

ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«УНІВЕРСИТЕТ ЕКОНОМІКИ ТА ПРАВА «КРОК»
Кафедра журналістики

Пяткевич Андрій Валерійович

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА
«Вплив пропаганди на міжнародну журналістику: сучасні тенденції та
виклики »
(тема)

061 Журналістика
(шифр і назва спеціальності)

«Міжнародна журналістика»
(освітня програма)

Подається на здобуття освітнього ступеня магістр

Магістерська кваліфікаційна робота Пяткевича Андрія Валерійовича містить результати власних доробок. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

А.В. Пяткевич
(підпис, ініціали та прізвище здобувача)

Науковий керівник:

Ткач Дмитро Іванович, доктор політичних наук, професор

Київ-2025 р.

ЗМІСТ

ВСТУП	2
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ПРОПАГАНДИ ТА ЕВОЛЮЦІЯ ЇЇ ІНСТРУМЕНТІВ.....	6
1.1. Історичний розвиток пропаганди у журналістиці.....	6
1.2. Пропаганда vs. журналістика в умовах цифрової доби.....	15
1.3. Технологічний прорив: поява ШІ як нового інструменту впливу.....	22
РОЗДІЛ 2. ПРАКТИКИ ВИКОРИСТАННЯ ШІ ДЛЯ ПРОПАГАНДИ В МІЖНАРОДНІЙ ЖУРНАЛІСТИЦІ	28
2.1. Генеративні моделі, deepfake та дезінформація.....	28
2.2. Алгоритми та соціальні мережі як провідники пропаганди.....	32
2.3. Кейс-стаді: використання ШІ в умовах інформаційних війн.....	37
РОЗДІЛ 3. ЗАГРОЗИ ТА ШЛЯХИ ПРОТИДІЇ ШІ-ПРОПАГАНДИ	45
3.1. Основні загрози для журналістики та демократії.....	45
3.2. Технології й інституційні механізми протидії.....	53
3.3. Роль медіаосвіти та журналістських стандартів.....	66
ВИСНОВКИ	73
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	77

ВСТУП

Актуальність теми пропаганди в міжнародній журналістиці полягає у розвитку інформаційних технологій та відповідного перенесення журналістики в цифрову площину. У ХХІ столітті медійний простір зазнав фундаментальних змін, пов'язаних із бурхливим розвитком цифрових технологій. Зокрема, стрімке впровадження штучного інтелекту (ШІ) у сферу масової комунікації започаткувало нову епоху формування інформаційної реальності. Журналістика, яка традиційно асоціювалася з пошуком істини, фактами та професійною етикою, стикається з безпрецедентним викликом: конкуренцією з симульованим, автоматизованим, але водночас надзвичайно переконливим контентом.

Алгоритмічні стрічки новин, генеративні мовні моделі, deepfake-відео, синтетичні голоси — ці інструменти ШІ дедалі частіше використовуються не лише для створення розважального чи навчального контенту, але й як потужна зброя у війнах за вплив на масову свідомість. Особливу занепокоєність викликає тенденція використання ШІ для поширення дезінформації, маніпуляцій, інформаційних атак і пропагандистських кампаній, які маскуються під журналістику і легко інтегруються в медійне середовище.

Міжнародна журналістика, як стратегічний сегмент глобального інформпростору, перебуває у центрі цієї трансформації. З одного боку, вона залишається інструментом міжкультурного діалогу, висвітлення геополітичних подій і формування глобального порядку денного. З іншого — стає полем битви, де новітні технології використовуються державами, політичними структурами та приватними гравцями для досягнення власних цілей.

На тлі війни в Україні, напруження між США та Китаєм, загострення на Близькому Сході, виборів у демократичних державах — інформаційні операції, засновані на ШІ, відіграють дедалі більшу роль. Вони не лише підмінюють реальність, а й змінюють саму суть сприйняття новин, підриваючи довіру до медіа як інституту.

У цьому контексті дослідження ролі ШІ як інструменту пропаганди є надзвичайно актуальним. Необхідно не просто фіксувати наявність нових технологічних явищ, а глибоко проаналізувати їхні механізми, вплив, ризики, а також потенційні шляхи протидії. Це завдання є особливо важливим для України, яка перебуває в умовах гібридної війни й щодня стикається з дезінформацією та маніпуляціями у медіапросторі.

Метою дослідження є комплексний аналіз використання технологій штучного інтелекту як інструменту пропаганди в міжнародній журналістиці, визначення ключових загроз та розробка рекомендацій щодо протидії цим викликам у контексті збереження демократичних стандартів та журналістської етики.

Завдання дослідження:

- Дослідити етапи трансформації пропаганди в історичному контексті від класичних форм до цифрової доби.
- Визначити роль ШІ у зміні механізмів комунікаційного впливу та створенні симулятивного контенту.
- Проаналізувати практичні кейси використання ШІ як пропагандистського інструменту в Росії, Китаї, США, Україні та інших країнах.
- Оцінити вплив ШІ-пропаганди на стан міжнародної журналістики, медіадовіру, аудиторну поведінку та демократію загалом.
- Розглянути існуючі технологічні й правові механізми протидії поширенню ШІ-дезінформації.

-Розробити пропозиції щодо вдосконалення політик, освітніх ініціатив та редакційних стандартів, які забезпечують стійкість медіа до впливу ШІ-пропаганди.

Об'єкт дослідження – Процеси створення, поширення та впливу пропагандистського контенту в міжнародному інформаційному просторі в умовах застосування технологій штучного інтелекту.

Предмет дослідження – Особливості використання інструментів штучного інтелекту для формування, трансляції та масштабування пропаганди в міжнародній журналістиці, а також механізми протидії цим явищам.

Методи дослідження

- Аналіз наукової літератури — для вивчення теоретичних підходів до розуміння пропаганди, журналістики та штучного інтелекту;
- Контент-аналіз — для дослідження прикладів медіаматеріалів, створених або змінених за допомогою ШІ;
- Кейс-стаді — для розгляду конкретних прикладів використання ШІ в інформаційних кампаніях (Росія, Китай, США, Україна);
- Порівняльний аналіз — для зіставлення моделей регуляції ШІ-пропагандив різних країнах;
- Метод експертного оцінювання — для аналізу думок журналістів, дослідників і фактчекерів щодо ролі ШІ у зміні стандартів медіа;
- Метод синтезу та узагальнення — для формування висновків і рекомендацій щодо протидії ШІ-пропаганді.

Структура та обсяг роботи: Магістерська робота складається зі вступу, трьох основних розділів, висновків, списку використаних джерел та додатків.

Загальний обсяг роботи становить приблизно 70–80 сторінок друкованого тексту, включаючи таблиці, приклади кейсів, інфографіку та джерела. Робота містить понад 130 наукових і публіцистичних джерел, з яких значна частина — актуальні західні дослідження, платформи аналітики, нормативні документи та приклади використання ШІ в медіа

Розділ 1. Теоретичні засади пропаганди та еволюція її інструментів

1.1. Історичний розвиток пропаганди у журналістиці

Поняття пропаганди та його еволюція. Пропаганда є формою комунікації, що полягає у цілеспрямованому поширенні ідей, поглядів, фактів або чуток з метою впливу на суспільну думку та поведінку людей на користь певної справи чи ідеї. У науковій літературі пропонуються різні визначення цього поняття – від нейтральних до негативних. У нейтральному розумінні пропаганда – це будь-яка кампанія переконування, незалежно від того, правдиву чи викривлену інформацію вона використовує. Натомість у негативному сенсі пропагандою називають маніпулятивні інформаційні кампанії, які навмисно вводять аудиторію в оману. Насправді пропаганда може мати як негативний, так і позитивний зміст: наприклад, пропаганда здорового способу життя переслідує суспільно корисну мету і не вважається аморальною. Отже, ключовою ознакою пропаганди є не тематика повідомлень, а спосіб впливу – цілеспрямований, емоційно забарвлений та часто однобічний подання інформації для досягнення заздалегідь визначеної реакції аудиторії.

Термін «пропаганда» походить від латинського *propagare* – «поширювати». Він закріпився у мовах після заснування у 1622 році у Ватикані організації *Congregatio de Propaganda Fide* («Конгрегація поширення віри»), що займалася розповсюдженням католицької віри. Спершу слово «пропаганда» не мало негативного забарвлення і позначало будь-яке розповсюдження поглядів. У XIX столітті, особливо після Французької революції, поняття набуло політичного характеру та частково негативної конотації – пропагандою почали називати однобічну агітацію в інтересах влади або партій. У XX столітті, з появою тоталітарних режимів, термін «пропаганда» остаточно утвердився в його сучасному негативному значенні. Зокрема, в СРСР існували відділи агітації та пропаганди, а в нацистській Німеччині діяло ціле Міністерство пропаганди – ці

структури відкрито використовували мас-медіа для ідеологічного контролю суспільства.

Пропаганда далеко не монолітна: її поділяють на види за різними критеріями. Наприклад, за джерелом і правдивістю інформації виділяють «білу», «сіру» та «чорну» пропаганду. «Біла» пропаганда ведеться відкрито – аудиторія знає, хто звертається, при цьому повідомлення в цілому правдиві, хоча подаються однобічно (у вигідному світлі для пропагандиста). «Сіра» – це вплив, де джерело не афішується, а достовірність інформації частково сумнівна; вона балансує на межі правди і неправди. «Чорна» пропаганда цілком приховує справжнє джерело і зазвичай складається з брехливих відомостей; часто її поширюють від імені протилежної сторони, аби дискредитувати її (класичний приклад – підроблені листівки чи документи, нібито створені ворогом, що виставляють його у негативному світлі). Крім того, пропаганду класифікують за тематикою (політична, воєнна, релігійна, комерційна тощо) та за каналами (вербальна – мовленнева або текстова, візуальна – плакати, кіно, меми, аудіо – радіо, музика). Незалежно від різновиду, пропаганді притаманний набір типових прийомів: спрощення складних проблем до зрозумілих гасел, апеляція до емоцій (страху, гніву, гордості), повторення ключових слоганів багато разів, використання стереотипів і ярликів («ворог», «зрадник», «герой»), відбір тільки зручних фактів (решта замовчується), посилення на авторитети або вигадані статистичні дані для надання ваги своїм твердженням тощо. Такі прийоми виявлені і описані дослідниками пропаганди ще в середині ХХ століття, і відтоді арсенал мало змінився – змінилися переважно носії та охоплення аудиторії, але не психологічна природа впливу.

Пропаганда і журналістські стандарти. На відміну від звичайного інформування або дискусії, пропагандистське повідомлення, як правило, має декілька характерних рис. По-перше, воно носить маніпулятивний характер: пропагандисти можуть вдаватися до перебільшень, спотворення фактів або

апеляції до емоцій, аби схилити аудиторію на свою сторону. По-друге, пропаганді притаманна однобокість висвітлення – вона подає лише ту інформацію, яка підтримує потрібні тези, замовчуючи або дискредитує альтернативні погляди. По-третє, пропаганда діє систематично і масово: вплив здійснюється не разово, а повторювано, через різні канали, охоплюючи якомога ширшу аудиторію. Нарешті, пропаганда зазвичай орієнтована на чітко визначену цільову аудиторію, і меседжі підлаштовуються під цінності та настрої певних груп. У професійній журналістиці, навпаки, цінується об'єктивність, баланс думок і перевірка фактів. Журналістський стандарт вимагає подання різних точок зору і достовірної інформації, тоді як пропаганда ставить за мету переконати будь-якою ціною – навіть якщо доводиться порушувати принципи правдивості або етики. Таким чином, історично журналістика й пропаганда розвивалися поруч, нерідко використовуючи одні й ті самі медіа-інструменти, але з різною метою та методами.

Виникнення масової пропаганди: епоха друку. Хоча елементи пропаганди існували з давнини (усні заклики ораторів, релігійні проповіді, площові агітатори), справжнього масового характеру пропаганда набула з розвитком преси. Винахід книгодрукування у XV столітті зробив можливим швидке тиражування текстів, що одразу почали використовувати для ідеологічного впливу. Уже в період Реформації та Контрреформації релігійні та політичні сили активно застосовували друковані памфлети й листівки для пропагування своїх ідей. У наступні століття, із зростанням грамотності населення, друковані газети й брошури стали головним каналом донесення політичних меседжів до масової аудиторії. З'явилися перші приклади партійної преси – видань, що відкрито обстоювали позицію певної політичної сили. Журналістика того часу часто була поєднана з пропагандою: газети нерідко слугували «рупором» урядів або рухів. Наприклад, під час Американської та Французької революцій численні відозви, памфлети і революційні газети мобілізували громадськість, формуючи потрібну

підтримку народних мас. Таким чином, уже на ранніх етапах розвитку журналістики друковане слово було зброєю пропаганди, здатною як надихнути на соціальні зміни, так і розпалити ворожнечу.

У другій половині XIX століття поширився феномен «жовтої преси» – сенсаційних масових газет, які в гонитві за увагою читачів вдавалися до перебільшень і перекручувань. Це стало передвісником використання медіа у пропагандистських цілях на комерційній основі. Яскравим прикладом була преса у США кінця XIX ст., коли окремі видання, змагаючись за тиражі, свідомо драматизували події та навіть поширювали неперевірені чутки. Така сенсаційність не лише приносила видавцям прибуток, а й впливала на суспільні настрої, іноді підштовхуючи до політичних рішень. Відомо, що перебільшені газетні репортажі про воєнні інциденти могли створювати у читачів войовничі настрої – фактично слугувати пропагандою, замаскованою під новини. Це свідчить, що ще до появи електронних медіа журналістика могла бути інструментом маніпуляції масовою свідомістю.

Перша світова війна і поява інституційної пропаганди. Початок XX століття відзначився небаченим раніше масштабом пропагандистських кампаній, що супроводжували світові конфлікти. Під час Першої світової війни воюючі держави вперше створили державні органи для інформаційного впливу на населення та війська. Так, у Великій Британії діяв Бюро воєнної пропаганди, у США – Комітет громадської інформації (так званий Комітет Джорджа Кріла). Ці структури координували випуск агітаційних плакатів, патріотичних статей у газетах, кінохронік та інших матеріалів, покликаних підтримувати бойовий дух і формувати потрібне ставлення до ворога. У ході війни пропаганда стала невід’ємною частиною воєнної стратегії: з’явилися перші теоретики пропаганди, як-от Гарольд Лассуелл, які досліджували методи масового впливу через слово та зображення. Саме тоді сформувалося розуміння пропаганди як потужної сили, здатної мобілізувати націю або деморалізувати противника.

Набутий під час війни досвід масового маніпулювання суспільною думкою почав застосовуватися і у мирний час. У 1920-х роках політичні лідери та навіть бізнес-структури активно переймають методи пропаганди для власних потреб – у передвиборчих кампаніях, рекламі, кампаніях з формування громадської думки. З’являються нові дослідники й практики переконування мас: приміром, американський експерт Едвард Бернейс, натхненний воєнними пропагандистськими успіхами, став одним із піонерів сфери *public relations* (зв’язків з громадськістю). Він відверто називав PR «пропагандою в мирний час» і застосовував психологічні прийоми впливу на аудиторію у комерційній рекламі та політичному консультуванні. У цей період слово «пропаганда» на Заході починає набувати негативних конотацій – настільки, що його уникають у публічному вжитку, замінюючи більш нейтральними термінами на кшталт «інформаційна кампанія» чи той-таки PR. Проте зміст явища не змінився: технології масового переконування продовжили розвиватися, дедалі ширше використовуючи новітні медіа.

У 1920-х роках набуває популярності радіо – новий медіум, що відкрив ще ширший доступ до аудиторії.

Пропаганда в епоху радіо. Радіомовлення стало справжнім технологічним проривом у комунікації: голоси політиків та дикторів почали одночасно звучати в тисячах домівок, долаючи відстані й неграмотність слухачів. Не дивно, що радіо швидко взяли на озброєння пропагандисти різних країн. У 1930-х роках авторитарні режими активно використовували радіоефір для утвердження своєї ідеології. В нацистській Німеччині Йозеф Геббельс – міністр пропаганди – називав радіо «восьмим великим мистецтвом» і сприяв масовому поширенню дешевих радіоприймачів, щоб кожна німецька сім’я могла слухати промови фюрера та офіційні повідомлення. У СРСР радіо також стало важливим каналом агітації: передачі з Москви транслювали партійні наративи багатьма мовами, а в містах і селах встановлювали гучномовці, щоб охопити навіть тих, хто не мав

власного приймача. Радіо дозволяло владним структурам негайно доносити потрібні меседжі і формувати громадську думку майже в режимі реального часу.

Завдяки радіо пропаганда перестала залежати від уміння людини читати – достатньо було слухати. Під час Другої світової війни радіопропаганда досягла апогею: по обидва боки конфлікту існували потужні радіостанції, які передавали новини та спеціально створені програми для деморалізації ворога або піднесення власного населення. Наприклад, нацистська Німеччина вела передачі численними мовами для інших країн, а союзники запускали контрпропагандистські радіохвилі на окуповані території. У роки війни поняття «пропагандистський фронт» стало таким же реальним, як і фронт військовий, а журналісти і диктори перетворилися на «бійців» інформаційного фронту.

У повоєнні десятиліття радіо залишалося важливим засобом інформаційного протистояння, особливо в умовах «холодної війни». Західні радіостанції (наприклад, «Голос Америки», «Радіо Вільна Європа») вели мовлення на аудиторію по той бік «залізної завіси», пропагуючи демократичні цінності та надаючи альтернативну інформацію громадянам соціалістичних країн. У відповідь радянський блок глушив ці передачі та створював потужний потік власної радіопропаганди, яка інтерпретувала події світової політики крізь призму ідеології марксизму-ленінізму. Таким чином, радіоэфір фактично став ще одним фронтом глобального ідеологічного протистояння, де журналістика і пропаганда знову тісно перепліталися.

Разом із радіо у ХХ столітті для пропаганди активно використовувалося і кіно. Візуальні образи здатні викликати сильні емоції та глибоко закарбовуватися в пам'яті, тому режими приділяли велику увагу кінематографу і фотомистецтву як засобам впливу. У нацистській Німеччині створювалися масштабні документально-постановочні фільми на кшталт «Тріумф волі» (1935) режисерки Лені Ріфеншталь – вони прославляли гітлерівський режим, використовуючи естетику масових ритуалів і героїчних образів. У Радянському Союзі

кінематограф теж став частиною державної пропаганди: кінострічки показували ідеалізоване життя радянських людей, історичні фільми переписували минуле згідно з партійною лінією, а військові драми звеличували подвиги Червоної армії. В демократичних країнах часів Другої світової війни уряди також залучали кіно: випускалися патріотичні новинні хроніки, навчальні фільми і плакати, що пояснювали населенню необхідність тих чи інших дій (економії, евакуації тощо) та демонізували ворога. Фотографія і плакат доповнювали кінематограф: відомі воєнні плакати зі слоганами на кшталт «*Loose Lips Sink Ships*» («Бездумні розмови топлять кораблі») нагадували громадянам про обережність, а карикатури висміювали лідерів противника. Таким чином, пропаганда використовувала цілий комплекс медіа-засобів, щоб оточити людину з усіх боків потрібними повідомленнями – у текстах, звуках, зображеннях.

Телебачення як інструмент впливу. Після війни наступним етапом еволюції медіа стало телебачення. У 1950–1960-х роках телебачення перетворилося на головне джерело новин та розваг для широких мас. Візуальна природа телебачення ще більше посилила вплив медіа на аудиторію, адже «картинка» здатна викликати сильні емоції та створювати ефект присутності. Держави зрозуміли значення телебачення як інструменту контролю над інформаційним простором. У країнах із тоталітарними режимами телебачення перебувало під жорстким контролем цензури і повністю підпорядковувалося цілям пропаганди. Наприклад, радянське центральне телебачення подавало відфільтровану картину реальності, покликану зміцнювати довіру до влади та формувати образ «щасливого життя» в СРСР. У демократичних суспільствах телебачення теж не уникло впливу пропагандистських тенденцій: політики навчилися використовувати теледебати, інтерв'ю та рекламні ролики для просування своїх меседжів. Зародився феномен політичного іміджмейкінгу – створення привабливого образу кандидата через телекамеру. Водночас розслідувальна журналістика на телебаченні продемонструвала, що вільні медіа здатні кидати

виклик офіційній пропаганді: репортажі з В'єтнамської війни наприкінці 1960-х, показуючи реальні жахи конфлікту, підірвали урядову версію подій і сприяли антивоєнним настроям у США. Цей випадок продемонстрував двоїстість телебачення: з одного боку, влада прагнула використовувати його як канал впливу, з іншого – сміливі журналісти могли перетворити його на інструмент викриття.

У другій половині XX століття, з розвитком супутникового мовлення та появою перших глобальних телеканалів (CNN та ін.), інформація почала миттєво долати кордони держав. Це означало, що монополія влади на пропаганду у межах однієї країни послабилася: аудиторія отримала доступ до альтернативних джерел інформації. Проте водночас уряди та різні групи навчилися використовувати і глобальні медіа для просування своєї точки зору на міжнародній арені. Телебачення стало ареною інформаційних війн: так, під час «холодної війни» західні та радянські теленовини давали діаметрально протилежні інтерпретації одних і тих самих подій, формуючи у глядачів двох блоків кардинально різні картини світу.

Перехід до цифрових форм впливу. Кінець XX – початок XXI століття приніс чергову революцію у сфері медіа – стрімкий розвиток цифрових технологій та інтернету. Журналістика поступово мігрувала в онлайн-простір, а разом із нею туди перемістилася й пропаганда. Якщо друк, радіо і телебачення дозволяли впливати переважно за схемою «від одного – багатьом» (коли один мовник або видання транслює повідомлення масовій аудиторії), то інтернет створив середовище комунікації «багато – багатьом». Соціальні мережі, блоги, форуми дали змогу будь-кому поширювати інформацію практично без контролю з боку традиційних інститутів. Це відкрило нові можливості для пропагандистів, але й поставило нові виклики перед журналістикою.

На початку цифрової доби існував оптимізм щодо демократизуючого потенціалу інтернету: вважалося, що вільний обмін інформацією сприятиме

плюралізму та прозорості. Частково так і сталося – отримати доступ до різних точок зору стало легше. Проте дуже швидко з'ясувалося, що ті ж самі інструменти онлайн-комунікації можуть використовуватися для масового поширення дезінформації, теорій змови та мови ворожнечі. Пропаганда адаптувалася до нових умов: почали з'являтися фейкові новинні сайти, що імітують справжні ЗМІ, фабрики тролів у соціальних мережах, координовані кампанії з поширення вигідних наративів через тисячі бот-акаунтів. Інтернет-простір переповнився контентом, і в цій перенасиченості пропагандистські меседжі можуть непомітно вбудовуватися у загальний потік інформації. До того ж, алгоритми платформ (про які докладніше йтиметься далі) навчилися підбирати користувачам контент на основі їхніх вподобань, що створило феномен інформаційних «бульбашок». У таких умовах відрізнити якісну журналістику від маніпулятивного контенту стає дедалі складніше.

Особливо виразно роль пропаганди проявляється у XXI столітті під час геополітичних конфліктів та криз: інформаційні атаки, кампанії дезінформації і масові «вкиди» в соціальних мережах стали невіддільною складовою так званих гібридних воєн. Це підкреслює, що еволюція пропаганди далека від завершення – навпаки, у поєднанні з новітніми технологіями вона набуває дедалі витонченіших форм. Відповідно, перед міжнародною спільнотою постає завдання вироблення ефективних механізмів протидії інформаційним маніпуляціям і захисту простору правди.

Отже, історичний розвиток пропаганди тісно пов'язаний зі зміною медіаінструментів. Кожна нова епоха – друкована, радіо-, телевізійна, цифрова – не лише не витіснила пропаганду, а навпаки, дала їй нові форми та канали впливу. Журналістика, яка зародилася як пошук правди та інформування громадськості, на кожному етапі змушена була співіснувати з пропагандою та протистояти їй, виробляючи професійні стандарти і етику. У сучасній цифровій добі ці стандарти проходять чергове випробування під натиском еволюціонованих

пропагандистських інструментів – від персоналізованої реклами до мережових ботів та штучного інтелекту.

1.2. Пропаганда vs. журналістика в умовах цифрової доби

У сучасному цифровому медіапросторі питання відмінності між пропагандою та журналістикою набуло особливої гостроти. Обсяги інформації, з якими щоденно стикається аудиторія, зросли багаторазово, і в цьому потоці справді важко відрізнити неупереджене повідомлення від маніпулятивного. Журналістика традиційно прагне до об'єктивного інформування: її метою є донести до суспільства факти, перевірені дані і різні точки зору, щоб громадяни могли робити власні висновки. Пропаганда ж натомість ставить за мету переконати аудиторію у заздалегідь заданій позиції або спонукати до певної дії; для цього вона не гребує селективним добором фактів чи навіть відвертою неправдою. Якщо журналістика апелює передусім до раціонального аналізу та суспільної значущості інформації, то пропаганда апелює до емоцій, стереотипів і інстинктів, часто спрощуючи складні явища до зрозумілих гасел.

Відмінною є і функція: журналістика у демократичному суспільстві виконує роль «сторожового пса», контролюючи владу та викриваючи зловживання, тоді як пропаганда зазвичай є інструментом влади чи зацікавлених груп для контролю суспільства. Інакше кажучи, журналістика прагне зробити громадян більш обізнаними і здатними критично мислити, а пропаганда – більш слухняними і схильними приймати нав'язані рішення. Наприклад, журналіст стримано повідомить: «Унаслідок сутички загинули 5 осіб», тоді як пропагандист із тією ж подією використає емоційно заряджені формулювання: «Ворог жорстоко вбив 5 безневинних людей». У першому випадку слова добираються обережно і нейтрально, щоб не нав'язувати оцінок, тоді як у другому вжито лексику, що викликає обурення і однозначно маркує винного – саме так пропаганда підкорює емоції аудиторії.

В умовах цифрової доби межа між журналістським контентом та пропагандистським стала розмитішою, ніж будь-коли. По-перше, кожен користувач інтернету може виступати творцем і розповсюджувачем інформації. Це означає, що на одній платформі (наприклад, у стрічці соціальної мережі) поруч можуть з'являтися як матеріали авторитетного новинного видання, так і дописи анонімних блогерів чи пропагандистські мему. Зовні вони можуть виглядати подібно – заголовки, фото, відео – і пересічному читачеві дедалі складніше відразу зрозуміти, кому довіряти. По-друге, пропаганда маскується під журналістику. Створюються псевдо-новинні сайти, які імітують формат новинних порталів, але поширюють дезінформацію або тенденційну інтерпретацію подій. Такі ресурси нерідко використовують зовні нейтральний тон і атрибути журналістики (інтерв'ю, аналітичні статті), проте їхній зміст слугує чийось політичним або ідеологічним інтересам. Як наслідок, аудиторія, не озброєна медіаграмотністю, може приймати пропагандистський продукт за незалежну журналістику.

Відмінності у принципах та етиці. Журналісти у своїй діяльності керуються певними професійними принципами: достовірність, перевірка фактів, баланс думок, відокремлення фактів від коментарів, відповідальність перед аудиторією. Натомість пропагандисти керуються передусім ефективністю впливу – важливо не *як* подано, а *чого* досягнуто. У цифрову епоху це протиріччя проявляється особливо виразно. Наприклад, етичний журналіст, дізнавшись новину, прагне спершу перевірити її джерела та контекст, тоді як пропагандист може негайно запустити неперевірену інформацію, якщо вона вписується в потрібну йому картину світу. Швидкість обігу новин онлайн створює тиск і на традиційні медіа: гонитва за оперативністю і кліками іноді спонукає навіть поважні видання публікувати матеріали з неповною перевіркою або з надто гучними заголовками. Це розмиває межу між сумлінною журналістикою та елементами пропаганди, адже маніпулятивні прийоми (на кшталт clickbait-заголовків) стали

використовувати і деякі легітимні ЗМІ для привернення уваги аудиторії. Таким чином, економічні та технологічні реалії цифрового ринку інформації ставлять журналістів перед етичним вибором: дотримуватися високих стандартів і ризикувати втратити охоплення, або ж підлаштовуватися під масові смаки і алгоритми, ризикуючи втратити якість та довіру.

Дискредитація журналістики та “постправа”. Складність ситуації посилюється тим, що сучасна пропаганда веде цілеспрямовану атаку на самі засади незалежної журналістики. Популярною тактикою стало звинувачення провідних медіа у брехливості чи упередженості – риторика про «фейкові новини» застосовується, парадоксально, самими дезінформаторами, щоб посіяти недовіру до об’єктивних джерел. Коли аудиторії нав’язують думку, що «усі брешуть» і «правди ніде немає», значна частина людей може втратити орієнтири. Це і є підґрунтя феномену епохи постправа, коли суспільство втомлюється шукати об’єктивність і більше реагує на емоційно привабливі заяви, ніж на реальні факти. У такій атмосфері пропагандистам значно легше просувати свої наративи, адже контраргументи журналістів автоматично сприймаються частиною публіки як «також пропаганда з іншого боку». Для демократії це становить серйозну проблему: якщо громадяни не довіряють жодним медіа, поле для раціональної дискусії звужується, поступаючись місцем теоріям змови та популістським гаслам.

Ще одним тривожним явищем у цифровому середовищі стало цілеспрямоване переслідування журналістів онлайн. Пропагандистські мережі нерідко організовують кампанії тиску на незалежних репортерів через соціальні платформи: від масових хвиль образ і погроз у коментарях до поширення наклепів і зламу особистих сторінок. Мета таких атак – дискредитувати журналіста в очах публіки, психологічно на нього тиснути й змусити відмовитися від гострих тем. Особливо часто цього зазнають журналісти-розслідувачі та ті, хто критикує авторитарні режими. Реагуючи на ці виклики, журналістська

спільнота та правозахисні організації виробляють механізми підтримки – від юридичної допомоги і кібербезпеки до публічної солідарності з постраждалими. Проте проблема залишається актуальною: інформаційні війни ведуться не лише на рівні контенту, а й через персональні атаки, що підривають свободу слова.

Алгоритми як нові “редактори”. Однією з ключових особливостей цифрової медіадоби є алгоритмічна фільтрація інформації. Великі платформи – такі як Google, Facebook, Twitter, YouTube – використовують алгоритми для персоналізації стрічки новин під кожного користувача. Ці алгоритми аналізують поведінку, вподобання, пошукову історію і на основі цього вирішують, який контент показати, а який – ні. З одного боку, це робить споживання інформації зручнішим: читач бачить здебільшого те, що йому цікаво. З іншого боку, виникає явище “фільтраційних бульбашок” – інформаційних ізоляцій, коли людина отримує лише ті новини і погляди, які відповідають її попереднім вподобанням. У такій бульбашці користувач майже не стикається з альтернативними точками зору, а отже, не має нагоди перевірити або скоригувати свої переконання. Для пропаганди це сприятливий ґрунт: алгоритми самі підсовують людині контент, що підтверджує її світогляд. Якщо цей світогляд сформований хибними наративами, то алгоритмізоване медіаполе лише посилює оману, замість того щоб її розвінчати.

Окрім пасивного формування «інформаційних бульбашок», персоналізовані алгоритми відкрили шлях до мікротаргетингу – надточного налаштування пропагандистських повідомлень під конкретні групи людей. Сучасні платформи дозволяють рекламодавцям (у тому числі й політичним партіям) показувати інформацію лише певним сегментам аудиторії, відібраним за інтересами, демографією чи моделями поведінки онлайн. Це означає, що пропагандист може транслювати різним групам різні меседжі, спеціально адаптовані під їхні вподобання та страхи. Подібна практика утруднює викриття маніпуляцій: широка громадськість може навіть не підозрювати, які тези

навіюють окремим спільнотам. Відомі міжнародні скандали (наприклад, випадок з компанією *Cambridge Analytica*) продемонстрували, що мікротаргетована політична пропаганда в соцмережах здатна впливати на результати виборів та референдумів, залишаючись водночас поза полем зору традиційних регуляторів і спостерігачів.

Наслідком роботи алгоритмів з пріоритизації контенту є і те, що емоційно забарвлені, скандальні матеріали часто отримують більше видимості, ніж спокійні аналітичні репортажі. Механізм простий: алгоритм відбирає те, на що користувачі активніше реагують лайками, коментарями, поширеннями. А з досвіду відомо, що люди бурхливо реагують на контент, який шокує або обурює. Тому сенсаційні новини, теорії змови чи провокаційні заяви збирають високі показники взаємодії – і платформи ще більше їх підносять у стрічках, щоб утримати увагу користувачів. Так виникає порочне коло: клікбейт-заголовки і спрощені, поляризовані наративи отримують алгоритмічне підживлення, тоді як ґрунтовні журналістські матеріали можуть лишитися на узбіччі уваги. Це ставить журналістику перед моральною дилемою: чи варто підлаштовуватися під алгоритми, випускаючи більш «вірусний» контент (ризикуючи скотитися до табloidності), чи все-таки триматися стандартів якості, сподіваючись, що аудиторія цінуватиме глибину і правдивість?

Проблема правди та довіри. В алгоритмізованому медіаполі підтримання правдивості інформації стало спільним викликом для журналістів, технологічних компаній і суспільства загалом. Журналісти змушені шукати нові підходи, щоб достукатися до аудиторії, яка розпорошена по “бульбашках” та перевантажена потоками даних. Одним зі шляхів стало активне впровадження фактчекінгу – перевірки фактів, спростування популярних фейків, маркування неправдивої інформації. Провідні світові медіа та незалежні організації запускають проекти з боротьби з дезінформацією (наприклад, українська ініціатива *StopFake* з 2014 року викриває та спростовує неправдиві повідомлення в медіа, а міжнародні

проекти на кшталт *PolitiFact* чи *Snoopes* займаються перевіркою фактів у політичній риторичі), створюють розділи, де аналізують і розвінчують пропагандистські вкиди. Також виникає нова галузь – медіаграмотність: освітні програми і кампанії, що навчають аудиторію розрізняти надійні джерела, критично оцінювати контент і протистояти маніпуляціям.

Окремо слід згадати про роль глобальних інтернет-платформ у цьому контексті. Упродовж тривалого часу такі соціальні мережі, як Facebook, Twitter, YouTube, декларували нейтралітет щодо змісту: мовляв, вони лише технічні платформи, а не медіа, і не несуть відповідальності за правдивість того, що публікують користувачі. Такий підхід фактично зняв з них обов'язок фільтрувати пропаганду. Це створило сприятливі умови для дезінформаторів, адже їхній контент поширювався мільйонам людей без будь-якого редакційного контролю. Лише після низки резонансних випадків – від втручань у вибори через соцмережі до випадків масового насильства, спровокованих фейковими новинами – технокомпанії почали визнавати необхідність принаймні часткової модерації. Зараз тривають дискусії про те, чи повинні платформи нести відповідальність за поширення неправдивої інформації так само, як традиційні медіа відповідають за свій контент. Це етичне і правове питання залишається відкритим, але його вирішення суттєво вплине на баланс між свободою слова та інформаційною безпекою.

Показовим прикладом стала «інфодемія» часів пандемії COVID-19: соціальні мережі переповнилися фейковими порадами щодо лікування, теоріями змови про походження вірусу та антивакцинаторською пропагандою. Ці вкиди інколи набували більшого розголосу, ніж спростування від науковців, і спонукали людей відмовлятися від засобів захисту. Журналісти й медіаорганізації були змушені екстрено налагоджувати фактчек і просвітницьку роботу, співпрацювати з платформами для маркування неправдивого контенту. Ситуація з «інфодемією» показала, що помилкова інформація в

алгоритмізованому середовищі може поширюватися швидше, ніж правдиві роз'яснення, створюючи реальну загрозу здоров'ю суспільства.

Водночас самі цифрові платформи поступово визнають відповідальність за якість інформаційного середовища. Після низки скандалів, пов'язаних із поширенням дезінформації (від втручання у вибори до небезпечних порад під час пандемії), деякі соцмережі почали змінювати алгоритми та правила модерації. Наприклад, впроваджується позначення сумнівного контенту, зменшення охоплення відверто фейкових постів, пріоритизація достовірних джерел у видачі. Однак ці кроки викликають суперечки щодо цензури та свободи слова: деякі користувачі звинувачують платформи в упередженості, коли ті блокують пропагандистські матеріали. Таким чином, баланс між свободою вираження і захистом правди став центральною етичною проблемою сучасного медіаполя.

Отже, журналістика і пропаганда в цифрову епоху стикаються у складному двобої. Журналісти намагаються зберегти вірність фактам та етиці, тоді як пропагандисти використовують всі переваги технологій для посилення свого впливу. Алгоритми та нові медіа не можна розглядати як нейтральне тло – вони самі стали активними чинниками, що впливають на розподіл правди і брехні в інформаційному просторі. У цих умовах суспільству вкрай важливо підтримувати якісну журналістику, розвивати критичне мислення і вимагати прозорості від технологічних гігантів, адже від цього залежить, чи зможе публіка отримувати правдиву картину світу, чи тонути в океані перекручених дзеркал пропаганди.

1.3. Технологічний прорив: поява ШІ як нового інструменту впливу

Стрімкий розвиток технологій штучного інтелекту (ШІ) наприкінці 2010-х – початку 2020-х років відкрив для медіасфери як нові можливості, так і нові загрози. Штучний інтелект почав застосовуватися у журналістиці та масових комунікаціях на різних рівнях – від автоматизації рутинних процесів до

створення повністю нового типу контенту. Цей технологічний прорив значно посилив арсенал інструментів, доступних і журналістам, і пропагандистам.

III-технології в медіа: Одним із перших напрямів використання штучного інтелекту в медіагалузі стало автоматичне генерування текстів на основі даних. Такі системи, відомі як *робожурналістика*, вже зараз пишуть біржові огляди, звіти про спортивні матчі чи прогнози погоди, підставляючи актуальні цифри у заздалегідь задані шаблони. Великі інформаційні агенції впровадили алгоритми, що здатні за лічені секунди створювати короткі новинні замітки на базі статистичної інформації чи офіційних пресрелізів. Наприклад, агентства *Associated Press* та *Reuters* експериментують з генерацією фінансових новин за допомогою III, щоби оперативніше охоплювати рутинні теми. Хоча такі тексти поки що досить шаблонні, їхнє використання дозволяє журналістам зосередитися на складніших аналітичних завданнях. Водночас виникає питання: чи не відкриває це скриньку Пандори для пропаганди? Адже якщо алгоритм може автоматично продукувати новини, його можна налаштувати і на продукування маніпулятивних повідомлень у масових масштабах.

Дійсно, сучасні генеративні моделі III (такі як просунуті мовні моделі GPT-типу) здатні створювати розлогі тексти, стилістично не відрізнити від написаних людиною. Це означає, що поширення пропаганди може бути поставлене на конвеєр: програма вмиє згенерує сотні варіантів статті або допису в соціальній мережі, які підтримують потрібний наратив, вносячи невеликі зміни, щоб обійти фільтри. Боти, керовані III, можуть масово постити такі повідомлення, удаючи активність тисяч «користувачів». Таким чином створюється *ілюзія консенсусу* – у людей складається враження, що «усі так думають». До того ж, III дозволяє кожному такому боту швидко реагувати на новини дня, перекручуючи або спотворюючи їхню інтерпретацію. Така автоматизація пропаганди радикально підвищує її масштаб та ефективність у порівнянні з минулими епохами, коли для створення дезінформації була потрібна

значна участь людей. Практичні експерименти вже підтверджують дієвість такої автоматизованої пропаганди. За даними досліджень у 2024 році, тексти, згенеровані великою мовною моделлю на підтримку певної тези, виявилися в середньому не менш переконливими для читачів, ніж аналогічні тексти, написані людьми. Державні органи деяких країн також активно випробовують ШІ для інформаційних операцій: зокрема, у ЗМІ повідомлялося, що китайські структури генерували тисячі соцмережових дописів через ШІ з метою вплинути на громадську думку за кордоном.

Штучний інтелект також зробив можливим появу принципово нового типу маніпулятивного контенту – так званих діпфейків. Йдеться про фото-, аудіо- та відеоматеріали, створені за допомогою нейромереж, які надзвичайно реалістично імітують реальних людей та події. Діпфейк може показувати політика, який нібито говорить те, чого він насправді не говорив, або «кадри» події, якої насправді не було. Раніше для викриття фотомонтажу чи підробленого відео достатньо було знайти грубі візуальні сліди редагування, але сучасні алгоритми генерують зображення з нуля, без явних артефактів. Вже зафіксовані випадки застосування діпфейків у пропагандистських цілях: від фальшивих відеозвернень від імені державних лідерів до псевдосвідчень «очевидців» катастроф. Зокрема, під час війни 2022 року було виявлено діпфейкове відео, де президент України нібито оголошує капітуляцію – очевидно, створене ворогом з метою деморалізації. Хоч цей ролик оперативно викрили та видалили, сам його факт продемонстрував небезпечний потенціал таких технологій у веденні інформаційної війни. Такий інструмент здатен підірвати довіру до будь-якого візуального доказу – адже глядач більше не впевнений, чи не є побачене на екрані майстерною підробкою. Журналістика змушена відповідати на цей виклик розвитком методів перевірки автентичності контенту, залучаючи ті самі ШІ-технології для детекції аномалій, аналізу метаданих тощо. Проте це гонка

озброєнь: як тільки створюються засоби розпізнавання діпфейків, розробники навчають моделі обходити ці фільтри.

Алгоритми, що керують увагою. Штучний інтелект проявляє свій вплив не лише у створенні контенту, а й у його розповсюдженні та відборі. Як вже зазначалося, платформи використовують алгоритми машинного навчання, щоб вирішувати, який допис показати користувачеві першим, а який можливо і не показувати взагалі. Ці алгоритми виконують роль невидимих редакторів, оптимізуючи стрічку під інтереси кожного – але й під бізнес-моделі компаній, що прагнуть максимально утримати увагу. Саме вони породили такі явища, як фільтраційні бульбашки, коли люди живуть у медіасвіті, що підтверджує їхні упередження, та культуру клікбейту, коли заголовки зумисне роблять гучними і провокативними заради залучення кліків. Персоналізація контенту за допомогою ШІ двосічна: з одного боку, вона підвищує релевантність новин для читача, з іншого – звужує кругозір і може посилювати радикальні настрої. Наприклад, відомо, що алгоритми відеоплатформ можуть «затягувати» глядача дедалі більш екстремальним контентом, варто лише почати з невинного пошуку – система на основі попередніх патернів інших користувачів припускає, що радикальніші відео довше утримають вашу увагу, і підсовує їх. Таким чином, технології ШІ мимоволі можуть ставати своєрідними «прискорювачами» пропаганди, якщо їхня метрика успіху – це просто максимальний час, проведений користувачем онлайн.

Персоналізація та мікропропаганда. Застосування ШІ для глибокого аналізу великих даних аудиторії привело до появи надзвичайно точних інструментів мікротаргетингу, про який згадувалося вище. Нині пропагандист може не тільки загально розраховувати на сприятливу «бульбашку», але й формувати окремі повідомлення під аудиторію з дуже специфічними характеристиками – аж до рівня окремої людини. Алгоритми аналізують історію пошуків, вподобання, соціальні зв'язки користувача і прогнозують, який

аргумент чи образ найімовірніше вплине саме на нього. На практиці це означає, що двоє людей, сидячи поруч, можуть бачити в своїх стрічках абсолютно різні новини та повідомлення, спеціально підібрані під їхні профілі. Один, наприклад, бачитиме переважно новини, що вселяють страх і підтверджують певну конспірологічну теорію, а інший – оптимістичні повідомлення на ту ж тему, які закликають до підтримки певних дій. Кожному демонструється своя «реальність». Така фрагментація інформаційного поля, керована ШІ, ставить перед суспільством важке питання: як досягти спільного розуміння фактів, якщо різні групи населення отримують різні версії картини світу?

Проте слід мати на увазі, що алгоритми ШІ не є повністю нейтральними чи об'єктивними. Вони вчаться на великих обсягах існуючого контенту, який часто містить і певні упередження, і неточності. Таким чином, штучний інтелект може ненароком тиражувати та підсилювати вже присутню в інформаційному полі упереджену перспективу. Якщо, скажімо, онлайн переважають якісь хибні чи екстремістські твердження з певної теми, модель, натренована на цих даних, схильна сприймати їх як «нормальні» й продукувати текст у тому самому ключі. Це виклик для розробників: необхідно впроваджувати в ШІ фільтри та етичні обмеження, щоб запобігти автоматичному розмноженню пропагандистських штампів. Деякі великі компанії вже намагаються розв'язати цю проблему, навчаючи моделі розпізнавати токсичний чи неправдивий контент і уникати його поширення. Але повністю виключити людський нагляд поки що неможливо – завжди лишається ризик, що складний алгоритм знайде обхідний шлях і видасть небажаний результат.

Звичайно, ШІ-технології можуть використовуватися і на благо журналістики та суспільства. Вони вже допомагають у моніторингу величезних масивів даних, виявленні трендів, перевірці фактів (алгоритми можуть автоматично співставляти твердження політика з наявними базами даних і сигналізувати про ймовірну брехню). ШІ здатен відфільтровувати коментарі з мовою ворожнечі,

підтримувати модерацію дискусій, перекладати інформацію різними мовами в реальному часі, що розширює доступ аудиторії до контенту. Приміром, великі редакції використовують інструменти з елементами ШІ, щоб швидко сортувати тисячі документів у журналістських розслідуваннях (як це було зі «зливами Panama Papers»), або аби аналізувати супутникові знімки при розслідуванні воєнних злочинів. Ці досягнення демонструють конструктивний потенціал штучного інтелекту для підсилення якісної журналістики.

Однак ті самі технології у руках недобросовісних гравців перетворюються на зброю масового дезінформаційного ураження. Світ вже зіткнувся з прикладами масштабних кампаній, де боти та фейкові акаунти, підживлювані ШІ-контентом, впливали на громадську думку цілих країн. Штучний інтелект став у пригоді й для цензури та інформаційного контролю. Держави з авторитарними режимами впроваджують алгоритми для моніторингу та фільтрації онлайн-контенту: машини аналізують мільйони повідомлень у соцмережах і миттєво блокують ті, що містять «небезпечні» для влади ключові слова або зображення. Таким чином, ШІ не лише допомагає поширювати потрібні наративи, але й пригнічує альтернативні голоси. У результаті суспільство отримує спотворену картину реальності не лише через присутність пропаганди, а й через відсутність критичних чи опозиційних повідомлень, які були відфільтровані алгоритмами цензури. Подібна технологічно підсилена цензура вже діє, за повідомленнями, у ряді країн, створюючи безпрецедентний за масштабом і точністю апарат придушення свободи слова.

Відтак, постає необхідність етичного та правового регулювання застосування ШІ у інформаційній сфері. Деякі країни та соцмережі починають позначати контент, створений ШІ, розробляються міжнародні кодекси та стандарти щодо протидії дезінформації, підтримуваної ШІ. Вчені працюють над алгоритмами, що могли б автоматично виявляти пропагандистські повідомлення

(аналізуючи лексику, логічні хиби, неприродну активність акаунтів тощо). Але поки що ці зусилля лише наздоганяють реальність.

ШІ як виклик і шанс. Поява штучного інтелекту як інструменту впливу – це водночас виклик і нова можливість. З одного боку, пропаганда отримала у своє розпорядження неймовірно потужні технології, що дозволяють обманювати, навіювати та розколювати суспільство з небаченою швидкістю і масштабом. З іншого боку, ті ж технології можуть послужити людству – допомагаючи виявляти неправду, просвіщати аудиторію, створювати більш персоналізований, але чесний медіапростір. Ключовим фактором залишаються людські рішення: як саме ми застосуємо ШІ. Якщо суспільство виробить дієві механізми контролю і етики для ШІ в медіа, то цей інструмент стане союзником, а не ворогом журналістики. В противному разі ризикуємо опинитися в реальності, де розрізнити правду від імітації буде практично неможливо, а інформаційний хаос підірве самі основи громадянського життя. Таким чином, еволюція інструментів пропаганди досягла якісно нового рівня – і відповідь на це має бути такою ж новаторською та скоординованою.

РОЗДІЛ 2. Практики використання ШІ для пропаганди в міжнародній журналістиці

2.1. Генеративні моделі, deepfake та дезінформація

Штучний інтелект (ШІ) стрімко став інструментом, що здатен створювати переконливий фейковий контент у різних формах – від текстових «новин» до фотографій і відео. За останні п'ять років (2020–2025) генеративні моделі на кшталт GPT-3/4, неймережі Midjourney, DALL-E, а також алгоритми для дипфейк-відео стали доступними широкому загалу. Це означає, що не лише

державні агенції, а й окремі особи можуть генерувати правдоподібні матеріали, вводячи аудиторію в оману. Застосування таких технологій у пропаганді докорінно змінює інформаційний ландшафт міжнародної журналістики, адже межа між реальними та штучно створеними повідомленнями швидко стирається. Вже сьогодні існують цілі мережі веб-сайтів, де «новини» пишуться не журналістами, а мовними моделями – причому без перевірки фактів. Незалежні експерти фіксують стрімке зростання кількості таких ресурсів: за даними моніторингової компанії, станом на початок 2025 року у світі виявлено понад 1000 веб-сайтів, контент на яких повністю або переважно генерується ШІ. Багато з них спеціалізуються на сенсаційних і пропагандистських матеріалах, маскуючись під звичайні медіа.

Генерація фейкових новин і зображень. Сучасні генеративні моделі здатні писати статті, що виглядають як репортажі авторитетних видань, або створювати фейкові дописи у соціальних мережах від імені вигаданих «очевидців». Наприклад, мовні моделі можуть згенерувати детальний текст «новини» про вигадані події, підкріпивши її сфабрикованими «коментарями очевидців» та псевдо-фактами. Одночасно графічні нейромережі генерують фотореалістичні зображення: від «документальних» фото бойових дій чи протестів, яких насправді не було, до вигаданих портретів громадських діячів. Відомим став випадок, коли у 2023 році мережу сколихнуло фото Папи Римського Франциска у модному білому пуховику – зображення було створене нейромережею Midjourney, але мільйони людей прийняли його за справжнє. Подібно, в інтернеті вірусно поширювалися «фотографії» нібито з арештом Дональда Трампа, згенеровані за допомогою ШІ: їхня висока якість та драматичність змусили багатьох користувачів повірити у реальність побаченого. Отже, штучний інтелект дозволяє пропагандистам масово тиражувати зображення, що викликають потрібні емоції і формують альтернативну версію реальності.

Варто зауважити, що поняття «deepfake» виникло лише наприкінці 2010-х років, коли ентузіасти вперше продемонстрували можливість реалістично замінювати обличчя на відео за допомогою нейромереж. Спершу такі технології використовувалися у розважальному або навіть злочинному контексті (наприклад, для скандальних фальшивих відео за участі знаменитостей), але згодом проникли і в політичну сферу. Уже в 2022 році аналітики застерігали, що світ ще не бачив найгірших проявів епохи фейкових новин – і подальший прогрес ШІ, що зробив генерацію контенту масовим явищем, підтвердив ці побоювання.

Діпфейк-відео як інструмент дезінформації. Окремо варто відзначити феномен діпфейків – відео або аудіо, де обличчя та голос людини замінено ШІ на основі інших даних. З 2020 року такі технології стали настільки доступними, що їх почали застосовувати в інформаційних війнах. Найбільш резонансний приклад – підроблене відеозвернення президента України Володимира Зеленського у березні 2022 року. Хакери поширили ролик, у якому нібито Зеленський закликає українських військових скласти зброю і здатися; це відео було згенероване ШІ та розміщене на зламаному новинному сайті. Хоча «Зеленський» виглядав дещо неприродно і говорив з невластивим йому акцентом, діпфейк міг дезорієнтувати частину глядачів, особливо якщо вони побачили його побіжно. Українська влада оперативно спростувала цей фейк, а сам Президент записав звернення з поясненням, що жодної капітуляції не буде. Цей випадок став першим підтвердженим використанням діпфейк-технології для впливу на реальний перебіг війни, наочно продемонструвавши новий рівень інформаційної загрози. Невдовзі після цього інформаційного інциденту почали з'являтися і інші тривожні сигнали: так, у 2022–2023 роках у соцмережах циркулювали фальшиві «заяви» від імені лідерів держав про нібито укладення миру чи, навпаки, про ескалацію конфліктів – переважно це були діпфейк-відео невисокої якості, які втім могли ввести в оману необізнаних громадян. Показовий приклад – у 2022 році мера Берліна ошукали через відеодзвінок із «відомим співрозмовником»:

вона поспілкувалася з особою, яка дуже переконливо представилася мером Києва Віталієм Кличком, але згодом з'ясувалося, що це була майстерна підробка. Таким чином зловмисники тестують межі довіри: навіть високопосадовці стають жертвами технологічно витончених розіграшів, які межують з пропагандистськими операціями.

Крім того, останніми роками фіксуються й інші способи застосування дідфейків для маніпуляцій. Напередодні виборів у різних країнах з'являлися відеоролики, де кандидати нібито роблять скандальні заяви або «підловлені» на порушеннях – згодом викривалося, що ці матеріали штучно створені, аби очорнити репутацію. У США в 2024 році виборці побачили передвиборчі відео, повністю згенеровані нейромережею: окремі партії використали ШІ для ілюстрації негативних сценаріїв за певного розвитку подій, намагаючись вплинути на емоції виборців (детальніше про це – у підрозділі 2.3 щодо США). Хоча багато таких матеріалів зрештою викриваються фактчекерами чи журналістами-розслідувачами, навіть короткочасний ефект від них може бути значним – від очорнення опонента до породження загальної невпевненості.

Психологічний вплив та дезорієнтація аудиторії. Реалістичність, із якою ШІ створює фейки, має потужний психологічний ефект. Люди традиційно довіряють побаченому на власні очі чи почутому від, здавалося б, надійного джерела. Тому добре зроблений фейковий репортаж або світлина може викликати у глядача сильні емоції – шок, обурення, страх – і закарбуватися в пам'яті раніше, ніж прийде спростування. Пропаганда цілеспрямовано експлуатує такі реакції: наприклад, сфабриковані ШІ фото начебто «жертв» чи «звірств» у ході конфлікту здатні розпалити ненависть до однієї зі сторін і схилити громадську думку. Так, під час війни між Ізраїлем та угрупованням ХАМАС у 2023 році соцмережами масово ширилися зображення поранених дітей і зруйнованих міст, частина яких виявилася згенерованою нейромережею – такі кадри підігрівали гнів аудиторії, хоча реальний контекст міг бути спотвореним.

Брак навичок критичного мислення у значної частини аудиторії призводить до того, що багато хто приймає побачені фейки за чисту монету. У результаті дезінформація, підсилена ШІ, здатна дезорієнтувати суспільство: люди починають сумніватися в достовірності будь-яких новин, або навпаки – беззастережно вірять емоційним вкидам, які відповідають їхнім упередженням. Виникає небезпечне явище, коли неправдива історія виглядає настільки правдоподібно, що навіть після її спростування частина аудиторії все одно залишається впевненою в її правдивості. Дослідники медіа вже відзначають цей ефект: глядачі, які ознайомилися з діпфейк-відео, надалі починають менше довіряти навіть справжнім новинам. Більш того, поширення діпфейків та інших ШІ-викривлень підриває саму концепцію доказів. В інформаційному дискурсі з'явився термін «дивіденд брехуна»: тепер політики чи публічні особи, проти яких оприлюднюють компрометуючі відео чи аудіо, можуть зручним чином заявити, що це «підробка, згенерована комп'ютером». Така риторика додатково розмиває межу між правдою й вигадкою – аудиторія часто вже не знає, кому і чому вірити. Протистояти цьому психологічному ефекту надзвичайно складно: освітні програми з медіаграмотності та попередження від експертів дають певний результат, але не можуть миттєво охопити всіх.

У підсумку, активне застосування генеративного ШІ для створення фейкового контенту породжує безпрецедентний виклик для міжнародної журналістики. З одного боку, нові технології дають пропагандистам потужні інструменти для масового і навіть персоналізованого впливу на свідомість, шляхом створення ілюзорної реальності. З іншого боку, суспільство ризикує втратити довіру до медіа як джерела об'єктивної інформації, якщо підробки стануть невідрізними від справжніх матеріалів. На думку автора, найбільш підступним наслідком цієї тенденції є саме ерозія довіри: громадяни опиняються у стані, коли будь-яка новина сприймається з підозрою, а це зрештою грає на руку пропагандистам. Таким чином, проблема діпфейків та ШІ-генерації в інформаційному просторі –

це не лише технологічний виклик, але й психологічна та етична дилема сучасності, що потребує уваги та нових підходів у перевірці інформації.

2.2. Алгоритми та соціальні мережі як провідники пропаганди

В сучасному медіапросторі соціальні платформи перетворилися на ключових посередників інформації, і алгоритми цих платформ нерідко сприяють поширенню пропагандистських наративів. YouTube, Facebook, TikTok, X (Twitter) мають мільярдну аудиторію і потужні рекомендаційні системи, що визначають, які новини та дописи бачить користувач. Такі алгоритми оптимізовані під залучення уваги – тобто, вони віддають перевагу контенту, що викликає емоції та затримує глядача. На практиці це означає, що провокаційні, шокуючі або спірні матеріали (якими часто є пропагандистські повідомлення) отримують більше охоплення.

Наприклад, YouTube тривалий час був відомий тим, що його алгоритм рекомендацій міг «радикалізувати» глядачів, непомітно ведучи їх від нейтрального контенту до більш екстремістського або конспірологічного – включно з державними пропагандистськими відео. Хоча згодом компанія змінила політику, аби стримати поширення крайнощів, проблема повністю не зникла. Facebook зі свого боку надає пріоритет дописам, які активно коментують і поширюють; дослідження показали, що алгоритм Facebook непропорційно часто просуває дописи, що викликають обурення чи страх. Цим користувалися політичні маніпулятори: зокрема, у 2020–2021 роках викрили кілька скоординованих кампаній, коли через мережу Facebook-сторінок і груп масово розганялися фейки – від неправдивих «новин» про пандемію та вибори до цілеспрямованих наклепів на етнічні групи або країни. Відомо, що у М'янмі пропаганда в Facebook сприяла розпалюванню насильства проти мусульман-рохінджа (цей кричущий випадок стався напередодні 2020-х, але продемонстрував небезпеку неконтрольованого алгоритмічного поширення ненависті). TikTok зі своїм форматом коротких відео та загадковим алгоритмом

персоналізації також став платформою для блискавичного розповсюдження наративів. Варто людині затриматися на одному відео з певною тематикою, як стрічка незабаром наповнюється подібним контентом. Це середовище активно використовувалося пропагандою: аналітики назвали війну в Україні 2022 року «першою війною TikTok», адже застосунок заповнили як реальні, так і постановочні кадри бойових дій та емоційних звернень, які часом важко відрізнити одне від одного. Крім того, у TikTok регулярно фіксують тренди із сотень відео, що висміюють або демонізують певну сторону конфлікту – такі ролики часто виявляються організованими акціями, підтриманими зацікавленими акторами. Twitter (X) традиційно відігравав роль платформи для політичних дискусій і новинних оновлень у реальному часі, але і там алгоритмічна стрічка та тренди можуть бути викривлені штучним втручанням. Поширена тактика – використання ботів і координованих акаунтів для розкрутки певного гештегу чи теми, щоб вона потрапила у розділ трендів і привернула увагу журналістів та масового користувача. Наприклад, дослідники виявляли, що значна частина твітів під деякими політичними гештегами походила від новостворених акаунтів зі схожою поведінкою – явна ознака «ботової» активності. У 2023–2024 роках, після змін у політиці модерації X, спостерігалось зростання присутності акаунтів, пов'язаних з державними пропагандистськими мережами: послаблення контролю дало їм більше простору для маневру, а вилучення позначок «державне медіа» з профілів пропагандистських каналів (здійснене у 2023 році) зробило їх менш помітними як упереджені джерела.

Алгоритмічні механізми соціальних мереж сприяють формуванню так званих «фільтр-бульбашок» – ситуації, коли користувач бачить переважно той контент, який відповідає його вподобанням і поглядам. З одного боку, це створює комфортну інформаційну зону, з іншого – ідеальні умови для пропаганди. Людина, що вже схиляється до певної точки зору, буде отримувати ще більше подібних повідомлень, які зміцнюють її переконання, практично не стикаючись

з альтернативними поглядами. У міжнародній журналістиці це явище особливо небезпечно: авторитарні режими та екстремістські рухи можуть непомітно підживлювати алгоритм «зручними» для себе наративами, і ті розповсюджуватимуться майже автоматично. Достатньо запускати інформаційні вкиди через декілька популярних сторінок або акаунтів – далі механізми платформи самі донесуть їх до широкої аудиторії, якщо ті будуть достатньо резонансними. При цьому аудиторія навіть не усвідомлює, що її інформаційне меню звужене – відтак пропаганда діє ефективніше, адже альтернативні дані просто не проникають у бульбашку.

Політична реклама, таргетинг, інфлюенсери, бот-мережі. Окрім непрямого впливу через алгоритми, пропаганда використовує й інші цифрові інструменти – більш точкові та керовані. Один з них – таргетована політична реклама в соцмережах. Платформи на кшталт Facebook чи YouTube дозволяють рекламодавцям надзвичайно точно визначати свою аудиторію за географією, інтересами, демографією. Цим скористалися як легальні політичні кампанії, так і зовнішні гравці, щоб поширювати пропагандистські меседжі під виглядом звичайної реклами. Скандально відомим був випадок із компанією Cambridge Analytica (ще у 2016–2018 рр.), коли дані десятків мільйонів користувачів Facebook використовувалися без належної згоди для мікротаргетингу виборців специфічними маніпулятивними повідомленнями. У наступні роки технології таргетингу лише вдосконалювалися. З 2020 по 2024 роки у різних країнах фіксувалися спроби закупівлі політичної реклами з неправдивим або провокаційним змістом, спрямованої на вузькі групи – наприклад, національні меншини, конкретні вікові категорії, професійні спільноти. ШІ допомагає автоматизувати створення сотень варіацій таких оголошень, пристосованих під вподобання підгруп аудиторії, що підвищує ефективність пропаганди. Варто зазначити, що великі платформи після суспільного розголосу почали вводити обмеження та прозорість для політичної реклами (наприклад, Facebook

запровадив бібліотеку реклами, де можна побачити, хто і на яку аудиторію таргетує оголошення). Проте зловмисники часто обходять правила, маскуючи політичну агітацію під звичайний контент або використовуючи фіктивні сторінки для розміщення оголошень.

Ще один інструмент – «інфлюенсерська» пропаганда, тобто залучення популярних осіб у соцмережах для просування потрібних наративів. У 2020-х роках сформувався прошарок блогерів і лідерів думок, які свідомо або несвідомо ретранслюють меседжі державної пропаганди. Деяких із них прямо фінансують або координують урядові структури. Так, Китай відомий створенням мережі проурядових англомовних відеоблогерів, що через YouTube та Twitter поширюють «альтернативну» версію подій (наприклад, заперечують порушення прав людини чи вихваляють успіхи Компартії). Росія також адаптувалася: окрім офіційних каналів RT і Sputnik, Кремль робить ставку на західних ультраправих та ультралівих інтернет-активістів, які поширюють проросійські меседжі серед своїх підписників під виглядом «власної позиції». Зафіксовано випадки, коли російські структури намагалися непомітно спонсорувати популярних західних блогерів, аби ті озвучували кремлівські тези (наприклад, про «необхідність» скасувати санкції чи про «нацистів у Києві»). Водночас уряди демократичних країн теж освоїли цей інструмент – під час виборчих перегонів штаби кандидатів співпрацюють з впливовими блогерами в Instagram, TikTok чи YouTube, аби ті ненав'язливо формували позитивний образ свого кандидата або критикували опонентів. Аудиторія при цьому може навіть не усвідомлювати, що отримує інформацію з прихованим агітаційним ухилом, довіряючи улюбленому ютуберу чи тіктокеру.

Нарешті, невід'ємною частиною цифрової пропаганди стали бот-мережі. Йдеться про сотні й тисячі автоматизованих акаунтів у соцмережах, запрограмованих на поширення певних тез і наративів. У міжнародній журналістиці вже задокументовано чимало випадків, коли цілі «ферми тролів»

керували арміями ботів у Facebook чи X для масового репосту потрібних гасел, нападок на опонентів або створення ілюзії консенсусу. З появою більш розвиненого ШІ ці боти стали ще правдоподібнішими: нейромережі навчилися генерувати реалістичні профілі з фотографіями неіснуючих людей, писати пости, що важко відрізнити від дописів живих користувачів, і навіть підтримувати дискусії в коментарях. Як показали розслідування, у 2023–2024 роках Росія запустила кілька прихованих інформаційних операцій за кордоном, використовуючи AI-ботів, що видавали себе за звичайних громадян західних країн. Ці боти поширювали проросійські погляди щодо війни в Україні та інших тем, намагаючись вплинути на суспільну думку. Подібно діють і інші державні гравці – від Ірану до Венесуели – формуючи мережі фейкових акаунтів для просування своїх інтересів. Аналітики оцінюють, що значна частина всього контенту у Twitter про міжнародні події може генеруватися або підсилюватися ботами. Так, за різними оцінками, від 10% до 20% активних акаунтів у цій мережі є несправжніми, попри зусилля модераторів з блокування спаму.

Використання бот-мереж у поєднанні з алгоритмами створює ефект «ехо-камери» в інформаційному полі. Пропагандистське повідомлення, підхоплене тисячами скоординованих ботів, миттєво множиться і виходить у топ-результати, після чого його починають обговорювати реальні люди – часто не підозрюючи про штучне походження початкового тренду. Автор цих рядків спостерігав ситуації, коли сумнівний вклад, розкручений ботами до трендової теми, за лічені години опинявся в ефірі міжнародних ЗМІ – просто тому, що алгоритми розцінили його як «гарячу новину» з численними згадками. Таким чином, цифрові технології дозволяють пропаганді працювати одночасно і вертикально (через авторитетних інфлюенсерів чи офіційні канали), і горизонтально (через масову активність ботів серед пересічної аудиторії). Це серйозний виклик для міжнародної журналістики, яка змушена функціонувати в умовах перенасичення інформаційного простору скоординованими, штучно посиленими

повідомленнями. На міжнародному рівні регулятори також намагаються протидіяти цим явищам: зокрема, Євросоюз у 2022 році запровадив оновлений Кодекс поведінки щодо дезінформації, який зобов'язує великі платформи виявляти та позначати штучно створений контент (як-от дїпфейки) і видаляти мережі ботів. Однак ці вимоги поки діють переважно на добровільних засадах, і ефективність їхнього впровадження залишається обмеженою.

2.3. Кейс-стаді: використання ШІ в умовах інформаційних війн

У різних країнах та конфліктах технології штучного інтелекту проявляють себе по-різному як інструмент пропаганди. Розглянемо декілька показових кейсів.

РФ: дезінформація про війну в Україні

Російська Федерація протягом останніх років продемонструвала агресивне використання ШІ для проведення інформаційних операцій, особливо після початку повномасштабної війни проти України у 2022 році. Одним із перших сигналів стала згадана раніше поява дїпфейк-відео із «капітуляцією» президента Зеленського. Ця спроба посіяти паніку серед українців відкрила еру застосування глибинних фейків на війні. Надалі російська пропаганда вдалася до цілої низки прийомів на основі ШІ. Зокрема, фіксували випадки створення фейкових новинних сюжетів та сайтів, стилізованих під західні медіа, що поширювали кремлівські наративи (наприклад, про «біолабораторії» в Україні чи про те, що українське керівництво «втекло з країни»). Для ілюстрації таких фейкових новин все частіше використовували генеративну графіку: у мережі з'являлися нібито фотографії, що «підтверджують» неправдиві історії (скажімо, сфабриковані знімки руйнувань або «жертв», які насправді були згенеровані нейромережею). Нерідко пропагандисти підробляють і образи українських військових – наприклад, створюють фейкові фото, де солдати ЗСУ нібито мають нацистську символіку чи знущаються з цивільних, аби викликати обурення.

Ще небезпечнішим інструментом стали бот-мережі, підсилені ШІ. Російські «фабрики тролів», які діяли ще з середини 2010-х, у 2020-х отримали нове «озброєння»: алгоритми, що автоматично генерують пости кількома мовами і створюють цілі легіони облікових із людськими аватарками. У 2023–2024 рр. було розкрито декілька таких мереж, які працювали на міжнародних платформах. Зокрема, Міністерство юстиції США повідомило, що зірвало операцію, в якій російські фахівці створили тисячі фейкових акаунтів американців, використовуючи ШІ для контенту та фото профілів. Ці «боти-привиди» поширювали проросійські тези про війна, намагаючись вплинути на громадську думку в США та Європі. У самій Росії ШІ теж застосовується для пропаганди: державні медіа впроваджують автоматизовані системи модерації коментарів і навіть генерації новин під потрібним кутом, а соцмережі наповнюють боти, які атакують опозиційних блогерів чи іноземні медіа критичного спрямування.

Не менш показовою є історія з дідфейками проти українців за кордоном. Виявилось, що російські спецслужби створювали фейкові відео та аудіо, щоб дискредитувати українських біженців і військових. У 2023 році в Молдові російська пропаганда поширила начебто репортаж ВВС, де «українські біженці» звинувачувалися в інсценуванні трагедій – як з'ясувалося, сюжет був повністю підроблений за допомогою ШІ. Такі кейси свідчать про намагання Кремля посіяти розбрат між Україною та її союзниками, використовуючи найсучасніші технології для надання фейкам переконливості. На думку автора, російська інформаційна війна проти України сьогодні фактично стала полігоном для обкатки ШІ-інструментів пропаганди, ефекти від яких відчуються по всьому світу.

Китай: контроль і просування державного порядку денного

Китайська Народна Республіка підходить до технологій ШІ з власною стратегією: поєднуючи тотальний внутрішній контроль інформації з тонкою зовнішньою пропагандою. В середині країни Пекін активно використовує

алгоритми для цензури та управління порядком денним. Соцмережі і платформи (WeChat, Weibo, Douyin та інші) працюють під наглядом державних систем штучного інтелекту, що фільтрують заборонені теми і просувають пропагандистські меседжі. Уявний «Великий китайський фаєрвол» у 2020-х став ще більш розумним: влада за допомогою машинного навчання миттєво виявляє вірусний контент, який суперечить лінії партії, і негайно його блокує або «топить» у потоці схвальних для режиму постів.

Паралельно Китай експортує свій наратив назовні, дедалі частіше застосовуючи генеративний ШІ. Один із трендів – поява віртуальних пропагандистів. У 2023 році викрито мережу відео в соцмережах, де англомовні «журналісти» (як з'ясувалося, їхні обличчя та голоси були створені нейромережею) хвалили досягнення Китаю і критикували західні країни. В одному такому відео від імені вигаданого каналу “Wolf News” двоє штучних ведучих повідомляли, що у США буцімто провалюється боротьба зі злочинністю, зате Китай успішно проводить міжнародні саміти – чиста пропаганда, замаскована під новини. Крім того, напередодні виборів на Тайвані 2024 року китайські кібер-групи, за даними аналітиків, запустили хвилю дівфейк-відео, щоб підірвати довіру до про-тайванських політиків. Зокрема, було розповсюджено ролик з AI-генерованим диктором, який глузливо висловлювався про тайванську президентку Цай Інвень, порівнюючи її з «порожньою зеленою рослиною» – очевидна спроба вплинути на виборців методом очорнення через штучне відео.

Китайська держава інвестує значні ресурси в розробку власних ШІ для медіа. Уже кілька років на державних телеканалах працюють AI-ведучі новин – цифрові аватари, які цілодобово зачитують підготовлені редакцією тексти з незмінно проурядовою тональністю. Хоча спершу аудиторія сприймала їх як курйоз, поступово технологія вдосконалюється, і такі віртуальні ведучі можуть з'являтися в різних мовних версіях, транслюючи вигідну Пекіну точку зору для

глядачів за кордоном. У 2022–2023 роках Китай також прийняв законодавчі обмеження щодо «гіперреалістичного» контенту: провайдерів зобов'язали чітко маркувати матеріали, створені ШІ, щоб запобігти паніці або шкоді від дідфейків. Показово, що навіть авторитарна влада побоюється безконтрольного використання таких технологій всередині країни. На переконання автора, китайський підхід демонструє, наскільки потужним інструментом у руках держави може бути ШІ: він одночасно виступає «охоронцем» внутрішнього інформаційного порядку і «агітатором» назовні, причому масштаби обох процесів вражають. Таким чином, Пекін використовує ШІ двома шляхами: для тотальної внутрішньої цензури і для витонченої зовнішньої пропаганди, створюючи глобальний інформаційний імідж, бажаний для китайської влади.

США: маніпуляції в політичних кампаніях

Сполучені Штати Америки опинилися на передовій фронті протистояння із загрозою ШІ-дезінформації, особливо у контексті виборів. Хоча сама американська держава не займається пропагандою у традиційному сенсі, різні політичні сили та сторонні гравці активно використовують ШІ для маніпуляцій громадською думкою. Уже під час президентських виборів 2020 року фіксувалися приклади масованих онлайн-кампаній, де боти та фейкові акаунти (багато з яких курувалися з-за кордону) поширювали дезінформацію про кандидатів, процедуру голосування тощо. Проте справжній вибух AI-технологій припав на цикл виборів 2024 року. Американські виборці зіткнулися з новим явищем: передвиборча агітація, створена нейромережами.

У 2023 році широкий резонанс отримав ролик, опублікований одним з партійних штабів, який повністю складався зі згенерованих зображень і відеофрагментів. У цьому відео змальовувалися гіпотетичні «жахливі наслідки» перемоги опонента – кадри хаосу на вулицях, економічного краху тощо – всі ці картини були зроблені ШІ для нагнітання страху у виборців. Інший приклад – команда одного з кандидатів у праймеріз поширила в соцмережах фото, де

конкурент обіймається з особою, яка заплямувала себе скандалом; згодом з'ясувалося, що це зображення теж згенеровано, аби викликати негативну асоціацію. Подібні прийоми раніше вимагали майстерності графічних редакторів, а тепер стали доступними натисканням кнопки.

Водночас на рівні інтернет-дискурсу в США почали з'являтися фейкові аудіо та текстові матеріали, що впливають на політичні настрої. В месенджерах циркулювали записи, де, наприклад, голосом, схожим на когось із лідерів, озвучуються провокаційні заяви – часто такі «зливи» викривають як підробку, але вони встигають потрапити у новинні цикли й посилити поляризацію. У соцмережах множилися блоги та акаунти, нібито пов'язані з активістами чи журналістами, які насправді генерувалися ШІ й поширювали тенденційні матеріали (наприклад, масово продукували наклепи на одного кандидата під виглядом «розслідувань»). У травні 2023 року стався показовий інцидент: у Twitter розлетілося фото вибуху біля Пентагону, яке спричинило короткочасну паніку на біржі – і лише згодом експерти помітили, що зображення було синтезоване нейромережею. Цей випадок продемонстрував, наскільки реалістичні AI-фейки можуть миттєво впливати на суспільство і навіть економіку.

Американські інституції вимушені реагувати на нові виклики. Напередодні виборів 2024 року влада відкрито попереджала, що зафіксовано спроби зарубіжних гравців (у тому числі Росії) використати ШІ для втручання – зокрема, для створення контенту, який сіє розбрат серед виборців США. Одночасно і самі технологічні компанії запровадили політику маркування синтетичного контенту та вдосконалюють засоби виявлення дідфейків. Ба більше, в окремих штатах – наприклад, у Техасі та Каліфорнії – були прийняті закони, що забороняють публікацію оманливих дідфейків про кандидатів незадовго до виборів (за кілька тижнів). Втім, реалізація таких заборон стикається з труднощами через питання свободи слова. З погляду автора, виборча кампанія 2024 року стала своєрідним

випробуванням: рівень недовіри до медіа й політичної реклами зріс, адже багато виборців усвідомили, що яскравий ролик чи скандальний аудіозапис можуть виявитися продуктом алгоритму, а не реальним свідченням. Вибори 2024 року стали показовими не лише для США – схожа тенденція помітна і в інших країнах. Приміром, в Індії кандидати відкрито застосовували масові діпфейк-ролики зі своїми виступами різними мовами, аби охопити максимальний електорат; це спричинило дискусію про етичність таких методів у демократичному процесі.

Ізраїль / Палестина: генерація відео-реакцій у TikTok

Збройний конфлікт між Ізраїлем та палестинським угрупованням ХАМАС у 2023–2024 роках став ще одним фронтом, де ШІ використовувався для інформаційного впливу. Особливу роль відіграла платформа TikTok, на якій молодіжна аудиторія з усього світу активно стежила за розвитком подій через короткі відео. У цій мережі можна було побачити безліч «реакційних» відеороликів: користувачі, часто молоді ізраїльтяни чи палестинці, емоційно коментували новини про черговий ракетний удар чи трагедію. Проте експерти помітили, що дехто з цих «очевидців» могли бути не зовсім тим, за кого себе видають. Виявилося, що частина відео-реакцій створювалася з використанням ШІ – від підроблених голосових доріжок до цифрових аватарів людей, яких не існує. Таким чином пропагандисти з обох сторін конфлікту намагалися маскувати агітаційні відео під справжні емоційні звернення звичайних людей, щоб здобути більше довіри і співчуття у глядачів.

Також в розпал протистояння з'явилися і більш витончені інформаційні провокації. Прикладом, у соцмережах поширили діпфейк-відео, де відома модель Белла Хадід (в реальному житті вона підтримує палестинців) нібито записує звернення на підтримку Ізраїлю. Цей ролик швидко викрили як фальшивку, але він встиг викликати плутанину серед її шанувальників і потрапив у заголовки новин. Крім того, у соцмережах масово циркулювали синтетичні матеріали, що розпалювали ненависть. З ізраїльського боку траплялися випадки поширення

графічних зображень насилля, приписаних ХАМАС (як-от жахливі фото поранених дітей), які на перевірку виявлялися або вирваними з контексту старими кадрами, або створеними нейромережею «для наочності». З палестинського боку теж використовувалися перебільшені або вигадані картини страждань у Газі, щоб шокувати міжнародну спільноту – частина таких відео могла бути відредагована або згенерована. Алгоритми TikTok підсилювали ці емоційні матеріали: відео із шокуючим змістом набирали мільйони переглядів і тисячі коментарів, через що платформа ще активніше їх рекомендувала іншим користувачам. Ба більше, старі дипфейки часом спливають у новому контексті: у жовтні 2023 року на TikTok раптово набуло популярності змонтоване відео з американським президентом Джо Байденом, де той начебто оголошує примусовий призов до війська – хоча насправді цей фальшивий ролик було створено раніше і не мав стосунку до реальних подій, його поширили тільки для нагнітання паніки під час близькосхідної кризи.

Окремим явищем стали віртуальні інфлюенсери і боти у контексті цього протистояння. За короткий час після початку війни у жовтні 2023 року на TikTok та X виникли нові акаунти, що позиціонували себе як незалежні репортери на місці подій. Деякі з них щоденно викладали «ексклюзивні» відео із зони боїв – але розслідування показали, що кілька таких профілів були фейковими, а обличчя «репортерів» створені ШІ. Це свідчить про високий рівень технологічної підготовки інформаційних воєн у регіоні: сторони конфлікту не тільки воюють фізично, а й змагаються за віральність своїх наративів у цифровому просторі, використовуючи найновіші інструменти. Одним з найбільш вірусних символів цієї війни стало зображення з написом «All Eyes on Rafah» («Усі погляди на Рафах»), яке начебто складало білосніжні намети біженців – насправді цей вражаючий кадр зворушливої солідарності було повністю згенеровано нейромережею. Втім, менш ніж за дві доби він поширився десятками мільйонів переглядів через Instagram, X та особливо TikTok. Аудиторія захопилася ідеєю,

не надто переймаючись тим, що зображення нереальне. Особисто спостерігаючи за стрічкою TikTok у ті дні, автор був вражений хаотичним змішанням правди, напівправди і відвертих фейків: від щирих людських переживань до явно постановочних сцен. Цей кейс став наукою для міжнародних журналістів: тепер будь-який контент із зони конфлікту доводиться перевіряти в рази ретельніше, адже за кадром може стояти не камера репортера, а алгоритм.

РОЗДІЛ 3. Загрози та шляхи протидії ШІ-пропаганді

3.1. Основні загрози для журналістики та демократії

Стрімкий розвиток технологій штучного інтелекту у сфері медіа приніс не лише нові можливості, а й безпрецедентні загрози для сучасної журналістики та демократичних суспільств. Пропаганда, підсилена ШІ, здатна впливати на масову свідомість більш приховано й масштабно, ніж традиційні методи. Розглянемо ключові виклики останніх років: пониження довіри до медіа, посилення поляризації і радикалізації аудиторій та фактичну підміну професійної журналістики так званими “медіасимулякрами” – штучно створеними імітаціями медіаконтенту.

Втрата довіри до медіа

Довіра до засобів масової інформації є фундаментом демократії: громадяни мають вірити, що отримують правдиві факти і неупереджений аналіз. Проте поширення ШІ-пропаганди серйозно підриває цю довіру. Синтетичні медіа – від підроблених новинних текстів до реалістичних згенерованих фото, аудіо чи deepfake-відео – створюють ситуацію, коли споживачам дедалі важче відрізнити правду від фальсифікації. Внаслідок цього виникає феномен “інфляції істини”: люди починають сумніватися у достовірності будь-якої інформації, навіть з авторитетних джерел.

Останні п’ять років позначилися резонансними випадками, що похитнули довіру аудиторії до медіа. Наприклад, у березні 2022 року під час розпалу війни РФ проти України у мережі було поширено deepfake-відео з нібито зверненням президента Володимира Зеленського, який закликав українських військових скласти зброю. Ця підробка, створена із використанням нейромереж для підміни обличчя і голосу, мала на меті деморалізувати українське суспільство. Хоча фейк швидко викрили і видалили, сам його факт продемонстрував, що тепер навіть відеозвернення лідера країни не гарантує автентичності. Аналогічно, у травні 2023 року соціальні мережі облетіло зображення вибуху біля будівлі Пентагону,

яке виявилось згенерованим ШІ. Ця фальшивка спричинила паніку на фінансових ринках, доки не була оперативно спростована. Подібні інциденти небезпечні тим, що підривають сам принцип доказовості фактів: фотографія чи відео, які традиційно слугували надійним свідченням подій, тепер можуть виявитися цифровою ілюзією.

Наслідком є ефект, який умовно можна назвати “дефіцитом правди” у суспільстві. Глядач або читач, раз обманутий правдоподібним фейком, надалі менш схильний довіряти навіть правдивим повідомленням. Це живить загальну атмосферу скептицизму і теорій змови. Ба більше, сама наявність технології deepfake дає підстави зловмисникам дискредитувати реальні викриття: політик, спійманий на скандальному відео, може заявити, що це фальшивка, згенерована ШІ, і частина публіки йому повірить. Таким чином, ШІ-пропаганда працює у двох напрямках: вона не лише вводить в обіг неправдивий контент, але й сіє сумнів щодо справжнього. У результаті страждає репутація журналістики як інституту – адже якщо аудиторія більше не вірить медіа, ґрунт для демократичного обговорення і прийняття рішень руйнується.

Втрата довіри до медіа прямо загрожує демократії. Дезінформація, посилена ШІ, підточує консенсус щодо базових фактів, необхідний для здорового суспільного діалогу. Люди починають жити в різних інформаційних реальностях і можуть відкидати навіть офіційні спростування, вважаючи, що “все навколо – пропаганда”. Це веде до цинізму та апатії: частина громадян узагалі перестає споживати новини, не знаючи, кому довіряти, тоді як інша частина радикалізується навколо альтернативних, часто конспірологічних медіа-джерел. У таких умовах авторитарні сили отримують перевагу: їм легше просувати свої наративи, коли дискредитовано саму ідею об’єктивної правди. Таким чином, відновлення довіри до медіа стає першочерговим завданням, адже без цього інші засоби протидії пропаганді матимуть обмежений ефект.

Збільшення поляризації та радикалізації

Наступний великий ризик, пов'язаний із ШІ-пропагандою, – посилення суспільної поляризації та радикалізації груп населення. Алгоритми штучного інтелекту, що лежать в основі соцмереж і медіаплатформ, налаштовані на максимально довге утримання уваги користувачів. Для цього вони формують індивідуальні стрічки новин, підлаштовуючись під вподобання людини. Непомітно для самих користувачів виникають так звані «фільтраційні бульбашки»: кожен отримує контент, що підтверджує його погляди, і майже не бачить альтернативної точки зору. Пропагандисти успішно експлуатують цей механізм, впорскуючи в інформаційний потік дезінформацію, яка резонує з емоціями та упередженнями певної аудиторії.

У результаті суспільство фрагментується на ворогуючі табори, кожен зі своїм набором “фактів” і наративів. За останні роки у багатьох країнах спостерігаємо зростання політичної та ідеологічної поляризації, яке співпало в часі з бурхливим розвитком соціальних мереж і алгоритмічної персоналізації контенту. ШІ-пропаганда виступає каталізатором цього процесу. Наприклад, в США зовнішні актори (насамперед російські “фабрики тролів”) під час виборчих кампаній активно таргетували різні групи населення: окремо чорну спільноту, окремо ультраправих, сіючи розбрат на расовому та політичному ґрунті. За допомогою бот-акаунтів і автоматизованого розповсюдження контенту, підсиленого ШІ, вони одночасно поширювали провокативні меседжі для протилежних сторін – наприклад, розпалюючи недовіру між прихильниками різних партій. Це класична тактика “розділяй і володарюй”, що у цифрову епоху отримала нове життя.

Крім того, системи рекомендацій контенту (такі як алгоритми YouTube чи TikTok) мають тенденцію пропонувати дедалі більш радикальні матеріали, щойно користувач зацікавився якоюсь темою. Це відбувається тому, що шокуючий чи крайній контент часто краще привертає увагу, а отже – підвищує

залученість. Дослідження показували, що користувач, переглядаючи політичне відео поміркованого змісту, може вже за кілька кроків автоматично отримати в рекомендаціях більш екстремістські ролики. Так виникає “радикалізаційний лабіринт”: людина непомітно для себе сповзає від поміркованих поглядів до крайнощів, підживлювана алгоритмом. Особливо вразливою є молодь, яка споживає короткі відео в сервісах на кшталт ТікТок: емоційно заряджені кліпи, що містять маніпулятивні меседжі, миттєво ширяться цими мережами і можуть формувати світогляд цілої генерації.

Поляризація небезпечна тим, що підточує основу демократичного дискурсу – здатність різних груп шукати компроміс і спільне розуміння реальності. Коли суспільство розділене на “інформаційні гетто”, де кожна сторона бачить іншу через спотворювальну призму пропаганди, зростає ворожнеча і нетерпимість. В екстремальних випадках це веде до насильства. Наприклад, теорії змови та дезінформація, поширені онлайн, сприяли радикалізації частини американців, що врешті призвело до штурму Капітолію 6 січня 2021 року – безпрецедентної події, коли натовп, підбурений фальшивими наративами про “вкрадені вибори”, вдався до атаки на інститути влади. Аналогічні тенденції помітні і в Європі: дезінформація щодо пандемії COVID-19 чи теми міграції, масово розганялася ботами та троями, розколола суспільну думку та посилила вплив крайніх рухів. У східній Європі, зокрема, російська пропаганда свідомо поляризує громади за мовною, етнічною, релігійною ознаками, щоб дестабілізувати країни та послабити їхню згуртованість перед зовнішньою агресією.

ШІ-пропаганда також може прямо радикалізувати окремих осіб через персоналізований підхід. Завдяки можливостям мікротаргетингу (точкового націлювання рекламних чи інформаційних повідомлень), пропагандисти здатні виявити уразливі категорії людей – наприклад, тих, хто переживає соціальні труднощі чи має екстремістські схильності – і “бомбардувати” їх специфічним

контентом. Алгоритми аналізують поведінку користувача і підсовують йому саме ті повідомлення, що найбільш ймовірно викличуть лють або страх. Такі емоції притупляють критичне мислення і штовхають до радикальних дій. На сьогодні зафіксовано випадки, коли терористичні та екстремістські угруповання використовували соцмережі та месенджери з елементами ШІ для вербування прихильників. Наприклад, “Ісламська держава” ще кілька років тому активно поширювала професійно змонтовані пропагандистські відео на YouTube, а нині з’явилися дані, що ІДІЛ випробовує створення відеозвернень за допомогою ШІ-аватарів, які виступають від імені “військових командирів”. Отже, технології дозволяють ворожій пропаганді масштабувати радикальний вплив і роблять його більш персоніфікованим.

Підсумовуючи, можна стверджувати, що без належної протидії ШІ-пропаганда здатна дедалі більше розколнувати суспільство та радикалізувати окремі групи. Це пряма загроза для демократії, адже конструктивний громадський діалог стає неможливим на тлі ворожнечі та недовіри. Тому серед інших завдань критично важливо зменшити поляризаційний ефект інформаційних алгоритмів і штучно згенерованого контенту – зокрема, шляхом коригування роботи платформ та активної просвіти населення (про що йтиметься далі).

Підміна журналістики “медіасимулякрами”

Окремою, системною загрозою є явище поступової підміни реальної журналістики так званими “медіасимулякрами” – тобто штучно створеними медіа-продуктами, що імітують форму і стиль журналістики, але не несуть її цінностей та відповідальності. Завдяки ШІ сьогодні можна згенерувати практично будь-який інформаційний контент: віртуальні диктори читають новини з екрану, роботи-автори пишуть статті, фейкові експерти коментують події у соцмережах. Зовнішньо все це дуже схоже на справжні журналістські

матеріали, проте насправді перед нами – симуляція, інформаційна декорація, створена для маніпуляції, а не для об’єктивного інформування.

Останніми роками з’явилося чимало прикладів таких медіасимулякрів на міжнародній арені. У 2022–2024 роках дослідники фіксували випадки, коли в соцмережах масово поширювалися відео з псевдо-новинними ведучими, згенерованими ШІ. Зокрема, китайські прокремлівські мережі створили вигаданого мовника під назвою “*Wolf News*”, ролики якого вели реалістичні аватари-новинарі. В одному відео такий “ведучий” звинувачував уряд США у нездатності приборкати насильство із застосуванням зброї, в іншому – вихваляв роль Китаю на міжнародному саміті. Глядач, побіжно переглядаючи стрічку, міг сприйняти ці сюжети за випуски реальних новин. Натомість це були цілеспрямовано сконструйовані пропагандистські меседжі, озвучені віртуальними дикторами.

Подібні технології вперше були обкатані в авторитарних державах на внутрішню аудиторію. Китай, наприклад, ще з 2018 року впроваджував у державних медіа цифрових телеведучих – програмно згенерованих людей, що читали новини з екрану. Тепер ця технологія “експортується” для зовнішньої пропаганди. У 2024 році з’ясувалося, що проурядові групи поширюють у західних соціальних мережах відео з ведучими, створеними на основі deepfake. Ці аватари, наче справжні журналісти, доносять до глядача позицію Пекіна з різних питань – від оцінок політичних лідерів Тайваню до виправдання дій росіян у війні проти України. Хоча іноді штучність таких роликів вдається розпізнати (наприклад, за неприродною інтонацією чи мімікою), тенденція очевидна: державна пропаганда отримала універсальний інструмент маскуванню під незалежну журналістику. Замість того щоб витратити час і ресурси на виховання справжніх пропагандистів, режими можуть штучно “тиражувати” їх десятками, контролюючи кожне слово.

Ще один вимір проблеми – створення фальшивих медіа-майданчиків, які повністю складаються з ШІ-контенту. Недобросовісні актори можуть запустити мережу сайтів, стилізованих під новинні портали чи аналітичні видання, наповнивши їх матеріалами, згенерованими нейромережею. Такі статті можуть не мати реальних авторів або підписуватися вигаданими іменами. У 2022–2023 рр. викрито кілька подібних операцій: зокрема, велика прокремлівська мережа сайтів під назвою «Правда», про яку повідомив проект NewsGuard. Ця мережа з квітня 2022 року наплодила сотні веб-ресурсів різними мовами, де щодня публікувалися тисячі статей з кремлівськими дезінформаційними наративами. Значну частину цих текстів, як припускають, писали саме алгоритми ШІ за заданими тезами. Для недосвідченого читача такі сайти виглядали як звичайні новинні агенції – з тією різницею, що їхній контент був однобоким і брехливим. Більш того, медіасимулякри цього типу можуть взаємно цитувати одне одного, штучно підвищуючи свою видимість у пошукових системах. Відтак, якщо людина шукає певну тему, висока ймовірність, що вона натрапить на матеріали з мережі цих фейкових ЗМІ, навіть не підозрюючи про їхнє походження.

Також фіксуються випадки створення фейкових персоналій журналістів або експертів. У 2020 році була розкрита історія з вигаданим автором “Олівером Тейлором” – молодим журналістом, який публікувався в західній пресі з викривальними статтями про певних осіб, але згодом з’ясувалося, що його фотографія була згенерована нейромережею, а біографія – вигадана. Це означає, що за ним стояли ляльководи, які використовували ШІ, щоб створити ілюзію реальної особи і надати своїм пропагандистським текстам більшої легітимності. Подібна тактика може застосовуватися і у ширших масштабах: армія “віртуальних інфлюенсерів” чи “ботів-коментаторів”, наділених штучними фотографіями та іменами, які ретранслюють потрібні меседжі з вигляду як думки звичайних людей чи незалежних оглядачів.

Наслідки проникнення медіасимулякрів в інформаційний простір руйнівні. По-перше, вони витісняють якісний журналістський контент, забираючи увагу аудиторії та рекламні ресурси. Автоматизована фабрика новин може засипати соціальні мережі та пошукові видачі дешевим контентом у величезних кількостях, тоді як робота справжнього журналіста потребує часу на перевірку фактів і аналіз. У конкуренті за увагу перемагає той, хто публікує більше і швидше – а це, на жаль, часто означає перевагу для алгоритмів над людьми. По-друге, симулякри підривають самі стандарти правдивості: якщо у стрічці поряд існують реальні новини і “штучні” новини, аудиторія поступово звикає до сенсаційного, емоційно забарвленого стилю, який нав’язує пропаганда, і може перестати цінувати фактажність та об’єктивність. Це створює хибний попит на жовтуватий, маніпулятивний контент і тисне на традиційні медіа, змушуючи їх теж спрощувати та драматизувати подачу матеріалу, аби не втратити читача. Таким чином, межа між журналістикою і пропагандою розмивається – саме чого і прагнуть автори дезінформації.

Зрештою, масове вторгнення ШІ-симулякрів у медіасередовище загрожує *легітимністю* журналістики. Якщо громадяни не можуть відрізнити справжній репортаж від згенерованої псевдо-новини, вони почнуть трактувати всі медіа як ненадійні. Це вигідно пропагандистам: дискредитувавши незалежну пресу, вони усувають конкурента і можуть транслювати свої меседжі напряду. Такий сценарій вже частково реалізується в умовах авторитарних режимів, де державна пропаганда намагається повністю замінити собою незалежні ЗМІ. Однак виклики постають і перед демократіями – тепер атака йде не лише через політичний тиск чи цензуру, а й через високотехнологічне імітування інформаційної реальності.

3.2. Технології й інституційні механізми протидії

У відповідь на зазначені загрози світова спільнота напрацьовує різноманітні підходи протидії ІІІ-пропаганді. Вони охоплюють як технологічні рішення, що дозволяють виявляти і фільтрувати фейки, так і інституційні заходи – від роботи незалежних організацій з фактчекінгу до створення нових законів і правил регулювання цифрового простору. Жоден із цих інструментів не є панацеєю, але в комплексі технології та політики можуть значно знизити ефективність ворожої пропаганди. Розглянемо основні механізми протидії, що застосовуються останніми роками: розвиток AI-детекторів неправдивого чи синтетичного контенту, посилення фактчекінгу та використання журналістики даних, а також нормативне врегулювання діяльності ІІІ та онлайн-платформ, започатковане в ЄС, США та інших регіонах.

AI-детектори фейків та синтетичного контенту

Один із найочевидніших кроків – використати силу штучного інтелекту для боротьби з породженими ним же викривленнями. Мова про розробку і впровадження спеціалізованих детекторів фейків, здатних автоматично розпізнавати синтетичний контент: підроблені зображення, змонтовані аудіозаписи, текст, написаний нейромережею, тощо. У світі інтенсивно ведуться дослідження в цій галузі, і за останні кілька років з'явилося чимало прототипів та комерційних продуктів, призначених для виявлення deepfake та інших видів дезінформації.

Детектори зображень і відео. Для розпізнавання deepfake-відео та згенерованих зображень використовують алгоритми комп'ютерного зору, що шукають характерні артефакти або невідповідності. Ранні підходи спиралися на те, що синтетичні обличчя можуть мати неприродні деталі – наприклад, ненормальне моргання очей, неприродну текстуру шкіри або “пливучі” контури при рухові. Сучасні детектори пішли далі: вони аналізують на піксельному рівні шум і структуру зображення, намагаючись знайти сліди роботи генеративних

моделей (наприклад, невідповідності світлотіні або викривлення фону). Є підходи, що відстежують мікросигнали людського тіла – скажімо, зміну тону шкіри від серцебиття: такі дрібні фізіологічні ознаки поки що важко імітувати ШІ достовірно. Відеодетектори теж можуть перевіряти аудіодоріжку на наявність ознак синтезу голосу (несинхронність руху губ із мовленням, частотні особливості звуку). Деякі великі технологічні компанії інвестують у цю сферу: так, Microsoft у 2020 році презентувала інструмент Video Authenticator, здатний оцінити ймовірність того, що відео є дипфейком. Meta (Facebook) організувала конкурси для розробників детекторів дипфейків, щоб стимулювати прогрес у цій царині. У цілому, ідея полягає в тому, щоб створити “антивірус” для фейкових медіа – автоматизовану систему, яка б попереджала модераторів платформи чи навіть самих користувачів, якщо контент виглядає підозрілим.

Детектори тексту. Окремий напрям – виявлення текстів, написаних моделями на кшталт ChatGPT. Спочатку передбачалося, що нейромережевий “стиль” можна вчислити статистично: моделі часто генерують більш однотипні, шаблонні фрази, уникають орфографічних помилок, можуть допускати граматичні особливості. Створено програми, що аналізують вірогідність того, що даний текст згенерований ШІ (наприклад, OpenAI пропонувала свій класифікатор). Проте завдання виявилось складним: автори швидко навчилися *підстроювати* стиль машинних текстів під людський, додаючи помилки або варіюючи лексику, і прості детектори втрачають ефективність. Більш перспективним шляхом є впровадження «водяних знаків» – особливих міток у генерованому тексті, невидимих для читача, але помітних для програми. Наприклад, модель могла б навмисно добирати синоніми за певним алгоритмом, що видасть її “почерк”. Такі водяні знаки нині експериментально впроваджуються: їх розробляють, зокрема, в OpenAI та інших лабораторіях, щоб допомогти освітянам і редакторам розпізнавати машинні тексти. Втім,

універсального вирішення поки нема, і творці генеративних моделей визнають, що відрізнити їхній продукт від написаного людиною дедалі важче.

Превентивні засоби. Паралельно з детекцією фейків після їх появи, існує підхід превентивний – ідентифікація достовірного контенту. Якщо довести справжність джерела, то все інше можна трактувати як потенційно підроблене. На цьому принципі базується ініціатива Content Authenticity Initiative (CAI), започаткована компанією Adobe та партнерами. Вона пропонує стандарти, за якими цифрові зображення та відео матимуть захищені метадані про походження: яка камера чи пристрій їх зробив, чи редагували файл і як саме. Наприклад, сучасний фотокамери можуть одразу підписувати знімки цифровим підписом, а будь-яка зміна пікселів знімає цей підпис. Якщо така система стане масовою, новинні редакції зможуть швидко відрізнити оригінальні фото від відредагованих. Аналогічно, у соцмережах або месенджерах може відобразитися позначка про перевірене джерело контенту (скажімо, “це відео отримано від прес-служби і не змінювалося”). Вже зараз Twitter (X) та Facebook експериментують із позначками для маніпулятивного медіа: наприклад, у Twitter до зміни власника діяла політика, за якою явні дипфейки або сильно відредаговані відео маркувалися попередженням для користувачів. Деякі держави навіть зобов’язують платформу робити це законодавчо (про що далі). Таким чином, технологічна протидія йде двома шляхами – знаходити підробки і підтверджувати оригінали.

Варто зазначити, що “гонка озброєнь” між генераторами і детекторами фейків триває. Щоразу, коли з’являється новий ефективний детектор, розробники генеративних моделей вдосконалюють свої методи, щоб обходити перевірки. Приміром, якщо детектор орієнтується на певні артефакти зображення, алгоритм генерації можна модифікувати, аби прибрати ті артефакти. Це вимагає постійного оновлення засобів захисту. Ба більше, з відкриттям доступу до потужних моделей ШІ, зловмисники й самі можуть використовувати їх для

“знешкодження” детекторів – наприклад, автоматично підбирати такі параметри фейку, щоб він не видавав статистичних ознак підробки. Тому фахівці сходяться на думці, що чисто технологічними заходами проблему не вирішити – детектори мають доповнюватися іншими формами протидії.

Попри обмеження, AI-детектори вже приносять користь. Великі платформи застосовують системи машинного навчання для фільтрації відвертих фейків: наприклад, Facebook повідомляла, що виявляє і блокує мільйони підроблених акаунтів і ботів, у тому числі з поведінкових ознак, які “вчить” ШІ. YouTube використовує алгоритми для автоматичного видалення відео, які вже були позначені незалежними фактчекерами як фейкові (наприклад, повторні завантаження небезпечної дезінформації). Ці заходи не завжди помітні користувачу, але вони створюють перший бар’єр на шляху масового вкидання пропаганди. Перспективним напрямом є вбудовування детекторів прямо у пристрої: скажімо, смартфон в майбутньому міг би попереджати свого власника, якщо отримане зображення ймовірно штучне.

Отже, розвиток AI-детекторів є необхідною складовою відповіді на нові загрози. Уявімо, якби кожна людина мала своєрідний “фільтр правди”, що в режимі реального часу сигналізує про можливий обман – це суттєво ускладнило б життя пропагандистам. До такої мети ще далеко, але поступ поступовий: наукова спільнота, бізнес і навіть військові відомства інвестують у цю сферу, розуміючи стратегічну важливість технологічної оборони проти інформаційних атак. Втім, самі по собі інструменти ШІ не можуть врахувати весь контекст і етичну складову, тому надзвичайно важливо поєднувати їх із людським фактором – зокрема, з роботою професійних фактчекерів та журналістів-розслідувачів.

Фактчекінг і журналістика даних

Другий фронт боротьби з ШІ-пропагандою – це діяльність професійної спільноти журналістів, аналітиків, незалежних організацій, які займаються

фактчекінгом (перевіркою фактів) та розслідуванням дезінформаційних кампаній. Попри автоматизацію багатьох процесів, людський розум і критичний аналіз залишаються ключовими в протистоянні маніпуляціям. Останні п'ять років фактично створили нову галузь у журналістиці – журналістику протидії дезінформації, котра поєднує традиційні розслідувальні методи з використанням сучасних даних та інструментів.

Розвиток фактчекінгових ініціатив. У всьому світі набули поширення незалежні фактчекінгові проекти, завдання яких – оперативно перевіряти сумнівні заяви та інформаційні вкиди і публічно спростовувати фейки. Приклади включають такі організації, як Snopes, PolitiFact, FactCheck.org у США, Full Fact у Британії, StopFake в Україні, EU vs Disinfo при ЄС та десятки інших. Ці команди складаються з журналістів та аналітиків, які моніторять інформаційний простір (особливо соціальні мережі) на предмет вірусних фейків. Коли виявляють підозрілий контент, вони використовують різні методи перевірки: шукають оригінальні джерела, звертаються до експертів, застосовують OSINT-інструменти (відкриті джерела розвідки) для аналізу фото і відео. Результати перевірки публікують у вигляді статей-спростувань, часто з поясненням, як саме був зроблений фейк. Наприклад, у випадках із deepfake-відео фактчекери можуть демонструвати кадр за кадром невідповідності або наводити дані експертів з кібербезпеки про походження файлу.

Фактчекери дедалі частіше беруть на озброєння і самі технології ШІ. Значний обсяг рутинної роботи – моніторинг тисяч повідомлень – тепер полегшують алгоритми, які виявляють потенційно вірусні фейки ще на ранніх стадіях. Соцмережі надають фактчекерам спеціальні інструменти: приміром, Facebook співпрацює з мережею незалежних фактчекерів і передає їм на перевірку ті пости, які автоматичні системи позначили як можливий обман (на основі попередніх шаблонів або повідомлень від користувачів). Деякі редакції використовують мовні моделі для попередньої перевірки фактів: такі моделі

можуть сканувати тексти політичних заяв чи статей і порівнювати їх із наявними базами даних, щоб видати аналітику, які твердження виглядають неправдивими. Хоча остаточне рішення завжди залишає за собою людина, ШІ може зменшити навантаження і пришвидшити реагування.

Журналістика даних та OSINT. В протидії складним, прихованим дезінформаційним кампаніям особливо допомагають методи журналістики даних – тобто аналіз великих масивів інформації та візуалізація прихованих зв'язків. Наприклад, аби викрити мережу ШІ-ботів, що просувають певний хештег у Twitter, журналісти-аналітики збирають дані про сотні акаунтів: хто що пише, коли створено сторінку, які є спільні риси. За допомогою спеціалізованих програм (таких як Graphika, NodeXL тощо) можна побудувати граф зв'язків між цими акаунтами і побачити, що вони координуються з єдиного центру. Так були неодноразово викриті операції російських “тролеферм” у соцмережах, китайських інформаційних кампаній, спрямованих на Захід, тощо. Міжнародні колективи журналістів, такі як Bellingcat, продемонстрували силу OSINT (розвідки з відкритих джерел) у розслідуванні фейків. Вони використовували аналіз супутникових знімків, метаданих фото, геолокацію відео для того, щоб встановити правду в резонансних подіях – від збиття малайзійського Boeing на Донбасі до хімічних атак у Сирії. В епоху ШІ ці ж інструменти стають у нагоді для розвінчання міфів, створених пропагандою. Наприклад, коли у мережі поширюють фото нібито “з місця події”, журналіст-розслідувач може за лічені хвилини знайти оригінал цього зображення через зворотний пошук Google або визначити, що фото генеровано (виявивши явні цифрові артефакти чи абсурдні деталі на зразок викривлених написів). Таким чином, поєднання людської уважності з цифровими інструментами дозволяє швидко нейтралізувати чимало вкидів.

Співпраця з платформами. Значну роль у масштабній протидії дезінформації відіграють самі медіаплатформи. Facebook, Twitter (X), YouTube,

ТікТок протягом останніх років впровадили програми співпраці з фактчекерами та маркування перевіреної інформації. Наприклад, якщо незалежні перевіряльники встановили, що певна новина є фейком, Facebook знижує її видимість в стрічках користувачів та додає попередження “Перевірка фактів: інформація спростована”. У YouTube під відео на спірні теми (COVID-19, вибори, війна) стали з’являтися плашки з посиланням на офіційні джерела або Вікіпедію, щоб глядач міг ознайомитися з контекстом. Twitter до зміни керівництва також практикував позначки для дезінформації, а у 2021 році запусив функцію Community Notes (раніше “Birdwatch”), коли самі користувачі можуть додавати примітки до сумнівних твітів з поясненнями щодо їх правдивості. Хоча ці заходи не ідеальні, вони створюють ще один рівень захисту: пропагандистському контенту стає важче маскуватися під правдивий, коли він помічений як спірний.

Приклад України та міжнародні кейси. В українському медіапросторі з 2014 року (початку агресії РФ) з’явилося багато ініціатив, спрямованих на викриття саме російської дезінформації. Окрім вже згаданого StopFake, активно працюють проекти Інформаційного спротиву, VoxCheck, Детектор Медіа та інші. Вони спростовують фейки про війну, які генеруються і розганяються з використанням як традиційних методів (телебачення, заяви офіційних осіб РФ), так і сучасних (мережа ботів, анонімні телеграм-канали, постановочні відео). В 2022–2023 рр. під час повномасштабного вторгнення, українські фактчекери неодноразово ловили росіян на використанні ШІ-графіки: наприклад, фотографії “усміхнених українських солдат-нацистів”, що знуцаються з цивільних – ці зображення мали явні ознаки генерації (симетричні обличчя-копії, нереалістичні деталі одягу тощо). Оприлюднення таких викриттів заважає ворогу легітимізувати брехню. Подібні практики важливі й глобально: коли в М’янмі в 2021–2022 роках почала ширитися етнічна ненависть, міжнародні журналісти даних змогли відстежити джерела масових фейків у Facebook і домогтися

блокування аккаунтів, причетних до підбурювання геноциду. Це доводить, що професійна перевірка фактів – не марна справа, вона реально рятує життя і стабільність.

На противагу блискавичній швидкості ІІІ-генерації контенту, фактчекінг часто є ретроспективним (спростовує вже після появи вкиду) і не завжди встигає запобігти первинному ефекту дезінформації. Проте його роль полягає у відновленні правди і підтриманні інформаційної гігієни в довгостроковій перспективі. Спростування, надане авторитетним джерелом, формує у частини аудиторії імунітет проти аналогічних фейків у майбутньому. До того ж, системна робота фактчекерів виявляє тренди пропаганди – які теми найбільш атаковані, які нові методи застосовуються (наприклад, поява чергового дипфейку одразу стає сигналом для медіаспільноти бути обережнішими з контентом від певних джерел). Це знання дозволяє проактивно готувати контрзаходи, зокрема проводити “пребанкінг” (попереджувальне інформування). Пребанкінг означає випередити брехню: наприклад, якщо очікується хвиля фейків про вибори, медіа публікують матеріали “Чого очікувати і як розпізнати дезінформацію під час виборчої кампанії”. Така освіта аудиторії значно знижує шанси, що люди поведуться на пропаганду.

Отже, фактчекінг і журналістика даних є невід’ємною частиною відповіді на виклики ІІІ-пропаганди. Це той людський фактор, який може критично оцінювати контент, чого не здатна зробити машина автономно. Поєднання зусиль журналістів і технологій дає синергію: людина вказує машині, що шукати, а машина допомагає людині швидше аналізувати. В підсумку, прозорість і підзвітність інформації зростає, а маніпуляторам стає важче діяти безнаказанно.

Регуляція (ЄС, США): закони про ІІІ та цифрові послуги

Третій компонент протидії – регуляторний і політичний. Усвідомлюючи загрозу, яку несе неконтрольоване поширення ІІІ-контенту, демократичні держави останніми роками почали розробляти законодавчі та нормативні акти, покликані

обмежити шкідливий вплив ШІ-пропаганди. Особливо активним є Європейський Союз, який у 2022–2025 рр. ухвалив або підготував до ухвалення низку важливих документів, серед яких Акт про цифрові послуги (DSA) та Акт про штучний інтелект (AI Act). У США дискусія довкола законодавчого регулювання ШІ також поживалася, хоча комплексного закону поки немає; натомість діють точкові закони на рівні штатів, а також використовуються існуючі норми (виборче, медійне, кримінальне право) для протидії дезінформації. Розглянемо ключові підходи регуляції: від вимог прозорості щодо ШІ-контенту до відповідальності платформ за розповсюдження пропаганди.

Європейський Союз: прозорість та відповідальність платформ. ЄС традиційно займає проактивну позицію у сфері цифрового регулювання. Акт про цифрові послуги (Digital Services Act, DSA), що набув чинності у 2023 році, встановлює нові правила для великих онлайн-платформ (соцмереж, пошуковиків, маркетплейсів) щодо боротьби з незаконним і шкідливим контентом, включаючи дезінформацію. Зокрема, DSA зобов'язує найбільші платформи проводити оцінку ризиків впливу їхніх сервісів на суспільство – і одним із таких ризиків прямо названо дезінформацію. Платформи мають впроваджувати заходи для зменшення цих ризиків: удосконалювати модерацію, співпрацювати з фактчекерами, прозоро пояснювати роботу алгоритмів рекомендацій. DSA також вимагає більшої прозорості алгоритмів: компанії повинні розкривати ключові параметри, за якими відбувається відбір і ранжування контенту, а дослідники отримують доступ до даних для незалежного аудиту (що допомагає виявляти, наприклад, непропорційне поширення пропагандистських матеріалів). Окремо DSA забороняє використання темних патернів і фейкових акаунтів: платформи повинні виявляти ботів і чітко маркувати, якщо користувач взаємодіє з нелюдською системою (наприклад, чат-ботом служби підтримки). Ці норми спрямовані на те, щоб знизити анонімність і маскування, якими користується пропаганда.

Ще одним важливим кроком ЄС став розвиток *спеціального законодавства щодо ШІ*. Акт про штучний інтелект (AI Act) – всеосяжне регулювання ШІ, що наразі перебуває на фінальній стадії ухвалення. Хоча цей акт стосується ШІ загалом (його застосування у різних галузях), у ньому є положення безпосередньо пов'язані з темою пропаганди. Зокрема, планується впровадити вимоги прозорості для систем, що генерують аудіовізуальний контент, здатний ввести в оману щодо своєї природи (тобто deepfake). Постачальники таких систем будуть зобов'язані вживати заходів, аби результат їхньої роботи був позначений як штучно створений. Наприклад, якщо поширюється згенероване відео з реалістичним обличчям, на ньому має явно вказуватися, що це не реальна людина. Винятки можуть робитися для контенту, створеного в художніх, сатиричних чи дослідницьких цілях, але принцип у тому, що споживач інформації має право знати, коли перед ним ШІ-імітація. За недотримання цієї норми передбачені значні штрафи. Таким чином, ЄС прагне на рівні закону забезпечити розпізнаваність ШІ-контенту, що значно ускладнить масове застосування дипфейків у політичних чи медійних маніпуляціях.

Крім того, у Європі діють нормативні акти та органи, спрямовані саме на боротьбу з іноземною пропагандою і дезінформацією. Ще у 2018 році ЄС запусив Кодекс практик щодо дезінформації – добровільну угоду з великими інтернет-компаніями, де вони взяли на себе зобов'язання протистояти фейкам (видаляти фальшиві акаунти, обмежувати рекламу на сайтах-розповсюджувачах дезінформації, співпрацювати з перевіряльниками фактів). Після російського вторгнення в Україну 2022 року Євросоюз перейшов і до жорсткіших кроків: з міркувань безпеки на території ЄС було заблоковано мовлення відверто пропагандистських ресурсів RT (Russia Today) та Sputnik. Хоча це радикальний захід і він викликав дискусії про цензуру, його обґрунтували тим, що ті канали є інструментами війни, а не медіа у класичному розумінні. Урок тут у тому, що в

надзвичайних ситуаціях демократії готові застосувати й заборонні інструменти проти інформаційної агресії.

США: свобода слова vs безпека виборів. У Сполучених Штатах, де свобода слова захищена Першою поправкою, регуляція медіаконтенту – завжди тонке питання. Проте загроза дипфейків змусила законодавців на рівні окремих штатів ухвалити специфічні закони. Так, штат Каліфорнія у 2019 році прийняв закон, що забороняє розповсюдження сфабрикованих медіа за участю кандидата за 60 днів до виборів, якщо вони зроблені з наміром нашкодити кандидату або ввести виборців в оману. Подібний закон є в Техасі. Ідея цих актів – не допустити ситуації, коли прямо перед голосуванням в інтернет “викидають” дипфейкове відео (скажімо, де кандидат робить шокуюче зізнання), яке може спотворити волевиявлення. порушникам загрожують цивільні позови від постраждалих кандидатів. Втім, практичне застосування таких законів наштовхується на питання доказовості й меж свободи слова: нещодавно суд призупинив дію каліфорнійського закону, пославшись на його можливу неконституційність (обмеження висловлювань у передвиборчий період). Це показує складність балансування між протидією фейкам і захистом конституційних прав. Федерального закону, що прямо б регулював deepfake-дезінформацію, досі немає, хоча у Конгресі періодично пропонують законопроекти на цю тему (наприклад, DEEP FAKES Accountability Act, який поки що не ухвалений).

Натомість на рівні виконавчої влади США роблять акцент на співпрацю з технологічними компаніями та підвищення готовності суспільства. У 2023 році Білий дім ініціював зустрічі з керівниками провідних AI-компаній (Google, OpenAI, Microsoft тощо) і ті добровільно погодилися впроваджувати безпекові заходи, зокрема систему watermarking для генерованих зображень та відео. Крім того, американські відомства, відповідальні за кібербезпеку та вибори, випускають рекомендації для штатів і виборчкомів – як розпізнавати і реагувати на дипфейки, що можуть вплинути на перебіг виборів чи громадський порядок.

Наприклад, Федеральна виборча комісія восени 2023 обговорювала, чи можна кваліфікувати певні AI-матеріали як “шахрайські” за чинним виборчим законодавством. Хоча чітких правил не встановили, сам факт уваги означає, що проблема визнається на високому рівні.

Також варто згадати застосування існуючих законів у боротьбі з пропагандою. Міністерство юстиції США в 2022–2023 роках проводило операції проти російських дезінформаційних мереж: зокрема, було викрито й закрито десятки сайтів, замаскованих під справжні новинні портали, що фінансувалися Кремлем. Ці дії вчиняли на основі законів про іноземних агентів, відмивання грошей, шахрайство тощо. Деяких осіб, причетних до “фабрики тролів” (наприклад, бізнесмена Євгена Пригожина до його загибелі), оголошено в розшук чи під санкції за втручання у вибори. Таким чином, США використовують правові механізми покарання організаторів пропаганди, що теж є елементом стримування.

Глобальні ініціативи та інші країни. Проблема ШІ-дезінформації – глобальна, тому й регуляторні відповіді шукають на міжнародному рівні. Організація Об’єднаних Націй у вересні 2023 року провела окрему сесію, присвячену перспективам регулювання ШІ, де обговорювалася і тема deepfake. Низка держав підписала етичні рекомендації UNESCO щодо ШІ, які містять принцип відповідального використання технологій без шкоди для прав людини і демократії. У рамках “Групи семи” (G7) у 2023 році започатковано “Хіросімський процес” – міжнародні консультації про регулювання штучного інтелекту, включно з його впливом на інформаційний простір. Хоча поки це не перейшло в площину договорів, очевидно, що формується спільне бачення загрози і бажання узгодити підходи.

Цікавий кейс Китаю: влада КНР, сам будучи активним користувачем ШІ для цензури та пропаганди, одночасно запровадила жорсткі правила для приватного сектору. З січня 2023 року у Китаї діє регуляція, що вимагає: весь

згенерований аудіо-відео контент, який може ввести в оману, має містити чітку позначку; заборонено створювати дипфейки реальних людей без згоди – за це передбачено покарання. Фактично, Пекін прагне монополізувати інструменти впливу: держава може застосовувати їх у своїх цілях, а от неконтрольоване використання громадянами чи компаніями – під заборонаю. Для демократій цей досвід неоднозначний, але технічні методи (наприклад, авто-додавання водяного знаку “AI Synthetic” на контенті) привертають увагу фахівців на Заході як можливий приклад для власних регуляторних рішень, тільки у м’якшій формі.

Отже, регуляція у сфері ШІ-пропаганди поки що перебуває на етапі становлення. Європейський Союз закладає важливий прецедент, запроваджуючи право на правду – вимогу до прозорості алгоритмів і позначення штучного контенту. Сполучені Штати, хоча діють обережніше, теж рухаються до усвідомлення: ігнорувати проблему не можна, інакше під загрозою опиняються виборчі процеси та національна безпека. В цілому, у світі формується консенсус, що великі онлайн-платформи та розробники ШІ несуть відповідальність за використання їхніх інструментів у деструктивних цілях. У найближчі роки слід очікувати появи нових норм – можливо, і на рівні міжнародних домовленостей – які встановлять “червоні лінії” для застосування ШІ в інформаційному просторі. При цьому важливо, щоб регуляція не стала надмірною цензурою: завдання полягає у тому, щоб знайти баланс між свободою вираження та захистом суспільства від масованого обману, посиленого машинами.

3.3. Роль медіаосвіти та журналістських стандартів

Поряд із технологічними та правовими засобами, вирішальну роль у протидії ШІ-пропаганді відіграє людський фактор – рівень медіаграмотності суспільства та дотримання високих професійних стандартів самою медіаспільнотою. Якщо перші дві групи заходів можна порівняти зі щитом і мечем (технічний захист і правове примушення), то просвіта і саморегуляція – це

фундамент, на якому будується довгострокова стійкість до інформаційних загроз. Навіть найдосконаліші детектори чи суворі закони не застрахують, якщо громадяни легко піддаються маніпуляціям, або якщо журналісти в гонитві за сенсаціями нехтують правдою. Тому дуже важливо інвестувати в медіаосвіту, розвиток критичного мислення аудиторії та відродження базових принципів журналістики – об’єктивності, точності, прозорості.

Критичне мислення як бар’єр проти маніпуляцій

Медіаосвіта населення – тобто навчання розуміти і критично оцінювати інформацію – є визнаним у світі способом підвищити колективний імунітет до пропаганди. Коли людина обізнана про те, як працюють медіа, як створюються новини і як їх можуть перекручувати, їй значно складніше “втравити” відверту брехню чи змусити поширювати фейки. У контексті ШІ-дезінформації медіаграмотність набуває нових вимірів: громадян необхідно ознайомити з існуванням технологій генерації контенту і навчити деяким базовим прийомам викриття. Наприклад, вже зараз в багатьох країнах шкільна програма включає обговорення теми deepfake: школярам показують приклади відеопідробок, пояснюють, чому не можна автоматично довіряти “очевидному” і які є способи перевірити відео чи зображення (пошук оригіналу, звернення до офіційних джерел, критичні питання щодо деталей). Ці знання діти несуть у сім’ї, таким чином просвіта охоплює ширші верстви.

Критичне мислення – це, по суті, здатність ставити під сумнів і аналізувати отриману інформацію, перш ніж робити висновки або поширювати її далі. Щоб стимулювати таку навичку, уряди та громадські організації останніми роками започатковують різноманітні програми. Наприклад, в Україні у відповідь на російську гібридну агресію активно діє кампанія “Фільтр”, орієнтована на молодь: це цикл лекцій, відеороликів, онлайн-курсів про те, як відфільтровувати інформаційне сміття. Подібні тренінги проводить і Міністерство освіти, і профільні NGO (IREX, Академія української преси тощо). Тематика обов’язково

включає сучасні прийоми обману: мова йде і про ботоферми, і про фейкові сайти, і про синтетичні медіа. Таким чином, українське суспільство, переживаючи постійний потік ворожої пропаганди, поступово підвищує свою резистентність – люди швидше впізнають шаблонні кремлівські наративи, уже насторожено ставляться до занадто емоційних “сенсаційних” повідомлень.

Світовим взірцем у медіаграмотності часто називають Фінляндію: там ще з 2010-х років на державному рівні впровадили навчання інформаційної гігієни в школах і для різних груп населення. Коли прокотилася хвиля російської дезінформації в Європі, фіни виявилися найменш вразливими: за дослідженнями Єврокомісії, рівень довіри до відвертих фейків у Фінляндії був чи не найнижчий серед країн ЄС. Це підтверджує, що критичне мислення – потужний бар’єр: навіть якщо людина побачить дипфейк чи прочитає брехливу новину, вона швидше за все зреагує здоровим сумнівом і перевірить додатково, замість того щоб негайно ділитися й обурюватися.

До медіаосвіти належить і концепція “превентивного щеплення” проти дезінформації – так званий *prebunking*. Суть його в тому, щоб заздалегідь інформувати аудиторію про типові техніки маніпуляторів, аби коли люди зіштовхнуться з ними, вони вже мали “антитіло” у вигляді знання. В останні роки Google та університети провели експерименти з показом коротких відео-уроків перед роликами на YouTube: в цих міні-кліпах наводився приклад фейку і пояснювалося, як його створено (наприклад, використано емоційну мову, або підроблено фото, або перекручено статистику). Глядач, озброєний цими знаннями, потім менш охоче вірив у схожі дезінформаційні повідомлення. Такий підхід уже масштабують, зокрема у країнах Східної Європи перед виборами – своєрідна вакцинація від фейків.

Критично мисляче суспільство зменшує ефективність ШІ-пропаганди у рази. Адже навіть якщо пропагандисти поширять гучний дипфейк, люди не будуть реагувати панічно або агресивно – спершу вони пошукають

підтвердження, зважать різні точки зору. Звичайно, не всі громадяни стануть “детективами правди”, але підвищення середнього рівня грамотності створює середовище, де фейк не гуляє неконтрольовано. Громадський осуд поширювачів брехні, розвиток культури перевірки перед репостом – усе це виробляється з часом, і єдиний шлях тут – освіта і ще раз освіта.

Особливої уваги потребують вразливі групи – літні люди, які менше знайомі з новими технологіями, та підлітки, чия психіка лише формується. Для старшого покоління в деяких країнах проводять лекції з медіаграмотності в бібліотеках, університетах третього віку, поширюють друковані пам’ятки про типові ознаки фейкових новин (на кшталт “сенсаційні заголовки з багатьма знаками оклику”, “джерело невідоме” тощо). Молодь же, навпаки, часто технічно підкована, але через брак життєвого досвіду може не розуміти, як глобально працює пропаганда. Тому акцент в школах зміщується з просто цифрових навичок до етики споживання інформації: дітям прищеплюють розуміння цінності правди, навчають відповідальності за поширення відомостей. У деяких країнах вже вводять курси “логіки і критичного мислення” в старших класах, де розбирають логічні хиби, риторичні прийоми, маніпуляції з даними – і роблять це на прикладах актуальних фейків, щоб було наочно.

Отже, критичне мислення широких мас – це та невидима, але надміцна стіна, об яку може розбитись найвитонченіша ШІ-брехня. Інвестувати в його розвиток – стратегічно правильний шлях для демократичних держав. Медіаграмотна громада не тільки менш вразлива до зовнішньої пропаганди, а й внутрішньо стійкіша: у ній нижчий градус ненависті, більше культури дискусії, краще працюють інститути, бо їхні рішення спираються на факти, а не емоційні міфи. Це, по суті, вклад у національну безпеку і здоровий розвиток суспільства в умовах інформаційної доби.

Повернення до принципів об'єктивності й прозорості

Поки аудиторію навчають не вестися на маніпуляції, самі медіа та журналісти теж мають зробити “домашнє завдання” – підвищити стандарти своєї роботи, щоби виправдати довіру громадян. Адже недарма пропагандисти прагнуть виставити традиційну журналістику у непривабливому світлі: часом і самі медіа дають до того привід, опускаючи планку. У епоху постправди й ШІ-вкидів професійна спільнота мусить згуртуватися довкола базових принципів: перевірка фактів, неупередженість, чітке відокремлення фактів від думок, прозорість джерел, етичність методів. Лише демонструючи ці цінності на практиці, журналістика зможе знову виділитися на тлі інформаційного шумовиння та завоювати авторитет у публіки.

Об'єктивність та неупередженість. Хоча в теорії медіа давно дискутують, чи досяжна повна об'єктивність, зараз мається на увазі прагнення журналіста подавати матеріал збалансовано і чесно, без свідомого викривлення. В умовах політично поляризованого суспільства важливо, щоб провідні новинні організації дотримувалися нейтрального тону, подаючи всі релевантні сторони історії. Це противага тенденції соціальних мереж, де виграє найгучніший і найсуб'єктивніший голос. Якщо люди бачитимуть, що традиційні медіа висвітлюють події комплексно – наприклад, і про успіхи, і про проблеми уряду, – у них буде менше причин шукати “альтернативну правду” в сумнівних джерелах. Особливо критично це під час криз та конфліктів: медіа, що зберігають холодний розум і перевіряють кожну новину з поля бою, протистоять паніці та чуткам. Приміром, під час військових дій в Україні багато хто цінує ті ЗМІ, що не поширюють неперевірені повідомлення з Telegram, а чекають офіційного підтвердження. Такі редакції можливо і “програють” в швидкості, зате виграють у довірі. Це і є свідоме дотримання стандарту точності попри тиск емоційного фону.

Перевірка інформації перед публікацією. Золотим стандартом журналістики завжди було знайти підтвердження з двох незалежних джерел. На жаль, в еру інтернету цей принцип іноді поступався гонитві за кліками. Зараз є розуміння, що треба повернутися до ретельної верифікації, особливо якщо йдеться про контент з соцмереж або від неперевіраних авторів. Багато редакцій вводять посади фактчекерів всередині команди: їхнє завдання – перед випуском матеріалу пройтися по кожному факту і цифрі, звірити з наявними базами, подзвонити експертам для уточнення. Це уповільнює процес виробництва новин, але значно знижує ймовірність, що медіа ненароком стане ретранслятором чужої пропаганди. Особливо це стосується візуального контенту: у світлі випадків з дипфейками тепер обов'язковим протоколом має бути перевірка будь-якого “сенсаційного” відео чи фото. Великі агентства (Reuters, AP) розробили внутрішні інструкції з перевірки користувацьких відео з місця подій: журналісти зобов'язані проаналізувати метадані, геолокацію, знайти автора, по можливості отримати оригінал файлу. Це захищає не лише від дипфейків, а й від банального виривання з контексту. Адже пропаганда часто бере справжнє зображення, але додає хибний підпис – журналіст має виявити й це.

Прозорість джерел і методів. Щоб повернути довіру, медіа повинні стати більш відкритими перед аудиторією. Зараз багато видань практикують публікацію бекграунду матеріалу: в кінці новини можуть зазначити, звідки взяли дані (наприклад: “за інформацією звіту ООН”, “повідомив такий-то чиновник у Twitter о хх:уу годині”). Деякі онлайн-ресурси навіть надають гіперпосилання на оригінали документів або інтерв'ю. У телевізійних новинах все частіше показують фрагменти першоджерел – щоб глядач бачив, що слова не вирвані з контексту. Така прозорість позбавляє пропагандистів можливості звинуватити медіа в “безпідставності” чи перекрученнях – кожен може перевірити сам. Крім того, низка видань започаткувала практику коригування і вибачень: якщо помилилися, вони не ховають це, а відкрито публікують спростування і

вибачаються перед читачами. Цей чесний підхід, хоч і важкий репутаційно в короткостроковій перспективі, довгостроково зміцнює авторитет: люди бачать, що перед ними відповідальна організація, яка піклується про правдивість.

Етика використання ШІ в журналістиці. Цікавий новий аспект – як самі медіа інтегрують ШІ у свою роботу, і як це впливає на стандарти. Деякі редакції вже експериментували з генеративним ШІ для написання новин або довідкових матеріалів. Були випадки, коли це призводило до помилок і скандалів. Наприклад, відомий випадок у США: сайт CNET у 2023 році використовував нейромережу для створення фінансових статей, і спочатку не повідомив про це читачам. Коли з'ясувалося, та ще й виявили багато неточностей у цих статтях, репутація видання постраждала. Висновок простий: прозорість в тому, де і як застосовано ШІ, є обов'язковою. Якщо журналіст використовував нейромережу для збору даних або генерації чорнового тексту – варто про це згадати, принаймні редакції внутрішньо слід запровадити політику перевірки подібних матеріалів особливо ретельно. Деякі медіа вже офіційно прописали в редакційних політиках: “Ми використовуємо ШІ як допоміжний інструмент, але публікуємо лише після вичитки людиною, і ніколи не генеруємо вигаданих фактів”. Такі стандарти важливі, бо якщо ЗМІ почнуть перенасичувати стрічку автоматичними текстами, вони ризикують самі перетворитися на фабрику симулякрів і втратити унікальну цінність – людський аналіз і перевірку.

Коллективна протидія дезінформації. Журналістські стандарти включають і солідарність перед обличчям загрози. Великі редакції та медійні організації все частіше об'єднуються в коаліції для боротьби з пропагандою. Яскравий приклад – міжнародний проект First Draft News (нині Foreword): він зібрав десятки ЗМІ і університетів для спільного моніторингу та розвінчування онлайн-фейків. Під час виборів або криз такі мережі обміну інформацією дозволяють швидко реагувати: якщо десь з'являється новий вид дипфейку, про це одразу дізнаються журналісти по всьому світу і можуть застерегти свою аудиторію. Крім того,

професійні спілки журналістів проводять тренінги, де навчають колег методам цифрової безпеки, перевірки контенту тощо. Така самоосвіта спільноти гарантує, що стандарти не залишаться теорією, а втілюватимуться в нових умовах.

На завершення слід підкреслити: повернення до основ журналістики – це не крок назад, а необхідна умова виживання професії у XXI столітті. Об’єктивність, достовірність, прозорість, відповідальність – ці “старомодні” слова стають актуальними як ніколи, бо вони відрізняють справжні новини від штучно згенерованої “інфошуми”. Якщо медіа будуть вірні цим принципам, вони зможуть поступово відновити свою роль надійного джерела інформації. А маючи надійні джерела, суспільство значно легше протистоятиме будь-яким пропагандистським атакам – хоч за допомогою ІІІ, хоч без нього.

ВИСНОВКИ

У XXI столітті світ вступив у нову фазу інформаційної еволюції, у центрі якої перебуває штучний інтелект (ШІ) — не просто як технологія, а як інструмент, що змінює саму структуру інформаційного обігу, принципи комунікації та глобальну журналістику. У результаті проведеного дослідження було з'ясовано, що ШІ не лише доповнює можливості медіа, а й перетворюється на окремий гравець, здатний створювати, поширювати і трансформувати пропагандистські наративи з безпрецедентною швидкістю та ефективністю.

Розгляд історичного розвитку пропаганди у журналістиці засвідчив, що пропаганда завжди існувала паралельно з інформаційними технологіями: від гравюр і памфлетів до радіо й телебачення, і зрештою — цифрових платформ. У кожному епоху пропаганда використовувала найбільш передові на той момент канали поширення інформації. Саме тому поява інтернету, а згодом — соціальних мереж і генеративних ШІ, природно відкрила нові можливості для впливу на масову свідомість.

У XXI столітті, коли основну частину новинного контенту споживачі отримують через платформи з алгоритмічною подачею — TikTok, YouTube, Facebook, X (Twitter) — зникає чітке розмежування між журналістикою і пропагандою. На перший план виходить не достовірність інформації, а її емоційний ефект, віральність та здатність викликати реакцію.

Це призвело до утворення нового типу “медіа-середовища”, в якому класичні стандарти журналістики більше не є домінуючими, а пропагандистські меседжі часто мають більшу видимість та вплив, ніж перевірені факти. Алгоритми ШІ створюють інформаційні бульбашки, підсилюючи погляди користувачів і сприяючи радикалізації — процесу, який колись потребував років, тепер реалізується за лічені дні.

Аналіз практичного застосування ШІ показав, що ключовими загрозами є генеративні моделі (текстові й візуальні), які здатні продукувати фейкові новини,

зображення та відео (deepfake) на професійному рівні. Такі матеріали можуть мати не лише пропагандистський, а й стратегічний вплив у політичних, військових та соціальних конфліктах. Відео з “виступами” політиків, яких насправді не було, або фейкові кадри з місця подій — уже сьогодні змінюють хід публічних дебатів, підривають довіру до ЗМІ та створюють плутанину в момент критичних подій.

Окрему загрозу становлять алгоритми рекомендацій, які визначають, який саме контент бачить користувач. Це призводить до формування викривленої картини світу: людина бачить лише ті погляди, які збігаються з її попередніми вподобаннями, що формує “інформаційні гетто”.

Відзначимо також конкретні кейси, які ілюструють механізми та масштаби використання ІІІ у пропаганді:

- **РФ** активно застосовує deepfake у зовнішній пропаганді (наприклад, фейкові звернення українських військових або політиків), а також для деморалізації суспільства. Російська пропаганда створює цілі “фабрики контенту”, які автоматизовано поширюють наративи через TikTok та Telegram.
- **Китай** комбінує контрольовану журналістику з алгоритмами ІІІ для внутрішнього контролю й міжнародного впливу (наприклад, проєкти на кшталт CGTN). Протести в Гонконгу чи інформація про уйгурів подається за допомогою глибоко модерованого ІІІ-контенту.
- **США** стикаються з використанням ІІІ у політичній рекламі. Приклад — deepfake-ролики про Джо Байдена під час президентської кампанії, створені невідомими обліковками, які отримували мільйонні перегляди до моменту видалення.
- **Ізраїль / Палестина** — платформи TikTok використовуються обома сторонами для мобілізації молоді, інтерпретації подій у реальному часі.

Генеровані відео впливають на міжнародну думку, часто не маючи жодного відношення до реальних фактів.

Як творець подібного контенту, можу ствердно сказати: масштаб та вплив відео, які виглядають правдоподібно, але є синтетичними, сьогодні значно перевищує вплив класичних медіа. Такі продукти швидко розповсюджуються через соцмережі, проникають у журналістські сюжети, а іноді навіть у політичні виступи. Цей досвід підтверджує — ШІ перетворюється на глобального актора пропаганди.

Пропаганда, посилена штучним інтелектом, створює численні ризики для журналістики та демократії:

- Втрата довіри до інформаційного середовища.
- Поява феномену «постправди», коли відчуття правдивості важливіше за факти.
- Збільшення поляризації, радикалізація, втягування користувачів у конфлікти.
- Створення медіасимулякрів — фіктивної журналістики, яка імітує об'єктивність.

Проте й існують важливі механізми протидії. На технологічному рівні — це AI-детектори deepfake та синтетичного контенту. На інституційному — регулювання з боку Європейського Союзу (Digital Services Act, AI Act), а також ініціативи типу Partnership on AI чи Global AI Accord.

Ще одним ключовим елементом протидії є **медіаосвіта**. Вміння критично мислити, перевіряти джерела, розуміти логіку алгоритмів — усе це необхідно включати до програм шкільного й університетського навчання. Журналісти мають повернутися до стандартів прозорості, верифікації та пояснення, як було отримано інформацію.

Як автор, який працює з ШІ та глибоко розуміє його можливості, можу зазначити: саме етичні стандарти мають визначати межу між творчим інструментом і зброєю впливу. Власний досвід демонструє: легко перетнути цю межу, якщо не усвідомлювати відповідальності перед аудиторією.

Штучний інтелект не просто доповнює пропаганду — він її заново створює. Ми живемо у світі, де новини, емоції, події можуть бути створені штучно, згенеровані без участі людини. Це кидає виклик класичній журналістиці, але й дає шанс переосмислити її роль.

Майбутнє інформаційного простору залежатиме від того, як ми зуміємо поєднати технічний прогрес із гуманітарними цінностями. Якщо ми навчимося не лише виявляти фейки, але й пояснювати їх природу, не лише розпізнавати ШІ-контент, а й не піддаватись емоційному шантажу — ми зможемо створити нову, відповідальну модель інформаційної екосистеми.

Саме зараз визначається, чи стане ШІ союзником правди, чи зброєю дезінформації. І це — вибір не технології, а людини.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Pawelec M. (2022). *Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions*. *Digital Society*, 1(2), 19.
2. Momeni M. (2024). *Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions Through Misinformation*. *Journal of Creative Communications*, 19(1), 41–56.
3. Barman D., Guo Z., & Conlan O. (2024). *The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination*. *Machine Learning with Applications*, 16, 100545.
4. Raman R., Nair V. K., Nedungadi P., та ін. (2024). *Fake news research trends, linkages to generative artificial intelligence and sustainable development goals*. *Heliyon*, 10(3), e024727.
5. Diel A., Lalgi T., Schröter I. C., & MacDorman K. F. (2024). *Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers*. *Computers in Human Behavior Reports*, 9, 100538.
6. Nasiri Sh. & Hashemzadeh A. (2025). *The Evolution of Disinformation from Fake News Propaganda to AI-driven Narratives as Deepfake*. *Cyberspace Studies*, 9(1), 229–250.
7. de Nadal Ll. & Jančárik P. (2023). *Beyond the deepfake hype: AI, democracy, and “the Slovak case”*. *HKS Misinformation Review* (Harvard Kennedy School).
8. Goldstein J. A. & Sastry G. (2023). *The Coming Age of AI-Powered Propaganda: How to Defend Against Supercharged Disinformation*. *Foreign Affairs*, April 7, 2023.

9. Easterly J., Schwab S., & Conley C. (2024). *Artificial Intelligence's Threat to Democracy: How to Safeguard U.S. Elections From AI-Powered Misinformation and Cyberattacks*. Foreign Affairs, Jan 3, 2024.
10. Hwang Y., Ryu J. Y., & Jeong S. H. (2021). *Effects of disinformation using deepfake: The protective effect of media literacy education*. (Дослідження, цит. за Hwang & Ryu, 2021 у Sage Journals**).**
11. Groh M. et al. (2022). *Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds*. *Proceedings of the National Academy of Sciences (PNAS)*, 119(1), e2123000119.
12. Kietzmann J. et al. (2020). *Deepfakes: Trick or Treat?* *Business Horizons*, 63(2), 135–146.
13. Vaccari C. & Chadwick A. (2020). *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*. *Social Media + Society*, 6(1).
14. Mavchak R. & Panasiuk O. (2023). *Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets relating to deepfakes and the Russo-Ukrainian war*. *Harvard Kennedy School Misinformation Review*, 4(3).
15. Torres A. et al. (2023). *The global emergence of deepfakes in news and political communication: A review*. *New Media & Society*, 25(5), 1340–1357.
16. Schick N. (2022). *The World Needs Deepfake Experts to Stem This Chaos*. *Wired (Opinion)*, April 2022.
17. Newton C. (2024). *What Happened in Slovakia Will Happen Everywhere*. *Platformer (Substack)*, Oct 2, 2024.
18. Tucker J. (2023). *AI-Generated Propaganda in Russia's War on Ukraine*. *Lawfare (blog)*, May 5, 2023.
19. Subrahmanian V. (2023). *Can Democracy Survive the Disruptive Power of AI?* *Carnegie Endowment (blog)*, March 6, 2023.

20. Wheeler T. (2023). *This Is Not a Drill: Generative AI's Real-World Impact on Information Operations*. Brookings TechStream, July 17, 2023.
21. Hwang T. (2020). *Digital Footprints: Propaganda in the Era of Artificial Intelligence*. RAND Blog, Oct 2020.
22. Dufour G. (2023). *AI-generated propaganda floods TikTok – and young voters may not care*. Rest of World (Independent media), Nov 1, 2023.
23. Політична тека (2023). *Російська “мислевірусологія”: як Кремль тренує нейромережі на пропаганду*. (Блог на Medium, автор – український дослідник)
24. Орлов М. (2025). *ШІ-революція у пропаганді: як GPT навчився писати “темники”*. (Substack-блог українського журналіста-розслідувача).
25. Carvin A. (2024). *This Is What Russian Propaganda Looks Like in 2024*. NPR, Oct 12, 2024.
26. Li B. et al. (2024). *Social Media Manipulation in the Era of AI*. RAND Corporation, Aug 29, 2024.
27. NATO StratCom COE (2021). *Deepfakes: Primer and Forecast*. NATO Strategic Communications Centre of Excellence, Riga.
28. NATO StratCom COE (2019). *The Role of Deepfakes in Malign Influence Campaigns*.
29. Hybrid CoE (2022). *AI-based Technologies in Hybrid Conflict: The Future of Influence Operations*. The European Centre of Excellence for Countering Hybrid Threats, Paper 14, June 23, 2022.
30. EU DisinfoLab (2023). *Generative AI and Disinformation: Recent Advances, Challenges, and Outlook*. (Brussels, EDMO report).
31. Brennan Center for Justice (2023). *Regulating AI Deepfakes and Synthetic Media in the Political Arena*. (Authors: Weiner D. & Norden L., Published Dec 5, 2023).

32. US Office of the Director of National Intelligence (ODNI, 2023). *Annual Threat Assessment of the U.S. Intelligence Community (Excerpt on AI)*. Feb 2023.
33. Europol (2023). *Innovation Report: Law Enforcement Tech Series – Facing Reality? Law Enforcement and the Challenge of Deepfakes*. European Union Agency for Law Enforcement, June 2023.
34. The Aspen Institute (2021). *AI in Disinformation: The Next Frontier*.
35. Центр стратегічних і міжнародних досліджень (CSIS, 2023). *Distrust Everything: Misinformation and AI*. (Transcript of panel, June 2023).
36. Reuters (2022). *Deepfake footage purports to show Ukrainian president capitulating*. Reuters News, March 16, 2022.
37. Reuters (2024). *Russia using generative AI to ramp up disinformation, says Ukraine minister*. Oct 16, 2024.
38. The Atlantic (2022). *We Haven't Seen the Worst of Fake News*. The Atlantic, Dec 20, 2022 (by Matteo Wong).
39. The Atlantic (2024). *AI's Fingerprints Were All Over the Election*. The Atlantic, Nov 15, 2024 (by Matteo Wong).
40. The New York Times (2023). *As AI Tools Spread, Governments Seek to Curb 'Disinformation-for-Hire'*. NYTimes, Aug 5, 2023.
41. Washington Post (2023). *A 'deepfake' was used to spread disinformation in a tense town election*. WP, May 16, 2023.
42. BBC News (2023). *AI deepfake bots 'weaponised' in conflict zones, experts warn*. BBC, Oct 30, 2023.
43. Associated Press (2023). *Misinformation 2.0: AI deepfakes trigger a rise of deception online*. AP News, July 30, 2023.
44. Politico (2023). *Deepfakes, distrust and disinformation: Welcome to the AI election*. Politico (EU), Oct 10, 2023.
45. Wired (2023). *Brace Yourself for the 2024 Deepfake Election*. Wired, Nov 2023.

46. Wired (2024). *Indian Voters Are Being Bombarded With Millions of Deepfakes. Political Candidates Approve*. Wired, May 20, 2024.
47. ABC News (2023). *Deepfake 'news anchors' appear in pro-China footage on social media*. ABC Australia, Feb 2023.
48. Foreign Policy (2021). *The Grim Future of Realistic Political Deepfakes*. Foreign Policy Magazine, Feb 2021.
49. Financial Times (2023). *AI and the new age of propaganda*. FT Special Report, Sept 18, 2023.
50. NBC News (2023). *Israel-Hamas war sees surge of AI-generated fake news, researchers say*. NBC, Nov 2, 2023.
51. Український кризовий медіа-центр (UCMC, 2023). *Deepfakes – AI in the Hands of Propaganda*. UCMC, 14 червня 2023 (англомовна аналітика).
52. Detector Media (2025). *Росія використовує дінфейки проти українських біженців, — розслідування*. Детектор медіа, 2 лютого 2025 (авторка Ольга Кацімон).
53. MediaSapiens (Detector Media, 2024). *Росія використовує ШІ для посилення кампаній з дезінформації проти України, — дипломат*. MediaSapiens, 17 жовтня 2024 (авторка Юлія Поліковська).
54. Texty.org.ua (2024). *Як же нами легко маніпулювати. Аналіз коментарів до жартівливих дінфейків у ТікТоці*. Тексти, 19 вересня 2024 (Юлія Дукач, Марія Ревук).
55. Центр стратегічних комунікацій (SPRAVDI, 2025). *РФ за допомогою ШІ генерує дінфейки про «невдячних» українських біженців*. 3 лютого 2025.
56. Інститут масової інформації (ІМІ, 2023). *У тіктоку шириться хвиля російських дінфейкових відео з українськими політиками*
57. ЦЕДЕМ (2023). *Дезінформація та штучний інтелект: (не)видима загроза інформаційній безпеці*. (Блог Центру демократії та верховенства права, 5 квітня 2023).

- 58.Ukrinform (2023). *Штучний інтелект і дезінформація: можливості та ризики в умовах війни*. Укрінформ, 5 квітня 2023.
- 59.Детектор медіа (2023). *«Репортери без кордонів» опублікували рекомендації щодо боротьби з дінфейками*. 12 квітня 2023.
- 60.AIN.ua (2024). *Понад 20% коментаторів повірили у відео з дінфейковими політиками — дослідження Texty.org.ua*. AIN, 21 вересня 2024.
- 61.Liga.net (2023). *Шахраї використовують ШІ для клонування голосу: як не стати жертвою нового виду афери*. Ліга.Tech, 18 травня 2023.
- 62.StopFake (2022). *“Це фейк”: як розпізнати та протидіяти дінфейкам*. StopFake, 30 березня 2022.
- 63.VoxCheck (2023). *Дінфейки: нова зброя російської пропаганди. Розбір VoxCheck*. Vox Ukraine, 11 липня 2023.
- 64.Інтерфакс-Україна (2023). *РНБО: Росія готує хвилю дінфейків проти вищого керівництва України*. Інтерфакс, 15 березня 2023.
- 65.Кореспондент.net (2023). *Фейкове звернення Зеленського: як технологія deerfake стала зброєю пропаганди*. Кореспондент, 16 березня 2023.
- 66.Детектор медіа (2022). *YouTube блокуватиме дінфейки, створені штучним інтелектом*. 14 березня 2022.
- 67.Texty.org.ua (2023). *У Німеччині серед молоді у TikTok процвітають російські фейки та пропаганда – дослідження*. Тексти, 6 жовтня 2023.
- 68.Суспільне Новини (2023). *Дінфейк – що це таке і як вберегтися від брехні? (роз’яснювальний сюжет)*
- 69.СтратКом Україна (Facebook, 2024). *Дінфейк — що це таке і як вберегтися від брехні?* (пост від 7 листопада 2024).
- 70.Міністерство цифрової трансформації (2023). *Кіберзлочинці все частіше використовують ШІ для фейків*

71. Pawelec M. (2022). *Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions*. *Digital Society*, 1(2), 19.
72. Momeni M. (2024). *Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions Through Misinformation*. *Journal of Creative Communications*, 19(1), 41–56.
73. Barman D., Guo Z., & Conlan O. (2024). *The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination*. *Machine Learning with Applications*, 16, 100545.
74. Raman R., Nair V. K., Nedungadi P., та ін. (2024). *Fake news research trends, linkages to generative artificial intelligence and sustainable development goals*. *Heliyon*, 10(3), e024727.
75. Diel A., Lalgi T., Schröter I. C., & MacDorman K. F. (2024). *Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers*. *Computers in Human Behavior Reports*, 9, 100538.
76. Nasiri Sh. & Hashemzadeh A. (2025). *The Evolution of Disinformation from Fake News Propaganda to AI-driven Narratives as Deepfake*. *Cyberspace Studies*, 9(1), 229–250.
77. de Nadal Ll. & Jančárik P. (2023). *Beyond the deepfake hype: AI, democracy, and “the Slovak case”*. *HKS Misinformation Review* (Harvard Kennedy School).
78. Goldstein J. A. & Sastry G. (2023). *The Coming Age of AI-Powered Propaganda: How to Defend Against Supercharged Disinformation*. *Foreign Affairs*, April 7, 2023.
79. Easterly J., Schwab S., & Conley C. (2024). *Artificial Intelligence’s Threat to Democracy: How to Safeguard U.S. Elections From AI-Powered Misinformation and Cyberattacks*. *Foreign Affairs*, Jan 3, 2024.

80. Hwang Y., Ryu J. Y., & Jeong S. H. (2021). *Effects of disinformation using deepfake: The protective effect of media literacy education*. (Анотація дослідження, цит. за Hwang & Ryu, 2021 у Sage Journals).
81. Groh M. et al. (2022). *Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds*. Proceedings of the National Academy of Sciences (PNAS), 119(1), e2123000119.
82. Kietzmann J. et al. (2020). *Deepfakes: Trick or Treat?* Business Horizons, 63(2), 135–146.
83. Vaccari C. & Chadwick A. (2020). *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*. Social Media + Society, 6(1).
84. Mavchak R. & Panasiuk O. (2023). *Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets relating to deepfakes and the Russo-Ukrainian war*. Harvard Kennedy School Misinformation Review, 4(3).
85. Torres A. et al. (2023). *The global emergence of deepfakes in news and political communication: A review*. New Media & Society, 25(5), 1340–1357.
- RAND Corporation. Artificial Intelligence, Deepfakes, and Disinformation: A Primer. (2021)
86. Freedom House. Freedom on the Net 2023: The Repressive Power of Artificial Intelligence. (04.10.2023)
87. Detector Media. Росія використовує дипфейки проти українських біженців — розслідування. (02.02.2025)
88. The Guardian. How China is using AI news anchors to deliver its propaganda. (18.05.2024)
89. The Record (Recorded Future). Deepfake news anchors spread Chinese propaganda on social media. (09.02.2023)
90. EUvsDisinfo. What we have learnt about FIMI after three years of full-scale war in Ukraine. (2023)

- 91.Detector Media. ЦПД: Пропагандисти поширили дипфейкове відео з актором Віллемом Дефо, який нібито ображає Україну. (07.01.2025)
- 92.NATO StratCom COE. Virtual Manipulation Brief 2024/1: Hijacking reality. (2024)
- 93.Reuters. Deepfaking it: America's 2024 election collides with AI boom. (30.05.2023)
- 94.FedScoop. Bipartisan House bill seeks labeling and disclosures for AI deepfakes. (2024)