

Näin ymmärrät tekoälyllä tuotettuja ja digitaalisesti muokattuja sisältöjä

Helppokäyttöiset tekoälymallit voivat auttaa ihmisiä saamaan tietoa ja tuottamaan sisältöjä, mutta ne voivat myös **lisätä riskejä, joita dis- ja misinformaatio** aiheuttavat avoimille yhteiskunnille ja demokraattiselle keskustelulle. On tärkeää estää yhteisten informaatioympäristöjen täyttyminen tekoälyllä tuotetusta ja digitaalisesti muokatusta dis- ja misinformaatiosta.

Uudet teknologiat, kuten sisällön alkuperään keskittyvät hankkeet ja tunnistustyökalut, ovat osa ratkaisua. **Teknologiset ratkaisut eivät kuitenkaan ole läheskään täydellisiä, ja tarvitaankin itsenäisiä faktantarkistajia, jotta yhteiskunta saa käyttöönsä yhteisiä tarkistettuja faktoja.**

Tässä on nopea katsaus siihen, miten itsenäiset ammattifaktantarkistajat tunnistavat ja tutkivat tekoälyllä tuotettua mis- ja disinformaatiota ja mitä heidän työstään voi oppia.

Tekoälyllä tuotetut sisällöt lisääntyvät

Tekoälyllä tuotettu mis- ja disinformaatio muodostaa nykyään pienen osan kaikista väitteistä, joita itsenäiset ammattifaktantarkistajat tutkivat. Digitaalisesti muokatut sisällöt ovat yleisempiä faktantarkistajien työssä.

EFCSN:n jäsenille tehdyssä sisäisessä kyselyssä **suurin osa faktantarkistajista oli kuitenkin sitä mieltä, että tekoälyllä tuotetut ja digitaalisesti muokatut sisällöt muuttuvat entistä tärkeimmiksi tulevaisuudessa.** Ennustetta tukevat viimeaikaiset, EU-vaaleihin liittyvät [esimerkit](#).

HUOMIOI: *Digitaalisesti muokatulla* tarkoitetaan kaikkia sisältöjä, joita on muokattu (mukaan lukien tekoälytyökaluilla) merkittävästi alkuperäisen viestin manipuloimiseksi tai muuttamiseksi. Tähän eivät kuulu selkeyden tai laadun vuoksi tehdyt muokkaukset.

Tekoälyllä tuotetulla tarkoitetaan kaikkia sisältöjä, jotka on tuotettu tekoälyjärjestelmällä.



Teknologia muuttuu nopeasti, mutta siihen ei voi luottaa ainoana keinona.

Tekoälyasiantuntijat ja ammattifaktantarkistajat ovat samaa mieltä siitä, että **tekoälyn tunnistustyökalut eivät yksinään riitä tekoälyllä tuotetun tai digitaalisesti muokatun sisällön tunnistamiseen tai tarkistamiseen.**

Asiantuntijat suosittelevat tutustumaan tekoälygeneraattoreihin ja tunnistustyökaluihin ennen jälkimmäisten käyttämistä. Mallien oppimisprosessien ja tilastojen ymmärtäminen auttaa faktantarkistajia tunnistamaan työkalujen vahvuuksia ja heikkouksia sekä ymmärtämään, kuinka hyvin ne toimivat. **Työkalut voivat kuitenkin toimia hyvänä aloituspisteenä.**

Sisällön alkuperään keskittyvät hankkeet, kuten C2PA Specifications, voivat auttaa lähteen ja mediasisällön historian tarkistamisessa, mutta vesileimat ja varmennus eivät ole virheettömiä.

Kuinka tekoälyllä tuotettu disinformaatio vaikuttaa ihmiseen

“Joka kerta, kun reagoi vaistomaisesti, ohittaa harkinnan.”

– Christine Dugoin*

PSYKOLOGIA:

Informaatiovaikuttamisessa hyödynnetään usein ajatusvinoumia.

Omien ja oman yleisön vinoumien ymmärtäminen voi auttaa disinformaation torjumisessa.

TAVOITTEET: Minkä vuoksi pahansuovat toimijat käyttävät tekoälyä disinformaation tuottamiseen tai levittämiseen? Mitä he tavoittelevat?

- Vaikutusvallan laajentamista toiseen maahan tai yhteisöön?
- Kiinnijäämisen välttämistä tai faktantarkistajien hukuttamista useisiin samanlaisiin väitteisiin?
- Ajatuksiin tai uskomuksiin vaikuttamista rakentamalla uskottavuutta valetiliverkostojen kautta?

* Christine Dugoin tutkii informaatiovaikuttamista La Sorbonnessa.

Tarkistaminen vaatii monipuolista lähestymistapaa ja yksityiskohtien ymmärrystä

Jos tunnistustyökalut eivät toimi, mikä toimii? On yhtä tärkeää ymmärtää väitteen asiayhteys kuin sen sisältö. Ammattifaktantarkistajat ovat tarvittavien tutkimustaitojen asiantuntijoita. Alla on muutamia vinkkejä.

“Tunnistustyökalut eivät koskaan toimi täydellisesti – enkä usko, että ne tulevat koskaan toimimaan.”
– Henk van Ess**



TARKASTELE LÄHDETTÄ: Voitko vahvistaa lähteen henkilöllisyyden? Mistä hän puhuu ja mitä hän jakaa? Ketkä ovat tekemisissä hänen sisältöjensä kanssa? Mitä vaikutuksia sisällöillä voi olla lukijoihin?



ARVIOI USKOTTAVUUTTA: Vahvista informaation paikkansapitävyys riippumattomasti käyttämällä luotettavia lähteitä kuten asiantuntijoita, joilla on konkreettista kokemusta alasta. Onko kuvatussa asiassa järkeä omien tietojesi perusteella?



Käytä **TEKNISIÄ TUTKIMUSKEINOJA** perinteisen tutkivan journalismin ja asiakirjojen tarkastelun lisäksi. Esimerkkejä ovat tiedonharavointi, paikannusmenetelmät, biometrinen tunnistus ja datan analysointi.



OPI & MUKAUDU: Tekoälyllä tuotetun mis- ja disinformaation tekijät mukautuvat jatkuvasti. Mukauta lähestymistapasi vastaamaan muuttuvia vaatimuksia.

TUTKIMUKSEN JAKAMINEN

Asiantuntijat suosittelevat esittämään analyysin läpinäkyvästi ja linkittämään lähteisiin tarkistetun väitteen ohessa. Tämä auttaa lukijoita seuraamaan tutkimusta ja ymmärtämään sen eri yksityiskohtia. Joissakin tapauksissa itse tutkimus on tärkeämpi kuin se, onko sisältö tuotettu tekoälyllä.

** Henk van Ess on avointen tietolähteiden tiedustelun ja faktantarkistuksen asiantuntija.

Alla olevat vihjeet voivat viitata siihen, että sisältö on tuotettu tekoälyllä tai että sitä on muokattu digitaalisesti. Muiden tässä oppaassa mainittujen vinkkien (konteksti, tutkimuskeinot ja tunnistustyökalut) kanssa ne voivat auttaa ymmärtämään, mikä sisällössä on totta.

Teksti

- Usein (muttei aina) **kieliopillisesti virheettömämpää** kuin ihmisen kirjoittama teksti
- Sisältää todennäköisesti **liian virallista tai jäsenneltyä kieltä** erityisesti sosiaalisen median kontekstissa
- **Liikaa adverbejä tai adjektiiveja**
- Ei tunteita, huumoria, sarkasmia tai idiomaattisia ilmauksia
- **Ei välttämättä yksityiskohtia** (nimiä, päivämääriä, paikkoja) tai omaperäisiä ideoita
- Tärkeinta: ovatko tekstin faktat kunnossa?

Video

- Älä käytä videon pysäytyskuviin työkalua, joka on tarkoitettu tekoälyllä tuotettujen kuvien tunnistamiseen
- Tarkastele **kasvonilmeitä ja -liikkeitä**, kuten silmien räpyttelyä, ja sopivatko suun liikkeet ääneen
- Voi sisältää **äkillisiä siirtymiä ja leikkauksia**

Ääni

- **Vertaa epäilyttävää ääntä aitoon** käyttämällä työkaluja, jotka tunnistavat eroja puhe- ja hengitystavoissa, intonaatioissa jne.
- Käyttäessäsi tunnistustyökaluja vältä heikkolaatuisia äänitiedostoja, joissa on kohinaa tai taustamelua
- **Puhe voi olla epäluonnollista tai mekaanista**, tauot ja luonnollinen hengitys saattavat puuttua.

Kuvat

- Etsi **epäluonnollisia yksityiskohtia**: täydellistä ihoa, sumuisia taustoja, epäluonnollista kauneutta tai valoa sekä kummallisuuksia, kuten ylimääräisiä sormia
- **Etsi kuvageneraattoreiden vesileimoja**
- Kiinnitä huomiota yksityiskohtiin: ovatko ne loogisia? Sopivatko ne aiheeseen?
- Käyttäessäsi tunnistustyökaluja **valitse hyvälaatuinen versio kuvasta mieluummin kuin sellainen, jota on jaettu useaan otteeseen**