

Kā pamanīt un izvērtēt mākslīgā intelekta radītu vai digitāli pārveidotu saturu

Pieejami un ērti izmantojami mākslīgā intelekta (MI) modeļi var palīdzēt cilvēkiem mācīties un veidot saturu, bet tie arī palielina riskus, ko sabiedrībai un demokrātijai kopumā rada dezinformācija un maldinoša informācija. Svarīgi nepieļaut, ka mūsu visu izmantotā informācijas vide tiek pieblīvēta ar dezinformāciju, kas radīta ar MI vai ir digitāli pārveidota.

Daļēji to var risināt jaunās tehnoloģijas, piemēram, satura izcelsmes noteikšanas programmatūra. **Bet tehniskie risinājumi ne tuvu nav perfekti un mums ir vajadzīgs neatkarīgu faktu pārbaudītāju darbs, lai nodrošinātu sabiedrību ar pārbaudītu informāciju.**

Piedāvājam īsu ieskatu tajā, kā profesionāli un neatkarīgi faktu pārbaudītāji nosaka un pārbauda ar MI radītu maldinošu un nepatiesu informāciju. Un ko no viņiem varam mācīties.

Mākslīgā intelekta satura apjoms aug

Pašlaik ar MI radītas maldinošas un nepatiesas informācijas pārbaude ir maza daļa no faktu pārbaudītāju darba. Viņi biežāk saskaras ar digitāli pārveidotu saturu.

Taču EFCSN biedru iekšējā aptaujā lielākā daļa faktu pārbaudītāju atzinuši, ka ar MI radīts un digitāli mainīts saturs kļūs arvien nozīmīgāks. Par to liecina arī nesenie [piemēri](#) Eiropas Parlamenta vēlēšanu kontekstā.

Izziņai: *Digitāli pārveidots* attiecas uz jebkāda veida būtiski mainītu saturu, ja ir mainījies tā sākotnējais vēstījums. Tas var būt darīts arī ar MI rīkiem. Tas neattiecas uz izmaiņām, kas ziņu padara skaidrāku vai kvalitatīvāku.

Mākslīgais intelekts jeb MI attiecas uz jebkura veida saturu, kas radīts ar mākslīgā intelekta sistēmām.



Tehnoloģijas strauji attīstās, bet uz tām vien paļauties nevaram

Eksperti un faktu pārbaudītāji ir vienisprātis: **ar MI noteikšanas rīkiem vien nepietiek, lai identificētu un atspēkotu ar to radītu vai digitāli pārveidotu saturu.**

Eksperti iesaka pirms kādas programmas izmantošanas iepazīties ar MI ģenerēšanas un noteikšanas rīkiem. Ja faktu pārbaudītāji saprot, kā šie modeļi tiek "apmācīti", kā arī izmanto statistiku, viņi var atpazīt to stiprās un vājās puses.

Satura izcelsmes noteikšanas iniciatīvas, piemēram, C2PA, var palīdzēt norādīt uz mediju satura avotu, taču uz ūdenszīmēm un verifikāciju pilnībā paļauties nevar.

Kā cilvēkus ietekmē MI radīta dezinformācija?

"Instinktīva reakcija uz informāciju izslēdz jūsu kritisko domāšanu."

– Kristīne Duguāna*

Psiholoģija: Ietekmes operācijas bieži tiek veidotas tā, lai iedarbotos uz cilvēku aizspriedumiem.

Izpratne par saviem un savas auditorijas aizspriedumiem var palīdzēt vērsties pret dezinformāciju.

Mērķi: Kāpēc cilvēki izmanto mākslīgo intelektu, lai radītu un izplatītu dezinformāciju? Ko viņi cer panākt?

- Sasniegt jaunu auditoriju citā kopienā vai valstī?
- Palikt nepamanītam vai apbērt faktu pārbaudītājus ar darbu, vienu vēstījumu pasniedzot daudz dažādās versijās?
- Ietekmēt cilvēku uzskatus, izveidojot viltotu sociālo mediju profilu tīklu un tā panākot vēstījuma uzticamību?

* Kristīne Duguāna ir informācijas ietekmes pētniece Sorbonnas Universitātē.

Atmaskošanai vajadzīga daudzpusīga pieeja un niansēta izpratne

Ko darīt, ja MI noteikšanas rīki nestrādā? Svarīgi saprast vēstījuma kontekstu un saturu. Profesionāliem faktu pārbaudītājiem šādas prasmes piemīt. Piedāvājam dažus padomus, kā to darīt.

"Noteikšanas rīki nekad nestrādās par 100% – es pat negaidu, ka tas kādreiz notiks."

– Henks van Ess**



PIEVĒRS UZMANĪBU AUTORAM: Vai vari to identificēt? Kādu saturu tas publicē un ar ko dalās? Kas un kā uz to reaģē? Kādu iespaidu šis saturs var atstāt uz lasītājiem?



NOSAKI TICAMĪBU: Pārbaudi informāciju. Vai tas, ko lasi vai redzi, šķiet ticami?



Izmanto **KRIMINĀLISTIKĀ LIETOTĀS METODEDES**, piemēram, automatizētu informācijas iegūšanu, ģeolokāciju jeb vietas noteikšanu, biometrisku atpazīšanu, modeļu analīzi.



MĀCIES UN PIELĀGOJIES: Ar mākslīgo intelektu radīta nepatiesa un maldinoša informācija pastāvīgi mainās. Pielāgo tai arī savu pieeju.

DALIES AR SAVU DARBU

Eksperti iesaka strādāt atklāti – atspēkojot kādu informāciju, pievienot saites uz avotiem un parādīt, kā nonācāt līdz konkrētajiem secinājumiem. Tas lasītājiem palīdzēs saprast jūsu veikto pārbaudi un uztvert visas vēstījuma nianšes.

** Henks van Ess is OSINT un faktu pārbaudes metožu eksperts.

Kādām pazīmēm pievērst uzmanību?

Ir vairākas pazīmes, kas var liecināt par digitāli izmainītu vai MI radītu saturu. Pievērs tām uzmanību un, izmantojot arī pārējos šīs rokasgrāmatas ieteikumus (par kontekstu, izmeklēšanas metodēm, MI noteikšanas rīkiem), varēsi labāk saprast, kas patiesībā slēpjas aiz tā, ko redzi.

Teksts

- Pārāk formāla un strukturēta valoda – īpaši sociālajos medijos.
- **Pārmērīgs apstākļa un īpašības vārdu lietojums.**
- Nav cilvēcisku emociju, humora, sarkasma un konkrētajai valodai raksturīgu izteicienu.
- Var trūkt **būtisku detaļu** (vārda, datuma, vietas), un neparādās oriģinālas idejas.
- Svarīgākais: vai tekstā minētie fakti ir korekti?

Video

- MI noteicējā neizmanto stopkadru jeb ekrānuzņēmumu no apturēta video.
- Pievērs uzmanību sejas izteiksmēm un kustībām, piemēram, acu mirkšķināšanai, un tam, vai lūpas kustas atbilstoši runātajam.
- Video var būt pēkšņas pārejas, “norauti” kadri.

Audio

- **Aizdomīgo audio salīdzini ar īstu skaņas paraugu.** Izmanto rīkus, kas var noteikt atšķirības cilvēka runā, elpošanā, intonācijā.
- Pārbaudei neizmanto sliktas kvalitātes skaņas paraugus ar statisku vai fona troksni.
- Runa var būt **nedabiska vai mehāniska**, bez pauzēm un dzirdamas elpošanas.

Attēli

- Par MI izmantošanu var liecināt perfekta āda, miglains fons, nedabisks skaistums, savāda gaisma un dažādas citas dīvainības, piemēram, pārāk daudz pirksti cilvēka rokā.
- **Skaties, vai uz attēla nav ūdenszīmju.**
- Pievērs uzmanību detaļām: vai tās ir loģiskas?
- Pārbaudot bildi ar MI noteikšanas rīkiem, izmanto tādu, kurai ir labāka izšķirtspēja. Centies atrast pēc iespējas senāku tā versiju.