

Kaip pastebėti ir suprasti DI sugeneruotą arba skaitmeniniu būdu pakeistą turinį

Prieinami ir lengvai naudojami DI modeliai gali padėti žmonėms mokytis ir kurti turinį, tačiau jie taip pat **sustiprina grėsmes, kurias dezinformacija kelia** atviroms visuomenėms ir demokratiniam diskursui. Mes turime saugoti savo bendras informacijos erdves, kad jų neužlietu DI sugeneruota ir skaitmeniniu būdu pakeista dezinformacija.

Dalį sprendimo gali pasiūlyti naujos technologijos, pavyzdžiui, turinio kilmės nustatymo programos. **Tačiau technologiniai sprendimai toli gražu nėra tobuli, todėl mums reikia nepriklausomai tikrinti faktus, kad visuomenė gautų patikrintas žinias.**

Štai trumpa apžvalga, kaip profesionalios ir nepriklausomos faktų tikrinimo organizacijos identifikuoja bei demaskuoja DI sugeneruotą dezinformaciją, ir ko galite iš jų pasimokyti.

DI sugeneruoto turinio vis daugėja

Šiuo metu DI sugeneruota dezinformacija sudaro nedidelę procentinę dalį visų teiginių, kuriuos tiria profesionalūs, nepriklausomi faktų tikrintojai. Jų darbe dažniau pasitaiko skaitmeniniu būdu pakeisto turinio.

Tačiau: atlikus vidinę Europos faktų tikrinimo standartų tinklo (angl. EFCSN) narių apklausą, **dauguma faktus tikrinančių organizacijų sutiko, kad ateityje DI sugeneruoto ir skaitmeniniu būdu pakeisto turinio svarba tik didės.** Šią prognozę patvirtina ir naujausi pavyzdžiai Europos Parlamento rinkimų kontekste.

GERESNIAM SUPRATIMUI: *Skaitmeniniu būdu pakeistas turinys* – tai bet kokios formos turinys, kuris buvo pakankamai pakoreguotas, kad būtų galima juo manipuluoti arba pakeisti jo pradinę žinią. Tai apima ir redagavimą DI įrankiais. Tačiau tai neapima redagavimo, siekiant aiškumo ar kokybės.

DI sugeneruotas – tai bet kokios formos turinys, kurį sukūrė dirbtinio intelekto sistema.



Technologijos sparčiai tobulėja, tačiau vien jomis pasikliauti negalime.

DI ekspertai ir profesionalūs faktų tikrintojai sutinka: **vien tik DI identifikavimo priemonių nepakanka, kad galėtume aptikti ar demaskuoti DI sugeneruotą ar skaitmeniniu būdu pakeistą turinį.**

Prieš pradėdant naudotis detektoriumi, ekspertai rekomenduoja susipažinti su DI turinio generatoriais ir detektoriais. Suprasdami, kaip modeliai yra treniruojami, ir pasitelkdami šiek tiek statistikos, faktų tikrintojai pradeda atpažinti stipriąsias bei silpnąsias įrankio puses ir jo sėkmės tikimybę. **Tačiau net ir tokiu atveju šie įrankiai gali tapti naudingu atspirties tašku.**

Turinio kilmės

iniciatyvos, kaip svetainė „C2PA Specifications“, gali padėti nustatyti žiniasklaidos turinio šaltinį bei istoriją, tačiau žymėjimas vandenženkliais ir tikrinimas nėra tobuli.

Kaip DI dezinformacija daro poveikį žmonėms?

„Kiekvieną kartą, kai reaguojate per savo instinktus, tai aplenkia jūsų refleksiją.“

Christine Dugoin*

PSICHOLOGIJA: Per įtakos operacijas dažnai siekiama pasinaudoti irracionaliais žmogaus polinkiais.

Suprasdami savo ir savo auditorijos psichologinį šališkumą, galite kovoti su dezinformacija.

TIKSLAI: Kodėl blogi veikėjai gali naudotis dirbtiniu intelektu, kad sukurtų ar paskleistų dezinformaciją? Kokį poveikį jie trokšta padaryti realiaame pasaulyje?

- Praplėsti savo įtaką į kitą šalį ar bendruomenę.
- Išvengti identifikavimo arba užversti faktų tikrintojus milžiniškais kiekiais dezinformacijos, generuojant daugybę panašių teiginių variantų.
- Daryti įtaką mintims ar įsitikinimams, priverčiant patikėti žinutėmis per neautentiškų paskyrų tinklus.

* Christine Dugoin yra informacijos įtakos tyrinėtoja Sorbonos universitete.

Norint demaskuoti teiginius, reikia daugialypės prieigos ir išsamaus supratimo.

Taigi, jei identifikavimo įrankiai neveikia, tuomet kas veikia? Svarbu suprasti ne tik teiginių turinį, bet ir jų kontekstą. Profesionalios faktų tikrinimo organizacijos turi reikiamų tyrimo įgūdžių. Štai keletas patarimų.

“Identifikavimo priemonės niekada neveiks 100 proc. – nesitikiu, jog kada nors taip bus.”

– Henkas van Essas**



ĮVERTINKITE ŠALTINIŲ: Ar galite patvirtinti jo tapatybę? Apie ką jis kalba ir kuo dalijasi? Kas skaito jo turinį? Kokį poveikį šis turinys gali turėti skaitytojams?



NUSTATYKITE PATIKIMUMĄ: Nepriklausomai patikrinkite informaciją patikimuose šaltiniuose, pavyzdžiui, pas ekspertus, turinčius praktinės patirties šioje srityje. Ar tai, kas vaizduojama, jūsų manymu, yra reikšminga?



Naudokitės **ŽINIASKLAIDOS EKSPERTIZĖS** metodais, leidžiančiais papildyti tradicinius tiriamuosius reportažus ir dokumentinius tyrimus. Tarp šių metodų yra duomenų ištraukimas, geolokacija, biometrinių duomenų atpažinimas, modelių analizė ir kt.



MOKYKITĖS IR PRISITAIKYKITE: DI generuojamos dezinformacijos kūrėjai nuolat adaptuojasi. Bandykite prisitaikyti ir jūs prie besikeičiančio peizažo.

DALYKITĖS SAVO DARBU

Kartu su paneigtais pareiškimais ekspertai rekomenduoja pateikti skaidrią analizę ir nuorodas į šaltinius. Tai gali padėti skaitytojams sekti tyrimą ir suprasti niuansuotą naratyvą. Kai kuriais atvejais tyrimas yra svarbesnis nei tai, ar turinys yra parašytas dirbtinio intelekto.

** Henkas van Essas yra atvirų šaltinių žvalgybos ir faktų tikrinimo metodų ekspertas.

EFC SN Trumpas vadovas: ką daryti, ko nedaryti ir kelios užuominos

Čia pateikiamos užuominos, kurios gali rodyti, kad turinys yra sugeneruotas dirbtinio intelekto arba pakeistas skaitmeniniu būdu. Kartu su kitais šiame vadove paminėtais patarimais (kontekstas, tyrimo technikos ir identifikavimo įrankiai) jos gali padėti suprasti tiesą, slypinčią už to, ką matote.

Tekstai

- Dažnai (bet ne visada) yra **gramatiškai taisyklingesni** nei žmogaus rašyti tekstai.
- Vartojama **pernelyg formali ar struktūrizuota kalba**, ypač socialinės žiniasklaidos kontekste.
- **Itin daug prieveiksmių ar būdvardžių.**
- Trūksta žmogiškų emocijų, humoro, sarkazmo ar idiomų.
- Gali trūkti **konkrečių detalių** (vardų, datų, vietovių) ar originalių idėjų.
- Svarbiausias dalykas: ar tekste pateikti faktai yra teisingi?
-

Vaizdo įrašai

- Nenaudokite DI sugeneruotų vaizdų detektoriaus sustabdytiems vaizdo įrašų kadrams.
- Sekite **veido išraiškas ir judesius**, pavyzdžiui, mirksėjimą, ir stebėkite, ar burnos krutėjimas atitinka garso įrašą.
- Gali pasitaikyti **staigių perėjimų ar iškarpytų kadro.**

Garso įrašai

- **Įtartiną garso įrašą palyginkite su autentišku pavyzdžiu**, pasitelkdami įrankius, kurie gali aptikti kalbos, kvėpavimo ar intonacijos skirtumus.
- Naudodami detektorius, venkite prastos kokybės garso pavyzdžių su statišku ar foniniu triukšmu.
- **Nenatūralūs ar mechaniški kalbos modeliai**, nėra pauzių ar natūralaus kvėpavimo.

Nuotraukos

- Atkreipkite dėmesį į **nenatūralias detales**: tobulą odą, išplaukusį foną, nenatūralų grožį ar šviesą, taip pat tokias keistenybes kaip papildomi pirštai.
- **Ieškokite vandens ženklų**, kuriuos palieka populiarūs vaizdų generatoriai.
- Atkreipkite dėmesį į tai, kas pavaizduota: ar vaizdai logiški? Ar jie derami?
- Naudodami detektorius, **rinkitės didelės skiriamosios gebos arba anksčiausiai įkeltą nuotraukos versiją, o ne tą, kuria jau buvo pasidalyta daugybę kartų.**