Як выявіць кантэнт, што створаны ШІ або лічбава зменены

Даступныя і простыя для карыстання сістэмы ШІ дапамагаюць людзям вучыцца і ствараць кантэнт, аднак разам з тым для адкрытых грамадстваў і дэмакратый яны ўзмацняюць **рызыку распаўсюджвання дэз- і скажонай інфармацыі.** Важна перадухіліць захламленне нашых агульных інфармацыйных прастораў фальсіфікацыяй ды дэзынфармацыяй.

Частка рашэння — новыя тэхналогіі: софт ды ініцыятывы па выяўленні крыніцы кантэнту і фэйкаў. **Аднак тэхналагічныя рашэнні не дасканалыя, і, каб атрымаць агульнапрымальны набор правераных фактаў, грамадству патрэбная праца незалежных фактчэкераў.**

Вашай увазе кароткі агляд таго, што робяць незалежныя прафесійныя фактчэкеры, каб ідэнтыфікаваць ды абвергнуць створаную ШІ дэзынфармацыю ды фальсіфікацыю.

Згенераваны ШІ кантэнт набірае папулярнасць

Сёння згенераваная ШІ няпраўда ды дэзынфармацыя займае невялікую долю ад усіх фэйкаў, якія даследуюць незалежныя прафесійныя фактчэкеры. А дыгітальна змененыя матэрыялы сустракаюцца ў працы фактчэкераў часцей.

Аднак: паводле ўнутранага апытання сяброў Еўрапейскай сеткі фактчэкінгу, **большасць фактчэкераў прагназуе, што значэнне кантэнту, створанага ШІ, і дыгітальна змененых матэрыялаў будзе толькі расці**. Нядаўнія прыклады ў кантэксце Еўрапейскіх выбараў пацвярджаюць гэта.

ТЛУМАЧЭННЕ: *Дыгітальная (лічбавая) змена* — гэта значная трансфармацыя любога кантэнту, якую ажыццявілі з мэтай маніпуляцыі ці змены першапачатковага значэння, уключаючы рэдагаванне інструментамі ШІ. Гэта не распаўсюджваецца на праўкі для паляпшэння зразумеласці ці якасці.

"Створаны ШІ" азначае любую форму кантэнту, што быў згенераваны сістэмамі штучнага інтэлекту.

Інфармацыю для вас падрыхтавала **AI@EUElections,** што фінансуецца Meta і знаходзіцца пад кіраўніцтвам **European Fact-Checking Standards Network** ва ўзаемадзеянні з яе сябрамі і знешнімі экспертамі.

Тэхналогіі развіваюцца хутка, але разлічваць толькі на іх нельга.

Эксперты па ШІ і прафесійныя спецыялісты па праверцы фактаў сыходзяцца ў меркаванні: **адных інструментаў выяўлення штучнага інтэлекту недастаткова для выяўлення або развянчання кантэнту, створанага ШІ або зmененага ў лічбавым выглядзе.**

Перш чым выкарыстоўваць дэтэктар, эксперты рэкамендуюць пазнаёміцца з ШІ-генератарамі кантэнту і дэтэктарамі. З разуменнем таго, як трэніруюцца сістэмы, і крупляй статыстыкі, фактчэкеры могуць прадбачыць моцныя і слабыя бакі інструмента ды верагоднасць яго поспеху. **Тым не менш, дэтэктары могуць быць карысным стартавым пунктам.**

Ініцыятывы па выяўленні крыніцы інфармацыі, такія як C2PA Specifications, могуць дапамагчы пацвердзіць рэсурс і гісторыю медыя кантэнту. Але што датычыць вадзяных знакаў і верыфікацыі — яны не бездакорныя.

Як дэзінфармацыя штучнага інтэлекту ўплывае на людзей

> “Кожны раз, як вы рэагуеце, калі так можна сказаць, сваімі паджылкамі, вы абмінаеце асабістую рэфлексію.”
> — Крысціна Дугуэн *

ПСІХАЛОГІЯ: Аперацыі па ўплыве часта маюць на мэце выкарыстанне кагнітыўных скажэнняў.

Асэнсаванне сябе і разуменне вашай аўдыторыі дапамагаюць супрацьстаяць дэзынфармацыі.

МЭТЫ: Чаму зламыснікі спадзяюцца на ШІ для стварэння або распаўсюджвання дэзынфармацыі? Якія мэты яны маюць у рэальным свеце?

- ➔ Пашырыць свой уплыў на іншую краіну або супольнасць?
- ➔ Праз шматлікія варыянты тоесных сцвярджэнняў перагрузіць фактчэкераў або перадухіліць выяўленне фэйкаў?
- ➔ Уплываць на думкі і перакананні аўдыторыі, усталёўваючы давер з дапамогай сеткі неаўтэнтычных акаўнтаў?
- ➔

**Крысціна Дугуэн — даследчыца ў галіне інфармацыйнага ўплыву ў Сарбоне.*

Абвяржэнне патрабуе ўсебаковага падыходу ды разумення нюансаў

І калі інструменты выяўлення не працуюць, што ж дапамагае? Важна зразумець кантэкст сцвярджэння настолькі ж глыбока, як і яго змест. Прафесійныя фактчэкеры экспертна валодаюць неабходнымі даследчымі навыкамі. І вось некаторыя парады.

> *“Інструменты выяўлення ніколі не будуць працаваць на ўсе 100%. Я і не чакаю, што некалі пачнуць.”*
> – Хэнк ван Эс**

РАЗГЛЕДЗЬЦЕ КРЫНІЦУ: Ці можаце вы пацвердзіць яе ідэнтычнасць? Пра што яна расказвае і што распаўсюджвае? Хто ўзаемадзейнічае з яе кантэнтам? Які ўплыў можа мець яе змест на чытачоў?

ВЫСВЕТЛІЦЕ НАДЗЕЙНАСЦЬ: Праверце інфармацыю незалежна з надзейных крыніц, такіх як эксперты з практычным досведам у дадзенай галіне. Ці мае распаўсюджаная інфармацыя сэнс, зыходзячы з вашага досведу?

Выкарыстоўвайце тэхнікі **МЕДЫЯКРЫМІНАЛІСТЫКІ,** каб узбагаціць традыцыйныя метады журналісцкіх і дакументальных расследаванняў. Напрыклад: скрэпінг дадзеных, геалакацыя, біяметрычнае распазнаванне, аналіз мадэляў і іншае.

ВУЧЫЦЕСЯ ДЫ АДАПТУЙЦЕСЯ: Стваральнікі згенераванай ШІ дэзінфармацыі ды фальсіфікацый увесь час адаптуюцца. Прыстасуйце свой падыход да зменлівага ландшафту.

АБАГУЛЬНЯЙЦЕ СВАЮ ПРАЦУ

Абвяржэнне няпраўдзівай інфармацыі эксперты рэкамендуюць суправаджаць празрыстым аналізам ды спасылкамі на крыніцы. Гэта можа дапамагчы чытачам прасачыць за ходам расследавання і зразумець нюансы наратыву.
У некаторых выпадках само расследаванне важнейшае за тое, створаны кантэнт ШІ ці не.

** Хэнк ван Эс — эксперт OSINT і тэхнік фактчэкінгу.

Кароткі даведнік: Што можна, што нельга і падказкі

Далейшы спіс прыкмет дапаможа вызначыць, ці быў кантэнт створаны ШІ або лічбава рэдагаваны. Разам з іншымі парадамі, ўзгаданымі ў гэтым гайдзе — кантэкст, метады даследавання і інструменты выкрыцця, — гэта дасць вам зразуменне праўды па-за тым, што падаецца на першы погляд.

Тэкст

- ➔Часта (але не заўсёды) мае **лепшую граматыку**, чым чалавек..
- ➔Часцей будуць выкарыстоўваць **занадта фармальную або структурызаваную мову**, асабліва ў сацыяльных медыя.
- ➔**Празмерная колькасць прыслоўяў або прыметнікаў.**
- ➔Адсутнасць чалавечых эмоцый, гумару, іроніі ды ідыяматычных выразаў..
- ➔Можа быць **дэфіцыт канкрэтных дэталяў** (імёнаў, дат, лакацый) або арыгінальных ідэй.
- ➔Найважнейшае: ці з'яўляюцца выказаныя ў тэксце словы праўдзівымі фактамі?

Відэа

- ➔Не ўжывайце дэтэктар выяваў, згенераваных ШІ, у дачыненні да відэа-кадраў.
- ➔Назірайце **за выразамі твару і рухамі,** напрыклад, за міргannем, або ці супадае рух вуснаў з маўленнем.
- ➔Відэа можа мець адметнасці ў выглядзе **рэзкіх пераходаў ці нечаканай абрэзкі.**

Аўдыё

- ➔**Параўнайце аўдыё,** якое выклікае падазрэнні, **з аўтэнтычным узорам,** выкарыстоўваючы інструменты, якія могуць выяўляць адрозненні ў маўленні, дыханні, інтанацыі...
- ➔Пры выкарыстанні дэтэктараў пазбягайце нізкаякасных аўдыё са статычным або фонавым шумам.
- ➔Могуць быць мець адметныя **ненатуральныя або механічныя маўленчыя патэрны**, недахоп паўз ці жывога дыхання.

Выявы

- ➔Шукайце **штучныя дэталі:** ідэальная скура, размыты фон, ненатуральная прыгажосць ці святло, а таксама дзівацтвы, напрыклад, дадатковыя пальцы.
- ➔**Пашукайце вадзяны знак** распаўсюджаных ШІ-генератараў выяваў.
- ➔Звяртайце ўвагу на дэталі таго, што паказана на выяве: ці гэта лагічна? Ці гэта дарэчна?
- ➔Пры выкарыстанні дэтэктараў **аддавайце перавагу версіі выявы з высокім дазволам якасці ці больш ранняй загрузкі, а не той, якую некалькі разоў** перазалілі.