

ШІ-ДИПФЕЙКИ ТА ТИСК НА ЕМОЦІЇ

ЯК СОЦМЕРЕЖІ ПЕРЕТВОРИЛИСЯ НА ПЛАТФОРМИ З ПОШИРЕННЯ БРЕХНІ

Кожна з чотирьох великих соцмереж – фейсбук, інстаграм, телеграм і твіток – має свої особливості, але є те, що їх поєднує: спільні алгоритми в поширенні дезінформаційних повідомлень.

Соціальні мережі перестали бути просто майданчиками для спілкування, розваг чи обміну новинами. В умовах війни вони стали повноцінним інформаційним середовищем, де формуються настрої, страхи, уявлення про реальність і навіть рішення, що впливають на безпеку людей. Фейсбук, телеграм, твіток та інстаграм сьогодні не лише відображають події – вони впливають на те, як ці події сприймаються, інтерпретуються і запам'ятовуються.

Сучасна дезінформація вже не виглядає як примітивна вигадка. Вона маскується під особисті історії, емоційні відео, «інсайдерські» повідомлення, психологічні поради або нейтральний лайфстайл-контент. Алгоритми соціальних мереж підсилюють саме такий контент – той, що викликає сильну реакцію, провокує обговорення й змушує ділитися побоченим без паузи на перевірку.

У цій реальності фейки, маніпуляції та інформаційні атаки стають системними. Вони використовують людські емоції, втому, страх, довіру до «своїх» і технологічні можливості штучного інтелекту. Контент легко масштабується, адаптується під різні платформи й аудиторії та поширюється одночасно в кількох соцмережах, створюючи ілюзію масовості й правдивості.

ФЕЙСБУК: ЕМОЦІЙНІ ІСТОРІЇ, ГРУПИ Й АЛГОРИТМІЧНЕ ПІДСИЛЕННЯ

Фейсбук залишається однією з платформ для споживання новин, особливо серед людей середнього та старшого віку. Тут працює формат великих текстових дописів і «особистих історій» – нібито реальних свідчень, сповідей і життєвих ситуацій, які апелюють до співчуття, страху чи обурення.

Типовий приклад – допис від імені звичайної людини, яка розповідає про власний досвід: проблеми з виплатами, мобілізацією, допомогою держави, волонтерством. У тексті майже немає перевірюваних фактів, зате багато деталей, що створюють ефект автентичності: емоції, побутові дрібниці, розмовна мова. Такий контент викликає співчуття й бажання поділитися, навіть якщо він містить маніпулятивні або неправдиві висновки.

Особливо ефективно це працює у фейсбук-групах, де інформація сприймається як внутрішня й «для своїх». Часом їх створюють як нейтральні: про допомогу, життєві питання, побут, громаду. Протягом тривалого часу група може не містити жодного допису, який би стосувався політики. Але згодом у ній починають з'являтися дописи з чіткими тезами – наприклад, про «зраду», «приховування правди», «реальний стан справ, про який мовчать».

Алгоритми платформи реагують на будь-яку активність – лайки, коментарі та поширення. Навіть критичні реакції збільшують охоплення дописів, що робить фейсбук зручним середовищем для швидкого масштабування маніпуляцій.

Окрему роль у сучасних маніпуляціях у фейсбуці відіграє штучний інтелект. Він дає змогу масштабувати дезінформацію до рівнів, які раніше були неможливими: створювати десятки тисяч коментарів, що імітують «живу» дис-

кусію, моделювати правдоподібні акаунти зі штучною історією активності та персоналізувати повідомлення під різні групи користувачів. Завдяки цьому фейковий допис може виглядати як підтриманий «звичайними людьми», а маніпулятивний наратив – як панівна думка в спільноті.

ШІ використовується не лише для текстів. У фейсбуці дедалі частіше з'являються фейкові фото й відео, зокрема пов'язані з війною, а також синтетичні голоси, які імітують реальних людей. Генеративні системи на кшталт Sora дозволяють створювати реалістичні відеосцени, що виглядають як зняті очевидцями, при тому що подія ніколи не відбувалася. У поєднанні з алгоритмічним підсиленням фейсбуку й активністю ботомереж такий контент швидко набирає охоплення і сприймається як достовірний – особливо тоді, коли під ним уже є сотні «емоційних» коментарів.

ТЕЛЕГРАМ: АНОНІМНІСТЬ, АВТОМАТИЗАЦІЯ ТА МАСОВІ ВКИДИ

Під час повномасштабної війни телеграм став для багатьох головним джерелом новин і термінових повідомлень. Саме сюди люди заходять під час обстрілів і кризових ситуацій у пошуках швидкої інформації. Водночас ця оперативність зробила месенджер одним із найнебезпечніших майданчиків для поширення дезінформації, адже телеграм практично не має модерації, дозволяє повну анонімність і не обмежує створення мереж каналів.

Анонімні телеграм-канали стають першою точкою появи фейків і маніпулятивних вкидів. Саме тут запускають чутки, деморалізаційні повідомлення, «інсайдерські» нібито «зливи» про фронт, мобілізацію чи рішення влади. Далі цей контент підхоплюють більші канали, блогери й інші соцмережі, що створює ілюзію масового підтвердження. Через це навіть неправдива інформація починає сприйматися як факт.

Телеграм особливо зручний для інформаційних операцій через технічні можливості автоматизації. Відкрите API дозволяє безплатно керувати сотнями каналів, одночасно публікувати ідентичні повідомлення, накручувати перегляди й коментарі, запускати ботоферми. За допомогою штучного інтелекту такі повідомлення легко переписують у різних варіантах, адаптуючи тон і лексику під конкретну аудиторію, щоб вони виглядали як дописи «реальних людей».

Окрему загрозу становлять фейкові «офіційні» канали та псевдопатріотичні спільноти. Вони можуть місяцями публікувати правдиві новини, накопичувати довіру, а згодом почати поширювати маніпулятивні наративи – про «зраду», корупцію, втрати або нібито злочини українських військових. Такі канали часто маскуються під місцеві новинні пабліки або сторінки державних структур і виглядають легітимно для пересічного користувача. Книжки з історії

Маніпуляції в телеграмі часто будуються на емоціях і поділі суспільства. Поширюють контент, який протиставляє регіони, військових і цивільних, переселенців і місцевих, «своїх» і «чужих». Коментарі під тригерними темами підсилюють ботоферми, створюючи відчуття паніки або «народного обурення». У результаті користувачі реагують емоційно, не перевіряючи джерела, і самі стають частиною поширення фейків.

Телеграм залишається важливим каналом отримання інформації, але його не можна сприймати як нейтральний або безпечний простір. Відсутність контролю, автоматизація й використання штучного інтелекту зробили месенджер ін-

струментом для інформаційних атак. Саме тому будь-які важливі повідомлення з телеграму потребують додаткової перевірки через офіційні джерела та надійні медіа.

ТІКТОК: ВІДЕОФЕЙКИ, ЕМОЦІЇ ТА ІЛЮЗІЯ СВДЧЕННЯ

Тікток працює насамперед через відео. Короткі ролики створюють ефект присутності й «очевидця», навіть якщо подія є постановною або згенерованою.

Платформа підходить для поширення інсценованих відео, вирваних із контексту фрагментів, підміни озвучки та контенту, згенерованого штучним інтелектом. Алгоритми тіктоку просувають ролики з високою емоційною реакцією – страхом, гнівом, співчуттям або обуренням.

Окремий пласт такого контенту – антимобілізаційні відео. За допомогою генеративних інструментів створюють або підсилюють ролики, які нібито показують насильницьку мобілізацію, «полювання» ТЦК чи приниження цивільних. Частина таких відео є постановними, інша – змонтована з фрагментів чужих подій, а ще частина – повністю згенерована або доопрацьована штучним інтелектом. У тіктоку такі ролики стають вірусними, бо алгоритм просуває контент, що викликає емоційну реакцію.

ШІ також застосовується для створення деморалізаційних відео від імені нібито українських військових. Це можуть бути синтетичні персонажі або змонтовані образи реальних людей, яким за допомогою AI-генерованого голосу приписують заяви про «зраду», відсутність мотивації, втрати або «безглуздість спротиву». Такі ролики подають у форматі «щирого зізнання» або «правди з фронту» й особливо ефективно працюють на аудиторію, яка не має можливості перевірити джерело або контекст. Книжки з історії

Ще один напрям – абсурдні мемні відео, створені за допомогою штучного інтелекту. Формально вони можуть виглядати як гумор або сатира, однак їхнє завдання – дискредитувати українських військових, владу чи міжнародну підтримку. Через перебільшення, карикатурні образи та гротеск такі ролики знижують серйозність теми війни, нормалізують знецінення втрат і поступово формують цинічне ставлення до реальності. Навіть коли користувач сприймає такий контент як жарт, емоційний слід і потрібний наратив залишаються.

Алгоритми тіктоку підсилюють ефект: чим довше користувач дивиться ролик – навіть без лайка чи коментаря – тим більше подібного контенту він отримує. У результаті формується замкнена стрічка з відео, де ШІ-контент, постановка й маніпуляція змішуються з реальними подіями, а межа між правдою і вигадкою стає майже непомітною.

Окремим явищем стали так звані тікток-експерти – люди без фахової підготовки, які впевнено коментують складні теми. Через ефект близькості й схожості з глядачем їм часто довіряють більше, ніж офіційним джерелам.

Коментарі, ботоактивність і повторюваність подібних відео створюють ілюзію правдивості. Навіть сумнівний ролик, поширений без перевірки, швидко стає вірусним і формуює уявлення про реальність.

ІНСТАГРАМ: МАНІПУЛЯЦІЇ ПІД МАСКОЮ ЕСТЕТИКИ ТА ТУРБОТИ

Інстаграм уже вийшов за межі платформи для фото й особистих історій. Сьогодні це теж один із каналів для поширення маніпуляцій, дезінформації та прихованої пропаганди. Йдеться не про прямі фейки, а про м'які, емоційно упаковані повідомлення – через «щирі» зізнання, псевдопсихологічний контент, лайфстайл-історії, фейкові волонтерські збори та акаунти, що імітують реальних людей.

Алгоритми інстаграму підсилюють контент, який викликає емпатію та довіру: привабливі обличчя, естетичні кадри, розповіді з повсякденного життя. Ця атмосфера близькості знижує критичність сприйняття й створює відчуття безпечного простору, де інформація сприймається не як вплив, а як особиста думка або дружня порада. В умовах війни, стресу та інформаційної втоми це робить платформу вразливою до маніпуляцій.

Свою роль у цьому відіграє штучний інтелект. За його допомогою створюють або доопрацьовують тексти, зображення й відео, які виглядають правдоподібно й емоційно. Мета запроваджує маркування AI-контенту та нові правила для політичних наративів, однак ці обмеження легко обходять: достатньо мінімальної обробки файлів або пересилання через месенджери – і маркери походження зникають. Заблоковані акаунти замінюються новими, тому мережі маніпулятивного контенту зростають швидше, ніж встигає реагувати модераторія.

Маніпулятивні публікації зазвичай не містять фактів, які можна перевірити. Натомість використовують розмиті формулювання на кшталт «усі знають» або «зараз усі про це говорять», апелюють до сильних емоцій – страху, обурення, жалю. Емоційний ефект підсилюють плачем у відео або драматичною подачею, а критичні коментарі видаляють, залишаючи лише підтримку. Фотокоментарі при цьому може бути не підробленим, а вирваним із контексту – старі або чужі зображення подають як актуальні.

Окремий ризик становлять рилз і сториз, де поширюють постановні сцени або відео, згенеровані чи доопрацьовані ШІ. Людина говорить про мобілізацію, владу чи війну, а на тлі з'являються емоційні кадри, які не мають прямого стосунку до сказаного, але створюють потрібний настрій. Такий формат виглядає як особисте спостереження, а не маніпуляція, і саме тому працює особливо ефективно.

Інстаграм рідко використовують для відкритої пропаганди. Натомість тут поширюють повідомлення, замасковані під турботу, втому, гуманізм або особистий досвід. У цьому середовищі межа між правдою, емоцією та навмисним впливом стає майже непомітною, тому сприймати інстаграм як нейтральний простір небезпечно – критичне мислення тут потрібне не менше, ніж у новинній стрічці.

СПІЛЬНА ЛОГІКА ВСІХ ПЛАТФОРМ

Попри різні формати, всі соцмережі при поширенні маніпулятивного контенту або фейків працюють за схожою схемою:

емоційний тригер → алгоритмічне підсилення → ілюзія масовості → зниження критичного мислення.

Контент легко адаптується під різні платформи й аудиторії, повторюється в різних формах і створює відчуття «всюдисущої правди». Саме тому сучасні фейки майже ніколи не існують лише в одній соцмережі – вони живуть одразу в кількох екосистемах.

Інформаційні атаки не є хаотичними. Вони створюються цілеспрямовано – з урахуванням психології людей, алгоритмів платформ і технологічних можливостей штучного інтелекту.

Протидія їм починається з уповільнення: не реагувати автоматично, перевіряти джерела, аналізувати контекст і розуміти, чому саме цей контент з'явився у стрічці. У сучасних умовах це не просто навичка медіаграмотності, а елемент особистої та суспільної безпеки.

Марія МАРТИНЮК

<https://litgazeta.com.ua/>
12.10.2026