Визначення ключових термінів із презентації проф. д-ра Мартена Рісіуса

Визначення на основі інформації від Google AI та ChatGPT

Ехо-камери

Цифрові простори (напр., групи у соцмережах), де люди майже виключно бачать інформацію, що підтверджує їхні власні погляди; це підсилює дезінформацію. Ехо-камери складаються лише з людей зі схожими поглядами, що позбавляє обміну з «зовнішнім світом».

Доксинг (doxing)

Публікація в інтернеті приватних даних людини (напр., адреси, телефону), часто як засіб залякування або покарання.

Цифровий вігіліантизм (самосуд онлайн)

Самостійне «здійснення правосуддя» в інтернеті — коли користувачі карають інших самі (напр., масовими нападами, шеймінгом або мовою ворожнечі) замість звернення до легітимних інституцій.

Організовані атаки з ransomware

Атаки хакерських груп, які шифрують дані та вимагають викуп — також використовуються для саботажу медіа та інформаційних каналів.

Приховане блокування (shadow banning)

Невидиме обмеження облікових записів чи дописів без відома користувачів (напр., зменшення охоплення), щоб стримувати дезінформацію.

Чат-боти з контр-наративами

Чат-боти на основі ШІ, що відповідають на неправдиву інформацію контраргументами або фактчекінгом.

Детоксикація на основі LLM

Застосування великих мовних моделей (LLM) для виявлення й зменшення токсичного або маніпулятивного контенту або для його перефразування у нейтральнішому тоні.

Prebunking (попереджувальне викриття)

Превентивна освіта, яка заздалегідь інформує про типові стратегії фейків, щоб люди краще розпізнавали маніпуляції.

Фідуціарні (опікунські) AI-агенти

Системи штучного інтелекту, покликані діяти в інтересах користувача (напр., фільтри проти дезінформації), подібно до «цифрового довірителя».

Red herring (відволікальний маневр)
Відволікання в аргументації — вводиться несуттєва тема, щоб відвести увагу від справжнього питання або фейку.

Вибіркове підбирання (cherry-picking)
Навмисний добір лише найзручніших елементів із множини, ігноруючи решту.

Хибні дихотомії
Логічні хиби, коли ситуацію подають як вибір лише між двома крайніми, взаємовиключними варіантами, хоча існує більше можливостей і відтінків.

Ad hominem
Риторична тактика/хиба, коли атакують особу замість аргументу, щоб послабити позицію опонента і вплинути на аудиторію.

Поляризація
У політичному контексті — соціальне розділення, що веде до конфліктів, або посилення розбіжностей у думках. Часто обидва процеси йдуть разом.

Наслідування
Наслідування або копіювання поведінки, стилю чи контенту інших.

Аргументи «слизького схилу»
Риторичні твердження, що нібито безпечна початкова дія (A) неминуче веде до низки негативних наслідків, які завершуються морально неприйнятним станом (B).

Деконтекстуалізація
Навмисне вилучення інформації, текстів, зображень або відео з їхнього первісного контексту, щоб надати їм нового, часто оманливого значення.

Троль
Людина — або іноді програма (бот) — яка навмисно публікує провокативний, образливий чи оманливий контент, щоб зірвати обговорення та провокувати або переслідувати інших.

Боти
Автоматизовані комп’ютерні програми, що самостійно виконують повторювані завдання швидше за людину.

Клонування
Створення ідентичних копій.

Домени
Людиночитні унікальні назви в інтернеті, що слугують адресами для знаходження конкретних сайтів замість важких для запам’ятовування числових IP-адрес.

Gatekeeper (вартовий доступу)
Особа, яка контролює доступ до чогось (напр., охоронець на вході), або, ширше, контролює, хто отримує доступ до певної категорії чи статусу.