використовувати професійного перекладача, ШІ може допомогти в оперативній роботі: швидко перекласти питання для інтерв'ю електронною поштою, зрозуміти загальний зміст відповіді іноземного експерта, підготувати додаткові питання.

Адаптація власних матеріалів. Українські журналісти, що працюють для міжнародних видань або хочуть донести важливі історії до світової аудиторії, можуть використовувати ШІ для створення базової версії перекладу своїх матеріалів, яку потім відредагує носій мови.

Fact-checking. При перевірці фактів часто потрібно звернутися до іноземних джерел. ШІ дозволяє швидко перевірити, що саме говорить оригінальне джерело, а не покладатися на чужі переказ и.

Адаптація контенту для різних аудиторій. Сучасні ШІ-системи можуть не лише перекладати текст дослівно, а й адаптувати його для іншої культурної аудиторії. Це важливо, бо хороший переклад — це не просто заміна слів однієї мови словами іншої, а передача смислів, емоцій, інтенцій з урахуванням культурного контексту.

Культурна локалізація. ШІ може автоматично пояснювати культурні реалії, незнайомі цільовій аудиторії. Наприклад, при перекладі українського матеріалу англійською може додати коротке пояснення, що таке вареники чи Івана Купала. Або навпаки, при перекладі американського матеріалу українською — пояснити, що таке Thanksgiving або NCAA.

Адаптація прикладів. Замість американських доларів використовувати гривні для української аудиторії, замість миль — кілометри, замість незнайомих українцям брендів — місцеві аналоги. ШІ може робити такі заміни автоматично або пропонувати варіанти журналісту.

Тональна адаптація. Різні культури мають різні очікування щодо формальності, гумору, емоційності тексту. Те, що звучить нормально американською, може здатися занадто фамільярним українською. ШІ може адаптувати тон відповідно до культурних норм цільової аудиторії.

Формати та конвенції. Дати (день/місяць/рік проти місяць/день/рік), імена (ім'я-прізвище проти прізвище-ім'я), адреси, номери телефонів — все це може потребувати адаптації, яку ШІ здатна зробити автоматично.

Обмеження та ризики машинного перекладу. Незважаючи на вражаючий прогрес, машинний переклад все ще має значні обмеження, особливо критичні для журналістської роботи:

Нюанси та контекст. ШІ може пропустити тонкі нюанси значення, іронію, натяки, культурні підтексти. Наприклад, саркастичний коментар може бути перекладений дослівно, втративши іронічний тон. Політичні евфемізми можуть бути перекладені буквально, втративши важливий підтекст.

Омоніми та багатозначність. Слова з кількома значеннями можуть бути перекладені неправильно, якщо система не зрозуміла контекст. Класичний приклад: англійське "right" може означати "правий", "правильний" або "право" — неправильний вибір кардинально змінює зміст.

Ідіоми та метафори. Образні вирази часто перекладяться буквально, втрачаючи зміст або звучачи комічно. "Elephant in the room" перекладене дослівно як "слон у кімнаті" не передає ідею очевидної проблеми, яку всі ігнорують.

Термінологія. Спеціалізована лексика — юридична, медична, технічна — може перекладатися неправильно. Це особливо небезпечно для журналістики, де точність термінів критична. Неправильний переклад назви хвороби, юридичного статусу чи технічної специфікації може мати серйозні наслідки.

Імена та назви. Власні назви — імена людей, географічні назви, назви організацій — можуть транскрибуватися або перекладатися різними способами, що створює плутанину. Особливо проблемні мови з нелатинською абеткою.

Культурно чутливі теми. При перекладі матеріалів про релігію, політику, гендерні питання, етнічні відносини ШІ може не вловити

культурної чутливості та використати формулювання, що можуть образити або бути неприйнятними в цільовій культурі.

Для журналістики перекладацькі помилки можуть мати критичні та далекосяжні наслідки, виходячи далеко за рамки простого недорозуміння:

1. Спотворення цитат та поширення дезінформації. Неправильно перекладена цитата політика, експерта чи свідка події може кардинально змінити її зміст, привести до створення хибних наративів, міжнародних скандалів або навіть судових позовів про наклеп. Журналіст несе професійну та юридичну відповідальність за точність переданої інформації, навіть якщо вона отримана через переклад. Втрата довіри аудиторії в такому разі є неминучою.

2. Дипломатичні та міжнародні інциденти. Некоректний переклад офіційних заяв, документів або правових норм може спричинити серйозні міждержавні непорозуміння, звинувачення у порушенні угод або ескалацію напруженості. Слово, перекладене з відтінком агресії або зневаги, де його не було, може мати політичні наслідки.

3. Загроза громадській безпеці та здоров'ю. У контексті кризових повідомлень (про стихійні лиха, пандемії, техногенні аварії) будь-яка помилка в перекладі інструкцій, рекомендацій або попереджень може безпосередньо загрожувати життю та здоров'ю людей. Неправильно перекладена назва хімічної речовини, дози ліків або маршрут евакуації може призвести до трагічних наслідків.

4. Підрив авторитету видання та дискредитація професії. Систематичні перекладацькі помилки руйнують репутацію медіа як надійного джерела. Аудиторія починає ставити під сумнів усю роботу редакції, що веде до втрати авторитету, аудиторії та, як наслідок, економічних втрат. Це також наносить шкоду загальному іміджу журналістики як професії, заснованої на правді та точності.

Висновок: Робота з перекладом у журналістиці — це не технічне завдання, а критично важлива складова редакційної етики та фактчекінгу. Вона вимагає не лише лінгвістичної компетентності, але й глибокого розуміння контексту, культурної чутливості та предметної експертизи. Використання сучасних інструментів, включаючи штучний інтелект, має супроводжуватися обов'язковим контролем з боку кваліфікованих перекладачів та редакторів, які можуть виявити та виправити семантичні, культурні та термінологічні пастки.

Точність перекладу — це питання професійної відповідальності перед суспільством.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання розкривають сутність гібридної роботи:

1. Концепція доповнення (Augmentation): У чому різниця між *автоматизацією* (Automation), яка замінює працю, та *доповненням* (Augmentation), яка розширює можливості людини? Чому *The New York Times* використовує ШІ для допомоги редакторам, а не для їх звільнення?

2. Рутинні задачі: Які три типи завдань найчастіше делегують ШІ у редакції: *Транскрибація* (розшифровка інтерв'ю), *Саммаризація* (створення вижимок з довгих звітів) та *Версіонування* (адаптація однієї статті для Instagram, Twitter та LinkedIn)?

3. Проблема "Білого аркуша": Як генеративні моделі (ChatGPT, Claude) допомагають подолати творчий блок на етапі брейнштормінгу? Чи етично використовувати ідеї заголовків, запропоновані машиною?

4. Синтез даних: Як інструменти на кшталт *Google Pinpoint* дозволяють журналістам-розслідувачам "спілкуватися" з гігантськими архівами документів (PDF, аудіо), знаходячи голку в стозі сіна за секунди?

5. Етичні гайдлайни: Яке золоте правило провідних медіа (AP, BBC, The Guardian) щодо використання генеративного контенту? (Підказка: Відповідальність завжди несе людина, а читач має бути поінформований, якщо роль ШІ була значною).

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання моделюють реальний робочий процес "Кіборг-журналіста":

Завдання 1. AI як субредактор (Editing Assistant)

• Вхідні дані: Візьміть свій старий текст (статтю або есе), бажано довгий і складний.

• Завдання: Використовуйте LLM (ChatGPT/Gemini) для виконання трьох редакційних завдань:

1. *"Скороти цей текст на 50%, зберігаючи ключові факти та цитати."*

2. *"Запропонуй 5 варіантів клікабельних (але не оманливих) заго-*

ловків."

3. *"Знайди в тексті логічні протиріччя або складні речення, які важко читати."*

• Аналіз: Оцініть якість роботи ШІ. Де він впорався краще за вас, а де втратив нюанси?

Завдання 2. Інтерв'ю з PDF (Аналітична робота)

• Матеріал: Завантажте довгий, нудний офіційний звіт (наприклад, звіт міської ради на 100 сторінок).

• Інструмент: ChatPDF або будь-яка LLM з функцією аналізу файлів.

• Завдання: "Допитайте" документ:

- *"Які три головні ризики згадуються у звіті?"*

- *"Знайди всі згадки про бюджетні витрати на освіту і склади таблицю."*

• Мета: Зрозуміти, як ШІ економить години читання "води".

Завдання 3. Фактчекінг галюцинацій (Verification)

• Ситуація: Ви попросили ШІ написати біографічну довідку про маловідомого українського політика або діяча культури.

• Завдання:

1. Згенеруйте текст.

2. Проведіть фактчекінг кожного речення вручну через Google.

3. Знайдіть хоча б одну "галюцинацію" (вигаданий факт, неправильну дату або неіснуючу посаду).

• Висновок: Чому правило "Trust but Verify" (Довіряй, але перевіряй) є критичним при роботі з LLM?

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА

1. Francesco Marconi. *Newsmakers: Artificial Intelligence and the Future of Journalism.* (Книга колишнього керівника R&D у WSJ, яка описує саме модель "асистування").

2. London School of Economics (LSE) JournalismAI. *Generating Change: A global survey of what news organizations are doing with AI.* (Найактуальніші звіти про стан індустрії).

3. AP Stylebook. *Artificial Intelligence guidelines.* (Стандарти використання ШІ в новинах).

4. Generative AI in the Newsroom. (Проєкт Nieman Lab з кейсами використання).

4. ГЛОСАРІЙ ТЕРМІНІВ

• Доповнена журналістика (Augmented Journalism): Підхід, при якому ШІ бере на себе рутинні, повторювані або обчислювально складні завдання (аналіз даних, транскрибація, тегування), звільняючи час журналіста для творчої роботи, інтерв'ю та етичних суджень.

• Транскрибація (Transcription): Автоматичне перетворення аудіозапису (інтерв'ю) в текст за допомогою ШІ (наприклад, Otter.ai, TurboScribe). Точність сучасних моделей досягає 95-99%.

• Саммаризація (Summarization): Здатність NLP-моделей створювати стислий виклад великого тексту, виділяючи головне.

• Промпт (Prompt): Текстовий запит або інструкція, яку журналіст дає генеративній моделі. Якість відповіді прямо залежить від якості промпту ("сміття на вході — сміття на виході").

• Галюцинація (Hallucination): Помилка генеративного ШІ, коли він впевнено подає вигадану інформацію як факт. Головний ризик для репутації журналіста, який використовує ШІ без перевірки.

6.3. Повністю автоматизоване виробництво

6.3.1. GPT-4, Claude та інші LLM для журналістики

Сучасний ландшафт автоматизації контенту розділений між двома принципово різними підходами. Якщо шаблонні системи Natural Language Generation (NLG) — це інженери-будівельники, які слідують точним кресленням-правилам, то Large Language Models (LLM), такі як GPT-4, Claude 3 чи Gemini, — це джазові музиканти світу тексту. Вони працюють не за детермінованою логікою, а на основі імпровізації та передбачення, що одночасно є їхньою велетенською силою і головною слабкістю у контексті серйозної журналістики.

1. Принцип роботи: передбачення наступного «токена» та парадокс «галюцинацій». В основі роботи будь-якої LLM лежить принцип статистичного передбачення. Модель не «знає» факти у людському розумінні, не звертається до бази знань. Натомість, аналізуючи колосальні обсяги тексту, зібраного з інтернету та інших джерел, вона навчається обчислювати, який фрагмент тексту (токен) з найбільшою ймовірністю має йти далі в заданій послідовності. Наприклад, побачивши фразу «Столиця України — …», модель з надзвичайно високою впевненістю виведе слово «Київ».

• Сила цього підходу: LLM володіє енциклопедічним охопленням тем і може генерувати текст, підтримувати діалог або створювати контент у практично будь-якій предметній області, демонструючи вражаючу вербальну гнучкість і креативність.

• Фундаментальна слабкість — «галюцинації»: Оскільки основна мета моделі — генерувати правдоподібний текст, а не пошук істини, вона схильна до фабрикації інформації. Коли модель стикається з запитом, на який у її тренувальних даних немає однозначної відповіді (наприклад, «Хто переміг на виборах у 2035 році?»), вона не каже «Я не знаю», а вигадує правдоподібну, але вигадану відповідь. Для журналістики, де точність факту є священною, це становить критичний ризик.

2. Огляд ключових інструментів: The Big Three. Сьогодні на ринку домінує трійка найпотужніших моделей, кожна з яких має власну спеціалізацію:

• GPT-4 (від OpenAI): Універсальний солдат. Ця модель є еталоном

потужності та гнучкості. Її сильні сторони — видатні логічні здібності, креативність і розуміння складних, багатоступеневих інструкцій. Вона ідеально підходить для брейнштормінгу, генерації десятків варіантів заголовків, креативного переписування (рерайту) текстів, написання есе чи сценаріїв. Основний недолік — обмеженість знань оновленням даних на певну дату (хоча платні версії отримали доступ до інтернету), а також іноді «лінива» відповідь на складні запити.

• Claude 3 (від Anthropic): Аналітик великих текстів. Модель, що спеціалізується на роботі з величезними обсягами тексту завдяки рекордному «контекстному вікну» — можливості одночасно «пам'ятати» і аналізувати тексти обсягом з цілу книгу. Вона пише стилем, який сприймається як більш «людський» і менш механічний. Claude менш схильний до надмірної цензури і часто згодний виконувати завдання, від яких відмовляється GPT. Це ідеальний інструмент для аналізу довгих PDF-звітів, суцільного читання стенограм інтерв'ю, написання комплексних лонгрідів на основі великої кількості джерел.

• Gemini (від Google): Дослідник з мультимедійним сприйняттям. Головною перевагою Gemini є глибока інтеграція з екосистемою Google, зокрема з пошуком у реальному часі. Функція перевірки фактів (Double-check) дозволяє моделі шукати підтвердження своїм твердженням в інтернеті. Крім того, Gemini демонструє найкращі серед конкурентів можливості в обробці та розумінні відео- та графічного контенту. Це робить її потужним інструментом для дослідницької журналістики, фактчекінгу та роботи з мультимедійними матеріалами.

3. Практичні сценарії використання в редакційній роботі. Великі мовні моделі не замінюють журналіста, але можуть стати потужним «силовим помножувачем» його продуктивності на різних етапах:

• Summarization (Стиснення та вижимка): LLM можуть миттєво аналізувати довгі документи (судові рішення, звіти, дослідження на 50+ сторінок) і формувати їх чіткі, лаконічні резюме у формі ключових пунктів або коротких абзаців, економлячи журналісту години читання.

• Repurposing (Перепакування контенту): Модель може швидко адаптувати одну базову статтю під різні формати та платформи: створити сценарій для пояснювального відео в TikTok, серію твітів

для Twitter (X), короткий текст для Instagram-сторіс або лист для розсилки.

• Ideation (Генерація ідей та заголовків): Коли настає творча криза, LLM може виступити як партнер для брейнштурму, запропонувавши десятки варіантів заголовків, концепцій для статей, ракурсів висвітлення події або питань для інтерв'ю, уникаючи шаблонних клікбейтних формулювань.

• Editing (Редактура та покращення тексту): Модель може виконувати роль помічника-редактора: знаходити та усувати тавтології, жаргон, робити текст більш лаконічним і ясним, перевіряти логічну послідовність або пропонувати альтернативні, більш яскраві формулювання.

Висновок: Великі мовні моделі відкривають нову еру доповненої журналістики (augmented journalism), знімаючи з журналіста рутинне навантаження та пришвидшуючи дописувальні процеси. Однак їхня сутність як статистичних генераторів тексту робить людський контроль, фактчекінг та редактуру обов'язковими заключними ланками у будь-якому робочому процесі. Успішне впровадження LLM у редакціях вимагає розуміння їхніх можливостей як потужного, але недосконалого інструменту, який працює під неухильним наглядом професійного журналіста-людини.

6.3.2. Генерація новин на основі даних у реальному часі.

Сучасна журналістика в умовах стрімкого інформаційного потоку все частіше будується на принципах Event-Driven архітектури, або журналістики, керованої подіями. Цей підхід передбачає, що автоматизована система знаходиться в режимі очікування до отримання чіткого сигналу-тригера від зовнішнього джерела структурованих даних. У момент надходження такого сигналу система миттєво активується, генеруючи та публікуючи контент. Це дозволило перетворити новину зі статичного, разового продукту на динамічну, «живу» сутність, що еволюціонує в реальному часі. Автоматизація дозволяє розвивати одну новинну тему через кілька послідовних стадій, які формують живий цикл статті. Перша стадія — миттєве сповіщення (Flash) — відбувається протягом 0-5 секунд після фіксації події. На цьому етапі система публікує один-два фактичні речення, мета яких — оперативно зайняти інформаційний простір, надіслати екстрене сповіщення аудиторії та заявити про

свою присутність як першоджерела. Протягом наступних 1-5 хвилин настає стадія оновлення (Update), коли система автоматично доповнює первинну зачіпку, підтягуючи додаткові структуровані дані, такі як історія подібних подій або географічний контекст, а генеративні моделі LLM можуть додати корисні пояснення чи поради. Нарешті, через 15-30 хвилин після події запускається стадія аналізу (Analysis), коли до процесу підключається людина-редактор, щоб додати авторський аналіз, коментарі експертів, свідчення очевидців та візуальний контент, перетворюючи автоматично створений каркас на комплексний матеріал.

Однак довіряти генерацію термінових, фактологічних новин чистій LLM не можна через ризик фабрикації інформації та відносну повільність у точній обробці цифр. Тому в промислових рішеннях використовується гібридна «сендвіч-модель», яка поєднує сили різних технологій. Її основу становить жорсткий, rule-based шаблон (скелет), запрограмований на безпомилкове заповнення ключових фактів: хто, де, коли і скільки. Цей скелет гарантує стовідсоткову точність цифр, імен та дат. Навколо нього генеративний штучний інтелект (LLM) будує «м'язи» — природномовні заголовок, лід та контекстуальні пояснення, додаючи потрібне стилістичне забарвлення. Для найкритичніших тем (повідомлення про військові дії, катастрофи) система не працює повністю автономно; вона включає контур людського контролю (Human-in-the-Loop), коли матеріал потрапляє в чернетку та чекає на фінальну перевірку та санкцію відповідального редактора перед публікацією.

Еталонним прикладом такої практики є Quakebot від Los Angeles Times. Його алгоритм працює так: отримавши автоматичний сигнал з структурованими даними від Геологічної служби США, скрипт парсить інформацію, вставляє ключові параметри в шаблон новини та створює чернетку в редакційній системі, після чого надсилає редактору сповіщення з проханням підтвердити публікацію. Цей процес дозволяє LA Times публікувати новини про землетруси всього за 2-3 хвилини після поштовхів, часто стаючи першим ЗМІ у світі, що повідомляє про подію. В Україні найяскравішим прикладом журналістики, керованої подіями, є автоматизовані системи сповіщень про повітряні тривоги, такі як «Мапа тривог». Ці системи, отримуючи офіційний сигнал від ДСНС, миттєво перетворюють

його на текстове попередження та візуальний маркер на карті, забезпечуючи миттєве, одночасне оповіщення мільйонів людей. Ця практика демонструє, як Event-Driven Journalism перетворюється з інструменту швидкої новини на критичну суспільну службу, де швидкість та точність безпосередньо пов'язані з безпекою життя людей. Таким чином, журналістика, керована подіями, представляє собою симбіоз надійної автоматизації, креативного потенціалу генеративних моделей та обов'язкового редакторського контролю, перетворюючи роль журналіста з первинного репортера фактів на куратора, аналітика та контролера високошвидкісних автоматичних систем.

6.3.3. Автоматизовані огляди та дайджести.

Здатність штучного інтелекту стискати та переформовувати великі обсяги інформації (Summarization) кардинально змінює ландшафт споживання новин і контенту. Ця технологія пропонує вихід з інформаційної перевантаженості, дозволяючи користувачеві замість читання десятка повних статей ознайомитися з одним чітким, згенерованим абзацем-висновком, який виділяє суть. В основі цієї трансформації лежать два принципово різні підходи до саммаризації, кожен з яких має свої переваги та обмеження.

Екстрактивна саммаризація (Extractive Summarization) працює за принципом «маркера». Алгоритм аналізує вихідний текст, оцінює важливість кожного речення (часто на основі частоти ключових слів, позиції в тексті або зв’язків з іншими реченнями) і просто копіює найбільш значущі фрагменти в підсумок, як правило, зберігаючи їхню оригінальну формулювання. Основним перевагою цього методу є гарантована точність і збереження авторського слова, оскільки система нічого не вигадує. Проте його недоліком часто стає рваний, незв’язний результат: підсумок може складатися з речень, взятих з різних частин тексту, без плавних логічних переходів, що ускладнює читання і розуміння загальної картини.

На противагу цьому, абстрактивна саммаризація (Abstractive Summarization) працює як досвідчений «рерайтер». Замість механічного копіювання, алгоритми на основі великих мовних моделей (LLM), такі як GPT-4 чи Claude, спочатку інтерпретують та усвідомлюють зміст і основну думку вихідного тексту, а потім переказують його новими, власними словами, створюючи абсолютно

новий текст. Це дозволяє отримувати плавні, зв'язні та стилістично вивірені підсумки, які читаються так, ніби їх написала людина. Однак цей метод несе в собі ключовий ризик «галюцинацій»: модель, прагнучи створити когерентний текст, може ненавмисно додати факт, висновок або відтінок, якого не було в оригінальному джерелі, що робить її непідходящою для саммаризації без наступної перевірки, коли критично важлива буквальна точність.

Автоматична саммаризація знаходить своє практичне втілення в кількох ключових форматах, які вже стали стандартом для сучасних медіа. Одним із найпопулярніших є формат TL;DR (Too Long; Didn't Read) — кнопка або блок на початку довгого матеріалу, який за запитом генерує стислу вижимку з 3-5 ключовими тезами (Key Takeaways). Цей підхід, популяризований діловими виданнями на кшталт Axios та Forbes, поважає час читача, дозволяючи йому швидко оцінити релевантність статті. Іншим перспективним напрямком є персоналізовані дайджести. Замість єдиної для всіх підбірки «Топ-5 новин тижня», система аналізує інтереси конкретного користувача (на основі історії читання) та автоматично генерує унікальну розсилку на кшталт: «Іване, поки ти спав, у світі технологій (твоя улюблена тема) сталося це…». Крім того, ШІ використовується для агрегації та зведення думок: прочитуючи десятки джерел про одну й ту саму подію (наприклад, саміт G7 чи парламентські дебати), модель може синтезувати єдиний огляд, що відображає спектр позицій та основних напрямків обговорення у різних ЗМІ.

Філософією, що лягла в основу багатьох сучасних автоматизованих дайджестів, є «Розумна стисліть» (Smart Brevity), концепція, яку успішно впровадило видання Axios. Її суть полягає в максимальній повазі до часу аудиторії та видаленні всього зайвого. Ідеальний ШІ-дайджест у цьому стилі будується за чіткою тричастинною структурою: 1) Що сталося? (одне речення з фактом); 2) Чому це важливо? (контекст і значення події); 3) Що буде далі? (прогноз наслідків або наступних кроків). Великі мовні моделі виявляються ідеальним інструментом для автоматичного переформатування будь-якого довгого тексту — від звіту до промови — у цей чіткий, інформаційно насичений і дієвий формат. Таким чином, автоматизована саммаризація не лише полегшує навігацію в інформаційному потоці, але й сприяє новому стандарту ясності та ефективності в комунікації.

6.3.4. Multimodal generation: текст + зображення + відео.

До недавнього часу сфера генеративного штучного інтелекту була фрагментованою: окремі моделі спеціалізувалися виключно на тексті (LLM), зображеннях (GAN, Diffusion) або аудіо. Проривом останніх років стало поширення мультимодальних моделей — систем, здатних сприймати, розуміти та генерувати дані різних типів (текст, зображення, відео, звук) в єдиному контексті. Ця технологія працює як універсальний перекладач, що руйнує бар'єри між форматами, відкриваючи нові можливості та виклики для контент-виробництва.

Злам бар'єрів: універсальність мультимодального ШІ. Мультимодальність реалізується через кілька ключових напрямів трансформації даних. Генерація зображень з тексту (Text-to-Image), започаткована такими інструментами, як DALL-E 3 (інтегрований у ChatGPT), Midjourney та Stable Diffusion, дозволила створювати візуальний контент на основі текстових описів. У медіа це знаходить застосування для ілюстрації складних абстрактних концепцій, які неможливо або дорого зняти на камеру: «економічна криза», «майбутнє штучного інтелекту», «емоційне вигорання». Ці інструменти також ефективні для швидкого створення інфографіки, карикатур або стилізованих ілюстрацій. Однак їхнє використання в журналістиці супроводжується суворим етичним табу: генерація фотореалістичних зображень реальних подій, місць або людей для новинних матеріалів розглядається як створення фейків і підриває довіру.

Наступним етапом стала генерація відео з тексту або зображень (Text-to-Video / Image-to-Video), представлена такими потужними системами, як Sora від OpenAI, Runway (Gen-2/Gen-3) та Pika Labs. Ця технологія відкриває революційні можливості для створення коротких відеофрагментів (B-roll) для доповнення сюжетів, анімації статичних графіків, створення експлейнерів для соціальних мереж (TikTok, Reels). Паралельно з цим зросла здатність ШІ аналізувати зображення та перетворювати їх на текст (Vision), реалізована в GPT- 4 Vision (GPT-4o) та Google Gemini. Ця функція, яку можна назвати «машинним зором», служить не лише для доступності (автоматична генерація текстових описів Alt-text для незрячих користувачів), але й стає потужним аналітичним інструментом. Журналіст може, наприклад, завантажити супутниковий знімок зони

конфлікту і отримати від ШІ кількісну оцінку зруйнованих будівель або виявлення техніки.

Концепція «синтетичного медіа-пакету»: нова парадигма виробництва.

Мультимодальність кардинально змінює організацію роботи. Раніше створення багатокомпонентного матеріалу вимагало координації цілої команди фахівців: автор тексту, дизайнер, відеооператор, монтажер. Сьогодні один фахівець (промпт-інженер), користуючись наборами мультимодальних інструментів, може створити чернетку повного медіа-пакету з єдиного запиту. Наприклад, за промптом «Створи матеріал про технологічні виклики колонізації Марса» система може послідовно згенерувати: детальний текстовий лонгрід (за допомогою LLM), низку стилізованих ілюстрацій марсіанської бази (за допомогою Text-to-Image моделі) та коротке відео симуляції посадки космічного корабля (за допомогою Text-to-Video інструмента). Це не усуває необхідність людської доробки та перевірки, але значно прискорює та демократизує процес прототипування контенту.

Проблема галюцинацій у візуальному контенті та застереження. Якщо текстові LLM схильні до фабрикації фактів, то мультимодальні моделі, особливо візуальні, схильні до створення абсурдних або технічно неправильних образів. Класичним прикладом є генерація зображень людських рук з шістьма пальцями, неправильною анатомією суглобів або невідповідністю перспективи. У відео ці проблеми посилюються: об'єкти можуть нелогічно змінювати форму, «плавитися», миттєво з'являтися чи зникати, порушуючи закони фізики та причинно-наслідкові зв'язки. Використання такого контенту без ретельної перевірки та редагування не лише викриває непрофесійність, але й може завдати серйозної шкоди репутації медіа, яке опублікувало явно штучний або технічно дефектний матеріал. Таким чином, мультимодальна генерація, попри свою революційність, не усуває, а навіть посилює потребу в критичному людському контролі, фактчекінгу та художньому редагуванні, щоб кінцевий продукт відповідав стандартам якості та достовірності.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю)

Ці питання перевіряють розуміння індустрії синтетичного

контенту:

1. Синтетичні медіа (Synthetic Media): Що це таке? Чому відео, створене ШІ (наприклад, звернення президента, якого не було), вважається синтетичним медіа, а звичайний CGI у фільмах Marvel — ні? (Ключ: високий ступінь автономності алгоритму при створенні).

2. Текст-у-Відео (Text-to-Video): Як технології типу *Sora, HeyGen* чи *Synthesia* змінюють економіку телебачення? Якщо створення новинного сюжету з "живим" диктором коштує $0 (тільки підписка на сервіс) і займає 2 хвилини, що буде з професією телеведучого?

3. Контент-ферми (Content Farms) та "MFA": Що таке сайти *Made-For-Advertising*? Як повна автоматизація призвела до засмічення інтернету тисячами статей низької якості ("AI Slop"), мета яких — просто показати рекламу випадковому відвідувачу?

4. Персоналізація відео на масштабі: Як бренди використовують автоматизацію для створення *індивідуальних* відеозвернень? (Кейс Cadbury, де ШІ генерував відео з індійською кінозіркою, яка називала ім'я конкретного магазину — зроблено тисячі версій одного відео).

5. Ефект "зловісної долини" (Uncanny Valley): Чому ідеально згенерований цифровий аватар часто викликає у глядача відразу або страх? Як цей психологічний бар'єр впливає на сприйняття повністю автоматизованих новин?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання дозволяють створити власний автоматизований пайплайн:

Завдання 1. Створення цифрового ведучого

• Інструмент: Synthesia, HeyGen або D-ID (тріал-версії).

• Завдання: Створіть 30-секундний новинний ролик.

1. Напишіть текст новини (можна згенерувати в ChatGPT).

2. Оберіть аватара (ведучого) та голос.

3. Згенеруйте відео.

• Аналіз: Оцініть результат. Чи помітно, що це робот? (Зверніть увагу на кліпання очей, дихання, інтонацію). Чи довіряли б ви такому ведучому?

Завдання 2. Детекція AI-контенту

• Ситуація: В інтернеті з'явилося фото/відео резонансної події (наприклад, арешт політика), яке виглядає підозріло.

• Завдання: Проведіть візуальний аналіз на наявність артефактів генерації:

- *Руки/Пальці:* (Чи правильна кількість? Чи природні фаланги?).

- *Текст на фоні:* (Чи читаються написи на вивісках, чи це "іншопланетна абетка"?).

- *Тіні та світло:* (Чи відповідають тіні джерелам світла?).

• Мета: Розвинути навичку візуальної верифікації (Media Forensics).

Завдання 3. Проєктування "Конвеєра" (Automated Pipeline)

• Легенда: Ви запускаєте YouTube-канал з прогнозом погоди для 100 міст України. Робити це вручну неможливо.

• Завдання: Намалюйте схему автоматизації.

- *Джерело даних:* (API погодного сайту).

- *Обробка:* (Скрипт перетворює "+20, Rain" у текст "У Києві сьогодні дощ...").

- *Візуалізація:* (Генерація слайдів з мапою).

- *Озвучка:* (TTS - Text-to-Speech сервіс).

- *Публікація:* (YouTube API).

• Питання: Де в цьому ланцюжку потрібна людина? (Тільки на етапі налаштування та моніторингу збоїв).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Nina Schick. *Deepfakes: The Coming Infocalypse.* (Книга про те, як синтетичні медіа змінять наше сприйняття реальності).

2. NewsGuard Reports. *Tracking AI-enabled Misinformation.* (Звіти про зростання кількості сайтів, повністю згенерованих ботами).

3. Synthesia Blog. *Case studies on AI avatars in corporate training.* (Позитивні приклади використання технології).

4. Samsung NEON project. (Приклад створення "штучних людей").

4. ГЛОСАРІЙ ТЕРМІНІВ.

• Синтетичні медіа (Synthetic Media): Будь-який медіаконтент (відео, аудіо, зображення, текст), який був повністю або частково згенерований за допомогою алгоритмів штучного інтелекту. Часто використовується як нейтральний синонім до слова "діпфейк".

• Цифровий аватар (Digital Avatar): Візуальне представлення людини (реальної або вигаданої) у цифровому середовищі, яке може

"промовляти" будь-який текст завдяки технологіям ліпсінку (синхронізації губ) та синтезу мовлення.

• TTS (Text-to-Speech): Технологія перетворення тексту в аудіо. Сучасні нейронні TTS (наприклад, ElevenLabs) здатні клонувати голос конкретної людини з емоційним забарвленням.

• MFA (Made-For-Advertising): Сайти-смітники, які автоматично генерують тисячі статей за допомогою ШІ на популярні теми, щоб залучати пошуковий трафік і заробляти на банерній рекламі. Якість контенту там зазвичай нульова.

• Зловісна долина (Uncanny Valley): Гіпотеза в робототехніці та комп'ютерній графіці: коли людиноподібний об'єкт виглядає майже як людина, але має дрібні дефекти (скляний погляд, неприродна міміка), він викликає у спостерігача почуття сильної відрази та страху.

6.4. Якість та достовірність

6.4.1. Проблема галюцинацій та фактичних помилок.

Галюцинація (Hallucination) у сфері штучного інтелекту — це критична слабкість, яка становить особливу небезпеку саме через свою впевненість. Вона визначається як генерація моделлю тексту, який є граматично правильним і логічно зв'язним, але фактично помилковим або таким, що не базується на наданих вхідних даних. Модель подає цю вигадану інформацію з тоном абсолютного експерта, ніколи не застосовуючи обережних формулювань на кшталт «можливо» або «я не впевнений». Ця ілюзія знання є фундаментальною проблемою для використання великих мовних моделей (LLM) в журналістиці, де точність факту є основою довіри.

Механізм виникнення: Чому статистична модель «бреше»?

Щоб зрозуміти природу галюцинацій, необхідно чітко усвідомити, що LLM — це не база знань (як енциклопедія), а скоріше високорівнева ймовірнісна машина. Її основна функція — передбачення наступного найбільш вірогідного слова (токену) у послідовності. Модель не «знає», що Київ — столиця України в семантичному розумінні. Вона лише обчислила на основі трильйонів прикладів з тренувальних даних, що після фрази «Столиця України —» слово «Київ» має надзвичайно високу ймовірність появи.

Галюцинації виникають саме з цієї ймовірнісної природи:

• Проблема «довгого хвоста»: Коли запит стосується рідкісної, специфічної або нової події, про яку в тренувальних даних мало інформації, модель опиняється на «хвості» розподілу ймовірностей. Їй недостатньо статистичних паттернів для точної відповіді. Однак, будучи налаштованою генерувати когерентний текст, вона починає «імпровізувати», обираючи комбінації слів, які звучать правдоподібно в заданому контексті, але не відповідають дійсності.

• Каскадне посилення помилки: LLM генерують текст послідовно, слово за словом, причому кожне наступне слово базується на попередніх, згенерованих нею ж. Якщо на початку відповіді виникає незначна неточність, модель будує на її основі подальший текст, що може призвести до каскаду галюцинацій і повністю вигаданого висновку.

Типи галюцинацій

Таблиця 6.1.

Класифікація галюцинацій за типом порушення

Тип галюцинації	Суть та приклад	Наслідки в журналістиці
Фактична фабрикація	Вигадування подій, цифр, тверджень, що не мають підтвердження в реальності. *Приклад: «Тарас Шевченко написав роман «Хіба ревуть воли...» (автор роману — Панас Мирний).*	Пряме поширення дезінформації, спотворення історичних або наукових фактів.
Невірна атрибуція	Правильна цитата або факт приписується неправильному джерелу чи особі. *Приклад: Модель наводить реальну цитату про цифровізацію, але приписує її міністру, який її ніколи не говорив.*	Підриває довіру до особи, спотворює контекст висловлювання, може призвести до скандалу.
Вигадані посилання та джерела	Модель створює описи неіснуючих наукових статей, судових рішень, книг або веб-сайтів. *Приклад: Відомий випадок, коли адвокат у США представив суду вигадані ChatGPT прецеденти (Varghese v. China Southern Airlines), що призвело до серйозних професійних санкцій.*	Повна втрата наукової та юридичної ваги матеріалу, дискредитація видання, правові ризики.

Складено автором. Джерело. "LLM Hallucinations Taxonomy: Types and Mitigation" // Towards Data Science (2025). https://towardsdatascience.com/llm-hallucinations-taxonomy-2025

Методи боротьби: від «заземлення» до людського контролю.

Боротьба з галюцинаціями — це комплексний процес, що поєднує технологічні рішення та редакційні процедури. Ключова стратегія — перетворити модель з «творчого письменника» на «уважного бібліотекаря».

Заземлення (Grounding) та RAG (Retrieval-Augmented Generation): Це найефективніший технічний підхід. Заземлення — це процес жорсткої прив'язки відповіді моделі до конкретних, перевірених джерел. Його практичною реалізацією є RAG. Замість того, щоб покладатися на внутрішні (часто неточні) знання моделі, система спочатку здійснює пошук у спеціально підготовленій базі надійних документів (закони, офіційні звіти, перевірені статті). Потім LLM отримує інструкцію: «Використовуючи виключно надані нижче

документи, дай відповідь. Якщо відповіді немає в документах, скажи «Не знаю»». Це радикально знижує рівень галюцинацій.

Партнерства з медіа для якісних даних: Провідні компанії, як-от OpenAI, укладають угоди з авторитетними новинними виданнями (The Financial Times, News Corp, Axel Springer), отримуючи доступ до їхніх архівів для навчання та покращення точності моделей. Це дозволяє «заземлити» ШІ на професійний журналістський контент.

Людський фактчекінг як неминуча ланка: Жодна технологія не може повністю усунути необхідність людської перевірки. Стандартні процедури фактчекінгу залишаються обов'язковими:

- Перевірка джерел: Кожне цитоване джерело, навіть надане ШІ, має бути знайдене та перевірене журналістом незалежно. Необхідно переконатися, що стаття чи документ існують, належать заявленому автору та містять саме ту інформацію, яку наводить модель.

- Верифікація цитат: Цитати, особливо від публічних осіб, слід шукати за допомогою пошукових систем у лапках та перевіряти в авторитетних ЗМІ, що безпосередньо висвітлювали подію.

- Відстеження змін у зображеннях та відео: Для візуального контенту обов'язковим є зворотний пошук за зображенням (наприклад, через Google Images або TinEye), щоб встановити оригінальне джерело, дату та контекст публікації, виявивши можливі маніпуляції чи неправильні підписи.

Практичний дороговід для редакції. Інтеграція LLM в редакційний процес вимагає чітких правил безпеки.

Висновок: Проблема галюцинацій не є технічним багом, який можна виправити, — це наслідок самої природи великих мовних моделей як статистичних генераторів тексту. Тому довіра до ШІ в журналістиці не може бути безумовною. Майбутнє полягає не в повній автоматизації, а в синергії: використанні технологій на кшталт RAG для підвищення точності, поєднаному з непохитною дисципліною фактчекінгу та критичним мисленням журналіста. Текст, згенерований ШІ, повинен проходити тиї ж, якщо не більш суворі, перевірки, що й текст, написаний людиною. У цьому контексті роль журналіста трансформується з первинного генератора слів у куратора, верифікатора та гаранта достовірності, що є найважливішою функцією в епоху генеративного штучного інтелекту.

Таблиця 6.2.

Стратегії боротьби з галюцинаціями на різних етапах роботи

Етап роботи з LLM	Ключові дії для запобігання галюцинаціям	Мета
Постановка завдання (Промпт)	Формулювати запити максимально конкретно. Використовувати інструкції: «Базуй відповідь тільки на наданих документах», «Якщо інформації недостатньо, так і скажи».	Мінімізувати простір для імпровізації моделі.
Технологічна підтримка	Впроваджувати архітектуру RAG, налаштовувати пошук в закритих, перевірених базах даних редакції (архіви, звіти, закони).	«Заземлити» модель на достовірні факти.
Фактчекінг та перевірка	Кожен факт, ім'я, дата, цитата та джерело, наведені ШІ, підлягають незалежній перевірці класичними журналістськими методами. Особливу увагу — на незвичайні або сенсаційні твердження.	Виявити та усунути галюцинації до публікації.
Людський контроль (Human-in-the-Loop)	Розглядати будь-який текст, згенерований ШІ, лише як чорновий матеріал або перший начерк. Фінальне рішення про публікацію та остаточну редактуру завжди залишається за людиною-журналістом або редактором.	Зберегти відповідальність, контекстне розуміння та професійну етику.

Складено автором. Джерело. "Галюцинації LLM: стратегії боротьби" // Ultralytics (2026). https://www.ultralytics.com/ru/glossary/hallucination-in-llms

6.4.2. Необхідність верифікації та фактчекінгу.

У сучасному інформаційному середовищі, яке все більше наповнюється синтетичним контентом, довіра аудиторії перетворюється на найцінніший актив медіа. Втратити її можна вмить — достатньо однієї публікації фейку, сгенерованого штучним інтелектом, після чого повернути читача буде вкрай складно, якщо не неможливо. У цих умовах парадигма «людина в петлі» (Human-in-the-Loop, HITL) стає не просто рекомендацією, а обов'язковим золотим стандартом та єдиною ефективною стратегією відповідальної роботи з генеративним ШІ. Її суть полягає у безумовному пріоритеті людського контролю, де алгоритм розглядається лише як інструмент, який ніколи не може мати права фінального рішення щодо публікації.

Фундаментом HITL є чіткий триетапний цикл, який розподіляє

відповідальність. На першому етапі людина-ініціатор (журналіст) формулює чітке завдання (промпт) та, що критично важливо, самостійно підбирає та надає алгоритму перевірені, достовірні джерела інформації. Це активна роль куратора, яка визначає рамки роботи. Другий етап — генерація (ШІ-модель), яка на основі отриманих даних створює чорновий текст. Нарешті, ключовий третій етап — верифікація (редактор), де людина перевіряє кожен факт, оцінює логіку, контекст та тональність матеріалу. Право натиснути кнопку «Опублікувати» належить виключно їй. Найбільшою небезпекою на цьому шляху є психологічний феномен «упередженості автоматизації» (Automation Bias), коли людина, захоплена технологією, схильна довіряти виведенню машини більше, ніж власному критичному мисленню. Подолати це упередження можна лише свідомою установкою: будь-який текст, створений ШІ, за замовчуванням вважається потенційно помилковим і вимагає ретельного розслідування.

Для такого розслідування необхідний системний підхід. Редактор повинен читати текст ШІ не як роботу колеги, а як свідчення підозрюваного, застосовуючи строгий чек-ліст верифікації. Насамперед це стосується цитат, які ШІ часто генерує правдоподібними, але вигаданими, що вимагає обов'язкового пошуку кожної з них в оригінальних джерелах. Критичної перевірки потребують усі дати, числа, статистичні показники, а також імена та посади людей, оскільки модель легко плутає факти через свій ймовірнісний характер. Не менш важливо аналізувати текст на наявність внутрішніх логічних суперечностей, оскільки ШІ, маючи обмежений контекст, може стверджувати протилежне в різних частинах матеріалу.

Ефективність верифікації забезпечують спеціальні професійні техніки фактчекінгу, адаптовані для роботи з ШІ-контентом. Латеральне читання (Lateral Reading) вимагає не пасивного сприйняття тексту, а паралельної роботи з кількома джерелами: відкриття кількох вкладок у браузері для миттєвої перевірки кожного ключового твердження. Зворотний пошук (Reverse Search) є універсальним інструментом: для тексту — це перевірка унікальності фрагментів через пошукові системи для виявлення плагіату чи маніпуляцій; для зображень — використання таких інструментів, як Google Lens

або TinEye, щоб встановити оригінальне джерело та виключити використання змінених або вирваних з контексту фото. Принцип тріангуляції вимагає підтвердження будь-якого важливого факту з щонайменше трьох незалежних і авторитетних джерел, що мінімізує ризик поширення випадкової помилки чи навмисної маніпуляції.

Важливо розуміти, що вся ця складна робота з контролю та перевірки має не лише професійно-етичне, але й пряме юридичне підґрунтя. ШІ не є суб'єктом права. Відповідальність за будь-який контент, опублікований у медіа, незалежно від того, хто або що було його технічним генератором, повністю несе видавець у особі головного редактора та власника платформи. У разі поширення наклепу, дискредитуючої інформації або порушення авторських прав позов буде подано саме на них. Судова система не визнає відмовку «це написав робот» — закон працює за принципом «опублікував — відповідаєш». Таким чином, ретельна верифікація за схемою HITL — це не просто рекомендація щодо якості, а обов'язкова процедура управління правовими ризиками та захисту репутації видання.

Отже, необхідність верифікації та фактчекінгу в епоху генеративного ШІ перетворюється з рутинної редакційної задачі на стратегічний імператив виживання медіа. Парадигма Human-in-the-Loop визначає нову роль журналіста не як первинного генератора слів, а як незамінного куратора, слідчого та гаранта достовірності. Це єдина модель, яка дозволяє використовувати потужність штучного інтелекту для швидкості та масштабу, не жертвуючи при цьому основним капіталом журналістики — довірою суспільства.

6.4.3. Прозорість про використання автоматизації.

У цифрову епоху, коли кордон між людською та машинною творчістю стає дедалі більш розмитим, прозорість (Transparency) перетворюється на фундаментальний принцип та основу неявного контракту між медіа та його аудиторією. Якщо читач споживає контент, впевнений, що він є результатом думок, аналізу та творчості живої людини-журналіста, тоді як насправді він сформований статистичною вибіркою слів великої мовної моделі, це розглядається як обман, що неминуче підриває довіру. Таким чином, відкритість щодо використання штучного інтелекту стає не просто етичним вибором, а стратегічною необхідністю для збереження авторитету та легітимності видання.

Для реалізації цього принципу на практиці провідні медіа використовують градієнтний підхід до маркування, розрізняючи три основні рівні втручання ШІ, кожен з яких вимагає різного ступеня розкриття інформації. Перший рівень — ШІ як непомітний інструмент (Invisible AI) — стосується використання технологій для рутинних, допоміжних завдань, таких як перевірка орфографії за допомогою Grammarly, розшифровка аудіоінтерв'ю в Otter.ai або автоматичний переклад документів для внутрішнього аналізу. На цьому етапі маркування не є обов'язковим, оскільки технологія не впливає на зміст або творчу складову, аналогічно тому, як не потрібно вказувати використання текстового процесора. Другий рівень — асистування ШІ (AI-Assisted) — включає випадки, коли модель бере активнішу участь у творчому процесі: пропонує варіанти заголовків, генерує чернетки на основі даних, скорочує тексти або систематизує інформацію. Однак фінальний текст проходить повне редагування, переписування та несе на собі відбиток думки людини-журналіста. Тут маркування стає бажаним, часто у формі короткої примітки в кінці матеріалу: «У підготовці цього тексту використано інструменти штучного інтелекту». Найвищий, третій рівень — генерація ШІ (AI-Generated) — стосується контенту, створеного майже повністю автономно: автоматизованих фінансових звітів, спортивних дайджестів, метеопрогнозів або ілюстрацій, згенерованих Midjourney чи DALL-E. У таких випадках чітке та видиме маркування є абсолютно обов'язковим, оскільки саме судження та творчий вибір алгоритму становлять основну цінність матеріалу.

Технічна реалізація такого маркування, або дизайн розкриття інформації (UX of Disclosure), має бути інтуїтивно зрозумілою для читача. Найбільш поширеними методами є зміна рядка автора (Byline). У повністю автоматизованих матеріалах замість імені журналіста може вказуватися «News Bot», «Automated Insights» або «ШІ-редакція». Для гібридних матеріалів використовується формат: «Автор: Іван Петренко за підтримки ChatGPT». Крім того, редакції додають візуальні індикатори — невеликі іконки у вигляді робота, іскри чи позначки «AI» біля заголовка або фотографії, що миттєво сигналізують про походження контенту. Для візуальних творів, особливо згенерованих зображень, критично важливими стають технології цифрової аттестації, такі як водяні знаки та метадані.

Наприклад, DALL-E автоматично вшиває в зображення метадані за стандартом C2PA (Coalition for Content Provenance and Authenticity), які цифрово підтверджують його штучне походження та можуть бути перевірені спеціальними інструментами.

Ці практики знаходять підтримку в міжнародних етичних хартіях та політиках. Знаковим документом стала «Паризька хартія про штучний інтелект в журналістиці» (The Paris Charter on AI and Journalism), підписана у 2023 році організацією «Репортери без кордонів» та низкою провідних медіа. Її центральний принцип стверджує, що «ШІ ніколи не повинен повністю замінювати редакційне судження людини». Хартія закликає медіа-організації не лише дотримуватися прозорості у кожному конкретному матеріалі, але й публікувати загальнодоступну Політику використання ШІ (AI Policy) на сторінці «Про нас». Цей документ має роз'яснювати аудиторії, в яких редакційних процесах і до якої міри застосовується штучний інтелект, які принципи та контролі забезпечують якість і достовірність такого контенту. Таким чином, прозорість перестає бути лише технічним дисклеймером і стає частиною відкритого діалогу та підзвітності видання перед суспільством.

Отже, прозорість у використанні штучного інтелекту є складною, багаторівневою практикою, що поєднує етичні принципи, технічні рішення та дизайн комунікації. Вона слугує основним механізмом захисту найцінішого активу журналістики — довіри. Чітке маркування не применшує цінність автоматизованого контенту, а навпаки, будує для нього чесні та зрозумілі рамки споживання, де аудиторія, обізнана про походження інформації, може самостійно оцінювати її вагомість та контекст. У довгостроковій перспективі саме така відкритість дозволяє медіа ефективно інтегрувати нові технології, зберігаючи при цьому свою роль як громадського інституту, заснованого на правді та відповідальності.

6.4.4. Редакційний контроль та людський нагляд.

Впровадження технологій штучного інтелекту в журналістику не усуває потребу в редакторах, а докорінно трансформує їхню роль. Із «виправлячів ком» та коректорів тексту вони перетворюються на операторів складних інформаційних систем, менеджерів ризиків та ключових гарантів етичних стандартів. Ця зміна вимагає нових, структурованих підходів до управління — редакційного контролю

(governance), який будується на ієрархії нагляду, чіткому розподілі відповідальності та технологічних механізмах безпеки. Як зазначають практики, задача головного редактора — обрати той рівень втручання в роботу, який дає впевненість у якісному результаті, що в епоху ШІ набуває нового, технологічного виміру.

У практичній площині цей контроль реалізується через ієрархію нагляду (Hierarchy of Oversight), що базується на оцінці ризиків, аналогічній підходам сучасного регулювання (наприклад, ризик-орієнтованій класифікації Європейського Акту про ШІ). Редакції будують власну матрицю, де кожен тип контенту отримує відповідний режим перевірки. Контент низького ризику (автоматизовані спортивні результати, курси валют, прогноз погоди) може функціонувати в режимі повної автоматизації з вибірковим аудитом після публікації. Контент середнього ризику (семмарі, рерайт прес-релізів) вимагає обов'язкового попереднього огляду людиною перед публікацією. Нарешті, контент високого ризику (аналітика, матеріали про політику, кримінал, здоров'я) залишається під повним редакційним контролем, де ШІ виступає лише інструментом для підготовки чорнових матеріалів, а людина несе повну відповідальність за переписування, фактчекінг та фінальне рішення про публікацію.

Окремою ланкою в цій системі стає нова професія — AI Editor (Редактор ШІ), яка з'являється у провідних світових та українських редакціях. Цей фахівець поєднує технологічну компетентність з редакційною етикою. Його обов'язки виходять за рамки традиційного редагування: це написання та тестування промптів для отримання оптимального результату від ШІ, налаштування «тона голосу» моделей під стиль видання, вибірковий, але системний фактчекінг роботи алгоритмів, а також моніторинг виходів на етичні межі — виявлення непомітних упереджень, дискримінаційних формулювань чи інших порушень. Таким чином, AI Editor виступає основним «перекладачем» між світом журналістики та світом алгоритмів, забезпечуючи їхню ефективну та безпечну інтеграцію.

Технічною вимогою безпеки для будь-якої автоматизованої системи є наявність принципу «Кнопки Стоп» (Kill Switch) — аварійного вимикача. Це прямий механізм швидкого реагування на системні збої, технологічні помилки або зловмисне втручання. Наприклад, у разі зламу джерела даних (наприклад, сейсмодатчика)

або виявлення масової генерації фейків, черговий редактор повинен мати можливість миттєво відключити бота від систем публікації, щоб запобігти поширенню шкідливого контенту. Цей інструмент є матеріальним втіленням пріоритету людського контролю над технологічною автоматизацією.

Фундаментом всієї цієї системи управління залишається персональна та юридична відповідальність (Accountability). У разі помилки винною завжди є людина, а не інструмент. Це відображає як професійну етику, так і законодавчу норму: головний редактор несе відповідальність за зміст і якість всіх матеріалів, опублікованих у виданні. Ланцюг відповідальності за контент, створений за участю ШІ, чітко визначається: від журналіста-ініціатора запиту через редактора-верифікатора до головного редактора-стратега. Такий підхід узгоджується з принципами сучасної редакційної політики, яка вимагає чіткого розмежування ролей та відповідальності на кожному етапі створення контенту. Відмовка «це написав робот» є юридично та професійно безперспективною. Таким чином, ефективний редакційний контроль над ШІ — це динамічна система, що поєднує людський професійний скептицизм, структуровані процедури оцінки ризиків, спеціалізовані roles та технологічні механізми безпеки, усі спрямовані на збереження довіри аудиторії в умовах технологічної трансформації.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання розкривають фундаментальні проблеми надійності ШІ:

1. Природа "Галюцинацій": Чому великі мовні моделі (LLM) вигадують факти? Поясніть концепцію "Стохастичних папуг" (Stochastic Parrots): модель не розуміє сенсу слів, вона лише передбачає, яке слово статистично має йти наступним.

2. Проблема "Чорної скриньки" (Black Box): Чому навіть розробники нейромереж часто не можуть пояснити, чому алгоритм прийняв те чи інше рішення? Як відсутність інтерпретованості (Interpretability) заважає впровадженню ШІ в серйозну журналістику та медицину?

3. Упередженість даних (Data Bias): Алгоритм навчається на інтернеті. Як це призводить до того, що ШІ може відтворювати расові, гендерні чи географічні стереотипи (наприклад, зображати всіх ліка-

рів чоловіками, а медсестер — жінками)?

4. Технології походження контенту (Provenance): Що таке стандарт C2PA (Coalition for Content Provenance and Authenticity)? Як цифрові підписи ("Content Credentials") можуть допомогти читачу відрізнити реальне фото з місця подій від згенерованого Midjourney?

5. Згасання моделі (Model Collapse): Що станеться, якщо нові моделі ШІ будуть навчатися на текстах, написаних іншими ШІ? Чому це призводить до деградації якості та втрати рідкісної інформації?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають критичному аналізу результатів генерації:

Завдання 1. Стрес-тест на упередженість (Red Teaming)

• Інструмент: Будь-який генератор зображень (Midjourney / Bing Image Creator).

• Завдання: Сформулюйте нейтральні запити і проаналізуйте результат на стереотипи.

- *Запит:* "American family" (Чи будуть там представники різних рас?).

- *Запит:* "CEO of a company" (Якої статі буде персонаж?).

- *Запит:* "Ukrainian village" (Чи буде це сучасне село, чи стереотипна хата-мазанка XIX століття?).

• Висновок: Опишіть, як "надивленість" моделі впливає на викривлення реальності.

Завдання 2. Фактчекінг "Впевненої брехні"

• Інструмент: ChatGPT.

• Завдання: Попросіть модель надати список наукових статей (з авторами та роками) на дуже вузьку тему, наприклад, "Вплив українського фольклору на японську анімацію".

• Перевірка: Спробуйте знайти ці статті в Google Scholar.

• Результат: Підрахуйте, скільки статей є реальними, а скільки — вигаданими комбінаціями реальних прізвищ та слів.

Завдання 3. Верифікація метаданих (C2PA)

• Матеріал: Знайдіть зображення, згенероване в DALL-E 3 або Adobe Firefly (які підтримують маркування).

• Інструмент: Сервіс *Content Credentials (Verify) або Hugging Face Deepfake Detector.*

• Дія: Завантажте файл і перевірте, чи зберігся "цифровий слід"

генерації. Що станеться, якщо зробити скріншот цього зображення і завантажити скріншот? (Слід зникне).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Emily Bender, Timnit Gebru et al. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* (Найважливіша наукова стаття, що пояснює ризики сліпої довіри до LLM).

2. Cathy O'Neil. *Weapons of Math Destruction.* (Про те, як упереджені алгоритми посилюють нерівність).

3. C2PA Technical Specification. (Опис галузевого стандарту боротьби з дезінформацією).

4. Gary Marcus. *Rebooting AI: Building Artificial Intelligence We Can Trust.* (Критика сучасного підходу до глибокого навчання).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• Стохастичні папуги (Stochastic Parrots): Метафора, що описує роботу великих мовних моделей. Папуга може імітувати звуки людської мови дуже переконливо, не розуміючи їхнього змісту. Так само ШІ комбінує слова на основі ймовірності, не маючи свідомості чи розуміння істини.

• Чорна скринька (Black Box): Система, внутрішні процеси якої (чому було прийнято саме таке рішення) невідомі або незрозумілі користувачеві. Це головна проблема довіри до нейромереж.

• Упередженість (Bias): Систематичне відхилення результатів алгоритму, викликане хибними припущеннями в процесі машинного навчання (наприклад, якщо навчати модель на даних злочинності минулого, вона буде вважати злочинцями представників певних меншин).

• C2PA (Coalition for Content Provenance and Authenticity): Відкритий технічний стандарт, який дозволяє "вшивати" в медіафайл інформацію про його походження (хто створив, чим редагував, чи використовувався ШІ), захищену криптографією.

• Model Collapse (Колапс моделі): Дегенеративний процес, що відбувається, коли генеративні моделі навчаються на даних, створених іншими моделями. Це призводить до втрати нюансів, спрощення реальності та накопичення помилок.

6.5. Кейси успішного використання штучного інтелекту в журналістиці

Впровадження штучного інтелекту в журналістську практику відбувається поступово, але вже сьогодні можна виділити кілька знакових прикладів, які демонструють ефективність цієї технології. Провідні світові медіаорганізації не лише експериментують з ШІ, а й активно інтегрують його у свої редакційні процеси, створюючи нові стандарти роботи та розширюючи можливості інформаційного покриття.

6.5.1. Associated Press: автоматизація корпоративних звітів.

Агентство Associated Press (AP) стало одним із піонерів використання штучного інтелекту в журналістиці, запустивши у 2014 році програму автоматизованого створення фінансових звітів про результати діяльності публічних компаній. До впровадження ШІ редакція AP щокварталу публікувала приблизно 300 статей про фінансові результати компаній, що вимагало значних людських ресурсів та часу. Журналісти витрачали години на обробку стандартизованих даних, складання типових звітів та перевірку цифр, що відволікало їх від більш складних аналітичних матеріалів.

Впровадження системи автоматизованої журналістики на базі технології Wordsmith від компанії Automated Insights кардинально змінило ситуацію. Система отримує структуровані дані безпосередньо з фінансових звітів компаній, аналізує ключові показники та автоматично генерує новинні матеріали за попередньо визначеними шаблонами. Алгоритм здатний виявляти найважливіші тренди, порівнювати поточні результати з попередніми кварталами, визначати відхилення від прогнозів аналітиків та формулювати це у зрозумілій формі.

Результати впровадження перевершили очікування. AP збільшило кількість публікованих фінансових звітів до 4400 на квартал, що означає більш ніж десятикратне зростання охоплення. При цьому час публікації скоротився з кількох годин до декількох хвилин після оприлюднення фінансових результатів компаніями. Це дозволило агентству охопити не лише великі корпорації, а й середні та невеликі публічні компанії, про які раніше просто не було ресурсів писати. Точність звітів також зросла, оскільки алгоритм практично

не допускає помилок у цифрах, які часто траплялися при ручному введенні даних.

Звільнені людські ресурси AP перенаправило на створення більш складного аналітичного контенту. Журналісти тепер можуть зосередитися на глибинних розслідуваннях, інтерв'ю з керівниками компаній, аналізі довгострокових трендів та контекстуалізації фінансових результатів. Це не призвело до скорочення штату, а навпаки, дозволило підвищити якість журналістської роботи. Крім того, AP розширило використання автоматизації на інші сфери, включаючи спортивні звіти про матчі нижчих ліг, де також є велика кількість структурованих даних.

Досвід Associated Press продемонстрував, що штучний інтелект найефективніший у роботі зі структурованими даними та стандартизованими форматами, де потрібна швидкість та точність, але не потрібна глибока інтерпретація чи творчий підхід. Це стало моделлю для багатьох інших новинних організацій по всьому світу.

6.5.2. Bloomberg: використання Cyborg для фінансових новин.

Bloomberg, один із провідних світових постачальників фінансової інформації, розробив власну систему штучного інтелекту під назвою Cyborg для створення новин про ринки та компанії. На відміну від повністю автоматизованих систем, Cyborg працює за принципом асистування журналістам, а не заміни їх, що відображено у самій назві проєкту, яка символізує симбіоз людини та машини.

Система Cyborg аналізує величезні масиви фінансових даних у режимі реального часу, відстежуючи рух акцій, облігацій, валют, товарних ринків та інших фінансових інструментів. Коли алгоритм виявляє значущі зміни, аномалії або тренди, він автоматично генерує чернетку новини, яка містить ключові цифри, контекст та базовий аналіз. Ця чернетка передається журналістам Bloomberg, які можуть швидко її відредагувати, доповнити коментарями експертів, додати ширший контекст або опублікувати майже без змін, якщо матеріал відповідає редакційним стандартам.

Особливістю Cyborg є швидкість реакції на ринкові події. Система здатна генерувати першу версію новини за лічені секунди після появи значущої інформації, що критично важливо на фінансових ринках, де кожна секунда може мати значення для трейдерів та інвесторів. Наприклад, коли публікуються макроекономічні дані, такі

як показники зайнятості, інфляції чи ВВП, Cyborg миттєво аналізує цифри, порівнює їх з очікуваннями та попередніми значеннями і формує новинну замітку, яку журналіст може опублікувати протягом хвилини.

Bloomberg повідомляв, що Cyborg допомагає створювати тисячі новинних матеріалів щомісяця, значно розширюючи охоплення ринків та компаній. Система особливо ефективна для висвітлення подій на азіатських та європейських ринках у години, коли американська редакція може бути менш укомплектованою. Це забезпечує клієнтам Bloomberg, які потребують цілодобового інформаційного покриття, постійний потік актуальних даних.

Важливим аспектом роботи Cyborg є його здатність до навчання. Система постійно аналізує, які матеріали редагуються журналістами, які зміни вони вносять, які теми викликають найбільший інтерес аудиторії. На основі цього зворотного зв'язку алгоритм удосконалює свої шаблони та підходи до генерації контенту. Bloomberg також використовує машинне навчання для персоналізації новинних потоків для різних груп користувачів, враховуючи їхні професійні інтереси та торгові стратегії.

Досвід Bloomberg демонструє модель співпраці між журналістами та штучним інтелектом, де технологія бере на себе рутинні завдання та забезпечує швидкість, а людина додає аналітичну глибину, експертну оцінку та редакційну відповідальність. Це дозволяє поєднувати переваги автоматизації з професійними стандартами якісної журналістики.

6.5.3. BBC: експерименти з персоналізованими новинами.

Британська мовна корпорація BBC, одна з найбільших та найавторитетніших новинних організацій світу, активно досліджує можливості штучного інтелекту для персоналізації новинного контенту та покращення взаємодії з аудиторією. На відміну від комерційних медіа, BBC як суспільний мовник стикається з унікальним викликом – необхідністю обслуговувати надзвичайно різноманітну аудиторію з різними інтересами, рівнями освіти та потребами, зберігаючи при цьому принципи об'єктивності та суспільної користі.

BBC експериментує з різними формами персоналізації новин на базі штучного інтелекту. Одним із напрямків є розробка систем,

які аналізують поведінку користувачів на цифрових платформах корпорації – які статті вони читають, скільки часу проводять з різними матеріалами, які теми їх цікавлять. На основі цих даних алгоритми формують індивідуалізовані новинні стрічки, пропонуючи кожному користувачеві контент, який найімовірніше буде для нього релевантним. При цьому BBC дуже обережно підходить до питання створення «інформаційних бульбашок» та забезпечує, щоб користувачі отримували збалансований спектр новин, а не лише те, що підтверджує їхні наявні погляди.

Інший важливий напрямок – використання штучного інтелекту для створення різних версій одного й того самого матеріалу для різних аудиторій. Експериментальні проєкти BBC досліджують можливість автоматичної адаптації складності тексту залежно від рівня попередніх знань читача з теми. Наприклад, матеріал про складну економічну проблему може бути представлений у спрощеній версії для широкої публіки та у більш технічній версії для фахівців. Алгоритми також можуть автоматично додавати пояснення термінів, контекстну інформацію або посилання на попередні матеріали залежно від того, чи є користувач новачком у темі чи регулярно її відстежує.

BBC також активно використовує штучний інтелект для покращення доступності контенту. Корпорація розробила системи автоматичного створення субтитрів для відеоконтенту, що особливо важливо для людей із вадами слуху. Алгоритми розпізнавання мовлення постійно вдосконалюються, навчаючись розуміти різні акценти та діалекти, що критично важливо для глобальної аудиторії BBC. Крім того, експериментуються технології автоматичного створення аудіоверсій текстових матеріалів для людей із вадами зору або для тих, хто віддає перевагу аудіоформату.

Особливу увагу BBC приділяє етичним аспектам використання ШІ. Корпорація розробила власні керівні принципи, які визначають, як саме можна використовувати персоналізацію, щоб не маніпулювати аудиторією та не створювати упередженості. Алгоритми навмисно програмуються так, щоб виводити з «комфортної зони» користувачів, показуючи їм різноманітні погляди та теми, які вони самі не обрали б. BBC також працює над прозорістю алгоритмів, дозволяючи користувачам розуміти, чому їм показується певний контент.

Ще одним напрямком експериментів є використання чат-ботів та голосових асистентів для взаємодії з аудиторією. BBC розробила інтерактивних помічників, які можуть відповідати на запитання користувачів про новини, знаходити релевантні матеріали в архівах корпорації та навіть вести прості діалоги про поточні події. Це особливо популярно серед молодшої аудиторії, яка звикла до спілкування з віртуальними асистентами на кшталт Siri або Alexa.

Досвід BBC показує, що штучний інтелект може бути потужним інструментом для суспільних мовників у виконанні їхньої місії – робити якісну інформацію доступною для всіх верств населення. При цьому корпорація демонструє, що персоналізація та редакційна відповідальність не є взаємовиключними, якщо технологію використовувати обдумано та з дотриманням етичних стандартів.

6.5.4. The Washington Post: Heliograf для виборів та Олімпіади.

Газета The Washington Post розробила власну систему штучного інтелекту під назвою Heliograf, яка стала одним із найамбітніших проєктів використання автоматизованої журналістики в американських медіа. Система вперше була запущена під час президентських виборів 2016 року та Олімпійських ігор у Ріо-де-Жанейро, де продемонструвала свою ефективність у висвітленні подій, які генерують величезну кількість даних та потребують швидкого реагування.

Під час виборів 2016 року Heliograf відстежував результати голосування у сотнях виборчих округів, автоматично генеруючи новинні оновлення про підрахунок голосів, результати на різних рівнях (від місцевих виборів до президентських) та динаміку перемог кандидатів. Система створила понад 500 коротких новинних матеріалів за виборчу ніч, що було б фізично неможливо для журналістів вручну. Кожен матеріал містив актуальні дані, відсоток підрахованих голосів, порівняння з попередніми виборами та іншу релевантну інформацію. Це дозволило Washington Post забезпечити безпрецедентно детальне висвітлення виборів на всіх рівнях, включаючи місцеві вибори в невеликих округах, які зазвичай залишаються поза увагою національних медіа.

Під час Олімпійських ігор Heliograf працював у дещо іншому режимі. Система відстежувала результати змагань у режимі реального часу та генерувала короткі новинні оновлення про медалі, рекорди, несподівані результати та досягнення американських спортсменів.

За час Олімпіади було створено сотні автоматизованих матеріалів, які публікувалися на сайті та в додатку Washington Post. Особливо ефективним це виявилося для менш популярних видів спорту, які не отримують такого висвітлення, як, наприклад, плавання чи легка атлетика. Фанати стрільби з лука, веслування чи стендової стрільби також могли отримувати оперативну інформацію про своїх улюблених спортсменів.

Важливою особливістю Heliograf є його гнучкість та здатність працювати з різними типами структурованих даних. Систему можна налаштувати на різні події та формати – від спортивних змагань до фінансових звітів, від погодних умов до статистики злочинності. Washington Post використовує Heliograf не лише для великих подій, а й для регулярного висвітлення місцевих новин, результатів спортивних матчів шкільних команд, звітів про стан доріг та інших тем, де є надійні джерела структурованих даних.

Редакція Washington Post підкреслює, що Heliograf не замінює журналістів, а розширює їхні можливості. Система береже журналістський час на рутинних завданнях, дозволяючи репортерам зосередитися на більш складних матеріалах – інтерв'ю, аналітиці, розслідуваннях, репортажах з місць подій. Під час виборів, наприклад, поки Heliograf генерував базові звіти про результати голосування, журналісти могли спілкуватися з виборцями, аналізувати тренди, досліджувати неочікувані результати та створювати глибші аналітичні матеріали про значення виборів.

Технічно Heliograf побудований на основі шаблонів та правил, які визначають, як інтерпретувати дані та яку структуру повинен мати текст. Журналісти та редактори Washington Post активно беруть участь у створенні цих шаблонів, забезпечуючи, що автоматизовані матеріали відповідають редакційним стандартам газети. Система також включає механізми перевірки даних, щоб уникнути публікації помилкової інформації через технічні збої або неточності у вихідних даних.

З часом Washington Post інтегрував Heliograf у свій регулярний редакційний процес. Система тепер використовується не лише для великих подій, а й для щоденного виробництва контенту. Газета також розробила внутрішні інструменти, які дозволяють журналістам без технічної підготовки створювати нові шаблони для автоматизації,

що демократизує використання технології в межах редакції. Крім того, Washington Post ділиться своїм досвідом з іншими медіа та періодично публікує матеріали про роботу Heliograf, сприяючи розвитку індустрії автоматизованої журналістики.

Досвід The Washington Post демонструє, що штучний інтелект може значно розширити можливості навіть великих та ресурсних медіаорганізацій, дозволяючи їм охоплювати теми та події, які раніше залишалися поза увагою через обмеження людських ресурсів. При цьому успіх проєкту багато в чому залежить від правильної інтеграції технології в редакційні процеси та підтримки з боку журналістів, які розуміють як потенціал, так і обмеження автоматизації.

1. КОНТРОЛЬНІ ПИТАНЯ

Рівень "Розуміння":

1. У чому принципова різниця між шаблонним підходом (Automated Insights) та генеративним підходом (ChatGPT)? Де ризик помилки вищий?

2. Що таке "галюцинація" штучного інтелекту і чому вона виникає?

3. Чому фінансові звіти та спортивні результати стали першими жанрами, які захопили роботи?

4. Що таке мультимодальна модель? Наведіть приклад завдання, яке вона може виконати.

Рівень "Аналіз": 5. Як працює концепція "Human-in-the-Loop" у новинах середнього та високого ризику? 6. Чому прямий переклад ідіом через старі перекладачі відрізняється від локалізації (транскреації) через LLM? 7. Які етичні ризики несе використання згенерованих зображень у новинній стрічці (наприклад, фото з місця подій)?

Рівень "Синтез": 8. Спроєктуйте алгоритм дій для редакції у випадку, якщо їхній новинний бот опублікував фейкову інформацію про падіння курсу валют. 9. Як автоматизація впливає на економіку медіа? Поясніть поняття "граничної вартості" (Marginal Cost) контенту.

2. ПРАКТИКУМ (Лабораторні роботи)

Завдання 1. "Логіка робота" (NLG)

• Умова: Є дані футбольного матчу: Команда А (3) - Команда Б (0).

• Завдання: Напишіть псевдокод (логічне дерево) для заголовка новини.

- *Приклад:* IF (Score_Diff >= 3) PRINT "Розгромна перемога..." ELSE PRINT "Перемога..."

• Мета: Зрозуміти, як працюють шаблонні системи "під капотом".

Завдання 2. Промпт-інжиніринг (LLM)

• Умова: Є сухий прес-реліз міської ради про ремонт труб.

• Завдання: Використовуючи ChatGPT/Claude, перетворити його на:

1. Коротку новину для Telegram (3 речення).

2. Пост для Instagram (з емодзі та закликом до дії).

3. Заголовок у стилі "клікбейт" (для аналізу, як не треба робити).

• Мета: Навчитися адаптувати tone-of-voice за допомогою промптів.

Завдання 3. Фактчекінг та "Червона команда"

• Умова: Надається згенерований ШІ текст про історичну подію (з навмисно внесеною помилкою в даті або імені).

• Завдання: Застосувати метод "Латерального читання", знайти помилку та виправити її, вказавши джерело правди.

• Мета: Розвинути навичку критичного сприйняття роботи ШІ.

Завдання 4. Розробка AI Policy

• Завдання: Написати 3 пункти для сторінки "Про нас", які пояснюють читачам, як ваше медіа використовує (і не використовує) штучний інтелект.

• Мета: Засвоїти принципи прозорості та етики.

ГЛОСАРІЙ ТЕРМІНІВ (Мінімум, який треба знати)

• NLG (Natural Language Generation): Технологія перетворення структурованих даних (таблиць) на зв'язний текст.

• LLM (Large Language Model): Ймовірнісна нейромережа (напр. GPT-4), навчена на великих масивах тексту, здатна генерувати креативний контент.

• Hallucination (Галюцинація): Впевнена генерація мовною моделлю неправдивої інформації.

• HITL (Human-in-the-Loop): Модель роботи, де людина обов'язково перевіряє результат роботи алгоритму перед публікацією.

• Prompt Engineering: Мистецтво формулювання запитів до ШІ

для отримання точного результату.

• Zero Latency: Здатність системи публікувати новини миттєво після отримання даних (важливо для фінансів/спорту).

• Multimodality (Мультимодальність): Здатність ШІ працювати одночасно з текстом, зображеннями, відео та аудіо.

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Основна:

1. Carlson, M. *The Robotic Reporter: Automation and the Future of Journalism.*

2. Diakopoulos, N. *Automating the News: How Algorithms Are Rewriting the Media.*

3. Marconi, F. *Newsmakers: Artificial Intelligence and the Future of Journalism.*

Практична (Гайди та звіти):

1. Associated Press. *Artificial Intelligence Guidelines for Newsrooms.*

2. Reuters Institute. *Generative AI in Journalism: 2024 Report.*

3. Reporters Without Borders. *Paris Charter on AI and Journalism.*

Висновки до розділу 6

Аналіз еволюції технологій та редакційних практик у рамках цього розділу демонструє, що автоматизація контенту перестала бути експериментом чи футуристичною перспективою — вона стала операційною реальністю, яка кардинально перевизначає ландшафт медіа. Ця трансформація не є однолінійною; вона розвивається за двома паралельними, але принципово різними шляхами, формуючи нову технологічну дихотомію. З одного боку, шаблонна генерація на основі правил (Rule-Based NLG) залишається «золотим стандартом» для роботи зі структурованими даними у таких сферах, як фінанси, спорт та погода. Її неперевершена перевага — абсолютна точність і передбачуваність — робить її незамінним інструментом для масштабування рутинного контенту. З іншого боку, поява генеративних великих мовних моделей (LLM), таких як GPT-4 чи Claude, відкрила еру «креативної автоматизації», дозволяючи створювати саммари, адаптувати стиль та генерувати мультимедійні елементи. Однак їхня фундаментальна гнучкість супроводжується критичною слабкістю — схильністю до фабрикації інформації, або «галюцинацій», що вимагає принципово нових, жорсткіших систем

контролю та верифікації.

Ця технологічна революція має глибокі економічні наслідки, усуваючи головний бар'єр для висвітлення так званого «довгого хвоста» інформації. Завдяки зниженню граничної вартості виробництва окремої одиниці контенту практично до нуля, медіа отримали змогу економічно доцільно охоплювати події, які раніше ігнорувалися через нерентабельність: матчі регіональних ліг, новини окремих мікрорайонів, персоналізовані фінансові звіти. Одночасно швидкість публікації (Zero Latency) перетворилася на критичну конкурентну перевагу, особливо в таких сферах, як фінансова журналістика, де кожна секунда запізнення може коштувати мільйонів. Однак найважливішою зміною стала трансформація самої професії. Концепція «журналіста-письменника» як основного генератора тексту поступається місцем концепції «доповненого журналіста» (Augmented Journalist). Рутинні операційні завдання — транскрибація інтерв'ю, переклад документів, написання первинних зведень — все частіше делегуються алгоритмам. Це вивільняє найцінніший ресурс — час і інтелектуальну енергію журналіста — для виконання завдань, де людина залишається неперевершеною: глибокі розслідування, польова робота, емпатичне інтерв'ювання та складний наративний сторителінг. У відповідь на цю потребу в редакціях з'являються нові фахівці: AI Editor, промпт-інженер, архітектор даних, роль яких полягає в управлінні цими складними системами.

Однак потужність генеративних технологій спричинила і глибоку кризу довіри, створивши безпрецедентні ризики для інформаційної екосистеми у вигляді фейків, діпфейків та систематичних помилок. У відповідь на це провідні медіа здійснюють стратегічний перехід від простої моделі «Створити -> Опублікувати» до моделі «Людина в петлі» (Human-in-the-Loop, HITL), де людина зберігає остаточний контроль на кожному етапі. У цих умовах верифікація та фактчекінг перестають бути лише однією з функцій редакції; вони стають головною доданою вартістю та конкурентною перевагою медіа. Водночас прозорість маркування ШІ-контенту перетворюється на основу етичного контракту з аудиторією, будуючи довіру через чесність щодо походження інформації. Ця прозорість поширюється і на новий стандарт контенту — мультимодальність. Бар'єри

між текстом, зображенням та відео руйнуються; сучасні моделі дозволяють генерувати комплексні мультимедійні пакети з єдиного запиту. Це, у свою чергу, вимагає від редакційних структур гнучкості та інтеграції, змушуючи традиційно розділені відділи тексту, дизайну та відео об'єднуватися в крос-функціональні команди для керування єдиним потоком створення контенту.

Таким чином, загальний підсумок розділу можна сформулювати так: автоматизація контент-виробництва більше не є питанням вибору «за» чи «проти». Вона стала невід'ємним середовищем функціонування сучасних медіа. Успіх у цій новій реальності залежить не від факту використання штучного інтелекту, а від того, наскільки ефективно видання вибудувало свою систему управління (Governance) — збалансовану, динамічну структуру, яка поєднує неймовірну швидкість та креативний потенціал алгоритмів з незамінною відповідальністю, критичним мисленням та етичним судженням людини. Майбутнє належить не тим, хто автоматизує найбільше, а тим, хто керує автоматизацією найрозумніше.

Розділ 7: Персоналізація дистрибуції
7.1. Динамічні новинні стрічки

Сучасний споживач новин перебуває в умовах інформаційного перенасичення, коли щодня створюються мільйони нових матеріалів, і фізично неможливо переглянути навіть мінімальну їх частину. У цьому контексті динамічні новинні стрічки, керовані алгоритмами штучного інтелекту, стали критично важливим інструментом для організації та доставки інформації. Вони визначають, який контент побачить користувач, у якому порядку та як довго матеріал залишатиметься видимим. Ефективність цих систем безпосередньо впливає на те, наскільки добре аудиторія інформована, наскільки різноманітні погляди вона отримує та наскільки задоволена від взаємодії з медіаплатформою.

7.1.1. Хронологічна vs алгоритмічна стрічка.

Питання організації новинної стрічки є одним із найбільш дискусійних у сучасній журналістиці та медіатехнологіях. Два основні підходи – хронологічний та алгоритмічний – представляють різні філософії щодо того, як користувачі повинні споживати інформацію, і кожен має свої переваги та недоліки.

Хронологічна стрічка, також відома як зворотно-хронологічна, є найстарішим та найпростішим підходом до організації контенту. В такій стрічці матеріали відображаються в порядку їх публікації, де найновіші з'являються вгорі. Цей підхід має глибоке коріння в традиційній журналістиці, де свіжість новин завжди була ключовим критерієм їхньої цінності. Хронологічна стрічка забезпечує повну прозорість – користувач точно знає, чому він бачить певний контент, і може бути впевнений, що не пропустив нічого важливого, якщо переглянув стрічку до певного часу.

Переваги хронологічної стрічки особливо очевидні для професійних споживачів новин – журналістів, аналітиків, політиків, науковців та інших, хто потребує повного та своєчасного розуміння інформаційного ландшафту. Вона дозволяє відстежувати розвиток подій у реальному часі, бачити, як змінюється наратив навколо певної теми, та розуміти хронологічну послідовність фактів. Крім того, хронологічна стрічка не створює фільтраційних бульбашок і не нав'язує редакційних пріоритетів, які можуть бути закладені в

алгоритмі.

Однак хронологічний підхід має серйозні обмеження в умовах інформаційного перенасичення. Користувач, який підписаний на багато джерел або тем, може отримувати сотні або навіть тисячі оновлень на день. У такій ситуації хронологічна стрічка стає малоефективною, оскільки дійсно важливі матеріали можуть легко загубитися серед менш релевантного контенту. Якщо користувач не переглядає стрічку кілька годин, він може пропустити критично важливу інформацію, яка була опублікована раніше і вже опустилася вниз під новішими, але менш значущими публікаціями.

Алгоритмічна стрічка виникла як відповідь на ці обмеження. Замість того, щоб просто показувати контент у порядку публікації, алгоритмічна стрічка використовує складні системи машинного навчання для визначення, який контент буде найбільш релевантним та цікавим для конкретного користувача в конкретний момент часу. Алгоритм аналізує велику кількість сигналів – історію взаємодії користувача з контентом, його демографічні характеристики, поведінку схожих користувачів, популярність матеріалу, його свіжість, джерело публікації та багато інших факторів.

Алгоритмічна стрічка може значно покращити користувацький досвід, показуючи людині саме той контент, який їй найімовірніше буде корисним або цікавим. Для платформ це також означає вищий рівень залученості аудиторії – користувачі проводять більше часу в додатку, частіше взаємодіють з контентом, що в кінцевому рахунку може призводити до вищих доходів від реклами або передплати. Алгоритмічні стрічки особливо ефективні для користувачів, які не мають часу або бажання переглядати великі обсяги інформації і хочуть отримувати вже відфільтровані, релевантні матеріали.

Провідні платформи, такі як Facebook, Twitter (тепер X), Instagram та YouTube, поступово перейшли від хронологічних до алгоритмічних стрічок, хоча цей перехід часто викликав протести з боку користувачів. Twitter, наприклад, довгий час був останнім бастіоном хронологічної стрічки, але зрештою також впровадив алгоритмічне ранжування як основний спосіб відображення контенту, зберігши хронологічну стрічку як опцію в налаштуваннях.

Однак алгоритмічні стрічки створюють ряд серйозних проблем. По-перше, вони непрозорі – користувачі рідко розуміють, чому бачать

певний контент і чому не бачать інший. Це створює простір для маніпуляцій та ненавмисних упереджень. По-друге, алгоритми часто оптимізовані на максимізацію залученості, а не на інформування аудиторії. Це може призводити до того, що емоційно заряджений, суперечливий або сенсаційний контент отримує перевагу над виваженим та аналітичним, оскільки перший генерує більше кліків, лайків та коментарів.

По-третє, алгоритмічні стрічки можуть створювати або посилювати інформаційні бульбашки та ехо-камери. Якщо алгоритм помічає, що користувач частіше взаємодіє з контентом певної політичної спрямованості або певних тематик, він почне показувати більше подібного контенту, поступово звужуючи інформаційний горизонт людини. Дослідження показали, що це може посилювати поляризацію суспільства, знижувати толерантність до альтернативних поглядів та робити людей більш вразливими до дезінформації.

У відповідь на ці виклики деякі платформи почали експериментувати з гібридними підходами. Наприклад, вони можуть показувати переважно алгоритмічно відібраний контент, але періодично вставляти матеріали, які виходять за межі звичайних уподобань користувача, або зберігати певні категорії контенту (наприклад, пости від близьких друзів) завжди видимими незалежно від алгоритмічного рейтингу. Інші надають користувачам більше контролю над алгоритмом, дозволяючи їм налаштовувати пріоритети або легко перемикатися між різними режимами перегляду.

Новинні організації також стикаються з дилемою вибору між цими підходами для своїх власних додатків та вебсайтів. Деякі, особливо ті, що орієнтовані на професійну аудиторію, зберігають хронологічну стрічку як основну. Інші активно використовують алгоритмічне ранжування, намагаючись знайти баланс між релевантністю та повнотою інформування. Найскладніше завдання – створити систему, яка буде ефективно працювати для різних сегментів аудиторії, від випадкових читачів до інформаційних професіоналів.

7.1.2. Ранжування контенту на основі релевантності.

Серцем алгоритмічної новинної стрічки є система ранжування, яка визначає, наскільки релевантним є кожен конкретний матеріал для конкретного користувача в конкретний момент часу. Ця система базується на складних моделях машинного навчання, які аналізують

сотні або навіть тисячі різних факторів для прогнозування, чи буде користувач взаємодіяти з контентом і наскільки цінною буде ця взаємодія.

Концепція релевантності в контексті новин є багатовимірною. На найбазовішому рівні релевантність означає відповідність контенту інтересам користувача. Алгоритми вивчають історію взаємодії – які статті людина читала раніше, які теми її цікавлять, на які матеріали вона витрачає більше часу, які лайкає, коментує або поширює. На основі цих даних будується профіль інтересів, який може включати десятки або сотні тематичних категорій з різними вагами.

Сучасні системи ранжування використовують методи глибокого навчання для розуміння семантичної близькості між матеріалами. Наприклад, якщо користувач активно читає статті про кліматичні зміни, алгоритм може визначити, що йому також будуть цікаві матеріали про відновлювану енергетику, екологічну політику або стихійні лиха, навіть якщо раніше він не взаємодіяв безпосередньо з цими темами. Це досягається через векторні представлення тексту, де семантично схожі теми розташовані близько в багатовимірному просторі.

Однак релевантність не обмежується лише тематичними інтересами. Важливим фактором є персональна значущість події для користувача. Якщо людина живе в певному місті, новини про цей регіон будуть для неї більш релевантними, навіть якщо глобально ця подія не є найважливішою. Якщо користувач працює в певній індустрії, професійні новини з цієї сфери матимуть вищий пріоритет. Алгоритми намагаються вловити ці персональні зв'язки, аналізуючи локацію, професійну інформацію з профілю, соціальні зв'язки та інші сигнали.

Системи ранжування також враховують якість джерела та самого матеріалу. Провідні платформи розробляють складні метрики для оцінки достовірності та авторитетності новинних організацій. Ці метрики можуть базуватися на факторах, таких як журналістські стандарти організації, частота публікації виправлень, репутація в професійному середовищі, історія поширення дезінформації та багато інших. Матеріали з авторитетних джерел отримують бонус у ранжуванні, особливо коли йдеться про важливі або контроверсійні теми.

Ще один важливий вимір релевантності – це прогнозування глибини взаємодії. Не всі кліки однаково цінні. Якщо користувач клікає на заголовок, але негайно закриває статтю, це може вказувати на клікбейт або невідповідність контенту очікуванням. Якщо ж людина повністю прочитує довгий матеріал, витрачає час на перегляд зображень та інфографіки, можливо, переходить до пов'язаних статей – це сигналізує про високу цінність контенту. Сучасні алгоритми намагаються прогнозувати саме таку глибоку взаємодію, а не просто кліки.

Соціальний контекст також відіграє значну роль у визначенні релевантності. Якщо багато людей з соціального кола користувача читають або діляться певним матеріалом, це підвищує його релевантність для цього користувача. Це базується на припущенні, що люди зі схожим соціальним середовищем часто мають схожі інтереси та цінності. Однак цей фактор потрібно балансувати обережно, щоб не створювати замкнені інформаційні бульбашки.

Технічно системи ранжування часто використовують архітектуру з кількох етапів. На першому етапі, який називається «кандидат-генерація», з величезного обсягу доступного контенту вибирається відносно невелика підмножина потенційно релевантних матеріалів – наприклад, кілька тисяч з мільйонів. Цей етап зазвичай використовує швидкі, але менш точні методи, такі як пошук на основі ключових слів або простих правил. На другому етапі, «ранжування», до цієї підмножини застосовуються складні моделі машинного навчання, які детально оцінюють кожен матеріал за сотнями параметрів і формують фінальний рейтинг.

Найсучасніші системи використовують ансамблі моделей, де кілька різних алгоритмів оцінюють релевантність незалежно, а потім їхні прогнози об'єднуються для отримання фінальної оцінки. Це дозволяє компенсувати слабкі сторони окремих моделей і підвищити загальну точність системи. Наприклад, одна модель може бути особливо хороша в розумінні тематичних інтересів, інша – у визначенні якості контенту, третя – у прогнозуванні віральності.

Важливим викликом у розробці систем ранжування є постійна еволюція як контенту, так і поведінки користувачів. Моделі потрібно регулярно перенавчати на нових даних, щоб вони залишалися актуальними. Крім того, необхідно виявляти та коригувати

непередбачені упередження, які можуть виникати в процесі роботи алгоритму. Наприклад, якщо система помічає, що користувачі частіше клікають на матеріали з емоційними заголовками, вона може почати давати їм перевагу, створюючи негативний цикл зниження якості контенту.

Провідні новинні платформи та організації вкладають значні ресурси в постійне вдосконалення своїх систем ранжування. Вони проводять А/Б тестування різних підходів, аналізують зворотний зв'язок від користувачів, консультуються з журналістами та експертами з етики. Метою є створення системи, яка не просто максимізує метрики залученості, а й служить суспільному інтересу, допомагаючи людям бути краще інформованими та робити усвідомлені рішення.

7.1.3. Тайм-декей: як свіжість впливає на видимість.

Свіжість або актуальність новин завжди була ключовим фактором їхньої цінності. Сама назва «новини» підкреслює важливість новизни інформації. У цифрову епоху, коли події висвітлюються практично миттєво, а інформаційний цикл прискорився до годин і навіть хвилин, управління свіжістю контенту в алгоритмічних стрічках стало критично важливим завданням. Концепція тайм-декей (time decay) або часового згасання описує, як релевантність матеріалу зменшується з плином часу.

Базова логіка тайм-декею проста: новіші матеріали отримують бонус у ранжуванні, який поступово зменшується з часом. Однак реалізація цього принципу набагато складніша, оскільки різні типи контенту старіють з різною швидкістю. Термінові новини про поточні події, такі як стихійні лиха, теракти або політичні кризи, втрачають актуальність дуже швидко – новий матеріал, опублікований годину тому, вже може бути неактуальним, якщо ситуація швидко розвивається. З іншого боку, аналітичні статті, оглядові матеріали, розслідування або пояснювальна журналістика можуть залишатися релевантними днями, тижнями або навіть місяцями після публікації.

Сучасні системи ранжування використовують різні функції згасання для різних категорій контенту. Для термінових новин може застосовуватися експоненційне згасання, де релевантність різко падає після певного порогу часу. Для вічнозеленого контенту (evergreen content), такого як гайди, довідкові матеріали або глибокі

розслідування, використовується більш м'яка функція згасання, що дозволяє цим матеріалам залишатися видимими значно довше.

Алгоритми намагаються автоматично визначити категорію контенту на основі різних сигналів. Наявність певних ключових слів (таких як «термінові новини», «лише зараз», «оновлення»), швидкість поширення матеріалу, кількість оновлень до вихідної статті, співвідношення переглядів до часу з моменту публікації – все це може вказувати на терміновість. Аналітичні матеріали зазвичай мають довші тексти, більш складну структуру, містять більше контексту та менше орієнтовані на конкретну поточну подію.

Важливим аспектом тайм-декею є поняття «періоду напівжиття» контенту. Це час, протягом якого матеріал зберігає половину свого початкового рейтингу в системі ранжування. Для новин з високою терміновістю період напівжиття може становити лише кілька годин. Для довгострокового контенту він може сягати кількох тижнів. Розуміння та налаштування цих періодів напівжиття для різних типів контенту є ключовим завданням дизайнерів алгоритмів.

Однак тайм-декей створює певні виклики для журналістики. По-перше, він може стимулювати надмірний акцент на швидкості на шкоду якості. Якщо алгоритм значно фаворизує найновіший контент, редакції можуть почуватися під тиском публікувати матеріали якомога швидше, можливо, жертвуючи ретельною перевіркою фактів або глибиною аналізу. Це особливо проблематично в епоху, коли неперевірена або навіть хибна інформація може швидко поширюватися.

По-друге, жорсткий тайм-декей може несправедливо штрафувати якісну журналістику, яка вимагає часу на створення. Глибокі розслідування, комплексні аналізи, детальні репортажі – все це може бути витіснено з новинних стрічок швидко виробленими, але поверхневими матеріалами. Це створює економічний тиск на новинні організації, які можуть вважати, що інвестиції в якісну журналістику не окупляються з точки зору видимості та залученості аудиторії.

Деякі платформи намагаються вирішити ці проблеми через більш витончені підходи до тайм-декею. Наприклад, вони можуть ідентифікувати матеріали, які демонструють сталий інтерес (користувачі продовжують їх читати та ділитися навіть через деякий час після публікації), і застосовувати до них менш агресивне

згасання. Або вони можуть створювати окремі секції для різних типів контенту – швидку стрічку для термінових новин та більш повільну стрічку для аналітики та довгочитів.

Іншим підходом є адаптивний тайм-декей, який враховує індивідуальну поведінку користувача. Якщо людина регулярно читає старіші, але якісні матеріали, алгоритм може застосовувати для неї менш агресивне згасання. Якщо ж користувач переважно взаємодіє з найсвіжішими новинами, система адаптується, показуючи йому переважно щойно опубліковані матеріали.

Технологічно реалізація тайм-декею часто включає створення часової мітки для кожного матеріалу та регулярний перерахунок рейтингів. На великих платформах, де публікуються тисячі матеріалів на годину, це вимагає значних обчислювальних ресурсів. Тому часто використовуються оптимізації, такі як дискретні часові вікна (наприклад, рейтинги оновлюються кожні 15 хвилин, а не безперервно) або пріоритизація перерахунку для найбільш видимих матеріалів.

Балансування між свіжістю та релевантністю залишається однією з найскладніших задач у дизайні новинних стрічок. Ідеальна система повинна забезпечувати, що користувачі отримують найактуальнішу інформацію про події, що розвиваються, але при цьому не пропускають важливі довгострокові матеріали, які можуть мати глибший вплив на їхнє розуміння світу. Різні платформи та організації знаходять різні точки цього балансу, відображаючи свої цінності та розуміння потреб своєї аудиторії.

7.1.4. Diversity injection для уникнення ехо-камер.

Однією з найсерйозніших критик алгоритмічних новинних стрічок є їхня тенденція створювати інформаційні бульбашки та ехо-камери, де користувачі переважно бачать контент, який підтверджує їхні існуючі переконання та інтереси. Це відбувається тому, що алгоритми, оптимізовані на максимізацію залученості, природно схильні показувати людям те, з чим вони, ймовірно, погодяться або що їх зацікавить на основі попередньої поведінки. Однак така фільтрація може мати негативні наслідки як для окремих людей, так і для суспільства в цілому, обмежуючи експозицію до різноманітних поглядів та сприяючи поляризації.

Diversity injection (ін'єкція різноманітності) – це набір технік

та стратегій, спрямованих на навмисне збільшення різноманітності контенту в новинних стрічках користувачів. Мета полягає не в тому, щоб замінити персоналізацію випадковим контентом, а в тому, щоб знайти баланс між релевантністю та експозицією до нового, несподіваного або альтернативного контенту.

Концептуально різноманітність у новинних стрічках може розглядатися в кількох вимірах. Тематична різноманітність означає показ матеріалів з різних сфер – якщо користувач зазвичай читає про технології, йому періодично показують матеріали про культуру, науку, політику чи спорт. Точкова різноманітність стосується представлення різних поглядів на одну й ту саму тему, особливо на контроверсійні питання. Різноманітність джерел передбачає показ матеріалів не лише з улюблених видань користувача, а й з інших авторитетних джерел. Форматна різноманітність означає чергування різних типів контенту – текстових статей, відео, подкастів, інфографіки тощо.

Технічна реалізація diversity injection зазвичай включає модифікацію базового алгоритму ранжування. Найпростіший підхід – це резервування певної частини місць у стрічці для «різноманітного» контенту. Наприклад, з кожних десяти матеріалів вісім можуть бути відібрані за стандартним алгоритмом персоналізації, а два – навмисно вибрані для збільшення різноманітності. Ці позиції можуть заповнюватися матеріалами з категорій, з якими користувач рідко взаємодіє, але які визначені як важливі редакцією або які популярні серед широкої аудиторії.

Більш витончений підхід використовує багатокритеріальну оптимізацію, де алгоритм одночасно максимізує кілька цілей: релевантність для користувача, різноманітність представлених поглядів, покриття важливих тем тощо. Це складніше з обчислювальної точки зору, але дозволяє знайти більш природний баланс. Наприклад, замість механічного резервування позицій, система може плавно регулювати вагу різноманітності залежно від контексту та індивідуальних характеристик користувача.

Важливим питанням є визначення того, що саме вважати «різноманітним» для конкретного користувача. Це вимагає розуміння профілю споживання контенту людини. Якщо хтось читає переважно ліберальні видання, різноманітність може означати показ матеріалів з центристських або консервативних джерел, і навпаки. Однак це

потрібно робити обережно – занадто різкий контраст може призвести до того, що користувач просто проігнорує або негативно відреагує на такий контент, що в довгостроковій перспективі може посилити, а не послабити поляризацію.

Деякі платформи використовують поступовий підхід до розширення горизонтів, показуючи спершу контент, який лише трохи відрізняється від звичайних уподобань користувача, а потім поступово збільшуючи ступінь відмінності, якщо людина позитивно реагує. Це можна порівняти з роллю, яку відіграють алгоритми рекомендацій музики, які допомагають людям відкривати нових виконавців, що схожі на їхніх улюблених, але не ідентичні їм.

Етичний виклик diversity injection полягає в балансуванні між патерналізмом та автономією користувача. З одного боку, є аргумент, що платформи мають відповідальність сприяти інформованому громадянству, що вимагає експозиції до різноманітних джерел та поглядів. З іншого боку, нав'язування контенту, який користувач не запитував і не хоче бачити, може сприйматися як маніпуляція або неповага до його вибору. Різні платформи знаходять різні точки цього балансу.

Деякі організації обирають підхід максимальної прозорості та контролю користувача. Вони можуть явно позначати в стрічці матеріали, які показані для збільшення різноманітності, та пояснювати, чому саме ці матеріали були обрані. Користувачам може надаватися можливість налаштовувати рівень diversity injection або повністю його вимикати. Такий підхід поважає автономію користувача, але може бути менш ефективним, оскільки люди, які найбільше потребують розширення своїх інформаційних горизонтів, найменш ймовірно самостійно оберуть цю опцію.

Дослідження ефективності diversity injection дають змішані результати. Деякі експерименти показують, що експозиція до різноманітних поглядів може зменшити поляризацію та підвищити політичну толерантність. Інші дослідження виявили ефект бумеранга, коли знайомство з протилежними поглядами насправді посилює існуючі переконання, оскільки люди активно шукають способи дискредитувати або відкинути незручну інформацію. Багато залежить від того, як саме представлений різноманітний контент, наскільки він якісний, та від індивідуальних характеристик користувачів.

Журналістські організації, які розробляють власні стрічки та алгоритми рекомендацій, часто інтегрують diversity injection як частину своєї редакційної місії. Вони можуть використовувати поєднання алгоритмічних методів та ручної курації для забезпечення балансу. Наприклад, редактори можуть визначати певні теми або події як критично важливі для суспільства, і система забезпечує, що всі користувачі отримують експозицію до цих матеріалів незалежно від їхніх персоналізованих профілів.

Майбутнє diversity injection, ймовірно, буде більш нюансованим та адаптивним. Системи штучного інтелекту стають здатними краще розуміти не лише що читає користувач, а й чому, які його глибинні цінності та переконання. Це дозволить більш тонко підбирати різноманітний контент, який розширює горизонти, але не відчужує. Крім того, зростає розуміння того, що різноманітність – це не самоціль, а засіб для сприяння інформованому, критичному та відкритому сприйняттю світу, що є фундаментом здорового демократичного суспільства.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання розкривають логіку формування стрічки:

1. Смерть "Головної сторінки": Статистика показує, що прямий трафік на головні сторінки медіа падає роками (Nieman Lab). Чому концепція "One size fits all" (одна верстка для всіх) стала неефективною в епоху мобільного споживання?

2. Пошук без запиту (Query-less Search): Як працює концепція *Google Discover*? Чому ця стрічка вважається "передбачувальним пошуком" (predictive search), адже вона дає відповідь ще до того, як користувач ввів питання в рядок пошуку?

3. Сигнали ранжування (Ranking Signals): Які три ключові фактори (спадщина Facebook EdgeRank) визначають порядок новин у стрічці?

- *Affinity (Близькість):* Як часто ви взаємодіяли з джерелом?
- *Weight (Вага):* Коментар "важчий" за лайк.
- *Time Decay (Часовий розпад):* Чому новина втрачає цінність з кожною годиною, і як алгоритми вирішують проблему "застарілого контенту"?

4. Вічнозелений контент (Evergreen) у динаміці: Чому в персона-

лізованій стрічці ви можете побачити статтю дворічної давнини? У яких випадках алгоритм вирішує, що "стара" стаття є релевантною "тут і зараз" (наприклад, історія про епідемії під час спалаху грипу)?

5. Гібридні моделі: Як сучасні додатки (BBC, NYT) поєднують редакційний вибір (Top Stories — однакові для всіх, важливі для суспільства) та алгоритмічний блок (Recommended for You — щоб утримати увагу)?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають аналізувати поведінку стрічок:

Завдання 1. "Препарування" Google Discover

• Інструмент: Ваш смартфон (додаток Google або свайп вправо на Android).

• Завдання: Проскрольте перші 10 карток. Для кожної визначте ймовірну причину появи (Reasoning):

- *Історія пошуку:* "Я гуглив ціни на iPhone, тому бачу огляд гаджетів".

- *Локація:* "Я в центрі міста, тому бачу новину про затори".

- *Схожа аудиторія:* "Я це не гуглив, але люди мого віку це читають".

• Мета: Зрозуміти, наскільки прозорою є логіка дистрибуції.

Завдання 2. Порівняння: Apple News vs Flipboard vs Google News

• Ситуація: Ви — медіаменеджер, який обирає платформу для дистрибуції.

• Аналіз: Порівняйте три агрегатори за критерієм "Втручання алгоритму".

- Де більше ручного курування (редактори-люди)?

- Де чиста математика?

• Висновок: Яка платформа краще підходить для *breaking news* (термінових новин), а яка — для *longreads* (лонгрідів)?

Завдання 3. Проєктування логіки "Morning Brief"

• Легенда: Ви створюєте новинний додаток. Вам треба спроєктувати алгоритм ранкового дайджесту (5 новин).

• Умова: У вас є користувач, який любить спорт, але в країні почалася війна/вибори.

• Дилема:

- Варіант А: Показати 5 новин спорту (те, що він любить ->

високий CTR).

- Варіант Б: Показати 5 новин про війну (те, що важливо -> громадянський обов'язок).

- Варіант В: Мікс.

• Завдання: Пропишіть пропорцію (ваги) для Варіанту В. (Наприклад: 1 "Must Read" від редактора + 3 "Personal Interest" + 1 "Serendipity").

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Google Search Central. *Discover and your website.* (Офіційна документація про те, як потрапити в стрічку Discover — "святий граль" трафіку).

2. Nieman Lab. *The Homepage is Dead.* (Серія статей про зміну патернів споживання).

3. Axel Bruns. *Gatewatching and News Curation.* (Академічний погляд на перехід від gatekeeping до gatewatching).

4. Facebook News Feed Publisher Guidelines. (Правила гри для медіа в соцмережах).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• Динамічна стрічка (Dynamic Feed): Потік контенту, який постійно оновлюється і формується алгоритмічно для кожного користувача індивідуально, на відміну від статичної верстки газети чи сайту.

• Time Decay (Часовий розпад): Параметр алгоритму ранжування, який зменшує "вагу" (шанс показу) контенту з плином часу. Забезпечує свіжість стрічки.

• Query-less Search (Пошук без запиту): Технологія дистрибуції інформації (як Google Discover), яка передбачає інтереси користувача на основі його минулої поведінки та контексту, не чекаючи від нього активного пошукового запиту.

• Агрегатор новин (News Aggregator): Платформа (Google News, Feedly, Flipboard), яка не створює власний контент, а збирає заголовки з різних джерел і структурує їх для користувача.

• Сигнал (Signal): Будь-яка одиниця інформації, яку алгоритм використовує для прийняття рішення про ранжування (тип пристрою, час доби, історія кліків, швидкість інтернету).

7.2. Персоналізовані email розсилки

Email розсилки залишаються одним із найпотужніших інструментів дистрибуції новинного контенту, незважаючи на зростання соціальних мереж та месенджерів. На відміну від платформ, де алгоритми третіх сторін контролюють видимість контенту, email забезпечує пряму комунікацію з аудиторією. Однак у світі, де середній професіонал отримує понад сотню електронних листів щодня, лише добре персоналізовані розсилки можуть привернути увагу та утримати читачів. Штучний інтелект та машинне навчання революціонізують email маркетинг для новинних організацій, дозволяючи створювати індивідуалізований досвід для кожного підписника.

7.2.1. Сегментація списків розсилки.

Сегментація списків розсилки є фундаментальною практикою, яка визначає, наскільки ефективною буде email комунікація. Замість відправки однакового листа всім підписникам, новинні організації розділяють свою аудиторію на групи зі схожими характеристиками, інтересами або поведінкою, створюючи для кожного сегмента релевантний контент. Традиційна сегментація базувалася на базових демографічних даних та інформації, яку користувачі самостійно надавали під час підписки. Сучасні підходи з використанням штучного інтелекту значно складніші та точніші.

Базова демографічна сегментація включає поділ аудиторії за віком, статтю, географічним розташуванням, рівнем освіти та професією. Ці дані можуть бути отримані безпосередньо від користувачів під час реєстрації або виведені з їхньої поведінки на сайті. Географічна сегментація особливо важлива для новинних організацій, оскільки дозволяє надсилати регіональні новини релевантній аудиторії. Наприклад, підписники з Києва можуть отримувати детальне висвітлення місцевих подій столиці, тоді як мешканці Львова – новини західного регіону.

Професійна сегментація дозволяє таргетувати контент відповідно до індустрії або сфери діяльності читача. Бізнес-видання можуть мати окремі розсилки для фінансистів, технологічних підприємців, маркетологів, HR-фахівців тощо. Кожна група отримує новини, аналітику та інсайти, найбільш релевантні для їхньої професійної діяльності. Це збільшує цінність розсилки та ймовірність того, що

читач відкриє наступний лист.

Поведінкова сегментація аналізує дії користувачів на сайті та у попередніх email кампаніях. Які статті читають підписники? Скільки часу проводять з кожним матеріалом? На які теми клікають найчастіше? Чи діляться вони контентом у соціальних мережах? Які email вони відкривають, а які ігнорують? Ця інформація створює детальний профіль інтересів кожного користувача, що дозволяє формувати високорелевантні сегменти.

Сучасні системи машинного навчання можуть автоматично виявляти приховані сегменти в аудиторії, які не очевидні при ручному аналізі. Алгоритми кластеризації, такі як k-means або DBSCAN, аналізують багатовимірні дані про поведінку користувачів та групують їх у кластери зі схожими патернами споживання контенту. Наприклад, система може виявити сегмент «ранкових читачів аналітики» – людей, які регулярно відкривають email між 6 та 8 ранку та віддають перевагу довгим аналітичним матеріалам над короткими новинними замітками.

RFM-сегментація (Recency, Frequency, Monetary) адаптована з електронної комерції для новинних видань. Recency аналізує, як давно користувач останній раз взаємодіяв з контентом. Frequency показує, як часто він це робить. Monetary для безкоштовних новин замінюється на поглиблену залученість – чи є підписником преміум-контенту, чи робить донати, чи поширює матеріали. На основі цих параметрів можна виділити сегменти: активні читачі, які потребують постійного потоку контенту; неактивні підписники, яких потрібно реактивувати спеціальними кампаніями; VIP-читачі, які заслуговують на ексклюзивний контент.

Психографічна сегментація враховує цінності, переконання та стиль життя читачів. Хоча ці дані складніше отримати, вони можуть бути виведені з аналізу контенту, який споживає користувач. Наприклад, хтось, хто регулярно читає матеріали про кліматичні зміни, екологічну політику та сталий розвиток, ймовірно, цінує екологічну відповідальність. Цій людині можуть бути релевантні email про зелені ініціативи компаній, новини про відновлювану енергетику або розслідування екологічних порушень.

Життєвий цикл підписника також формує важливий принцип сегментації. Нові підписники потребують вступного контенту, який

знайомить їх з виданням, його стилем та найкращими матеріалами. Довгострокові лояльні читачі можуть отримувати більш складний контент та ексклюзивні матеріали. Підписники, які давно не взаємодіяли з розсилкою, потребують стратегій реактивації – можливо, опитування про зміну інтересів або пропозиції змінити частоту розсилок.

Прогнозна сегментація використовує моделі машинного навчання для передбачення майбутньої поведінки. Наприклад, алгоритм може ідентифікувати користувачів, які з високою ймовірністю відпишуться від розсилки в найближчому майсяці, на основі патернів зниження залученості. До цього сегмента можна застосувати превентивні стратегії утримання – персоналізовані пропозиції, опитування задоволеності або тимчасове зменшення частоти розсилок.

Технічно сегментація реалізується через CRM-системи та платформи email маркетингу, які інтегруються з аналітичними інструментами сайту. Дані про поведінку користувачів постійно оновлюються, і сегменти перераховуються, часто щоденно або навіть у реальному часі. Це означає, що підписник може переміщатися між сегментами залежно від зміни його поведінки та інтересів.

Ефективна сегментація вимагає балансу між деталізацією та практичністю. Надто багато сегментів можуть ускладнити управління кампаніями та розпорошити зусилля контент-команди. Надто мало – призведуть до недостатньої персоналізації. Провідні новинні організації зазвичай працюють з кількома рівнями сегментації: широкі сегменти для основних розсилок та більш вузькі для спеціалізованих кампаній.

Важливо також пам'ятати про етику та приватність. Користувачі повинні мати контроль над тим, які дані збираються, та можливість керувати своїми налаштуваннями розсилки. Прозорість щодо використання даних для персоналізації будує довіру та покращує відносини з аудиторією. Деякі видання дозволяють підписникам самостійно обирати теми, які їх цікавлять, поєднуючи експліцитну сегментацію з алгоритмічною.

7.2.2. Динамічний контент у email.

Динамічний контент у email розсилках представляє наступний рівень персоналізації після сегментації. Замість створення різних версій листа для кожного сегмента, динамічний контент дозволяє

мати один шаблон, елементи якого автоматично змінюються для кожного отримувача відповідно до його профілю та поведінки. Це робить можливим персоналізацію на індивідуальному рівні для тисяч або навіть мільйонів підписників одночасно.

Найпростішою формою динамічного контенту є персоналізація імені отримувача в темі листа або привітанні. Хоча це базова техніка, дослідження показують, що навіть такий простий елемент персоналізації може підвищити показник відкриття email на 10-15%. Однак сучасні можливості виходять далеко за межі вставки імені.

Динамічні блоки контенту дозволяють показувати різні статті або розділи різним підписникам в одному й тому самому email. Наприклад, розсилка може мати блок «Рекомендовані для вас», де кожен читач бачить три статті, відібрані алгоритмом спеціально для нього на основі історії читання. Хтось побачить матеріали про технології та стартапи, інша людина – про культуру та мистецтво, третя – про політику та економіку. Всі отримують один і той же email, але з різним наповненням ключових блоків.

Геолокаційна персоналізація автоматично включає місцеві новини релевантні для регіону підписника. Розсилка національного видання може містити універсальний блок з головними новинами та динамічний блок «У вашому регіоні», який автоматично заповнюється новинами з відповідної області. Це особливо цінно для видань з широкою географією аудиторії, оскільки дозволяє кожному читачеві відчувати зв'язок з виданням через локальний контент.

Темпоральна персоналізація враховує час відкриття email. Якщо підписник відкриває ранкову розсилку ввечері, він може побачити оновлений контент з подіями, які відбулися протягом дня, замість тих, що були актуальні вранці. Це досягається через включення в email коду, який завантажує свіжий контент з сервера в момент відкриття, а не в момент відправки.

Поведінкові тригери визначають, який контент показати на основі останніх дій користувача. Якщо читач недавно читав розслідування про корупцію в певній галузі, наступний email може містити динамічний блок з оновленнями або пов'язаними матеріалами з цієї теми. Якщо хтось почав читати довгу статтю на сайті, але не дочитав, email може містити посилання «Продовжити читання» з цим матеріалом.

Персоналізація на основі пристрою адаптує контент залежно від того, де користувач зазвичай читає email. Якщо алгоритм визначає, що підписник переважно відкриває листи на смартфоні, контент може бути оптимізований для мобільного перегляду – коротші анонси, більші кнопки, вертикальні зображення. Для тих, хто читає на комп'ютері, можна включати більше тексту та горизонтальні макети.

Адаптивні рекомендації використовують алгоритми машинного навчання для підбору контенту в реальному часі. Коли система готує email для відправки, вона аналізує всі доступні матеріали, опубліковані за останній період, та для кожного підписника обирає найбільш релевантні на основі його унікального профілю. Це може враховувати сотні факторів: історію читання, час доби, тренди серед схожих користувачів, популярність матеріалу, тематичний баланс попередніх розсилок тощо.

Контентні варіації дозволяють тестувати різні підходи до презентації одного й того ж матеріалу. Для одних читачів стаття може бути представлена з акцентом на емоційну складову, для інших – на фактах та цифрах, для третіх – через призму експертних думок. Алгоритм визначає, який підхід найкраще резонує з кожним сегментом аудиторії.

Технічна реалізація динамічного контенту вимагає інтеграції email платформи з системою управління контентом та аналітичними інструментами. Дані про користувачів мають бути доступні в момент генерації email, а шаблони повинні підтримувати умовну логіку для відображення різних варіантів контенту. Деякі сучасні платформи використовують API для завантаження контенту в реальному часі в момент відкриття листа, що дозволяє показувати найактуальнішу інформацію.

Виклики динамічного контенту включають технічну складність, потребу в значних обсягах якісних даних та ризик помилок у логіці персоналізації. Якщо алгоритм неправильно визначить інтереси користувача, нерелевантний контент може призвести до розчарування та відписок. Тому важливо регулярно тестувати та валідувати системи персоналізації, збирати зворотний зв'язок від підписників та надавати їм можливість коригувати свої налаштування.

Етичні міркування також важливі. Надмірна персоналізація може створювати відчуття, що організація «знає занадто багато» про

читача, що викликає дискомфорт. Балансування між релевантністю та приватністю, прозорість щодо використання даних та надання користувачам контролю над персоналізацією є ключовими принципами відповідального використання динамічного контенту.

7.2.3. Оптимізація теми листа та часу відправки.

Тема листа та час відправки є критичними факторами, що визначають, чи буде email відкритий взагалі. Навіть найкращий контент залишиться непрочитаним, якщо тема не привернула увагу або лист прийшов у невідповідний час. Штучний інтелект революціонізує підходи до оптимізації цих елементів, дозволяючи персоналізувати їх для кожного підписника.

Традиційно новинні організації використовували одну тему для всіх отримувачів, часто обирану редактором на основі інтуїції та досвіду. Сучасні підходи використовують A/B тестування та машинне навчання для створення та відбору найефективніших тем. Системи можуть автоматично генерувати десятки варіантів теми для одного email, тестувати їх на невеликій частині аудиторії та відправляти найуспішніший варіант решті підписників.

Алгоритми аналізують, які елементи тем корелюють з вищими показниками відкриття. Довжина теми має значення – занадто короткі можуть бути неінформативними, занадто довгі обрізаються на мобільних пристроях. Дослідження показують, що оптимальна довжина зазвичай становить 40-50 символів. Використання цифр, питальних знаків, емодзі, термінів терміновості («сьогодні», «зараз», «термінові новини») або персоналізації може впливати на результати, але ефективність цих елементів варіюється залежно від аудиторії та контексту.

Машинне навчання може передбачати, яка тема буде найефективнішою для конкретного підписника на основі його історії взаємодії з попередніми email. Деякі люди краще реагують на прямолінійні, інформативні теми («5 головних новин четверга»), інші – на більш інтригуючі або емоційні («Чому ця політична криза може змінити все»). Алгоритм може автоматично класифікувати підписників за їхніми перевагами та відправляти кожному відповідну версію теми.

Генерація тем на основі ШІ використовує моделі обробки природної мови для створення варіантів, які балансують

інформативність, привабливість та відповідність редакційному стилю видання. Системи можуть аналізувати тисячі попередніх успішних тем, вивчати патерни та генерувати нові варіації. GPT-подібні моделі здатні створювати теми, які звучать природно та професійно, враховуючи контекст конкретного матеріалу та аудиторії.

Емоційний аналіз тем визначає, який емоційний тон резонує з різними сегментами аудиторії. Деякі читачі краще реагують на нейтральні, фактологічні теми, інші – на ті, що викликають любопитство, занепокоєння або оптимізм. Алгоритми можуть автоматично налаштовувати емоційний тон теми відповідно до профілю підписника.

Оптимізація часу відправки є не менш важливою. Традиційно новинні розсилки відправлялися в один фіксований час для всіх підписників, зазвичай рано вранці. Однак різні люди мають різні ритми споживання інформації. Хтось перевіряє email о 6 ранку перед виходом на роботу, хтось – під час обідньої перерви, хтось – ввечері перед сном.

Аналіз індивідуальних патернів поведінки дозволяє визначити оптимальний час відправки для кожного підписника. Алгоритми відстежують, коли людина зазвичай відкриває email, коли найактивніша на сайті, коли ймовірність взаємодії з контентом найвища. На основі цих даних кожен підписник може отримувати розсилку у свій персональний оптимальний час, що може відрізнятися на години від інших читачів.

Predictive send time optimization використовує машинне навчання для прогнозування не лише коли користувач відкриє email, а коли він найімовірніше зробить цільову дію – прочитає статтю, поділиться нею, підпишеться на преміум або зробить донат. Це може не співпадати з часом найвищої ймовірності відкриття. Наприклад, хтось може відкривати всі email о 7 ранку, але читати довгі матеріали лише ввечері.

Часові зони та географія також враховуються. Для глобальних видань критично важливо відправляти email у відповідний час для кожного регіону. Користувач у Києві та користувач у Сан-Франциско повинні отримати ранкову розсилку вранці свого часу, а не одночасно за UTC.

Контекстуальні фактори, такі як день тижня, свята, важливі

події, можуть впливати на оптимальний час відправки. Дослідження показують, що показники відкриття можуть значно відрізнятися між буднями та вихідними, між звичайними днями та періодами великих новинних подій. Алгоритми враховують ці патерни при плануванні розсилок.

Технічна реалізація персоналізованого часу відправки вимагає складної інфраструктури, здатної обробляти відправку мільйонів email не одночасно, а протягом розширеного вікна часу. Це створює виклики для термінових новинних розсилок, де важлива синхронна доставка. Деякі організації використовують гібридний підхід: термінові новини відправляються всім одночасно, а планові розсилки – у персоналізований час.

Важливо також враховувати частоту розсилок. Занадто багато email можуть призвести до втоми та відписок, занадто мало – до втрати зв'язку з аудиторією. Алгоритми можуть автоматично налаштовувати частоту для різних користувачів на основі їхньої історії взаємодії. Активні читачі можуть отримувати щоденні розсилки, менш активні – тижневі дайджести.

Неперервна оптимізація через машинне навчання означає, що система постійно вчиться з кожної відправленої кампанії. Які теми працювали найкраще? Який час виявився оптимальним? Як змінюється поведінка користувачів з часом? Ці інсайти інтегруються назад у модель для покращення майбутніх кампаній, створюючи цикл постійного вдосконалення.

7.2.4. Triggered campaigns на основі поведінки.

Triggered campaigns або тригерні кампанії представляють найдинамічнішу форму email маркетингу, де розсилки автоматично ініціюються конкретними діями або подіями, пов'язаними з користувачем. На відміну від запланованих масових розсилок, тригерні email відправляються індивідуально у відповідь на поведінку конкретного підписника, що робить їх надзвичайно релевантними та своєчасними.

Базові тригери життєвого циклу включають welcome-серії для нових підписників. Коли хтось вперше підписується на розсилку, автоматично запускається серія листів, які знайомлять його з виданням, його основними рубриками, стилем матеріалів та найкращими публікаціями. Перший лист може прийти миттєво з

подякою та короткою інформацією про видання. Другий через кілька днів може представити різні типи контенту. Третій – запропонувати налаштувати персональні вподобання. Така серія значно підвищує залученість нових підписників та знижує ранні відписки.

Тригери взаємодії з контентом реагують на те, як користувач споживає матеріали на сайті. Якщо читач почав довгу статтю або розслідування, але не дочитав, через певний час йому може прийти email із нагадуванням та посиланням для продовження читання. Якщо хтось прочитав кілька матеріалів з певної серії або на певну тему, автоматично надсилається email з рекомендаціями подібного контенту або новими матеріалами з цієї теми.

Milestone-тригери відзначають важливі моменти в історії відносин читача з виданням. Email на ювілей підписки (рік з моменту реєстрації) може містити персоналізовану статистику – скільки статей прочитано, які теми були найпопулярнішими, найдовший матеріал, який людина осилила. Це створює емоційний зв'язок та нагадує про цінність, яку видання надає читачеві.

Тригери бездіяльності виявляють підписників, які перестали взаємодіяти з розсилками або відвідувати сайт. Якщо користувач не відкрив жодного email протягом місяця або не відвідував сайт певний час, запускається серія реактивації. Перший лист може запитати, чи все гаразд, чи змінилися інтереси. Другий може пропонувати найкращі матеріали за пропущений період. Третій може давати опцію змінити частоту розсилок або вибрати інші теми. Якщо й після цього немає реакції, система може автоматично перевести користувача в режим мінімальної комунікації, щоб не шкодити репутації відправника.

Event-based тригери реагують на зовнішні події. Якщо відбувається велика новинна подія в темі, яка цікавить конкретного користувача, він може автоматично отримати сповіщення з посиланням на відповідне висвітлення. Наприклад, хтось, хто регулярно читає про конкретну компанію, отримає email, коли ця компанія оприлюднить фінансові результати або зробить важливу заяву.

Тригери на основі соціальних дій активуються, коли контент набуває віральності або стає трендовим. Якщо стаття раптово отримує багато переглядів та поширень у соціальних мережах, алгоритм може автоматично відправити email підписникам, які цікавляться цією темою, але ще не бачили матеріал, з повідомленням

«Всі обговорюють цю статтю».

Персоналізовані news alerts дозволяють користувачам налаштувати ключові слова або теми, про які вони хочуть отримувати миттєві сповіщення. Коли видання публікує матеріал, що містить ці ключові слова, автоматично відправляється тригерний email. Це особливо цінно для професіоналів, які відстежують конкретні компанії, технології, законодавчі ініціативи або інші спеціалізовані теми.

Conversion-орієнтовані тригери спрямовані на перетворення безкоштовних читачів на платних підписників. Якщо користувач досяг ліміту безкоштовних статей на місяць (модель paywall), автоматично надсилається email з пропозицією передплати та поясненням її переваг. Якщо хтось регулярно читає преміум-контент в пробний період, за кілька днів до його закінчення приходить нагадування з спеціальною пропозицією.

Cart abandonment для журналістики адаптується як abandonment пропозиції передплати. Якщо користувач почав процес оформлення передплати, але не завершив його, через певний час приходить email із нагадуванням та можливо спеціальною пропозицією або допомогою з технічними питаннями, які могли завадити завершити покупку.

Технічна архітектура тригерних кампаній вимагає складної системи моніторингу подій та автоматизації. Кожна дія користувача на сайті або в додатку генерує події, які потенційно можуть ініціювати email. Система повинна в реальному часі оцінювати, чи відповідає подія умовам будь-якого з налаштованих тригерів, чи не отримав користувач вже забагато email за короткий період, чи релевантний цей конкретний email в поточному контексті.

Управління частотою та пріоритетом критично важливе для тригерних кампаній. Якщо користувач одночасно відповідає умовам кількох тригерів, система повинна вирішити, який email відправити, чи об'єднати кілька повідомлень в одне, чи відкласти менш терміновий тригер. Правила частоти забезпечують, що підписник не отримає занадто багато листів за короткий період, навіть якщо його поведінка активує багато тригерів.

Машинне навчання оптимізує тригерні кампанії, аналізуючи, які тригери найефективніші для різних сегментів аудиторії. Скільки часу варто чекати після певної дії перед відправкою тригерного email? Який контент включити? Яку тему використати? Алгоритми

навчаються на історичних даних і постійно налаштовують параметри для максимізації цільових метрик – відкриттів, кліків, конверсій, утримання.

Етичні міркування особливо важливі для тригерних кампаній, оскільки вони можуть відчуватися як надто інтрузивні, якщо виконані невдало. Баланс між корисністю та нав'язливістю, прозорість щодо того, чому користувач отримав конкретний email, та легкість керування налаштуваннями тригерів є ключовими для довіри та позитивного сприйняття.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання акцентують на відмінності спаму від корисної комунікації:

1. "Batch and Blast" vs. Персоналізація: Чому старий метод "Відправити один лист усій базі" (Batch and Blast) більше не працює? Як це впливає на репутацію домену та потрапляння в спам?

2. Динамічний контент (Dynamic Content): Як працює технологія, що дозволяє відправити 100 000 листів, але всередині кожного будуть різні блоки новин (наприклад, фанат спорту побачить футбол, а фанат культури — оперу), використовуючи лише один HTML-шаблон?

3. Тригерні розсилки (Triggered Emails): Чим тригерна розсилка (наприклад, "Ви не дочитали статтю") відрізняється від регулярного дайджесту? Чому Open Rate у тригерних листів зазвичай у 2-3 рази вищий?

4. Тема листа (Subject Line): Як використання імені ("Олено, для вас...") або даних про поведінку ("Знижка на те, що ви переглядали") у темі листа впливає на показник відкриття (Open Rate)?

5. Гігієна бази: Чому важливо видаляти підписників, які не відкривають листи (Inactivists)? Як "мертві душі" тягнуть донизу Deliverability (доставлення) для всієї розсилки?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають сценарному плануванню комунікації:

Завдання 1. Проєктування "Welcome-ланцюжка" (Onboarding Drip Campaign)

• Ситуація: Користувач підписався на платну версію вашого онлайн-журналу.

• Завдання: Розпишіть серію з 3-х автоматичних листів:

- *Лист 1 (Одразу): (Привітання + "Як користуватися сайтом").*

- *Лист 2 (Через 3 дні):* (Персональна підбірка "Найкращі статті розділу, який ви обрали при реєстрації").

- *Лист 3 (Через 7 днів):* (Опитування "Чи все вам подобається?" або пропозиція завантажити мобільний додаток).

• Мета: Зрозуміти логіку утримання (Retention) через email.

Завдання 2. Логіка Динамічного блоку

• Інструмент: Умовно (папір/схема).

• Завдання: Створіть структуру щотижневого дайджесту новин. У ньому є один "Змінний блок". Пропишіть умови (IF/THEN):

- *IF* Subscriber_Location = "Kyiv" *THEN* Show "Афіша подій Києва".

- *IF* Subscriber_Location = "Lviv" *THEN* Show "Афіша подій Львова".

- *ELSE* (якщо місто невідоме) *THEN* Show "Топ подій України онлайн".

Завдання 3. A/B тестування теми листа

• Контекст: Ви відправляєте розсилку про економічну кризу.

• Завдання: Сформулюйте дві гіпотези для Subject Line:

- *Варіант А (Раціональний):* "Огляд ринку: інфляція досягла 10%".

- *Варіант Б (Емоційний/Персоналізований):* "[Ім'я], ось як інфляція вплине на ваш гаманець".

• Метрика: Який показник ви будете порівнювати — Open Rate (відкриття) чи Click Rate (кліки всередині)? (Відповідь: Open Rate залежить від теми, Click Rate — від контенту всередині).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Chad White. *Email Marketing Rules.* (Настільна книга email-маркетолога. Охоплює все: від дизайну до законодавства).

2. Litmus Blog. (Провідний ресурс про дизайн та верстку листів, тренди адаптивності).

3. Really Good Emails. (Колекція найкращих прикладів дизайну та копірайтингу розсилок — для натхнення).

4. Mailchimp / Substack Resources. (Гайди для початківців про сегментацію).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• Drip Campaign (Крапельна кампанія): Серія автоматизованих листів, які надсилаються людям, що виконали певну дію (підписалися, купили), у визначеному порядку та з певними інтервалами.

• Open Rate (OR): Відсоток отримувачів, які відкрили лист. Головний індикатор привабливості теми листа та довіри до відправника.

• CTR (Click-Through Rate): Відсоток отримувачів, які клікнули хоча б на одне посилання всередині листа. Показник якості контенту.

• Unsubscribe Rate: Відсоток людей, які відписалися від розсилки після отримання листа. Якщо він > 0.5-1%, це сигнал тривоги.

• Hard Bounce vs Soft Bounce:

- *Hard Bounce:* Лист не доставлено, бо адреса не існує (треба видаляти з бази).

- *Soft Bounce:* Тимчасова проблема (переповнена скринька).

• Double Opt-in: Метод підписки, коли користувач має підтвердити свою пошту, клікнувши на посилання в першому листі. Юридичний стандарт якості бази.

7.3. Адаптивні вебсайти та додатки

Сучасні новинні вебсайти та мобільні додатки еволюціонували від статичних платформ, які показують однаковий контент усім відвідувачам, до динамічних, інтелектуальних систем, що адаптуються до кожного користувача. Штучний інтелект дозволяє створювати персоналізований досвід, який враховує інтереси, поведінку, контекст та потреби кожного читача, перетворюючи новинний сайт на індивідуального інформаційного помічника.

7.3.1. Персоналізація головної сторінки.

Головна сторінка новинного сайту традиційно була витриною редакційних пріоритетів – редактори вручну обирали найважливіші матеріали та розміщували їх у певному порядку, створюючи єдиний наратив для всіх відвідувачів. Цей підхід відображав класичну роль журналістики як gatekeepers – професіоналів, які визначають, що є важливим для суспільства. Однак у цифрову епоху, коли кожен користувач має унікальні інтереси та потреби, персоналізована головна сторінка стає стандартом, що дозволяє поєднувати редакційне керівництво з індивідуальною релевантністю.

Гібридна модель персоналізації є найпоширенішим підходом серед якісних новинних видань. Певні елементи головної сторінки залишаються універсальними для всіх користувачів – зазвичай це топ-1 або топ-3 найважливіших новин дня, які редакція вважає критично важливими для всіх громадян. Це гарантує, що навіть у персоналізованому середовищі всі читачі отримують експозицію до найзначущіших подій. Решта сторінки динамічно адаптується до кожного користувача, показуючи матеріали, які найімовірніше будуть для нього релевантними.

Алгоритми персоналізації аналізують багатовимірний профіль користувача для формування індивідуальної головної сторінки. Історія читання є базовим сигналом – які теми цікавили людину раніше, скільки часу вона проводила з різними матеріалами, які статті дочитувала до кінця, які лишала після кількох абзаців. Якщо хтось регулярно читає матеріали про технології та бізнес, головна сторінка автоматично підвищує видимість нових публікацій з цих категорій.

Поведінкові патерни доповнюють явні інтереси. Алгоритм помічає не лише що читає користувач, а й як – чи переходить

він до статті з головної сторінки чи через пошук, чи взаємодіє з мультимедійним контентом, чи віддає перевагу коротким замітках чи довгим розслідуванням, чи читає переважно з мобільного чи з комп'ютера. Ці патерни інформують не лише відбір контенту, а й його презентацію.

Контекстуальні фактори визначають, що саме показувати в конкретний момент. Час доби має значення – вранці люди можуть шукати швидкий огляд головних подій, ввечері – більш глибокі аналітичні матеріали. День тижня також впливає: у вихідні зростає інтерес до розважального та культурного контенту. Поточні події можуть тимчасово змінювати пріоритети – під час виборів політичні новини домінують навіть для тих, хто зазвичай їх уникає.

Collaborative filtering використовує інформацію про поведінку схожих користувачів для рекомендацій. Якщо люди зі схожим профілем інтересів активно читають певну статтю, вона отримує вищий пріоритет для показу цьому користувачу, навіть якщо він сам ще не взаємодіяв з цією темою. Це допомагає виявляти приховані інтереси та розширювати інформаційні горизонти природним чином.

Балансування свіжості та релевантності є постійним викликом. Новини за визначенням мають бути новими, але найсвіжіші матеріали можуть бути не найрелевантнішими для конкретного користувача. Системи персоналізації використовують складні формули, які враховують і час публікації, і відповідність інтересам. Стаття, опублікована годину тому на улюблену тему користувача, може отримати перевагу над щойно опублікованою статтею на малоцікаву тему.

Diversity injection інтегрується в персоналізацію головної сторінки для уникнення інформаційних бульбашок. Навіть якщо система добре знає інтереси користувача, вона навмисно включає певний відсоток матеріалів з інших категорій, різних поглядів або несподіваних тем. Це може бути реалізовано через виділені секції типу "Варто прочитати" або "Інша перспектива", де представлений контент, що виходить за межі звичайних уподобань.

Візуальна персоналізація адаптує не лише вибір контенту, а й спосіб його презентації. Для одного користувача стаття може відображатися з великим зображенням та коротким анонсом, для іншого – з детальнішим текстом та меншою картинкою, залежно

від того, що краще привертає їхню увагу. Розмір шрифтів, щільність інформації, кількість матеріалів на екрані можуть автоматично налаштовуватися відповідно до вподобань та можливостей користувача.

Progressive personalization поступово вдосконалює головну сторінку в міру накопичення даних про користувача. Для нового відвідувача показується базовий редакційний макет з найважливішими новинами. З кожним візитом, кожною прочитаною статтею система краще розуміє інтереси та налаштовує контент все точніше. Це створює плавний перехід від універсального до персоналізованого досвіду.

Real-time персоналізація означає, що головна сторінка може змінюватися навіть під час одного візиту. Якщо користувач прочитав кілька статей про конкретну тему, наступне оновлення головної сторінки може вже враховувати цей новий інтерес. Такий динамізм вимагає значних обчислювальних ресурсів, але створює відчуття "живого" сайту, який реагує на користувача.

Технічна реалізація персоналізованої головної сторінки використовує серверний та клієнтський рендеринг. Серверна персоналізація генерує унікальну версію сторінки для кожного користувача на основі його профілю, зберігаючи результат у кеші для швидкого завантаження. Клієнтська персоналізація завантажує базову структуру, а потім через JavaScript динамічно підвантажує персоналізований контент. Гібридні підходи поєднують обидва методи для балансу між швидкістю та персоналізацією.

A/B тестування постійно проводиться для оптимізації алгоритмів персоналізації. Різні варіанти логіки ранжування, макетів, способів презентації контенту тестуються на сегментах аудиторії для визначення, що краще працює. Метрики включають не лише кліки та час на сайті, а й довгострокове утримання, глибину залученості, конверсії в передплатників.

Етичні міркування вимагають прозорості та контролю. Користувачі повинні розуміти, що сторінка персоналізована, і мати можливість впливати на цей процес. Деякі видання надають налаштування, де можна явно вказати цікаві теми, відключити певні категорії контенту або повністю вимкнути персоналізацію, повернувшись до редакційної версії головної сторінки. Такий

контроль будує довіру та задоволеність користувачів.

7.3.2. Динамічні виджети та модулі.

Сучасна цифрова журналістика вийшла за рамки простого публікування лінійного тексту; вона все більше нагадує динамічне середовище, що реагує на присутність та інтереси користувача. Серцем цього середовища стали динамічні віджети — інтерактивні блоки на веб-сторінці або в додатку, вміст яких не є статичним, а генерується та підлаштовується автоматично для кожного користувача або групи в момент завантаження сторінки. На відміну від традиційного підходу, де редактор вручну закріплює певний контент (наприклад, «головну новину дня»), динамічні віджети працюють за принципом резервування місця, яке заповнюється найбільш релевантними даними, отриманими з алгоритмічних систем у реальному часі. Ця технологія є ключовим інструментом для побудови того, що називають «архітектурою уваги» — стратегічного проектування цифрового простору, спрямованого на максимальне залучення та утримання аудиторії.

Функціональність та цілі динамічних віджетів розрізняються залежно від їх логіки роботи. Один з найпоширеніших типів — це віджет «Читайте далі» (Read Next), який зазвичай розміщується під статтею або в її середині (in-article). Його алгоритм поєднує контекстуальну схожість (теми, ключові слова) з особистими інтересами користувача, вивченими з історії переглядів. Головна мета такого віджета — рециркуляція (recirculation), тобто переспрямування уваги читача на наступний матеріал, щоб продовжити сесію на сайті та запобігти швидкому закриттю сторінки. Окремим випадком є віджет «Популярне зараз» (Most Popular), що часто займає бокову панель (сайдбар). Його логіка заснована на трендах, проте в сучасній динамічній версії він рідко показує єдиний «топ сайту». Натомість, алгоритми можуть персоналізувати цей список, відображаючи «найпопулярніше в вашому регіоні» або «що обговорюють колеги з вашої галузі». Ще глибший рівень персоналізації пропонує віджет «Рекомендовано для вас» (Recommended for You), зазвичай розміщений на головній сторінці. Він аналізує довгострокову історію поведінки користувача, прагнучи передбачити його смаки подібно до того, як це робить Netflix або Spotify. Окремою категорією є товарні каруселі в e-commerce та медіа, де логіка будується на контексті

контенту: стаття про біг може супроводжуватися віджетом з пропозицією кросівок або спортивного одягу, що є основою нативної (органічної) монетизації контенту.

Ефективність динамічного віджета критично залежить від його розміщення на сторінці — так званих «зон уваги» (The Real Estate of Attention). Найпрестижнішою є позиція «Above the Fold» (перший екран без скролу). Вона забезпечує максимальну видимість, але ризикує викликати роздратування, якщо контент виявиться нерелевантним. Тут зазвичай розміщують блоки з терміновими новинами або найважливішими оновленнями. Одна з найефективніших зон для утримання уваги — це «In-Stream» (в потоці контенту), наприклад, між третім і четвертим абзацем статті. Користувач вже залучений у читання, що забезпечує високий рівень клікабельності (CTR) на пропоновані посилання. Традиційне розміщення «Bottom of Post» (підвал статті) працює для віджетів «Схожі новини», але його ефективність обмежена тим, що до кінця матеріалу дочитує лише 20-30% відвідувачів. Технічно складним, але ефективним рішенням є «Sticky Sidebar» (липка бокова панель) — віджет, який «пливе» за користувачем під час скролу, завжди залишаючись у полі зору.

Технологічною основою таких систем є асинхронне завантаження (Lazy Loading) та динамічна підміна контенту. Щоб не уповільнювати загальне завантаження сторінки, спочатку відображається основний текст статті. Паралельно, через невелику затримку (наприклад, 0.5 секунди), фоновий скрипт здійснює запит до платформи даних про клієнта (Customer Data Platform, CDP) — централізованої бази, що акумулює всю інформацію про користувача. Запит формулюється просто: «Хто зараз читає цю статтю?». Отримавши відповідь (наприклад, «це любитель футболу зі статусом преміум-підписника»), система відмальовує віджет, заповнюючи його контентом, підібраним саме під цей профіль — наприклад, останніми новинами про трансфери або анонсом матчу.

Найвищим проявом інтелекту в цій архітектурі є «розумний» динамічний пейвол (Dynamic Paywall Widget). Це еволюція примітивного «замка», що блокує контент. Такий віджет аналізує профіль користувача та контекст його поведінки, щоб згенерувати персоналізований заклик до дії. Наприклад, фанатові спорту він може запропонувати: «Підпишіться, щоб отримати ексклюзивний

репортаж зі стадіону перед фіналом», тоді як інвестору представити аргумент: «Доступ до повного аналізу ринку від наших експертів — у преміум-підписці». Більш того, деякі системи можуть динамічно змінювати умови, пропонуючи обмежену тимчасову знижку користувачам, які виявляють ознаки сумніву або ймовірності відписки (churn risk). Таким чином, динамічні віджети перетворюються з пасивних елементів інтерфейсу на активні інструменти комунікації, монетизації та утримання аудиторії, будучи ключовими компонентами архітектури, що не просто показує контент, а веде діалог з кожним окремим читачем.

7.3.3. Персоналізована навігація.

Традиційна архітектура новинного сайту ґрунтується на статичній, єдиній для всіх користувачів структурі деревоподібних категорій (Політика, Економіка, Спорт). Персоналізація навігації ламає цю жорстку, універсальну модель, перетворюючи цифровий продукт із пасивного каталогу на «живий організм», інтерфейс якого динамічно адаптується до індивідуальних інтересів та поведінкових паттернів кожного читача. Ця трансформація спрямована на те, щоб не тільки показувати релевантний контент, але й робити шлях до нього інтуїтивним і швидким, по суті, намагаючись передбачити бажання користувача ще до їх явного формулювання.

Одним із найпомітніших проявів цього підходу є адаптивне меню (Adaptive Menu). Проблема класичної навігації великих порталів полягає в їх перевантаженості: загальний список із 50 рубрик однаково складний для орієнтування як для професійного аналітика, так і для пересічного читача. Алгоритмічне рішення полягає в аналізі індивідуальної поведінки — які розділи користувач відвідує найчастіше — та автоматичному винесенні цих пунктів на найбільш видимі позиції (наприклад, у «гарячу зону» шапки сайту). Таким чином, для фаната футболу головне меню може набути вигляду: *Футбол* | *Спорт* | *Новини* | *Політика…*, тоді як для інвестора пріоритети змістяться: *Бізнес* | *Котирування* | *Економіка* | *Політика….* Однак така динаміка несе ризик порушення м'язової пам'яті та узгодженості UX: якщо елементи інтерфейсу постійно «стрибають», користувач втрачає орієнтацію. Тому передові практики передбачають зміну лише порядку або видимості певних пунктів при збереженні загальної структури та набору ключових,

загальнодоступних розділів.

Більш радикальним і повсякденним проявом персоналізації стала впроваджена під впливом платформ на кшталт TikTok та Apple News вкладка «Для вас» (The «For You» Tab). Вона формалізує поділ контентних потоків у додатках та на сайтах провідних медіа. Перший потік — «Headlines» або «Top Stories» — залишається сферою редакційного вибору та суспільного обов'язку медіа: тут публікуються новини, які важливо знати всім (військові події, вибори, масштабні катастрофи). Другий потік — «For You» або «My News» — повністю формується алгоритмом на основі інтересів конкретного користувача (наприклад, технології, здоровий спосіб життя, локальні новини його району). Стратегічно саме ця вкладка все частіше стає стартовою сторінкою для зареєстрованих користувачів, що значно збільшує глибину та тривалість сесії завдяки відчуттю максимальної релевантності.

Персоналізація проникає навіть у таку, здавалося б, об'єктивну функцію, як пошук (Personalized Search). Коли два користувачі вводять у рядок пошуку однаковий запит, вони можуть мати на увазі зовсім різні речі. Наприклад, запит «Динамо» для користувача А, який регулярно читає про спорт, дасть у результаті новини про футбольний клуб «Динамо Київ». Для користувача Б, який вивчає історію техніки, той самий запит пріоритетно відобразить матеріали про винахід динамо-машини. Алгоритми персоналізованого ранжування аналізують профіль інтересів, щоб піднімати вгору списку результатів найбільш релевантні статті, навіть якщо вони не є найсвіжішими хронологічно.

Окремої уваги заслуговує візуальна навігація у форматі «Stories» (кружечки вгорі екрану), який став мобільним стандартом. Його динаміка також підкоряється персоналізації: порядок розташування кружечків (найчастіше за темами чи авторськими колонками) не є хронологічним. Якщо користувач часто переглядає відеоогляди техніки, кружечок «Техно» займе перше місце. І навпаки, тема, яку він систематично ігнорує (наприклад, «Шоу-бізнес»), буде автоматично зсунута в кінець рядка або прихована, щоб не захаращувати інтерфейс. Таким чином, персоналізована навігація перестає бути лише технічним доповненням, перетворюючись на ключовий принцип побудови взаємодії з аудиторією, де кожен

елемент інтерфейсу прагне стати продовженням думки та інтересів конкретної людини.

7.3.4. Геотаргетинг та локалізація.

В умовах глобалізації цифрового простору ефективна комунікація вимагає не лише доступності контенту для міжнародної аудиторії, а й його глибокої інтеграції в локальний контекст. Геотаргетинг виступає ключовою технологією для досягнення цієї мети, представляючи собою метод автоматичної доставки та адаптації контенту на основі фізичного розташування користувача. Цей процес виходить далеко за рамки простого перекладу мови; це комплексна адаптація сенсу, формату та релевантності інформації до специфіки місцевого середовища, культурних особливостей та практичних потреб.

Точність та технології геолокації формують базис ефективного геотаргетингу, пропонуючи різні рівні деталізації. Найпоширенішим методом є визначення місцезнаходження за IP-адресою пристрою, що дозволяє з розумною точністю ідентифікувати країну та місто користувача. Ця технологія лежить в основі автоматичного перенаправлення на національні версії сайтів або відображення регіональних новин, незважаючи на певну похибку в точності. Для досягнення гіперлокального рівня необхідний доступ до GPS-даних мобільного пристрою, який надається лише за явного дозволу користувача. Така точність (до вулиці чи конкретної точки) відкриває можливості для гіперлокальної журналістики, що доставляє новини, релевантні саме для району або мікрорайону перебування людини. Найвищу точність (до декількох метрів) забезпечують технології на основі Wi-Fi-зондів або Bluetooth Beacon, що особливо цінно всередині великих приміщень: торгових центрів, стадіонів, аеропортів. Це дозволяє створювати найбільш контекстно-залежний досвід, наприклад, пропонуючи контент, прив'язаний до конкретного магазину чи виставкового стенду.

На основі точних геоданих будується потужна концепція «геопаркану» (Geofencing) — створення віртуального периметра навколо реальної географічної зони. Механізм працює за принципом тригера: коли смартфон користувача перетинає невидиму межу, система автоматично ініціює певну дію. У медіа ця технологія знаходить різнобічне застосування. Під час масових заходів (концертів, спортивних змагань) додаток може надсилати пуш-

повідомлення з актуальною інформацією про розклад, схеми проходу або екстрені оголошення безпосередньо при наближенні до місця події. У кризових ситуаціях, таких як пожежа, повінь або надзвичайний стан, геопаркан дозволяє надсилати таргетовані сповіщення з інструкціями з евакуації саме тим, хто опинився в небезпечній зоні. Розслідувальна журналістика також використовує цю технологію для персоналізації важливих матеріалів. Яскравим прикладом є проєкти, де читач, відкриваючи статтю про екологічне забруднення, отримує додаткову інформацію про те, наскільки близько від його фактичного місцезнаходження знаходяться виявлені джерела шкідливих викидів, що робить розслідування особисто значущим.

Сутністю глибокої локалізації є адаптація контенту, що охоплює куди більше, ніж лише мовний переклад. Це комплексна зміна контекстних даних, спрямована на створення відчуття природності та безпосередньої релевантності для користувача. Автоматично змінюється валюта (ціна товару в статті відображається в гривнях для українського користувача та в євро — для німецького), одиниці виміру (кілометри замість миль, градуси Цельсія замість Фаренгейта) та формати дат. Критично важливим є адаптація контенту за тематикою: глобальний спортивний портал може показувати на головній сторінці користувачеві з Італії новини про Серію А, користувачеві з України — про виступи українських гравців у європейських чемпіонатах, а користувачеві з Польщі — про успіхи місцевих команд та зірок.

Однак активне використання геоданих супроводжується суттєвими ризиками та викликами у сфері приватності. Технічним обмеженням є поширене використання VPN-сервісів, які «підміняють» IP-адресу, змушуючи систему визначати недостовірне місцезнаходження (наприклад, користувач у Києві може отримувати контент для Амстердама). Найбільшою проблемою є психологічний ефект «моторошності» (Creepy Factor). Коли медіа-додаток демонструє надто точне знання про місце перебування людини («Ви зараз біля певної аптеки на Хрещатику?»), це викликає тривогу та відчуття стеження, що призводить до втрати довіри. Саме тому більшість користувачів обережно налаштовані та часто забороняють новинним сайтам доступ до точних GPS-даних. Як компроміс, медіа-організації здебільшого покладаються на менш точний, але менш «нав'язливий» IP-таргетинг, знаходячи баланс між корисністю

персоналізації та недоторканністю приватного життя. Таким чином, успішна локалізація вимагає не лише технічної досконалості, але й глибокого розуміння етичних меж та потреби користувача відчувати контроль над своїми даними.

1. КОНТРОЛЬНІ ПИТАННЯ (для самоконтролю).

Ці питання розмежовують технічні підходи до дизайну:

1. Responsive vs. Adaptive: У чому різниця між RWD (Responsive Web Design — "гумовий" макет, що розтягується) та AWD (Adaptive Web Design — сервер визначає пристрій і віддає спеціальну версію сайту)? Чому для складних медіа порталів іноді краще мати окрему мобільну версію (https://www.google.com/search?q=m.site.com), ніж просто адаптивну?

2. Mobile-First Indexing: Як зміна алгоритмів Google (який тепер оцінює сайт насамперед за його мобільною версією) вплинула на пріоритети розробки? Чому "красивий десктоп" більше не рятує SEO, якщо мобільна версія повільна?

3. PWA (Progressive Web App): Що це за технологія, що дозволяє звичайному сайту працювати як додаток (зберігати іконку на екрані, надсилати пуші, працювати офлайн), але не вимагає завантаження з App Store? Чому медіа (наприклад, Forbes, Twitter Lite) масово переходять на PWA?

4. Ергономіка "Зони великого пальця" (Thumb Zone): Як змінилися діагоналі смартфонів за останні 10 років і як це вплинуло на розташування меню? Чому кнопка "Гамбургер" (меню) у верхньому лівому куті — це тепер поганий тон UX-дизайну?

5. Доступність (Accessibility/A11y): Чому адаптивність включає не лише розмір екрану, а й можливість споживання контенту людьми з вадами зору (Screen Readers)? Що таке семантична верстка і навіщо потрібні атрибути alt для картинок?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають оцінювати зручність інтерфейсів:

Завдання 1. Аудит "Зони великого пальця" (Thumb Zone Mapping)

• Інструмент: Ваш смартфон + скріншот сайту популярного українського медіа.

• Дія: Тримаючи телефон однією рукою, спробуйте виконати

ключові дії:

1. Відкрити меню рубрик.

2. Поширити статтю у Facebook.

3 Перейти на наступну новину.

• Аналіз: Які з цих зон є "зеленими" (легко дістати), а які "червоними" (треба перехоплювати телефон або задіяти другу руку)?

• Висновок: Чи адаптований цей сайт для читання в транспорті (однією рукою)?

Завдання 2. Перевірка швидкості (Core Web Vitals)

• Інструмент: Google PageSpeed Insights (або Lighthouse у Chrome DevTools).

• Об'єкт: Порівняйте сайт BBC та сайт районної газети.

• Показники: Зверніть увагу на:

- *LCP (Largest Contentful Paint):* Як швидко користувач бачить заголовок?

- *CLS (Cumulative Layout Shift):* Чи "стрибає" текст, коли довантажується реклама? (Це дратує і призводить до випадкових кліків).

• Мета: Зрозуміти, як технічна оптимізація впливає на лояльність читача.

Завдання 3. Концепт "Розумної адаптації" (Context-Aware UI)

• Завдання: Запропонуйте, як має змінитися інтерфейс новинного додатка в таких ситуаціях:

- *Ситуація А:* У користувача 5% заряду батареї. (Рішення: Автоматичний Dark Mode, вимкнення автоплею відео).

- *Ситуація Б:* Дуже повільний мобільний інтернет (Edge/2G). (Рішення: Завантаження лише тексту, картинки — за кліком).

- *Ситуація В:* Користувач швидко рухається (їде в авто). (Рішення: Збільшення шрифтів, пропозиція увімкнути аудіо-версію статей).

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Ітан Маркотт. *Responsive Web Design.* (Книга, з якої все почалося. Біблія адаптивного дизайну).

2. Люк Вроблевскі. *Mobile First.* (Філософія проєктування, де спочатку робиться мобільна версія, а потім вона "розширюється" до десктопу).

3. Google Developers. *Progressive Web Apps Training.* (Офіційні гайди про переваги PWA).

4. Nielsen Norman Group. *Mobile User Experience.* (Дослідження поведінки користувачів на смартфонах).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• RWD (Responsive Web Design): Підхід до веб-дизайну, який робить веб-сторінки такими, що добре відображаються на різноманітних пристроях і вікнах або розмірах екрана, використовуючи "гумові" сітки та медіа-запити.

• Mobile-First: Стратегія дизайну та розробки, при якій створення продукту починається з мобільної версії (найсуворіші обмеження), а потім адаптується для планшетів і десктопів.

• PWA (Progressive Web App): Веб-сайт, який використовує сучасні можливості браузера, щоб поводитися як нативний додаток (працювати офлайн, надсилати сповіщення, мати іконку), але не вимагає встановлення через App Store.

• Breakpoint (Точка злому): Значення ширини екрану (наприклад, 768px), при досягненні якого дизайн сайту змінюється (наприклад, зникає бокова колонка, меню перетворюється на "гамбургер").

• A11y (Accessibility - Доступність): Практика створення веб-сайтів, якими можуть користуватися люди з різними типами інвалідності. "11" означає кількість пропущених букв у слові Accessibility.

7.4. Push-повідомлення та алерти: Детальний посібник

7.4.1. Індивідуалізація тем та частоти.

Глибока сегментація користувачів. Поведінкові когорти. Створіть детальну систему класифікації користувачів на основі їхньої поведінки. "Суперактивні" користувачі (відкривають додаток 5+ разів на день) можуть отримувати до 8-10 повідомлень на день без негативного впливу на retention. "Помірні" користувачі (2-3 рази на день) оптимально реагують на 3-5 повідомлень. "Випадкові" користувачі (кілька разів на тиждень) потребують максимум 1-2 найрелевантніших повідомлення на день, інакше ви ризикуєте їх втратити.

Важливо також враховувати життєвий цикл користувача. Нові користувачі (перші 7 днів) потребують онбордингових повідомлень, які допомагають їм зрозуміти цінність продукту. Це можуть бути підказки про ключові функції, персональні рекомендації на основі їхніх початкових дій, заохочення до завершення профілю. Після онбордингу переходьте до звичайної каденції, але продовжуйте моніторити залученість протягом перших 30 днів.

RFM-аналіз для push-стратегії:

Адаптуйте модель RFM (Recency, Frequency, Monetary) для оптимізації push-повідомлень. Користувачі, які недавно були активні (Recency), більш схильні відкрити повідомлення. Ті, хто часто взаємодіє з додатком (Frequency), можуть отримувати більше повідомлень. Користувачі з високою цінністю (Monetary - для e-commerce, або engagement - для інших додатків) заслуговують на найбільш персоналізований контент.

Створіть матрицю 3×3×3, яка дає 27 різних сегментів. Наприклад, користувач з високою recency, високою frequency, але низькою monetary value може бути "активним браузером" - їм потрібні персоналізовані рекомендації та спеціальні пропозиції для конверсії. Користувач з низькою recency, але історично високою frequency та monetary - це "сплячий кит", якого варто реактивувати через win-back кампанії.

Система динамічних профілів. Створіть багатовимірний профіль кожного користувача, який постійно оновлюється:

• Часові преференції: Коли користувач найчастіше відкриває по-

відомлення? Аналізуйте не просто час надсилання, а час фактичного відкриття. Якщо ви надсилаєте в 9 ранку, а користувач відкриває о 14:00, можливо, варто надсилати пізніше.

• Тематичні інтереси: Використовуйте алгоритми колаборативної фільтрації. Якщо користувач часто взаємодіє з контентом категорії А і Б, а інші користувачі з подібним профілем також цікавляться категорією В, почніть тестувати повідомлення категорії В.

• Толерантність до частоти: Деякі користувачі відписуються після 3 повідомлень на день, інші не реагують негативно навіть на 15. Використовуйте градієнтний підхід - починайте з консервативної частоти (2-3 на день) і поступово збільшуйте, моніторячи метрики залученості.

• Чутливість до часу доби: Є користувачі-"жайворонки" (пік активності 6-10 ранку), "сови" (18-23 вечора) та "денні" (обідній час). Машинне навчання може автоматично визначити тип користувача за перші 2 тижні використання.

Прогресивна персоналізація тем. Не просто дозволяйте обирати теми - використовуйте інтелектуальну систему рекомендацій:

1. Початковий онбординг: Запропонуйте користувачу обрати 5-7 тем з 20-30 доступних. Використовуйте зрозумілі категорії з візуальними іконками.

2. Імпліцитне навчання: Відстежуйте, які теми користувач відкриває, читає повністю, на які реагує (лайки, шери, коментарі). Автоматично підвищуйте пріоритет цих тем, навіть якщо вони не були обрані явно.

3. Періодична рекалібрація: Раз на місяць показуйте користувачу персоналізований інсайт: "У січні ви найбільше цікавилися технологіями та спортом. Хочете отримувати більше повідомлень на ці теми?"

4. Гранулярний контроль: Для кожної теми дозволяйте налаштувати рівень: "Всі новини" (кожна публікація), "Важливі" (тільки топ-історії), "Дайджест" (зведення раз на день), "Вимкнено".

Адаптивні алгоритми частоти. Машинне навчання для оптимізації каденції:

Розробіть модель, яка прогнозує оптимальну кількість повідомлень для кожного користувача. Вхідні параметри можуть включати історичний open rate по годинах та днях тижня. Час між

повідомленням та відкриттям (латентність). Кількість повідомлень, після яких спостерігається падіння engagement. Тип контенту, який найкраще працює. Характеристики пристрою (iOS vs Android, модель телефону). Демографічні дані (якщо доступні).

Модель повинна щодня перераховувати оптимальну частоту, адаптуючись до змін у поведінці користувача. Використовуйте multi-armed bandit алгоритми для балансу між exploitation (надсилання перевірених типів повідомлень) та exploration (тестування нових підходів).

Правило затухаючої чутливості. Дослідження показують, що перше повідомлення дня має найвищий open rate (часто 15-25%), друге - вже нижчий (10-18%), третє - ще нижчий (7-12%), і так далі. Після певного порогу (зазвичай 4-6 повідомлень) кожне додаткове повідомлення не тільки має низький open rate, але й знижує показники попередніх.

Використовуйте цю криву для оптимізації:

• Перше повідомлення дня - найважливіший, найперсоналізованіший контент

• Друге та третє - важливі теми або high-intent повідомлення (наприклад, підтвердження замовлення)

• Четверте та далі - тільки якщо користувач показує високу толерантність, і тільки найрелевантніший контент

Система "розумних пауз". Якщо користувач ігнорує 3 повідомлення підряд одного типу, автоматично призупиніть цей тип на 48-72 години. Якщо користувач не відкриває жодних повідомлень протягом 5-7 днів, переведіть його в режим "мінімальної каденції" - тільки критичні повідомлення плюс один щотижневий дайджест найкращого контенту.

Коли користувач повертається до активності, не відразу повертайтеся до попередньої частоти. Застосуйте градуальне збільшення: 1 повідомлення в перший день, 2 в другий, 3 в третій, поки не досягнете оптимального рівня.

Контекстна частота. Частота повинна адаптуватися не лише до користувача, але й до контексту:

• Великі події: Під час важливих подій (вибори, спортивні чемпіонати, кризи) користувачі очікують більше оновлень. Можна тимчасово збільшити частоту на 50-100%.

• Вихідні vs робочі дні: Для B2B додатків частота повинна знижуватися у вихідні. Для розважальних - може підвищуватися.

• Сезонність: E-commerce додатки можуть надсилати більше повідомлень під час розпродажів (Чорна п'ятниця, Новий рік). Фітнес-додатки - в січні (новорічні обіцянки).

7.4.2. Контекстуальні повідомлення (час, місце).

Досконала часова оптимізація. Персональні хронотипи. Замість загальних "найкращих часів" для надсилання, визначте індивідуальний хронотип кожного користувача. Аналізуйте:

1. Перше відкриття дня: Коли користувач вперше бере телефон після пробудження? Це ідеальний момент для "доброго ранку" контенту.

2. Commute patterns: Багато людей активні в транспорті (7-9 ранку, 17-19 вечора). Це чудовий час для контенту, який можна споживати швидко - новини, короткі відео, розважальний контент.

3. Обідня перерва: Зазвичай 12-14 години. Час для більш глибокого контенту - довгі статті, огляди, рекомендації продуктів.

4. Вечірній downtime: 20-23 години - час релаксації. Ідеально для розважального контенту, персональних рекомендацій, контенту, який спонукає до дії наступного дня.

5. Вихідні патерни: У суботу та неділю звички кардинально змінюються. Користувачі прокидаються пізніше, мають більше вільного часу для тривалої взаємодії.

Predictive send time optimization. Використовуйте алгоритми машинного навчання для прогнозування найкращого часу надсилання кожному користувачу. Моделі можуть враховувати історичні дані про відкриття попередніх повідомлень. День тижня та сезонність. Погодні умови (люди більш активні в телефонах під час дощу). Тип контенту (термінові новини vs розважальний контент). Конкуруючі події (під час великих спортивних подій engagement може падати).

Платформи як Airship, Braze, CleverTap пропонують вбудовані AI-моделі для send time optimization. Якщо ви будуєте власне рішення, почніть з логістичної регресії, потім переходьте до більш складних моделей (gradient boosting, neural networks).

Часові зони та "подорожуючі" користувачі. Це критично для глобальних додатків. Система повинна:

1. Автоматично визначати часову зону користувача на основі IP

або GPS

2. Відстежувати зміни часової зони (користувач подорожує)

3. Коригувати розклад повідомлень відповідно до нової локації

4. Враховувати jet lag - перші 2-3 дні після зміни часової зони користувач може мати інший патерн активності

Respect mode та "розумне глушіння". Інтегруйтеся з системними налаштуваннями:

• iOS Focus modes: Визначайте, коли користувач у режимі "Не турбувати", "Сон", "Робота" і відкладайте неважливі повідомлення

• Android DND: Аналогічно для Android

• Календарні події: Якщо користувач надав доступ до календаря, уникайте повідомлень під час зустрічей

• Driving detection: Не надсилайте повідомлення, коли користувач за кермом (можна визначити через API руху або CarPlay/Android Auto активність)

Геолокаційна персоналізація. Рівні геолокаційної точності:

Різні use cases потребують різної точності:

1. Країна/регіон: Для локалізованого контенту, новин, пропозицій (точність ~100 км)

2. Місто: Для локальних подій, погоди, регіональних акцій (точність ~10 км)

3. Район: Для рекомендацій закладів, локальних сервісів (точність ~1 км)

4. Proximity: Для trigger-повідомлень біля конкретних точок (точність ~100 м)

Geofencing стратегії. Створюйте віртуальні "огорожі" навколо важливих локацій:

Retail geofencing: Коли користувач входить у радіус 500 метрів від магазину, надішліть персоналізовану пропозицію на основі його історії покупок. Коли він у магазині, можна надіслати нагадування про список бажань або поточні знижки в конкретних секціях.

Competitor geofencing: Етично сумнівна, але ефективна тактика - створення geofence навколо локацій конкурентів. Коли користувач відвідує конкурента, надішліть порівняльну пропозицію. Event-based geofencing: Під час концертів, спортивних подій, конференцій - активуйте спеціальні пропозиції для відвідувачів. Commute optimization: Визначте звичні маршрути користувача (дім-робота) і

надсилайте релевантний контент під час поїздки.

Технічні особливості геолокації:

• Background location tracking: Потребує явної згоди користувача і сильно впливає на батарею. Використовуйте обережно.

• Significant location changes: iOS та Android дозволяють отримувати оновлення тільки при значних змінах локації (зміна міста), що економить батарею.

• Region monitoring: Більш ефективна альтернатива постійному tracking - моніторинг входу/виходу з певних зон.

• Beacon technology: Для indoor позиціонування (всередині великих магазинів, аеропортів) використовуйте BLE beacons.

Privacy-first підхід до геолокації. Це критично важливо для довіри користувачів:

1. Явна згода: Чітко поясніть, навіщо потрібна геолокація і які переваги отримає користувач

2. Granular permissions: Дозвольте обирати між "завжди", "під час використання" і "ніколи"

3. Прозорість: Показуйте в налаштуваннях, які геолокаційні дані зібрані

4. Право на видалення: Дозволяйте очистити історію локацій

5. Анонімізація: Не зберігайте точні координати назавжди, агрегуйте в райони/міста через певний час

Контекстні тригери. Поведінкові тригери. Автоматичні повідомлення на основі дій користувача:

• Cart abandonment: Кошик залишений → через 1 годину нагадування → через 24 години знижка 10% → через 72 години остаточна пропозиція

• Browse abandonment: Переглянув продукт, але не додав у кошик → через 2 години персоналізована рекомендація з цим продуктом

• Search without results: Користувач шукав щось, чого немає → повідомлення, коли товар з'явиться в наявності

• Feature discovery: Користувач не використовував ключову функцію протягом 2 тижнів → навчальне повідомлення з demo

• Milestone achievements: Досягнення у додатку (1000 кроків, 10 тренувань, 100 покупок) → вітання та заохочення продовжувати

Погодні тригери. Інтеграція з погодними API дозволяє надсилати

контекстуальні пропозиції:

• E-commerce: "Дощ сьогодні вечері - замов парасольку з доставкою за 2 години"

• Доставка їжі: "Холодно надворі? Гарячий суп з безкоштовною доставкою"

• Туризм: "Чудова погода на вихідні - подивись пропозиції на походи"

• Страхування: "Попередження про ураган - перевір свою страховку"

Соціальні тригери:

• Friend activity: "3 твої друзі щойно замовили з нового меню - спробуєш?"

• Trending content: "Ця стаття вибухнула у твоєму місті - читай першим"

• Local events: "Концерт твого улюбленого артиста наступного тижня - квитки вже в продажу"

Lifecycle тригери:

• Anniversary: "Рік тому ти приєднався до нас - ось персональна знижка 20%"

• Inactivity: 7 днів без активності → "Сумуємо за тобою" + персоналізовані рекомендації

• Re-engagement: Користувач повернувся після тривалої відсутності → welcome back bonus

• Churn prevention: ML-модель прогнозує високу ймовірність відписки → спеціальна retention-кампанія

7.4.3. Оптимізація для максимізації open rate.

Ефективність пуш-сповіщень як інструменту взаємодії з аудиторією залежить від комплексної оптимізації декількох ключових параметрів. Це не лише мистецтво, а й наука, що базується на безперервному експериментуванні, глибокій персоналізації, врахуванні технічних обмежень платформ та стратегічному використанні мультимедіа. Щоб повідомлення не просто досягало екрану користувача, а спонукало до цільової дії, необхідно системно підходити до кожного елементу — від формулювання тексту до моменту його відправлення.

A/B тестування заголовків та контенту є фундаментальним методом об'єктивної оцінки ефективності. Це далеко не одноразова

акція, а система регулярних експериментів, яка дозволяє перейти від інтуїтивних припущень до даними підтверджених рішень. Регулярне порівняння різних варіантів текстів — коротких імпульсних проти довгих інформативних, питальних форм, що пробуджують цікавість, проти стверджувальних, які дають чітку вигоду, емоційно забарвлених проти суто фактологічних — дає змогу точно визначити, які механіки резонують із конкретною аудиторією. При цьому аналіз ефективності не повинен обмежуватися поверхневим Open Rate (коефіцієнтом відкриття). Справжню цінність показує відстеження подальшої поведінки користувача (conversion rate): чи перейшов він на сайт, чи дійшов до цільової сторінки, чи виконав бажану дію (покупка, реєстрація, читання статті). Лише такий глибинний аналіз дає змогу оцінити справжній вплив формулювання на досягнення бізнес-цілей.

Незважаючи на силу тестування, жоден загальний заголовок не може конкурувати з потужністю персоналізації змісту. Повідомлення, що містять ім'я одержувача або прямі згадки про його минулі взаємодії (переглянуті товари, прочитані статті, геолокація), демонструють значно вищі показники відкриття та клікабельності. Наприклад, формула "Олексію, курс з цифрового маркетингу, який ви вивчали, оновлено новим модулем" працює на порядок ефективніше за загальне "Нові курси на нашій платформі!". Така релевантність сигналізує про увагу до потреб користувача. Однак існує тонка межа між корисною персоналізацією та вторгненням у приватність. Надмірна деталізація або використання дуже конфіденційних даних без явної згоди може викликати занепокоєння та відштовхнути користувача. Ключ у досягненні балансу — персоналізація на основі публічно наданих або явно дозволених даних, яка сприймається як сервіс, а не як стеження.

Важливим драйвером ефективності є візуальне вдосконалення повідомлень. Rich Notifications (багаті сповіщення), доповнені якісними зображеннями-прев'ю, привертають значно більше уваги, ніж простий текст. Зображення товару, обкладинка статті або логотип акції миттєво передають контекст і підвищують впізнаваність бренду. Використання емодзі може розбити моноліт тексту, додати емоційного забарвлення та підкреслити ключову думку, проте важливо дотримуватися поміркованості. Один-два релевантних

символи, що відповідають тону повідомлення та цільової аудиторії, часто достатньо; надмірне використання може створити враження непрофесійності. Інтерактивні кнопки дії (наприклад, "Купити", "Переглянути", "Відкласти") надають сповіщенню функціональності, дозволяючи користувачеві виконати цільову дію безпосередньо з шторки сповіщень, що різко збільшує конверсію.

Успіх усіх цих стратегій нівелюється без дотримання технічних обмежень та принципів побудови тексту. Основні мобільні платформи мають різні ліміти: iOS рекомендує до 60 символів для заголовку та близько 240 для тіла повідомлення, тоді як Android дозволяє трохи більше. Проте найважливішим принципом є пріоритетність інформації. Оскільки повідомлення на екрані блокування або в шторці часто обрізаються, ключову пропозицію, вигоду або факт потрібно формулювати на початку тексту. Перші 40-60 символів повинні самостійно передавати всю суть і мотивацію для подальшої взаємодії. Також критично важливо враховувати таймінг відправлення; надсилання повідомлень під час нічного відпочинку або в години пік робочого дня може призвести до негативного сприйняття та швидкої відписки.

Таким чином, створення дієвих пуш-сповіщень — це стратегічний процес, що поєднує в собі безперервне навчання через A/B-тести, тактовну персоналізацію, привабливе оформлення та технічно грамотну реалізацію. Оптимальний результат досягається не максимізацією одного параметра, а злагодженою взаємодією всіх елементів, орієнтованою на конкретні потреби та поведінку кінцевого користувача.

7.4.4. Баланс між релевантністю та інвазивністю.

Ефективна робота з пуш-сповіщеннями в сучасних умовах — це мистецтво досягнення оптимального балансу. З одного боку, сповіщення мають бути максимально релевантними та своєчасними, щоб виконувати свою інформаційну та маркетингову функцію. З іншого — надмірна частота, невідповідність моменту або ігнорування уподобань користувача призводять до відчуття інвазивності, втоми від бренду та, як наслідок, відписок або видалення додатка. Отже, ключовим завданням стає побудова стратегії, яка ставить в центр не просто доставку повідомлення, а повагу до уваги та цифрового комфорту людини. Це досягається через систему інтелектуальних

пріоритетів, гнучкі налаштування користувача, технічні механізми зниження шуму та постійний моніторинг показників здоров'я каналу.

Фундаментом цієї стратегії є чітка система пріоритетів, яка класифікує всі сповіщення за ступенем терміновості та цінності для користувача.

• Критичні алерти (найвищий пріоритет): До цієї категорії належать повідомлення, що безпосередньо стосуються безпеки, фінансів або особистих даних користувача (наприклад, попередження про підозрілу діяльність, екстрені новини в його регіоні, зміна умов важливої послуги). Такі сповіщення можуть обходити певні тихі режими, щоб гарантовано бути поміченими, але їхнє використання має бути надзвичайно обережним і виваженим.

• Важливі транзакційні повідомлення (середній пріоритет): Сюди входять підтвердження замовлень, нагадування про заплановані події чи зустрічі, статуси доставки. Вони доставляються з обов'язковим дотриманням встановлених користувачем часових преференцій, як правило, не порушуючи періодів сну.

• Інформаційні та маркетингові повідомлення (низький пріоритет): Це рекомендації, новини, оновлення або рекламні пропозиції. Вони мають найнижчий пріоритет, їх доставка повинна бути максимально ненав'язливою — ідеальні кандидати для групування, відправлення у "тихий" режим або паузи під час годин пік.

Критично важливою практикою є повага до режимів "Не турбувати" (Do Not Disturb) та надання власних розширених налаштувань. Ідеальний додаток не лише автоматично припиняє сповіщення під час системного DND, але й пропонує власні, більш гнучкі інструменти контролю. Користувачам варто надати можливість самостійно встановлювати "тихі години" для робочих днів та окремо для вихідних, визначаючи, що в цей час проходять лише критичні алерти. Така гнучкість демонструє повагу до приватного часу та знижує ризик дратівливої реакції.

Ефективними технічними методами зменшення інвазивності є "тихі сповіщення" (Quiet Notifications) та групування (Notification Grouping). Тихі сповіщення доставляються без звукового сигналу та вібрації, тихо з'являючись в центрі сповіщень. Це ідеальний формат для інформаційних оновлень, які не вимагають миттєвої реакції. Групування ж дозволяє об'єднати кілька повідомлень одного типу

(наприклад, три нових коментарі під постом або кілька акційних пропозицій) в один компактний блок. Це запобігає переповненню шторки сповіщень та робить взаємодію значно чистішою.

Навіть при найточнішій настройці частина аудиторії може бажати скоротити кількість сповіщень. Тому механізми легкої відписки мають бути прозорими та інтуїтивними. Замість того, щоб приховувати налаштування в глибинах меню додатка, слід реалізувати превентивні та зручні опції. Наприклад, якщо користувач кілька разів поспіль проігнорував сповіщення певного типу (свайпнув, не відкриваючи), наступне таке сповіщення може містити безпосереднє питання: "Рідше сповіщати про такі новини?" з кнопками "Так" та "Ні". Даючи користувачу легкий контроль над частотою та типами сповіщень, ви запобігаєте радикальному рішенню — повністю відмовитися від усіх сповіщень або видалити додаток.

Об'єктивну картину ефективності стратегії дають метрики здоров'я системи сповіщень. Крім традиційних Open Rate і CTR, критично важливо уважно відстежувати:

• Показник відписки (Unsubscribe/Opt-out Rate): Це прямий індикатор незадоволеності. Стабільний рівень вище 2-3% на місяць — сигнал тривоги, що стратегія надто агресивна.

• Частка користувачів, що відключили сповіщення в налаштуваннях системи.

• Зв'язок між інтенсивністю сповіщень та показником видалення додатка (App Uninstall Rate).

Аналіз цих даних допомагає корегувати частоти та типи повідомлень у реальному часі.

Всі ці механізми ґрунтуються на принципах прозорості та довіри. З самого початку користувачеві має бути чітко пояснено, які типи повідомлень він буде отримувати та яку цінність вони несуть. Регулярні нагадування про можливості налаштування преференцій (наприклад, раз на квартал) є доброю практикою. Будь-який новий тип сповіщень, особливо маркетингових, повинен впроваджуватися через механізм opt-in (активної згоди), коли користувач сам дозволяє їх надсилання, а не через автоматичне підключення для всіх. Така чесна комунікація будує довгострокову лояльність, перетворюючи сповіщення з потенційного дратівника на корисний та очікуваний сервіс.

1. КОНТРОЛЬНІ ПИТАННЯ (для перевірки знань).

Ці питання допомагають зрозуміти технічні та психологічні нюанси каналу:

1. Психологія переривання: Push-повідомлення — це "зовнішній тригер" (External Trigger). Як працює механізм змінної винагороди (Variable Reward), коли ми чуємо звук сповіщення? Чому ми відчуваємо тривогу, якщо не подивимося на екран?

2. Web Push vs. Mobile Push:

- *Mobile App Push:* Мають доступ до функцій телефону (геолокація, вібрація), вищу стабільність.

- *Web (Browser) Push:* Працюють у Chrome/Safari без встановлення додатка. Чому їхній CTR (клікабельність) зазвичай нижчий, а відсоток відмов — вищий?

3. Стратегія Opt-in (Отримання дозволу):

- *iOS:* Користувач повинен явно натиснути "Дозволити". Як правильно обрати момент для цього прохання (не під час першого запуску, а після першої цінної дії)?

- *Android:* Дозвіл часто дається автоматично при встановленні (хоча нові версії ОС це змінюють). Як це впливає на статистику підписок?

4. Типи сповіщень:

- *Breaking News:* Термінові новини для всіх (високий ризик відписки, якщо новина не важлива).

- *Personalized Alert:* "Ваш улюблений автор опублікував статтю".

- *Passive/Silent Push:* Оновлення контенту у фоновому режимі без звуку, щоб додаток завантажувався швидше, коли ви його відкриєте.

5. Rich Push (Збагачені пуші): Як додавання зображень, емодзі та кнопок дій ("Зберегти", "Поділитися") прямо в тіло повідомлення впливає на залучення (Engagement)?

2. ПРАКТИЧНІ ЗАВДАННЯ.

Ці завдання навчають писати коротко і влучно (мікрокопірайтинг):

Завдання 1. "Битва за символи" (Micro-copywriting)

• Обмеження: Заголовок — 30 символів, Текст — 120 символів.

• Ситуація: Оголошено результати виборів мера. Переміг кандидат Іваненко.

• Завдання: Напишіть 3 варіанти пуша:

1. *Інформаційний (Сухий):* "Іваненко переміг на виборах з 55% голосів".

2. *Клікбейтний (Заборонений прийом):* "Шок! Ви не повірите, хто став новим мером..." (Поясніть, чому це вб'є довіру).

3. *Тизерний (Інтрига з користю):* "Іваненко — новий мер. Що він обіцяв змінити в перші 100 днів? Читайте розбір".

Завдання 2. Налаштування сегментації (Geo-fencing)

• Легенда: Ви працюєте в додатку "Міський транспорт". Сталася аварія метро на синій гілці.

• Дилема:

- Відправити всім 1 млн користувачів? (Ті, хто їде в авто або живе на іншому кінці міста, розлютяться).

- Відправити тільки тим, хто зараз поруч?

• Завдання: Опишіть алгоритм вибірки аудиторії.

- *Умова:* "Користувачі, чия геолокація зараз в радіусі 2 км від станцій синьої гілки" АБО "Користувачі, які часто їздять цим маршрутом у цей час".

Завдання 3. Аналіз "Кладовища сповіщень"

• Інструмент: Ваш власний телефон (Центр сповіщень в кінці дня).

• Аналіз: Подивіться на сповіщення, які ви не відкрили і просто змахнули.

• Рефлексія: Чому ви їх не вимкнули зовсім, але й не відкриваєте?

- *FOMO?* (А раптом щось важливе).

- *Лінь?* (Важко шукати налаштування).

- *Інформаційна самодостатність?* (Вам вистачило заголовка пуша, читати статтю не треба). Це феномен "Push as a Service".

3. РЕКОМЕНДОВАНА ЛІТЕРАТУРА ТА ДЖЕРЕЛА.

1. Nir Eyal. *Hooked: How to Build Habit-Forming Products.* (Розділ про тригери).

2. OneSignal / Airship Blogs. (Блоги провідних платформ розсилки пушів. Найактуальніша статистика по CTR та оптимальному часу відправки).

3. Google Material Design Guidelines. (Розділ Communication: як правильно проектувати сповіщення, щоб вони не дратували).

4. Columbia Journalism Review. *The Push Notification Report.* (Дослідження того, як медіа використовують пуші).

4. ГЛОСАРІЙ ТЕРМІНІВ.

• Opt-in Rate: Відсоток користувачів, які погодилися отримувати сповіщення при першому запиті. Для медіа норма — 40-60%.

• Churn Rate (через пуші): Кількість користувачів, які видалили додаток або вимкнули сповіщення після отримання конкретного пуша. Головна метрика "токсичності" розсилки.

• Deep Link (Глибоке посилання): Технологія, яка дозволяє при кліку на пуш відкрити не просто головну сторінку додатка, а конкретну статтю чи розділ всередині нього.

• Geo-fencing (Геофенсінг): Технологія, що дозволяє надсилати сповіщення, коли пристрій користувача перетинає віртуальний периметр (наприклад, "Ви проходите повз кав'ярню — ось вам знижка" або "Ви в районі перекриття руху").

• Frequency Capping (Обмеження частоти): Налаштування, яке обмежує максимальну кількість повідомлень одному користувачу на день (наприклад, "Не більше 3 пушів, якщо це не війна").

7.5. Платформи та інструменти

7.5.1. Parse.ly, Chartbeat для real-time аналітики.

Для редакції новин кожна секунда має значення, оскільки інтереси аудиторії та актуальність контенту змінюються миттєво. Спеціалізовані аналітичні платформи, такі як Parse.ly та Chartbeat, були створені саме для того, щоб забезпечити журналістів та редакторів пульсом їхньої аудиторії в реальному часі, відрізняючись від універсальних маркетингових інструментів на кшталт Google Analytics своєю філософією, метриками та інтерфейсом, орієнтованими саме на оперативну редакційну роботу.

Фундаментальною відмінністю цих платформ є фокус на реальному часі та якості взаємодії, а не на сукупних показниках. Якщо традиційні системи фіксують одноразову подію "перегляду сторінки", то Chartbeat, наприклад, постійно отримує від клієнта сигнали з інтервалом у кілька секунд, що дозволяє точно визначати, чи користувач активний на сторінці чи просто залишив вкладку відкритою. Це дає змогу вимірювати кількість одночасних користувачів (concurrents) — ключовий показник для новин, що відображає кількість людей на сайті "прямо зараз". Такий підхід забезпечує не тільки швидкість, а й набагато багатше розуміння поведінки аудиторії. Для зручності Chartbeat пропонує інструмент Heads-Up Display, який накладає статистику прямо на веб-сайт редакції, дозволяючи в режимі реального часу бачити популярність статей та приймати миттєві рішення щодо їхнього розміщення на головній сторінці.

Chartbeat здійснив революцію, зробивши центральною метрикою не кліки, а час залучення (Engaged Time). Цей показник обчислює не просто тривалість відкритої вкладки, а саме період, коли користувач активно взаємодіє з контентом через скрол, рух миші, клацання або перегляд відео. Такий вимір ефективності контенту більш точно відображає реальний інтерес читача. У редакціях часто можна побачити Big Board — великий екран, який показує ключові метрики сайту в реальному часі. Різке падіння показників на такому екрані може сигналізувати про технічні проблеми або про те, що опублікована новина не викликала інтересу, що змушує редакцію терміново реагувати .

Parse.ly, з іншого боку, глибше занурюється в аналітику самого контенту. Якщо Chartbeat — це "пульс", то Parse.ly — це "контекст" і "розуміння". Платформа автоматично аналізує та категоризує трафік за темами, авторами та розділами, даючи зрозуміти, не просто що популярне, а чому це популярне. Її системи передбачувальної аналітики можуть порівнювати поточні показники статті з історичними даними, вказуючи, що, наприклад, сьогоднішня стаття набирає на 20% більше переглядів, ніж зазвичай буває для подібних матеріалів у цей день тижня. Інтерфейс Parse.ly спеціально розроблений так, щоб бути зрозумілим для журналістів і редакторів, а не лише для аналітиків даних . Крім того, Parse.ly має розвинену систему сповіщень (alerts), які можна налаштувати для всього сайту, окремих розділів чи конкретних авторів. Ці сповіщення спрацьовують при незвичному спайку трафіку та можуть надсилатися прямо в Slack або на електронну пошту, що дозволяє команді швидко зреагувати на раптово популярну новину.

Для обох платформ критично важливою метрикою є рециркуляція (Recirculation). Цей показник відповідає на питання, який відсоток читачів, завершивши перегляд однієї статті, переходить до наступної на тому ж сайті. Висока рециркуляція означає ефективне утримання аудиторії, тоді як низький показник, особливо на популярній статті, свідчить про "тупик". Якщо стаття має сто тисяч переглядів, але рециркуляція дорівнює нулю, це сигнал для редактора про необхідність терміново додати в текст релевантні внутрішні посилання, віджет "Читайте далі" або інші елементи, що спонукають до подальшої взаємодії. Саме через такий фокус на глибині взаємодії та ланцюжковій поведінці аудиторії Parse.ly та Chartbeat стали незамінними інструментами для сучасних медіаредакцій, яким потрібно приймати швидкі та обґрунтовані рішення в умовах інформаційного потоку реального часу.

7.5.2. Optimizely, VWO для A/B тестування.

У сучасній цифровій журналістиці проектування інтерфейсу та користувацького досвіду перестало бути справою виключно дизайнерської інтуїції та редакційних здогадок. Воно перетворилося на точну науку, основану на даних та контрольованих експериментах. Центральним методом цієї науки є АБ тестування split testing процедура, в якій контрольна група користувачів взаємодіє з

оригінальною версією продукту варіант А, а тестова група отримує модифіковану версію варіант Б. Фінальна мета визначити, яка з версій ефективніше досягає ключової мети, наприклад, перетворює відвідувача на регулярного читача або платного передплатника. Після закриття безкоштовного Google Optimize у 2023 році ринок професійних інструментів для таких експериментів залишили дві провідні платформи Optimizely та Visual Website Optimizer VWO, кожна з яких пропонує унікальний підхід.

Optimizely позиціонує себе як рішення корпоративного рівня Enterprise, орієнтоване на складні технічні завдання. Його ключова перевага полягає в можливості проводити не лише фронтенд тести, які змінюють зовнішній вигляд елементів, але й серверні тести ServerSide Testing. Ця функція дозволяє експериментувати з глибинною бізнес-логікою продукту, наприклад, тестувати різні алгоритми персоналізованих рекомендацій контенту, варіювати умови та ціни підписок або покращувати роботу внутрішнього пошуку. Ще однією потужною технологією Optimizely є Feature Flags функціональні прапорці. Вони дозволяють впроваджувати нові функції, як-от темну тему сайту або новий блок віджетів, поетапно. Редакція може увімкнути новинку лише для 5 користувачів, щоб спочатку перевірити стабільність роботи та зібрати первинний зворотний зв'язок, а потім поступово збільшувати охоплення, мінімізуючи ризики.

VWO Visual Website Optimizer, з іншого боку, більше орієнтований на потреби маркетологів та редакторів, пропонуючи інтуїтивні інструменти для швидких візуальних експериментів. Його головна сила потужний візуальний редактор, що дозволяє створювати тестові варіанти сторінок методом перетягування елементів DragDrop, без необхідності залучення програмістів. Крім того, VWO має вбудовані інструменти аналітики, такі як теплові карти Heatmaps, які візуалізують точки найвищої активності користувачів на сторінці, показуючи, куди вони найбільше клацають або як далеко скролять. Це робить платформу ідеальною для швидкого тестування гіпотез щодо дизайну та розміщення елементів.

У контексті медіаіндустрії АБ тестування зосереджене на трьох критичних зонах, що безпосередньо впливають на ключові показники. Перша зона заголовки та зображення прев'ю, де основна мета

максимізація клікабельності CTR. Редакції тестують, чи привертає більше уваги заголовок-питання чи заголовок-констатація, чи ефективніше фото з обличчям людини крупним планом чи загальний план події. Для таких експериментів часто використовуються спеціальні інструменти, інтегровані в CMS, або платформи аналітики реального часу, як-от Chartbeat. Друга зона оптимізація пейволів та конверсії в підписки. Тут тестуються найважливіші елементи монетизації, порівнюючи ефективність різних призовів до дії, наприклад, емоційний заклик Підтримайте незалежну журналістику проти прямолінійного Купіть підписку. Також експериментують із стратегіями блокування контенту, визначаючи оптимальний момент для показу пейволу після першого абзацу чи після п'ятого, що дає користувачеві більше часу для занурення в матеріал. Третя зона загальний дизайн та користувацький досвід UX, що впливає на час залучення Engagement. Можна тестувати, чи збільшується тривалість читання, якщо прибрати відволікаючий бічний сайдбар, чи зручніше користувачам нова схема розміщення блоків статей на головній сторінці.

Будьякий АБ тест має ґрунтуватися на принципі статистичної значущості Statistical Significance. Це означає, що різниця в результатах між варіантами повинна бути достатньо великою, щоб виключати вплив випадковості. Наприклад, якщо варіант Б отримав 5 кліків, а варіант А 4, це ще не перемога. Такі інструменти, як Optimizely та VWO, автоматично розраховують рівень довіри Confidence Level, який показує ймовірність того, що отриманий результат є об'єктивним. Вважається, що тест можна завершити та робити висновки лише тоді, коли цей рівень досягає 95 і більше відсотків, що зазвичай візуально позначається зеленим індикатором в інтерфейсі платформи. Багатовимірне тестування Multivariate Testing MVT є розвиненою формою експерименту, коли одночасно тестується не один, а комбінація кількох елементів, наприклад, різні заголовки та різні зображення. Це дозволяє зрозуміти не лише ефект кожного окремого змінного фактора, але й їх взаємодію. Однак такі тести вимагають колосального трафіку для отримання статистично валідних результатів і, як правило, доступні лише найбільшим медіахолдингам. Таким чином, Optimizely та VWO, по суті, надають редакціям лабораторію для наукового підходу до продукту, де кожне

рішення про зміну інтерфейсу може бути перевірено, виміряно та обґрунтовано даними, а не припущеннями.

7.5.3. Mailchimp, Customer.io для email персоналізації.

Історія email-маркетингу в медіа пройшла шлях від примітивної стратегії "Batch and Blast" (завантажити базу контактів і відправити всім однакове повідомлення) до складних систем персоналізованої комунікації, що нагадують індивідуальну розмову. Сучасні інструменти трансформували email-розсилку з інструменту масового оповіщення в стратегічний продукт, який будує довгострокові стосунки з аудиторією, підвищує залученість та конверсію. Ця трансформація стала можливою завдяки появі платформ, що по-різному вирішують завдання: від простоти та доступності до глибокої інтеграції з поведінковими даними.

Mailchimp став своєрідними "вхідними дверима" в email-маркетинг для численних малих та середніх медіа завдяки своїй інтуїтивності та низькому порогу входу. Його головна сила — потужний візуальний редактор drag-and-drop, який дозволяє створювати естетичні листи-дайджести (наприклад, "Топ-5 статей тижня") без знань верстки чи програмування. Основний рівень персоналізації в Mailchimp базується на простих механізмах: це використання Merge Tags для вставки імені одержувача ("Привіт, |FNAME|") та ручне управління статичними списками (Lists), наприклад, окремо для донорів та звичайних підписників. Платформа також дозволяє налаштовувати базову автоматизацію, таку як серія листів "Welcome" для нових підписників. Однак Mailchimp демонструє свої обмеження, коли йдеться про складну сегментацію на основі реальної поведінки користувачів на сайті.

Саме на цій складності зосереджується Customer.io — інструмент, розроблений для "дорослих" медіа-продуктів з високими вимогами до персоналізації. На відміну від Mailchimp, Customer.io фокусується не на шаблонах, а на логіці даних. Він глибоко інтегрується з сайтом або додатком медіа, "бачачи" конкретні дії кожного користувача: які статті він читає, скільки часу проводить на сайті, що ігнорує. Це відкриває можливості для поведінкових та тригерних розсилок (Event-Based). Наприклад, система може автоматично надіслати лист з аналітичною статтею користувачеві, який прочитав три матеріали про політику за 24 години. Ключова відмінність — сегментація в

реальному часі. Замість ручного формування списків редактор задає правило: "Всі користувачі, які не відвідували сайт 30 днів". Як тільки користувач перетинає цей часовий поріг, він миттєво потрапляє в сегмент і отримує автоматичний лист-нагадування ("Ми за вами сумуємо. Ось що ви пропустили").

Для технічної реалізації глибокої персоналізації обидві платформи (але особливо потужно — Customer.io) використовують технологію динамічних контент-блоків (Dynamic Content Blocks). Вона дозволяє створювати один універсальний шаблон листа, окремі частини якого автоматично змінюються для різних одержувачів. Наприклад, основний блок із головною новиною дня залишається незмінним для всіх. А ось наступний блок може бути динамічним: для фаната спорту він відобразить огляд матчу, для любителя культури — рецензію на новий фільм, а для решти — прогноз погоди. Таким чином, немає потреби верстати десятки різних розсилок; достатньо одного інтелектуального шаблону, який алгоритм заповнює релевантним контентом.

Одним з найпотужніших застосувань автоматизації в email-маркетингу є "крапельні кампанії" (Drip Campaigns) — серії повідомлень, розподілених у часі за певною логікою. В медіа вони часто використовуються для онбордингу (Onboarding) нових підписників. Типова серія може виглядати так: одразу після підписки користувач отримує вітальний лист із коротким гідом по сайту; через три дні — підбірку найкращих лонгрідів видання за всю історію; через тиждень — нагадування про мобільний додаток; а через два тижні, якщо активність відсутня, — контрольне питання про доцільність продовження розсилки. Така спланована комунікація поступово залучає нового читача в інформаційний всесвіт бренду, підвищуючи його лояльність.

Таким чином, вибір між Mailchimp та Customer.io (або подібними платформами) визначається не технічним "прогресом", а стратегічною потребою медіа. Якщо завдання — ефективно та красиво інформувати широку аудиторію, достатньо інструментів на кшталт Mailchimp. Якщо ж мета — побудувати індивідуальний діалог, де кожен email є реакцією на конкретну дію користувача і кроком у довгостроковій взаємодії, то необхідні потужні системи поведінкової автоматизації, як Customer.io. Саме цей підхід перетворює email з каналу розсилки

в ключовий інструмент для збереження та поглиблення відносин з аудиторією в умовах інформаційної перевантаженості.

Таблиця 7.1.

Порівняння Mailchimp та Customer.io

Критерій	Mailchimp	Customer.io
Цільова аудиторія	Малі та середні медіа, стартапи.	Великі медіа-продукти зі складною логікою (The Atlantic, The Economist).
Основна філософія	Простота та доступність, красиві шаблони.	Глибока інтеграція даних та поведінкова автоматизація.
Ключові функції	Візуальний редактор (drag-and-drop), прості списки, базова автоматизація (Welcome-серії).	Тригерні розсилки на основі подій, сегментація в реальному часі, складні workflow.
Типова задача	Редакційний дайджест "Топ-5 статей тижня".	Автоматичний лист з аналітикою для користувача, який щойно прочитав 3 статті на одну тему.

Складено автором. Джерело. Customer.io vs Mailchimp Comparison // ForwardEmail (2026). https://forwardemail.net/en/blog/customer-io-vs-mailchimp-email-service-comparison

7.5.4. OneSignal, Firebase для push-повідомлень.

Push-повідомлення у сучасній медіаіндустрії стали інструментом стратегічного впливу, який за своєю ефективністю та інвазивністю часто порівнюють із "ядерною кнопкою" редакції. Цей канал забезпечує прямі, миттєві комунікації, що з'являються на екранах смартфонів (App Push) або комп'ютерів (Web Push). Його унікальність полягає в практично стовідсотковій видимості, оскільки повідомлення перериває будь-яку активність користувача, на відміну від електронного листа, який можна не помітити в папці "Вхідні" або "Спам". Однак саме через цю силу ціна редакційної помилки — нерелевантності, надмірності або технічного збою — є критично високою і може швидко призвести до масових відписок та втрати довіри аудиторії. Для управління цим потужним інструментом існує двоступенева екосистема технологій, де Firebase виступає технічним фундаментом, а OneSignal — пультом оперативного керування для редакторів.

Firebase Cloud Messaging (FCM) від Google є базовою, безко-

штовною інфраструктурою, "трубою" або протоколом, через який відбувається фізична доставка повідомлень з серверів видання на пристрої користувачів. Він надає розробникам необхідний набір API (інтерфейсів програмування) та SDK (наборів інструментів) для інтеграції функції пуш-сповіщень у мобільні додатки, забезпечуючи стабільний зв'язок як з пристроями на Android (напряму), так і з iOS (через Apple Push Notification service, APNs). Проте "гола" платформа FCM не призначена для повсякденної роботи контент-команд. Вона не має зручного графічного інтерфейсу, де можна було б швидко написати текст, обрати зображення, відібрати аудиторію за поведінковими ознаками та відправити кампанію. Кожна така операція потребує участі розробників для написання та розгортання відповідного коду, що робить FCM неефективним для динамічних потреб новинної редакції.

Саме для вирішення цієї проблеми існує OneSignal. Ця платформа є потужною надбудовою над технологічними "трубами" FCM та APNs, створеною спеціально для маркетологів, редакторів та аналітиків. OneSignal надає інтуїтивно зрозумілу адмін-панель, де журналіст може без участі технічних фахівців скомпонувати повідомлення, додати заголовок, основний текст, велике зображення для rich-пуша та інтерактивні кнопки дій. Головною силою OneSignal є просунута система сегментації аудиторії на основі даних про поведінку. Редактор може створювати динамічні групи користувачів за складними критеріями, наприклад: "користувачі, які не відвідували додаток протягом трьох днів", "користувачі, що прочитали понад п'ять статей із рубрики 'Політика' за останній місяць" або "підписники, які ніколи не клікали на новини про спорт". Крім того, OneSignal пропонує функцію Intelligent Delivery ("розумна" доставка), коли алгоритм аналізує історичну активність кожного окремого користувача та визначає оптимальний час для відправки (наприклад, коли людина зазвичай їде на роботу в метро й активніше перевіряє телефон). Таким чином, повідомлення надходить у найбільш вдалий момент, збільшуючи ймовірність відкриття та знижуючи ризик дратівливості.

У медіапрактиці push-повідомлення зазвичай класифікують за метою та аудиторією. Термінові новини (Breaking News) використовуються для подій суспільного масштабу (катастрофи, рішення влади, вибори). Такі пуши відправляються масово (broadcast)

усім підписникам, і головним пріоритетом для них є мінімальна затримка доставки. Нішеві або сегментовані пуши (Niche / Segmented) призначені для анонсування нового контенту в конкретній рубриці, наприклад, спортивних результатів або технологічних оглядів. Їхня аудиторія — це користувачі, які виявили попередній інтерес до відповідної теми (про що свідчать теги чи історія переглядів), що значно підвищує релевантність. Багаті пуши (Rich Push), оснащені великими зображеннями, GIF-анімацією або кнопками дій ("Читати", "Поділитися", "Зберегти на пізніше"), забезпечують значно вищий рівень взаємодії — їхній CTR (показник клікабельності) може на 20-30% перевищувати аналогічні показники простих текстових сповіщень.

Окремим, важливим каналом є Web Push-повідомлення. Ця технологія дозволяє медіа надсилати сповіщення прямо через браузери, такі як Chrome, Safari або Firefox, навіть якщо у користувача не встановлено фірмовий мобільний додаток. Головна перевага web push — легкість набору підписної бази, оскільки для підписки користувачеві достатньо один раз натиснути кнопку "Дозволити" при відвідуванні сайту, без необхідності завантажувати та встановлювати додаток. Однак цей канал має і суттєвий недолік — зазвичай нижчий рівень лояльності аудиторії та вищий поріг дратівливості. Користувачі часто підписуються на web push випадково або мимоволі, і при надмірній частоті відправлення або низькій релевантності контенту вони швидко блокують сповіщення або повністю відмовляються від підписки. Тому робота з web push вимагає ще більш обережного підходу до персоналізації, сегментації та контролю частоти, ніж робота з app push. Таким чином, ефективне використання push-сповіщень вимагає від медіа-організації не лише вибору технічної платформи (OneSignal як управлінського інтерфейсу поверх FCM), а й глибокого розуміння типів контенту, механізмів сегментації та необхідності суворого балансу між оперативністю та повагою до уваги користувача.

1. КОНТРОЛЬНІ ПИТАННЯ

Блок А. Розуміння (Теорія)

1. У чому полягає "Парадокс вибору" і чому він робить автоматичну персоналізацію ефективнішою за ручну кастомізацію?

2. Поясніть різницю між *Zero-Party Data* (дані, які користувач дає добровільно) та *Third-Party Data* (дані, куплені у брокерів). Чому другі вмирають?

3. Що таке "Filter Bubble" (Бульбашка фільтрів) у контексті новинної стрічки і які є механізми її розриву (Serendipity)?

4. Чим метрика *Recirculation* (Рециркуляція) відрізняється від *Bounce Rate* (Відмов)?

5. Як працює технологія Geofencing (Геопаркан) у мобільних додатках новин?

Блок Б. Застосування (Кейси) 6. Сайт має високий трафік на статтю, але низький час читання (10 сек). Про що це свідчить і як це виправити в реальному часі? 7. Ви запускаєте пейвол. Кому ви запропонуєте річну підписку одразу, а кому — пробний період за $1? Опишіть логіку алгоритму. 8. Чому розсилка з темою "Дайджест новин #54" має нижчий Open Rate, ніж розсилка з темою "Чому впав долар"?

2. ПРАКТИКУМ (Лабораторні роботи)

Завдання 1. Проєктування Onboarding-сценарію (Збір даних)

• Легенда: Ви запускаєте новий додаток про кіно.

• Завдання: Розробіть екран вітання для збору Zero-Party Data.

- *Питання 1:* Які жанри любите? (Теги).

- *Питання 2:* Як часто хочете отримувати пуші? (Частота).

- *Мотивація:* Що отримає користувач за відповіді? (Наприклад, "Ми приховаємо жахи, якщо ви їх боїтеся").

Завдання 2. Логіка тригерної розсилки (Email Automation)

• Інструмент: (На папері або в Miro) Симуляція Customer.io.

• Умова: Користувач прочитав 3 статті про "Інвестиції", але не підписаний на платний пакет.

• Завдання: Скласти ланцюжок дій (Workflow).

1. *Trigger:* Перегляд 3-ї статті категорії "Finance".

2. *Wait:* 2 години.

3. *Action:* Надіслати лист з темою "Хочете інвестувати професійно?".

4. *Content:* Добірка закритих статей + промокод на підписку.

Завдання 3. A/B Тестування заголовка (Optimizely)

• Легенда: Ви головний редактор.

• Гіпотеза: "Питальні заголовки працюють гірше за стверджувальні".

• Завдання: Сформулюйте умови тесту.

- *Варіант А:* "Чи піднімуть податки в Україні?"

- *Варіант Б:* "Нові податки в Україні: повний список змін".

- *Метрика успіху:* CTR (клікабельність).

- *Умова зупинки:* Досягнення 95% статистичної значущості.

Завдання 4. Аналіз дашборду (Chartbeat)

• Завдання: Проаналізуйте скріншот реального дашборду (надається викладачем).

- Яка стаття є "локомотивом" трафіку?

- Яка стаття має найвищий час залучення (Engaged Time)?

- Яке джерело трафіку переважає: Пошук, Соцмережі чи Direct?

- Рішення: Які зміни внести на Головну сторінку на основі цих даних?

3. ГЛОСАРІЙ ТЕРМІНІВ (Must-know)

• Churn Rate (Відтік): Відсоток підписників, які скасували підписку за певний період. Головний ворог платного медіа.

• Dynamic Paywall (Динамічний пейвол): Алгоритм, який вирішує, коли заблокувати доступ до контенту, базуючись на ймовірності того, що користувач заплатить.

• A/B Testing: Метод порівняння двох версій веб-сторінки/листа для визначення більш ефективної.

• Recirculation: Показник утримання аудиторії (перехід з однієї статті на іншу).

• Notification Fatigue: Втома користувача від надмірної кількості сповіщень, що веде до видалення додатка.

• Zero-Party Data: Дані, які користувач надає бренду свідомо та добровільно (анкети, налаштування).

4. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Книги та звіти:

1. Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You.* (Класика про ризики).

2. Nielsen Norman Group. *Personalization vs. Customization.* (База UX).

3. Reuters Institute. *Digital News Report.* (Щорічна статистика споживання новин).

Документація та блоги інструментів:

1. Chartbeat Blog. *Data Science of Engagement.*
2. Mailchimp Guides. *Email Marketing Benchmarks.*
3. Optimizely. *The Big Book of Experimentation.*

Висновки до розділу 7

Персоналізація дистрибуції: Від «Головної сторінки» до «Головного читача». Аналіз сучасних інструментів та стратегій дистрибуції демонструє глибоку трансформацію, яка перестала бути технологічною можливістю та перетворилася на обов'язкову операційну реальність. Ця зміна фокусу від монологу до діалогу, від масового мовлення до індивідуального сервісу формує п'ять ключових висновків, що визначають сьогоднішній та завтрашній день цифрових медіа.

Першим і найважливішим висновком є кардинальна зміна парадигми дистрибуції, що означає смерть класичної моделі Broadcast («мовлення для всіх»). Традиційна універсальна головна сторінка, яка була єдиною точкою входу та редакційним вибором для мільйонів, втрачає свою абсолютну роль. На її місце приходить модель Narrowcast («мова для кожного»). Сучасне медіа перетворюється зі статичної цифрової копії паперової газети на динамічний, адаптивний сервіс. Його основною продуктом стає не загальний сайт, а персоналізована стрічка новин («The Daily Me») для кожного окремого користувача, яка формується в реальному часі на основі його поведінкової історії, інтересів, геолокації та контексту взаємодії. Це фундаментальний перехід від редакційної гегемонії до суверенітету аудиторії.

Ця трансформація продиктована жорстким економічним імперативом, де релевантність стала валютою, якою користувач платить за контент своїм часом та грошима. У перенасиченому інформаційному просторі успіх визначається не обсягом виробленого контенту, а здатністю доставити саме те, що людина вважає цінним. Для рекламної моделі персоналізація безпосередньо впливає на ключові метрики життєздатності: збільшує час перебування на сайті, глибину перегляду та показник рециркуляції. Для підписної моделі, яка стає все більш критичною, саме динамічні пейволи, персоналізовані заклики до дії та індивідуальні email-кампанії є основним інструмен-

том боротьби з відтоком аудиторії. В економіці уваги контент, що не є релевантним, просто не існує.

Логічним наслідком цього є утвердження суверенітету даних як основи конкурентної переваги. Завершується ера залежності від сторонніх платформ дистрибуції (таких як соціальні мережі) та сторонніх даних (third-party cookies). Майбутнє належить медіа, які здатні будувати власні, замкнуті екосистеми збору даних, залучаючи користувачів до реєстрації та отримуючи від них zero-party data — інформацію, надану свідомо та добровільно. У такому контексті глибоке, багатовимірне знання про те, хто є читач, які його мотиви та звички, стає стратегічним активом редакції, що часто перевищує за цінністю окремі контентні матеріали.

Окрім «що» і «кому», радикальній переоцінці піддається питання «як» і «коли». Сучасна дистрибуція будується на принципі омніканальності, який можна виразити правилом «4R»: доставка Правильного контенту (Right Content) Правильній людині (Right Person) у Правильний час (Right Time) через Правильний канал (Right Channel). Вибір каналу — чи то web push, email-дайджест, сповіщення в додатку чи віджет на сайті — більше не є довільним рішенням редактора. Він диктується алгоритмічним аналізом звичок конкретного користувача: де, коли та на якому пристрої він найбільш схильний споживати новини. Дистрибуція стає гнучкою, адаптивною та багатоканальною.

Однак ця потужна гіперперсоналізація несе в собі серйозну етичну відповідальність та соціальний ризик. Алгоритми, заточені на максимізацію уваги, можуть створювати інформаційні «бульбашки» або «фільтри», що призводять до фрагментації суспільного діалогу, посилення поляризації та втрати спільного інформаційного простору. Тому відповідальні медіа зобов'язані інтегрувати в свої системи механізми серендипності (випадкового відкриття) та збалансованого редакційного курування. Завдання полягає в тому, щоб гарантувати, що поряд з контентом, який відповідає особистим інтересам, користувач отримує доступ до суспільно важливої інформації, розслідувань, новин, що формують громадянську думку та об'єднують суспільство.

Таким чином, загальний підсумок розділу полягає в тому, що дистрибуція остаточно перестала бути технічною, вторинною функцією публікації готового контенту. Вона трансформувалася в

органічну, невід'ємну частину самого журналістського продукту. Сьогоднішній медіаменеджер, редактор чи журналіст повинен мислити не лише категоріями заголовків та текстів, а й категоріями сегментів аудиторії, поведінкових тригерів, конверсійних воронок та багатоканальних стратегій. Майбутнє належить тим, хто розуміє, що в епоху персоналізації успішна доставка контенту — це і є його остаточне, найважливіше редагування.

ВИСНОВКИ

Перша частина навчально – методичного посібника: "Персоналізація медіа та алгоритмічна журналістика" представляє комплексний аналіз фундаментальних трансформацій медіаіндустрії, спричинених цифровою революцією, розвитком штучного інтелекту та зміною моделей споживання інформації. Розглянуті концепції формують теоретичну основу для розуміння сучасного медіаландшафту та закладають фундамент для подальшого вивчення технологій та практик персоналізації.

1. Парадигмальний зсув у медіаспоживанні. Від масової комунікації до індивідуалізованого досвіду. Історична еволюція медіаспоживання демонструє послідовний рух від моделі "один до багатьох" до моделі "багато до одного". Якщо традиційні медіа епохи мовлення характеризувалися централізованим виробництвом контенту та його масовою дистрибуцією за фіксованим розкладом, то цифрова ера принесла докорінно інші принципи організації медіапростору.

Цей перехід не був миттєвим — він відбувався поетапно, від появи кабельного телебачення з розширеним вибором каналів до інтернет-епохи з її практично необмеженим контентом. Кожен етап цієї еволюції розширював можливості вибору та контролю для аудиторії, поступово переміщуючи владу від виробників контенту до його споживачів.

Інтерактивність як нова норма. Цифрові технології трансформували пасивних споживачів у активних учасників медіапроцесу. Можливість коментувати, ділитися, створювати власний контент, взаємодіяти з авторами та іншими читачами радикально змінила динаміку медіакомунікації. Ця інтерактивність створила нові виклики для медіаорганізацій: необхідність управляти спільнотами, модерувати дискусії, реагувати на зворотний зв'язок у реальному часі.

Важливо, що інтерактивність стала не просто додатковою функцією, а базовою очікуванням аудиторії. Медіа, які не забезпечують можливості для взаємодії, сприймаються як застарілі та нерелевантні, особливо молодшими поколіннями споживачів.

Мобільна революція та контекстуальне споживання. Перехід до мобільних пристроїв як основного каналу доступу до медіа

створив нову реальність контекстуального споживання. Медіа тепер супроводжують людину протягом усього дня, в різних ситуаціях та контекстах. Це відкриває можливості для нових форм персоналізації, що враховують не лише інтереси користувача, а й його поточний контекст: місцезнаходження, час доби, тип пристрою, соціальну ситуацію.

Мобільність також змінила патерни споживання: від тривалих сесій до коротких, але частих взаємодій; від цілеспрямованого пошуку до випадкового просмотру; від запланованого споживання до спонтанного. Ці зміни вимагають нових підходів до створення та дистрибуції контенту.

2. Економіка уваги як фундаментальна зміна бізнес-моделі. Увага як найдефіцитніший ресурс. В умовах інформаційного вибуху традиційна економіка дефіциту інформації трансформувалася в економіку дефіциту уваги. Якщо раніше проблемою був обмежений доступ до інформації, то сьогодні критичним викликом стає відбір релевантного контенту з величезного обсягу доступної інформації.

Ця трансформація має глибокі наслідки для медіабізнесу. Час та увага аудиторії стали основною валютою, за яку ведеться жорстка конкуренція. Медіакомпанії конкурують не лише між собою, а й із соціальними мережами, відеоіграми, месенджерами, стрімінговими платформами — усіма цифровими активностями, що претендують на обмежений ресурс людської уваги.

Трансформація метрик успіху. Економіка уваги вимагає нових способів вимірювання успіху. Традиційні метрики — тираж, рейтинг, охоплення — хоч і залишаються важливими, поступаються місцем більш складним показникам залученості (engagement). Час перебування на сторінці, глибина прокрутки, кількість повернень, коефіцієнт завершення перегляду відео, соціальні взаємодії — ці метрики дають набагато глибше розуміння того, наскільки контент дійсно захоплює увагу аудиторії.

Важливо, що фокус зміщується від простої кількості до якості взаємодії. Один залучений, лояльний користувач, який регулярно повертається, проводить значний час з контентом та ділиться ним, може бути цінніший за десятки випадкових відвідувачів, які швидко залишають сайт.

Персоналізація як відповідь на дефіцит уваги. В контексті

економіки уваги персоналізація стає не просто конкурентною перевагою, а необхідною умовою виживання. Релевантний, персоналізований контент має значно вищі шанси привернути та утримати увагу в умовах нескінченного вибору альтернатив.

Алгоритмічна персоналізація дозволяє масштабувати індивідуальний підхід до кожного користувача, що було б неможливим для людських редакторів. Системи рекомендацій аналізують величезні обсяги даних про поведінку мільйонів користувачів, щоб знайти оптимальний контент для кожного конкретного споживача в конкретний момент часу.

3. Багатовимірність та складність персоналізації. Типологічне різноманіття. Персоналізація медіа — це не монолітне явище, а складна екосистема різноманітних підходів та технологій. Явна персоналізація, де користувач свідомо налаштовує свої переваги, співіснує з неявною, де система аналізує поведінку та робить висновки про інтереси. Колаборативна фільтрація, що знаходить схожих користувачів, доповнюється контент-орієнтованим підходом, що аналізує характеристики самого контенту.

Кожен підхід має свої переваги та обмеження. Колаборативна фільтрація ефективна для популярного контенту, але страждає від проблеми "холодного старту" для нових користувачів чи нового контенту. Контент-базовані методи можуть рекомендувати нові матеріали, але схильні до надмірної спеціалізації. Гібридні системи намагаються поєднати переваги різних підходів, мінімізуючи їхні недоліки.

Рівні персоналізації: від сегментації до генеративного контенту

Персоналізація медіа еволюціонувала від простої сегментації аудиторії до надзвичайно складних індивідуалізованих підходів. Сегментація, найпростіша форма, групує користувачів за спільними характеристиками та пропонує однаковий контент усередині сегмента. Це ефективно, але недостатньо точно.

Персоналізація на рівні індивідуального користувача враховує унікальну історію взаємодій, інтереси та поведінку кожної особи. Контекстуальна адаптація йде далі, враховуючи ситуаційні фактори: час доби, день тижня, місцезнаходження, тип пристрою, навіть погодні умови.

Передбачувальна персоналізація використовує машинне навчання

для прогнозування майбутніх потреб та інтересів користувача, пропонуючи контент до того, як людина усвідомила своє бажання його побачити. Найновіша генеративна персоналізація з використанням ШІ створює унікальний контент для кожного користувача — від персоналізованих новинних зведень до адаптованих стилів викладу.

Технологічна інфраструктура як основа. Ефективна персоналізація неможлива без складної технологічної інфраструктури. Системи збору даних — від традиційних cookies до складних систем поведінкової аналітики — формують базу для розуміння користувачів. Data Management Platforms (DMP) та Customer Data Platforms (CDP) консолідують інформацію з різних джерел, створюючи єдиний профіль користувача.

Рекомендаційні механізми, що працюють у реальному часі, аналізують мільярди можливих комбінацій користувач-контент, знаходячи оптимальні збіги. A/B тестування дозволяє експериментувати з різними підходами та вимірювати їхню ефективність. Усі ці компоненти мають працювати злагоджено, забезпечуючи швидкість, точність та масштабованість.

Баланс між персоналізацією та кастомізацією. Критичне питання — скільки контролю надати користувачеві, а скільки залишити алгоритмам. Повна автоматична персоналізація може бути ефективною, але позбавляє користувача відчуття контролю та може призвести до непередбачуваних результатів. Повна кастомізація надає максимальний контроль, але вимагає значних зусиль від користувача та не використовує потужність алгоритмічної оптимізації.

Оптимальний підхід часто полягає в гібридній моделі: алгоритми забезпечують базову персоналізацію, але користувачі мають можливість коригувати, налаштовувати та перевизначати автоматичні рішення. Така модель поєднує ефективність алгоритмів із збереженням автономії користувача, що критично важливо для довіри та задоволеності.

4. Алгоритмічна журналістика: трансформація професії. Переосмислення журналістської ролі. Алгоритмічна журналістика не є просто автоматизацією традиційних журналістських практик — це фундаментальне переосмислення ролі журналіста в інформаційній екосистемі. Традиційна модель журналіста як привілейованого посередника між подіями та аудиторією трансформується в більш

складну конфігурацію, де журналісти стають архітекторами інформаційних потоків, кураторами даних, верифікаторами алгоритмічних висновків.

Ця трансформація не означає зниження значущості журналістів — навпаки, їхня роль стає ще більш критичною, але якісно іншою. Якщо раніше цінність журналіста визначалася його доступом до інформації та здатністю її донести, то тепер ключовими стають навички критичного аналізу величезних масивів даних, розуміння алгоритмічних процесів, здатність ставити правильні питання до даних та інтерпретувати результати в журналістському контексті.

Спектр застосування: від аналітики до генерації. Алгоритмічна журналістика охоплює широкий спектр практик — від використання ШІ для аналізу даних та виявлення закономірностей до повністю автоматизованої генерації новинних текстів. Дата-журналістика використовує обчислювальні методи для аналізу великих масивів структурованих даних, виявлення трендів, аномалій та кореляцій, які були б недоступні традиційним методам.

Автоматизована генерація новин (robot journalism) створює текстові матеріали на основі структурованих даних — фінансових звітів, спортивних результатів, погодних даних. Алгоритмічне курування відбирає та ранжує контент для персоналізованих новинних стрічок. Верифікація та фактчекінг з використанням ШІ допомагає боротися з дезінформацією в масштабі, недосяжному для людей.

Нові компетенції та навички. Алгоритмічна епоха вимагає від журналістів нових компетенцій. Дата-грамотність стає базовою навичкою — розуміння принципів роботи з даними, статистичного мислення, здатності критично оцінювати дані та методи їх збору. Промт-інженерія — уміння ефективно взаємодіяти з великими мовними моделями, формулюючи запити так, щоб отримати оптимальні результати.

Розуміння принципів машинного навчання, обмежень алгоритмів, потенційних джерел упередженості стає необхідним для відповідального використання ШІ в журналістиці. Водночас не втрачають значущості традиційні журналістські навички — критичне мислення, етичне судження, здатність розповідати історії, розуміння суспільного контексту.

Етична відповідальність та прозорість. Використання алгоритмів у журналістиці піднімає складні етичні питання. Хто відповідає за помилки алгоритмічно згенерованих новин? Як забезпечити прозорість алгоритмічних рішень для аудиторії? Як уникнути відтворення та посилення упереджень, закладених у тренувальних даних?

Концепція "людини в циклі" (human-in-the-loop) передбачає, що критичні журналістські рішення завжди мають включати людське судження. Алгоритми можуть допомагати, прискорювати, масштабувати, але остаточна відповідальність за контент залишається за журналістами-професіоналами. Це вимагає нових редакційних процесів, процедур верифікації, стандартів прозорості.

5. Технологічна основа персоналізованих медіа. Екосистема збору та управління даними. Персоналізація неможлива без даних, а ефективне управління даними вимагає складної інфраструктури. Сучасні медіаорганізації збирають дані з безлічі джерел: поведінкові дані про взаємодії з контентом, демографічні характеристики, соціальні сигнали, контекстуальну інформацію про пристрої та місцезнаходження.

Технології відстеження — від традиційних cookies до сучасних систем поведінкової аналітики — фіксують кожну взаємодію користувача з контентом. Event tracking дозволяє моніторити специфічні дії: прокрутку, кліки, паузи у перегляді відео, взаємодії з інтерактивними елементами. Інтеграції з соціальними мережами та сторонніми платформами збагачують дані додатковим контекстом.

Системи управління даними консолідують цю розрізнену інформацію в єдині профілі користувачів, вирішуючи складні завдання ідентифікації користувачів через різні пристрої та канали, дедуплікації, узгодження даних з різних джерел. Customer Data Platforms (CDP) створюють єдине джерело правди про кожного користувача, доступне для всіх систем персоналізації.

Рекомендаційні системи як ядро персоналізації. Рекомендаційні механізми — це "мозок" персоналізованих медіа, що приймає рішення про те, який контент показати кожному користувачеві. Ці системи еволюціонували від простих правил до складних моделей глибокого навчання, здатних обробляти мільйони сигналів та знаходити неочевидні закономірності.

Колаборативна фільтрація знаходить користувачів зі схожими смаками та рекомендує контент, який сподобався цим схожим користувачам. Контент-базовані підходи аналізують характеристики контенту та збігають їх з профілем інтересів користувача. Матрична факторизація виявляє приховані латентні фактори, що визначають переваги користувачів.

Сучасні системи використовують глибоке навчання — нейронні мережі, здатні автоматично виявляти складні нелінійні залежності в даних. Контекстуальні бандити та reinforcement learning дозволяють системам навчатися в режимі реального часу, адаптуючись до змін у поведінці користувачів та динаміці контенту.

Real-time персоналізація та швидкість. Сучасні користувачі очікують миттєвої релевантності — персоналізація має відбуватися в режимі реального часу, враховуючи найсвіжішу інформацію про поведінку та контекст. Це вимагає інфраструктури, здатної обробляти величезні потоки даних з мінімальною затримкою.

Real-time персоналізація означає, що рекомендації оновлюються моментально після кожної взаємодії користувача. Система має миттєво реагувати на зміни контексту — наприклад, коли користувач переходить з мобільного пристрою на десктоп, змінює локацію, або його інтереси зміщуються протягом сесії.

A/B тестування та оптимізація. Персоналізація — це не разова налаштування, а постійний процес експериментування та оптимізації. A/B тестування та багатоваріантне тестування дозволяють порівнювати різні підходи до персоналізації, вимірювати їхній вплив на ключові метрики, приймати рішення на основі даних, а не інтуїції.

Ефективна культура експериментування передбачає систематичне тестування гіпотез про те, що працює для різних сегментів аудиторії. Це може включати тестування різних алгоритмів ранжування, форматів представлення контенту, дизайну інтерфейсів, копірайтингу заголовків та багато іншого.

6. Виклики та обмеження персоналізації. Проблема холодного старту. Однією з найскладніших проблем персоналізації є "холодний старт" — ситуація, коли система не має достатньо даних для ефективних рекомендацій. Це стосується як нових користувачів, про яких система нічого не знає, так і нового контенту, який ще не має історії взаємодій.

Для нових користувачів системи можуть використовувати демографічні дані, інформацію з соціальних мереж, явні переваги, зібрані під час онбордингу. Для нового контенту допомагає контент-базований аналіз, що дозволяє робити рекомендації на основі характеристик контенту, а не лише поведінкової історії.

Ехо-камери та фільтр-бульбашки. Персоналізація, оптимізована виключно на залученість, може призвести до створення ехо-камер — середовищ, де користувачі бачать лише контент, що підтверджує їхні існуючі погляди. Алгоритми, що рекомендують "більше того ж самого", поступово звужують інформаційний горизонт користувачів, обмежуючи їхнє знайомство з альтернативними точками зору.

Фільтр-бульбашки — персоналізовані інформаційні середовища, відокремлені від ширшого дискурсу — мають серйозні наслідки для демократичного суспільства. Коли люди живуть у різних інформаційних реальностях, ускладнюється формування спільної фактичної основи для суспільних дискусій.

Медіаорганізації все більше усвідомлюють цю проблему та експериментують з підходами, що балансують релевантність із різноманітністю — "diversity injection", експозицією до альтернативних точок зору, серендипністю, що дозволяє користувачам випадково відкривати щось несподіване.

Прозорість та пояснюваність. Алгоритми персоналізації часто є "чорними скриньками" — навіть їхні творці не завжди можуть пояснити, чому система прийняла конкретне рішення. Це створює проблеми довіри: користувачі не розуміють, чому бачать певний контент, і не можуть ефективно коригувати небажані рекомендації.

Рух до "explainable AI" в журналістиці передбачає створення систем, здатних пояснювати свої рішення зрозумілою людині мовою: "Ми показуємо це, тому що ви читали подібні статті", "Цей матеріал популярний серед людей з схожими інтересами". Така прозорість не лише будує довіру, а й дає користувачам інструменти для кращого контролю над персоналізацією.

Популярність vs релевантність. Багато рекомендаційних систем схильні до "popularity bias" — надмірної ваги популярного контенту. Це створює замкнене коло: популярний контент отримує більше експозиції, що робить його ще популярнішим, тоді як нішевий або новий контент важко "пробивається" до аудиторії.

Для медіаорганізацій це дилема: чи оптимізувати виключно на короткострокову залученість (що схиляє до популярного контенту), чи балансувати це з довгостроковими цілями — підтримкою різноманітного контенту, розвитком нових авторів, забезпеченням всебічного інформаційного раціону для аудиторії.

7. Приватність, етика та регуляції. Напруження між персоналізацією та приватністю. Ефективна персоналізація вимагає даних, але збір та використання персональних даних піднімає серйозні питання приватності. Користувачі хочуть релевантного контенту, але непокояться про те, скільки інформації про них збирають компанії та як вона використовується.

Регуляції як GDPR в Європі та CCPA в Каліфорнії встановлюють нові правила гри, вимагаючи явної згоди на збір даних, прозорості про їх використання, права на видалення даних. Технології як "privacy by design" намагаються вбудувати захист приватності в саму архітектуру систем персоналізації.

Етичне використання алгоритмів. Алгоритми можуть відтворювати та навіть посилювати упередження, присутні в тренувальних даних або закладені (свідомо чи ні) їхніми творцями. Якщо алгоритм тренувався на історичних даних, що відображають існуючі суспільні нерівності, він може продовжувати ці нерівності в своїх рекомендаціях.

Етичне використання алгоритмів вимагає постійної уваги до питань справедливості, інклюзивності, репрезентативності. Хто представлений у рекомендаціях? Чиї голоси ампліфікуються, а чиї маргіналізуються? Як алгоритми впливають на суспільний дискурс, демократичні процеси, формування громадської думки?

Consent management та прозорість. Сучасні медіаорганізації мають розвивати складні системи управління згодою користувачів — прозоро інформувати про збір даних, надавати зрозумілий контроль над налаштуваннями приватності, поважати вибір користувачів. Це не лише юридична вимога, а й питання довіри та довгострокових відносин з аудиторією.

Анонімізація та агрегація даних дозволяють використовувати інформацію для покращення сервісу, мінімізуючи ризики для приватності окремих користувачів. Диференційна приватність та інші сучасні технології дозволяють видобувати корисні insights з

даних, зберігаючи індивідуальну приватність.

8. Майбутнє персоналізації: тренди та перспективи. Від персоналізації до гіперперсоналізації. Розвиток технологій штучного інтелекту відкриває можливості для ще глибших рівнів персоналізації. Великі мовні моделі можуть не просто рекомендувати існуючий контент, а створювати унікальні матеріали, адаптовані до конкретного користувача — персоналізовані новинні зведення, резюме, адаптовані стилі викладу.

Мультимодальна персоналізація враховуватиме не лише текстові вподобання, а й переваги щодо формату (текст, відео, аудіо, інтерактив), стилю викладу, глибини деталізації, навіть емоційного тону контенту.

Баланс між технологією та людяністю. Попри всі технологічні можливості, майбутнє персоналізації не в заміні людського судження алгоритмами, а в знаходженні оптимального балансу між ефективністю технологій та незамінними людськими якостями — емпатією, етичним судженням, креативністю, розумінням культурного та соціального контексту.

Успішні медіаорганізації майбутнього будуть тими, хто зможе поєднати потужність алгоритмічної персоналізації з журналістською чесністю, етичною відповідальністю, та місією служіння суспільному благу. Технології — це інструменти, але цінності, що керують їх використанням, визначатимуть, чи стане персоналізація силою для позитивних змін чи джерелом нових проблем.

Фундаментальні концепції, розглянуті в першій частині підручника, формують теоретичну та контекстуальну основу для розуміння сучасного стану та майбутнього персоналізованих медіа. Наступні частини будуть поглиблювати ці знання, досліджуючи конкретні технології, інструменти, практики та кейси, що ілюструють реальне застосування персоналізації в медіаіндустрії.

ПЕРСОНАЛІЗАЦІЯ МЕДІА ТА АЛГОРИТМІЧНА ЖУРНАЛІСТИКА (ЧАСТИНА 1)

М.М. Марусинець, Д.І. Ткач, Д.К. Ткач

Київ - 2026

м. Київ, Україна
Підписано до друку 2026 р. Формат 60x84/16. Папір офсетний.
Друк офсетний. Гарнітура Times New Roman.
Ум. друк. арк. 3,96. Наклад 300 прим.
Зам. 231

Університет економіки та права «КРОК»
Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру ДК № 613 від 25.09.2001 р.

Університет економіки та права «КРОК»
місто Київ, вулиця Табірна, 30-32
e-mail: Print@krok.edu.ua