Аналітик-фактчекер про те, як розпізнати дипфейки: "Створити фейк легше, дешевше та швидше, ніж викрити його"

Читать на русском

Аналітик-фактчекер про те, як розпізнати дипфейки: "Створити фейк легше, дешевше та швидше, ніж викрити його"

Наприкінці лютого – на початку березня 2024 року користувачі українського сегменту тиктоку стали помічати відео, на яких нібито українські та російські солдати розповідають "правду" про війну, кажуть, що "між Путіним та Зеленським немає різниці", і підбурюють до так званого "Майдану- 3".

Однак якщо вдивитися у відео, то можна помітити, що обличчя людей на них часом виглядають плоскими, на них специфічні тіні, а голова начебто приклеєна до тіла. При цьому мова тих, хто говорить досить якісно записана, - зазначає видання " Сьогодні ". Ці відео – дипфейки (deepfake), створені за допомогою штучного інтелекту (ІІ).

Якість відео, створених штучним інтелектом, стрімко покращується. Так, у лютому 2024 року користувачів соцмереж здивувала презентована компанією OpenAI нейромережа Sora , яка генерує досить реалістичні відео високої якості за текстовим запитом. З фотографіями помітити підробку ще складніше. Перебуваючи в емоційному стані, особливо коли йдеться про повідомлення, пов'язані з війною, користувачі можуть не звертати уваги на те, чи реальні фото, які вони знаходять, і ділитися ними у своїх профілях.

Телеканал "Сьогодення" розповідає про те, якими бувають дипфейки, що допомагає звичайним користувачам не потрапляти на ці хитрощі і як соціальні мережі та ЗМІ можуть протидіяти поширенню фальшивих фото, аудіо та відео, створених за допомогою штучного інтелекту.

Дипфейки: користь і небезпека

Технології, що дозволяють створити фейкові відео, фото або аудіо, всупереч поширеній думці, приносять і користь: їх можна використовувати в освітніх цілях, культурі, науці. Проте значно частіше про дипфейки говорять у контексті поширення дезінформації та пропаганди. Одним із проектів, які викривають фейки в Україні, є проект VoxCheck, створений на базі аналітичного центру VoxUkraine. Він розпочинав роботу з класичного фактчекінгу: перевірки фактів та даних, оприлюднених політиками, а пізніше почав перевіряти також експертів та медіа. У 2020 році VoxCheck став одним із партнерів Facebook у фактчекінговій програмі для перевірки та оцінки точності контенту.

Редактор та старший аналітик проекту VoxCheck Кирило Перевощиков каже, що у своїй роботі все частіше стикається з дипфейками. Він ділить їх у три умовні групи.

Перша – це фейки, які використовуються для інформаційних атак, для інформаційно-психологічних операцій (ІПСО), щоб посіяти паніку чи викликати недовіру до влади в Україні. "З останніх можна згадати різні дипфейки як Залужного, і Зеленського. Що, мовляв, Залужний записав відео, де оголошує про держпереворот, говорить про те, що Зеленський нелегітимний", – згадує Перевощиков. Також він нагадує про дипфейку, в якому ведучий французького телеканалу нібито оголосив , що президент Франції скасував візит до Києва у лютому 2024 через те, що так званий режим Зеленського готував на нього замах.

За основу було взято реальний репортаж, до якого перезаписали аудіо та трохи підкоригували відео. "І судячи з останніх заяв української влади, зокрема розвідки, росіяни готують спецоперацію під назвою "Майдан 3", де саме "качати" ситуацію всередині України та шерити фейки про мобілізацію, про легітимність президента. Тобто розгойдувати ситуацію, щоб люди остаточно втомилися від війни та були готові до капітуляції у цій війні", – розповідає аналітик.

Другий підвид дипфейків – комерційні, їх публікують не для підриву довіри до влади, а для накрутки підписок чи клікбейту. "Часто у тиктоці, у ютубі поширюються фейки із символікою відомих телеканалів в Україні, наприклад, ТСН, ICTV, 24 канал. І там відомі ведучі з аудіо, згенерованим за допомогою ІІ, оголошують новини, що шокують", – пояснює аналітик. Такі відео зазвичай низької якості, їх легко викрити, але якась частина користувачів інтернету все одно продовжує споживати їх.

І, нарешті, у третю групу аналітик виділяє дипфейки, створені заради жарту, – сатиру чи гумор. Є мережа каналів, як українських, так і російських, яка використовує ІІ для гумору, для розваги аудиторії. Вони відкрито заявляють, що їхній контент – гумористичний, його не треба сприймати як серйозний.

Але при цьому часто [інші користувачі або канали] ці відео постають як серйозні новини". На початку цього року один із телеграм-каналів, який регулярно публікує дипфейки з українськими політиками чи військовими, записав відео з Василем Малюком, головою СБУ, в якому той нібито обіцяє приходити особисто додому до українців, які публікують фото чи відео роботи ППО. Далі цей фейк опинився у Z-каналах як реальна заява СБУ.

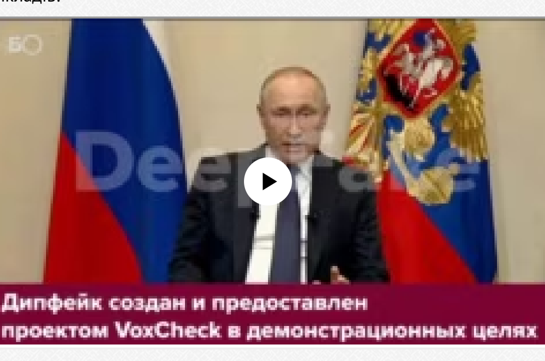
"Найнебезпечніші для українців, ну і, в принципі, для будь-якої аудиторії саме дипфейки, які поєднуються з ІПСО та створені для підриву підтримки України на Заході, для підриву довіри до влади, зниження підтримки армії або для розпалювання ворожнечі", – вважає Перевощиков .

Як розпізнати дипфейк

Тренд на дипфейки у пропаганді та дезінформації аналітик пов'язує з тим, що технології досить прості та доступні для пересічного користувача інтернету – не потрібно володіти навичками роботи у складних програмах, достатньо мати базову підписку на один із високотехнологічних сервісів за 1$.

Наприклад, працюючи в програмі ElevenLabs, яка добре сумісна з російською та українською мовами, можна генерувати досить якісні аудіо. Аналітик каже, що процес створення такої аудіодоріжки може зайняти 5-10 хвилин: достатньо знайти кілька прикладів голосу спікера з якісним звуком, щоб програма змогла скопіювати голос та запровадити текст, який потрібно цим голосом озвучити.

"Ми самі створювали дипфейки [з метою демонстрації]: для одного репортажу [про викриття дипфейків], наприклад, ми підробили голос Путіна, також у мене є приклад із Зеленським. І їхні голоси досить переконливо звучать, якщо не особливо вдаватися до деталей, прислухатися", – розповідає Кирило Перевощиков і демонструє один із прикладів:



Аудіо можна відредагувати, змінити акценти й надалі примонтувати до відео спікера, розповідає аналітик VoxCheck. Але працювати з відео вже складніше: "Часто пропагандисти трапляються саме на якості відео: там або міміка спікера відстає від звуку, або саме відео має погану якість, або є якісь артефакти на фоні, або губи якось дивно рухаються".

Ще одна важлива причина тренду на дипфейки у пропаганді – складність їхнього викриття. "Створити фейк легше, дешевше та швидше, ніж викрити його, тобто інструменти перевірки фейків, у тому числі і дипфейків, розвиваються повільніше, ніж інструменти для перевірки цих фейків. Ми як фактчекери, журналісти, будь-які працівники медіа завжди змушені запізнюватися, бо витрачаємо більше ресурсів, часу, щоб це викрити", – каже Перевощиков.

Він бачить причину успіху дипфейків ще й у тому, що аудиторія поки до них не звикла: "Якщо українці вже почали менше вживати російської дезінформації, менше дивитися російських каналів, то до таких технологій вони ще не звикли, і тому в них не вистачає знань, як цьому протидіяти".

Однак навіть звичайний користувач інтернету, який не має спеціальних інструментів, з ймовірністю може розпізнати дипфейк. "Головний інструмент, хоч би як пафосно це звучало, – це наша уважність", – каже Кирило Перевощиков.

Аналітик рекомендує для початку просто уважніше придивитися і прислухатися: "Якщо говорити про відео, ми звертаємо увагу на міміку спікера, чи вона збігається з аудіо, чи не відстає звук від картинки, яка якість самого відео чи аудіо. Сам голос може звучати неприродно, можуть бути якісь сторонні шуми, звуки – тобто ніби відео записане не в приміщенні, не на вулиці, а в студії, хоча для цього немає жодних передумов".

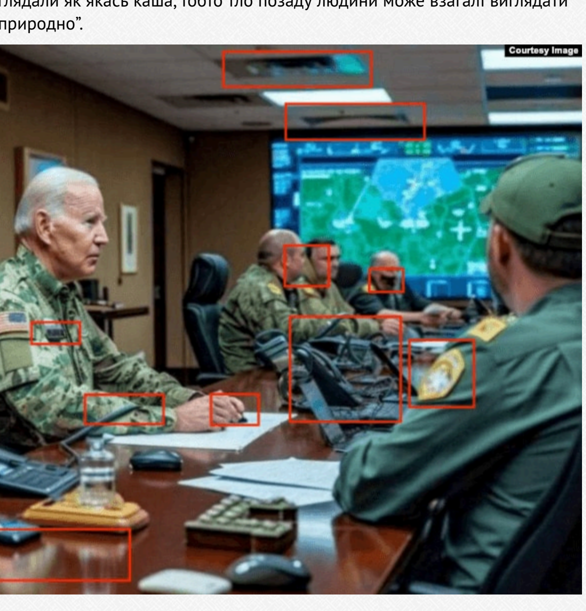
Крім того, недобре з відео може видати зацикленість рухів, повторення одних і тих же жестів того, хто говорить. Також аналітик VoxCheck рекомендує перевіряти першоджерело: якщо на відео чи фото задіяна якась відома людина (а більшість дипфейків, що розповсюджувалися в Україні, створені за участю відомих людей – військових, політиків, провідних, журналістів), завжди варто перевірити інформацію на офіційній сторінці або в перевірених та надійних ЗМІ.

"І якщо цієї інформації немає ні на офіційній сторінці, ні на сторінках відомих якісних перевірених ЗМІ, то тоді можна з упевненістю сказати, що це дипфейк", – констатує Кирило Перевощиков.

Розпізнати фото, згенероване ІІ, можливо, складніше. Однак вдивившись у картинку, можна помітити нестиковки: зайві або відсутні кінцівки або пальці, занадто глянсову шкіру, незвичайні символи. Коли генерують фото з групою людей, кілька людей можуть виглядати як близнюки. Крім того, якщо на підозрілому фото є відомі особи, варто знову ж таки перевірити першоджерело – чи це зображення публікувалося на офіційних каналах або в соцмережах цих людей.

"Наприкінці минулого року в українському сегменті фейсбуку часто розповсюджували фото начебто українських військових, які стоять на колінах та моляться перед боєм. І там, наприклад, на [формі] військової на передньому плані прапор України складався з трьох смуг – блакитна, жовта та блакитна. Тобто штучний інтелект якось дивно інтерпретував український прапор та створив новий символ. Більше того, самі військові були у формі, більше схожій на російську, вони не мали такого впізнаваного "пікселя".

До "проколів" штучного інтелекту, що часто зустрічаються, відносять дивно виглядаючий фон позаду зображених на картинці людей: "Нещодавно опублікували фото нібито Байдена у військовій формі, де він у командному центрі планує удари по Ємену. І всі дроти та девайси, які знаходилися на столі, дивно перемішалися і виглядали як якась каша, тобто тло позаду людини може взагалі виглядати неприродно".



Нерідко дипфейки використовуються для створення зображень про чудес природи. "Минулого року ми викривали фейк про те, що нібито моряки з Латинської Америки спіймали найбільшого у світі лобстера, і він виглядає за розміром як невеликий слон. Сам лобстер виглядає досить чітко, а люди позаду – замилені, весь обрій позаду них завалений. Усе виглядає максимально фальшивим на таких фото".

У викритті також може допомогти зворотний пошук зображень у будь-якому пошуковику (Google, Bing, Яндекс) – і система покаже картинки, схожі чи ідентичні до завантажених. В Яндексі краще шукати саме контент, створений російською мовою. Також звичайним користувачам може стати в нагоді вже згаданий сервіс для створення дипфейків ElevenLabs: у нього є своя програма AI Speech Classifier - вона визначає, чи створювали це відео або аудіо конкретно в цій програмі.

Оскільки ElevenLabs досить популярна серед українців та росіян, багато фейків можна виявити саме так. Крім того, фахівці згадують сервіси Hive Moderation або AI or not , але вони мають обмеження функцій, доступних безкоштовно. При цьому аналітики наголошують, що навіть спеціальні програми не можуть викрити усі дипфейки – і у програмах є похибки.

Крім того, є сервіси для перевірки, які доступні лише окремим фахівцям або ще перебувають у розробці та доступні великим компаніям, державним установам, але не широкій аудиторії.

"Тому ми наполягаємо на комплексному аналізі. Тобто не лише залили відео, перевірили в інструменті – все перевірка готова. Ні, спершу ми завжди перевіряємо першоджерела, завжди звертаємо увагу на якість фото, аудіо, відео, і вже після цього інструмент як бонус. Не основа нашої перевірки, а остаточний аргумент, що це не фейк", – ділиться тонкощами процесу перевірки контенту аналітик проекту VoxCheck Кирило Перевощиков.

Проте найкращий спосіб не потрапляти у пастку дипфейків – читати якісні джерела інформації, упевнений аналітик. І навіть у разі використання соцмереж та месенджерів, де може траплятися дуже різна і не завжди достовірна інформація, завжди можна передплатити якісні медіа, нагадує Перевощиков. "В Україні, наприклад, є білий список медіа, який складається Інститутом масової інформації. Там якраз загальноукраїнські медіа, які регулярно дотримуються стандартів. Також радимо читати україномовні версії міжнародних медіа – BBC, Deutsche Welle, Радіо Свобода. Тобто, якщо ви підписані на нормальні канали, якщо ви не читаєте анонімні канали, особливо масові в Україні, то у вас уже набагато менше шансів зіткнутися з дипфейком".

Дипфейки та відповідальність

На тлі все більшого поширення технології дипфейків компанія Meta (Facebook) оголосила , що в 2024 році, коли в усьому світі пройде низка важливих виборів (президентські вибори в США та Росії, вибори до Європарламенту та інші), вона маркуватиме будь-який контент, згенерований штучним інтелектом. Пости в інстаграмі та фейсбуці супроводжуватимуться повідомленнями про те, що зображення чи відео не справжні, а створені за допомогою ІІ. Також Meta планує маркувати ІІ-контент, створений не тільки в її сервісах, а й у продуктах інших компаній, наприклад, Google або Microsoft. Нововведення обіцяють втілити найближчими місяцями. Однак настільки серйозно до боротьби з фейками ставляться не всі соціальні мережі.

"Для українців найбільшу небезпеку становить телеграм, бо за опитуваннями він є головним джерелом інформації серед тих, хто читає інтернет-новини, соцмережі. І проблема телеграми в тому числі в тому, що топові телеграм-канали з великою кількістю передплатників – це частково або повністю анонімні канали, які не несуть жодної відповідальності за свій контент. Вони часто публікують і якісь брехливі новини, несумлінну рекламу, джинсу від політиків.

І мають мінімальні стандарти журналістики, вони кардинально відрізняються від традиційних ЗМІ. Українці почали читати менше новин із Росії та про Росію. Але саме проблема в тому, що навіть українські канали, які нібито й підтримують Україну, і можуть постити ті самі збори для військових, радіти проукраїнським новинам, публікують неперевірені новини.

У цьому якраз і пастка: ці канали вже після того, як набрали довіри, набрали передплатників, можуть постити дипфейки, навіть ту саму сатиру чи жарти. Але при цьому якщо вони не маркуватимуть це як несерйозний контент, деякі неуважні користувачі сприймуть їх як правду", – каже Перевощиков.

Він підкреслює, що при цьому в телеграмі немає інструментів для протидії фейкам: "Можна поскаржитися через насильство, поширення наркотиків, порнографії, але саме на фейки там не можна поскаржитися на відміну від, наприклад, фейсбуку. Або в тому ж твіттері є так званий інструмент Community Notes, тобто нотатки спільноти, і коли публікується якась дезінформація, користувачі можуть додати до цієї новини якийсь контекст: "Обережно! Тут наведені в повному обсязі факти, і у чому полягає правда". У телеграмі нічого подібного немає, і саме так, у цьому небезпека".

"Так само, як і з телеграмом, у ютубі, на жаль, немає інструментів для маркування фейкового контенту, є лише опція "скаржитися". І у нас поки що немає інструментів для обмеження хоча б їх поширення в соцмережах", – додає аналітик.

На тлі виборів до Європарламенту деякі ІІ-компанії пообіцяли, що співпрацюватимуть одна з одною, щоб люди отримували більш достовірні новини. Наскільки це може вплинути на поширення фейків, складно прогнозувати. Жодної конкретики про те, як саме вони протидіятимуть і якими інструментами, не повідомляли, наголошує аналітик. "Я бачу від них більше заяв, але менше справ саме для створення якогось продукту, наприклад, для тих самих фактчекерів", - каже аналітик проекту VoxCheck.

Що ж до регулювання дипфейків на державному рівні, то перший крок у цьому напрямку у березні 2024 року зробив Європарламент: він схвалив перший у світі закон, який регулює застосування штучного інтелекту державами. Він забороняє розпізнавання осіб у режимі реального часу, забороняє створення "соціальних рейтингів" та поліцейського "профілювання" з використанням ІІ, а також зобов'язує виробників контенту дотримуватись закону про авторське право та чітко маркувати все, що створено з використанням штучного інтелекту.

Автор: Любов Багацька

Джерело: Ніні

Теги: СБУ | VoxCheck | ІПСО | ИПСО | Перевощиков Кирилл | Фейк | Майдан-3 | Путін | Путин | Путин | Россия | Росія

Марія Войтенко
ОГЛЯДАЧ НОВИН

17 березня 2024 г., 21:09 | Переглядів: 1884 | Коментарі: 0

Роздрукувати | Надіслати товаришу

Коментарі:

comments powered by Disqus

Важливі новини
19.05.2026 Мільярди на крові під маскою меценатства: як власник «Трухи» Володимир Литвин легалізує дубайські ка... #Труха #Аваков #Urban One
13.05.2026 Розшукуваний бізнесмен Борис Ушерович через кіпрсько-іспанський ландромат... #РЖД #Кипр #Кіпр
ЦЕНЗУРА Як обійти блокування і читати наш сайт
Останні новини
По даті | По переглядам | По коментарям
23:11 Олександр Усик прибув на арену в Гізі перед першим боєм проти Ріко Вергувена
23:04 Росія атакувала інфраструктуру Одещини: серед дев'ятьох поранених є діти
22:55 Мільярд під землею: як три компанії з орбіти топпосадовців без конкуренції поділили будівництво підземних шкіл і дитсадків у Запоріжжі
22:47 Ситуація на фронті: Зафіксовано 173 боєзіткнення, ворог найактивніше атакує на Покровському та Гуляйпільському напрямках
22:39 Подоляк назвав можливі терміни завершення активної фази війни та скасування воєнного стану
Підписуйтесь на наш канал в Telegram. Оперативно про головне
22:31 РФ розгорнула на пускових позиціях понад 1000 безпілотників: попереджають про рух великої кількості БпЛА у напрямку Києва та області
22:23 Рівненський апеляційний суд оштрафував волинянина на 247 тисяч гривень за махінації з документами на Volkswagen
22:19 Схема провалилася: Одеський суд конфіскував раритетний Aston Martin через підробку документів і заниження вартості
22:15 Іран та США вийшли на фінальний етап підготовки меморандуму, – МЗС Ірану
22:09 Застосування газу та штовханина з перехожими: у Харкові місцеві жителі завадили затриманню чоловіка військовими
22:02 У Німеччині більше непрацюючих українців, ніж сирійців, - голова Офісу міграційної політики
21:55 Соліст Imagine Dragons під час концерту в Торонто розгорнув прапор України та побажав тривалого миру
21:49 На тлі переговорів та загрози відновлення ударів: Трамп опублікував карту Ірану в кольорах американського прапора
21:42 Попередження про масований удар: РФ може атакувати державні установи, повідомляють про ймовірний виліт стратегічної авіації РФ Ту-95
21:36 Сенатор Ліндсі Грем виступив проти можливої угоди США з Іраном, заявивши про загрозу балансу сил
21:29 Скандал у Татрах: українському блогеру Гавриліву заборонять в'їзд до Польщі на 5 років
21:21 Атака "шахедів": у Києві лунають вибухи
21:13 Перший бій у 2026 році: Усик б'ється за пояс WBC проти Вергувена
21:10 На війні в Україні загинув 23-річний громадянин Великої Британії Айртон Редферн, - BBC
21:06 Мобілізація в Києві: Конфлікт між перехожим і військовими ТЦК на Троєщині потрапив на відео
21:00 DeepState: Російські війська просунулися на південний схід від Слов'янська
20:51 Ветеран Мельник заявив, що у США виявляють українських військовозобов'язаних для повернення додому
Теги новин
COVID-19 агрессия России Атака Війна Война ВСУ Вторжение Дональд Трамп Донбасс ДТП Зеленский ЗСУ Киев Київ коронавирус Коррупция Напад Росії на Україну Нападение России на Украину оккупанты окупанти Порошенко Путин Росія Россия СБУ США Украина Україна ЧП Эпидемия коронавируса

Наші опитування
Чи вірите ви, що Дональд Трамп зможе зупинити війну між Росією та Україною?
Так, повністю зможе
Частково зможе, але не відразу
Ні, не зможе
Це залежить від дій інших сторін
Важко відповісти
Голосувати
Показати результати опитування
Показати всі опитування на сайті

2013-2026 © АНТИКОР — національний антикорупційний портал
Реклама на сайті • Наші партнери
Політика конфіденційності
Використання матеріалів сайту дозволено лише за наявності активного гіперпосилання на джерело. Усі права на тексти, зображення, фотографії та відеоматеріали належать їх авторам.
Facebook | Twitter | YouTube | RSS-підписка | Email-розсилка | Мобільна версія