

Како се забележува и разбира содржина генерирана со ВИ или дигитално изменета

Модели на Вештачка Интелигенција (ВИ) коишто се пристапни и лесни за употреба може да им помогнат на луѓето да научат и да создадат содржини, но тие исто така може да ги зголемат ризиците од мис/дезинформации што се закана по отворените општества и демократскиот дискурс. Важно е да се спречи нашите заеднички простори да се натрупаат со мис/дезинформации генерирани со ВИ или со дигитално изменети содржини.

Дел од решението е новата технологија како иницијативи за пронаоѓање на потекло и софтвер за откривање. Но, технолошките решенија се далеку од совршени и затоа потребна е работата на независните проверувачи на факти за да обезбедат општества со споделени верификувани информации.

Следува брз преглед на она што го прават професионалните и независни проверувачи на факти при идентификацијата и разоткривањето на мис/дезинформациите генерирани со ВИ и она што може да се научи од нив.

Сè повеќе ВИ-генерирани содржини

Денес, мис/дезинформации генерирани со ВИ претставуваат мал процент од сите анализирани тврдења на професионалните и независни проверувачи на факти. Позастапени во нивната работа се дигитално изменетите содржини.

Но: во внатрешно спроведена анкета на EFCSN, повеќето проверувачи на факти се согласуваат дека во иднина ќе се зголемува релевантноста на дигитално изменетата содржина и на онаа генерирана со ВИ. И [најновите примери](#) во контекст на европските избори одат во поддршка на таа прогноза.

СФАТИ: *Дигитално изменети* се сите форми на содржини значително изменети за да се манипулира или смени првичната порака, како и оние уредувани со ВИ-алатки. Ова не вклучува уредувања за јасност и квалитет.

Генерирани со ВИ се сите форми-содржини создадени од кој било систем на вештачка интелигенција.



Технологијата брзо напредува, но не може да се потпираме само на неа.

Експерти за ВИ и професионални проверувачи на факти се согласуваат: ВИ-алатките не се доволно за (раз)откривање дигитално изменети и ВИ-креирани содржини.

Пред користењето детектор, експерти препорачуваат да се запознаете со генератори и детектори на ВИ-содржини. Со разбирање на моделите и со малку статистика, проверувачите на факти може да почнат да ги согледуваат предностите и недостатоците на алатката, како и веројатноста на успех. Дури и тогаш алатките може да се добра појдовна точка.

Иницијативи за потекло како C2PA спецификации може да помогнат во сертификација на извор и историја на медиумска содржина, но водените знаци и верификацијата не се непобитни.

Како ВИ дезинформациите влијаат врз луѓето

„Секогаш кога се реагира со чувство - ако смеам така да кажам - се заобиколува умот“.

– Кристин Дугуан

ПСИХОЛОГИЈА: Операциите за влијание често се дизајнирани да ги злоупотребат психолошките предрасуди.

Разбирањето на сопствените и оние на публиката може да помогне во справувањето со дезинформации.

ЦЕЛИ: Зошто „нечесен играч“ би се потпрел на ВИ за креирање или ширење дезинформации? Што всушност сака да постигне со тоа?

- Да допре до друга држава или заедница?
- Да не биде откриен/да ги преплави проверувачите на факти со варијации од слично тврдење?
- Да влијае на мислење/убедување преку создавање доверба во неавтентички сметки на мрежи?

* Christine Dugoin е истражувач на влијанието на информациите при La Sorbonne.

Разоткривањето бара повеќестран приод и нијансирано разбирање

Ако не работат алатките за откривање, што работи? Разбирањето на контекстот е подеднакво важно колку и самата содржина. Професионалните проверувачи на факти се експерти за истражувачки вештини. Еве неколку совети. ПРАВОАГОЛНИКОТ-Хенк ван Ес

„Алатките за откривање не функционираат 100% - Не очекувам подобрување“.

– Henk van Ess**



ВОДИ СМЕТКА НА ИЗВОРОТ: Дали може да го потврдите идентитетот? За што зборуваат и што споделуваат? Кои реагира на нивната содржина? Каков ефект би имала содржината на читателите?



ИЗГРАДИ КРЕДИБИЛИТЕТ: Верификувај ја информацијата со независни и кредибилни извори како што се експерти со искуство од областа. Дали она што е прикажано е логично брз основа на вашето знаење?



Користи **ФОРЕНЗИЧКИ МЕДИУМСКИ** техники покрај традиционалното истражувачко известување и документарно истражување, како што се: извлекување податоци, геолокација, биометриско препознавање, шематска анализа и други.



УЧИ И АДАПТИРАЈ: Креаторите на мис/дезинформации генерирани со ВИ постојано се адаптираат. Адаптирај го твојот приод кон средината што постојано се менува.

СПОДЕЛИ ЈА СВОЈАТА РАБОТА

Експертите препорачуваат, заедно со разоктриените тврдења, да се споделуваат транспарентни анализи и линкови до изворите. Тоа им помага на читателите да разберат нијансирани наративи.

Во некои случаи, истражувањето е поважно отколку тоа што содржината е напишана со помош на ВИ.

** Henk van Ess е експерт во OSINT за техники за проверување факти.

Прирачник за брз упат: Прави, не прави и индиции

Следуваат сугестии коишто можат да индицираат за тоа дали содржината е дигитално изменета или е генерирана со ВИ. Заедно со другите индиции наведени во прирачничков (контекст, истражувачки техники и алатки за откривање), може да помогнат да ја сфатите вистината зад она што се гледа.

Текст

- Често (но не секогаш) има подобра граматика отколку просечен човек.
 - Во контекст на социјални медиуми, користи премногу формален и структуриран јазик.
 - Прекумерни прилози и придавки.
 - Недоволна човечка емоција, хумор, сарказам и изразување со идиоми.
 - Може да недостигнаат конкретни детали (имиња, датуми, места) или оригинални идеи.
- Најважно: точни ли се фактите наведени во текстот?

Видео

- Не користи детектор на слики генерирани со ВИ смрзнати од видео.
- Види ги изразите на лицето и движењата, како што се клепање, и дали движењето на устата се совпаѓа со звукот. Може да се забележат остри премини или сечења.

Аудио

Не користи детектор на слики генерирани со ВИ на исечоци од видео

Види ги изразите на лицето и движења како трепкање со очите, и дали движењето на устата се совпаѓа со збвукот.

Може да има неприродни или механички тек на говор, недоволно паузи или неприродно дишење.

Слики

- Барај неприродни детали: совршена кожа, замаглена заднина, неприродна убавина/светло и необични работи како повеќе прсти.
- Барај воден знак од најчести генератори на слики.
- Води сметка за деталите на сликата: дали е логично? Дали е соодветно?
- Кога користиш детектор, барај некој со висока резолуција или рана верзија на прикачената слика наместо некоја што била повеќепати споделувана