

Дефиниции на ключови термини от презентацията на проф. д-р Мартен Рисиус

Дефиниции въз основа на информация от Google AI и ChatGPT

Ехо камери

Дигитални пространства (напр. групи в социалните мрежи), в които хората виждат почти изцяло информация, която потвърждава собствените им убеждения; така се усилва дезинформацията. Ехо камерите се състоят само от хора със сходни възгледи, което лишава от обмен с външния свят.

Доксинг (doxing)

Публикуване в интернет на лични данни на човек (напр. адрес, телефон), често като средство за сплашване или наказание.

Дигитално самоуправие (vigilantism)

Самораздаване на „правосъдие“ в интернет — когато потребители наказват самостоятелно (напр. чрез масови нападки, засрамване или език на омразата), вместо да се обръщат към легитимни институции.

Организираните атаки с ransomware

Атаки от групи хакери, които криптират данни и искат откуп — използват се и за саботиране на медии и информационни канали.

Скрито ограничаване (shadow banning)

Невидимо ограничаване на акаунти или публикации без засегнатите да забележат (напр. намален обхват), за да се ограничи дезинформацията.

Чатботи с контра-наративи

Чатботове с изкуствен интелект, които отговарят на невярна информация, като добавят контрааргументи или проверки на факти.

LLM-базирана детоксикация

Използване на големи езикови модели (LLM) за откриване и намаляване на токсично или манипулативно съдържание или за пренаписването му в по-неутрален тон.

Prebunking (превантивно разобличаване)

Превантивно обучение, което информира предварително за типични стратегии на фалшиви новини, за да могат хората по-лесно да разпознават манипулации.

Фидуциарни (попечителски) AI агенти

Системи с изкуствен интелект, предназначени да действат в най-добрия интерес на потребителите (напр. филтри срещу дезинформация), подобно на „дигитален доверител“.

Red herring (отвличане на вниманието)

Разсейване в аргументацията — въвежда се несъществена тема, за да се отклони вниманието от същинския въпрос или фалшивата новина.

Избирателно подбиране (cherry-picking)

Умишлено избиране само на най-удобните елементи от даден набор, като останалото се игнорира.

Фалшиви дихотомии

Логически заблуди, при които ситуацията се представя като избор само между две крайни, взаимно изключващи се опции, въпреки че съществуват повече възможности и нюанси.

Ad hominem

Реторична тактика/заблуда, при която се атакува човекът вместо тезата, за да се отслабят аргументите му и да се повлияе на публиката.

Поляризация

В политически контекст — или социално разделение, водещо до конфликти, или засилване на различията в мненията. Често и двете вървят заедно.

Имитация

Подражаване или копиране на поведение, стил или съдържание на други.

Аргументи по „хлъзгавата плоскост“

Реторични твърдения, че уж безобидно начално действие (А) неизбежно води до верига от негативни последствия, завършващи с морално неприемлив краен резултат (В).

Деконтекстуализация

Умишлено изваждане на информация, текстове, изображения или видеа от оригиналния им контекст, за да им се придаде нов, често подвеждащ смисъл.

Трол

Човек — или понякога бот — който нарочно публикува провокативно, обидно или подвеждащо съдържание, за да разстройва дискусии и да провокира или тормози други.

Ботове

Автоматизирани компютърни програми, които самостоятелно изпълняват повтарящи се задачи по-бързо от човек.

Клониране

Създаване на идентични копия.

Домейни

Домейн е уникално, четимо от хора име в интернет, което служи като адрес за намиране на конкретен уебсайт вместо трудно запомняем числов IP адрес.

Gatekeeper (пазач на достъпа)

Лице, което контролира достъпа до нещо (напр. охрана на вход), или по-абстрактно — кой получава достъп до определена категория или статус.