

Sådan spotter du AI-genereret og digitalt modificeret indhold

Frit tilgængelige og letanvendelige AI-systemer kan hjælpe folk med at lære nye ting og hjælpe med at skabe indhold. Men de kan også **forstærke de risici, som mis- og desinformation** udgør for vores samfund og demokrati. Det er vigtigt at forhindre, at vores informationsrum bliver fyldt med AI-genereret og andet digitalt modificeret mis- og desinformation.

En del af løsningen er nye teknologier, såsom initiativer, der kan spore et stykke indholds oprindelse, og detektionssoftware, der kan udregne sandsynligheden for, at indhold er AI-genereret. **Men teknologiske løsninger er langt fra perfekte, og vi har brug for uafhængige faktatjekkeres arbejde for at give læsere adgang til verificeret fakta.**

Her er et kort overblik over, hvad professionelle, uafhængige faktatjekkere gør for at identificere og tjekke AI-genereret mis- og desinformation, og hvad du kan lære af det.

AI-genereret indhold er på fremmarch

I dag udgør AI-genereret mis- og desinformation en lille procentdel af alle påstande, der undersøges af professionelle, uafhængige faktatjekkere. Digitalt modificeret indhold er mere udbredt i faktatjekkerens arbejde.

Men: I en intern undersøgelse blandt EFCSN-medlemmer var **de fleste faktatjekkere enige om, at AI-genereret og digitalt modificeret indhold kun bliver mere relevant** i fremtiden. Og nyere eksempler i forbindelse med det europæiske valg understøtter denne prognose.

FORKLARING: *Digitalt modificeret* betyder enhver form for indhold, der er blevet ændret for at manipulere eller ændre det budskab, der oprindeligt blev formidlet. Dette inkluderer også AI-genereret indhold. Dog dækker det ikke over redigering for at øge kvaliteten eller tydeligheden af indhold.

AI-genereret betyder enhver form for indhold, der er blevet skabt af kunstig intelligens.



Teknologien er i konstant udvikling, men vi må ikke blive afhængige af kun den.

AI-eksperter og professionelle faktatjekkere er enige: **AI-detektionsværktøjer er ikke nok til at opdage eller faktatjekke AI-genereret eller digitalt modificeret indhold.**

Før du vælger at bruge et detektionsværktøj, anbefaler eksperter, at du får kendskab til AI-systemer, der kan generere og genkende AI-indhold. Med et kendskab til AI og en smule statistisk, kan faktatjekkere få forståelse for de styrker og svagheder, et AI-værktøj har. **Detektionsværktøjer kan dog stadig være et godt udgangspunkt.**

Indholds ophav

Initiativer såsom C2PA Specifications kan hjælpe med at verificere kilden og historien bag det indhold, du ser. At bruge et vandmærke til verificering er dog ikke uden for enhver tvivl.

Sådan påvirker AI-desinformation læseren

"Hver gang du reagerer med din mavefornemmelse, så tilsidesætter du din refleksionsevne."

- Christine Dugoin*

PSYKOLOGIEN BAG:

Påvirkningsoperationer er ofte lavet på en sådan måde, at de udnytter psykologiske *bias*.

Hvis du forstår dit publikums og dine egne *bias*, kan du bedre bekæmpe desinformation.

FORMÅL: Hvorfor bruger aktører AI til at lave eller sprede desinformation? Hvad forventer de at få ud af det?

- Udvide deres indflydelse til andre lande eller grupperinger?
- Undgå detektion eller overvælde faktatjekkere ved at lave mange variationer af de samme påstande?
- Indfluere holdninger ved at etablere troværdighed via et netværk af falske brugere?

* Christine Dugoin er researcher i informations påvirkning ved La Sorbonne.

Det kræver en nuanceret tilgang og forståelse at debunke

Hvis detektionsværktøjer ikke virker, hvad gør så? Det er lige så vigtigt at forstå konteksten af en påstand som dets indhold. Faktatjekkere er eksperter i de nødvendige teknikker, der bruges i undersøgelser. Her kommer nogle tips:

*“Detektionsværktøjer vil aldrig fungere 100% - det tror jeg ikke, at de nogensinde vil.”
– Henk van Ess***



HVEM ER KILDEN?: Kan du bekræfte deres identitet? Hvad skriver de om? Hvad deler de? Hvem interagerer med deres budskaber? Hvilken effekt kan deres budskaber have på læsere?



TROVÆRDIGE KILDER: Verificér informationen vha. troværdige kilder såsom eksperter med stor erfaring på området. Baseret på din viden giver påstanden så mening?



Brug **FORENSIC MEDIA** teknikker til at supplere traditionel efterforskning og research. Det er teknikker såsom: at høste data, geolokalisering, biometrisk genkendelse, mønsteranalyse m.v.



LÆR OG TILPAS: Aktører der skaber AI-genereret mis- og desinformation, tilpasser sig den teknologiske udvikling. Derfor er det vigtigt, at du også tilpasser din tilgang i takt med den teknologiske udvikling.

DEL DIT ARBEJDE

I takt med at du udgiver dit arbejde, anbefaler eksperter, at du også vedhæfter en transparent analyse og links til dine kilder. Dette kan hjælpe læsere med at følge med i en undersøgelse og forstå en nuanceret historie. I nogle tilfælde er indholdet af en undersøgelse vigtigere end at detektere AI-genereret indhold.

** Henk van Ess er ekspert i OSINT og faktatjek-teknikker.

Råd til mistænkeligt indhold: Dos, Don'ts og andre tips

Her kommer nogle tegn på, at det indhold, du ser, er AI-genereret eller digitalt modificeret. Kombineret med tidligere tips i denne guide (kontekst, undersøgelsesteknikker og detektionsværktøjer) kan disse tips hjælpe dig med at få en bedre forståelse af sandheden bag det, du ser.

Tekst

- Har ofte (men ikke altid) bedre grammatik end et menneske
- Bruger ofte et overdrevent formelt sprog, som ikke passer ind på sociale medier
- **Overdrevne adverbier eller adjektiver**
- Mangel på følelser, humor, sarkasme og hverdags vendinger
- **Kan mangle detaljer** (navne, datoer, steder) eller originale ideer
- Og vigtigst af alt: Er den angivne fakta korrekt?

Video

- Brug ikke detektionsværktøjer på stillbilleder fra videoer
- Kig efter **ansigtsmimik**, såsom at blinke, og om bevægelsen af munden matcher lyden i videoen
- Videoen kan indeholde **hurtige overgange eller klip**

Lyd

- **Sammenlign mistænkelige lydspor med et autentisk lydspor** vha. værktøjer, der kan detektere forskelle i tale- og åndedrætsmønstre, intonation m.v.
- Undgå lydspor af lav kvalitet ved brug af detektionsværktøjer
- Lydsporet kan indeholde **unaturlige eller mekaniske talemønstre** og mangle pauser eller åndedræt

Images

- Kig efter **unaturlige detaljer**: perfekt hud, sløret baggrund, unaturligt lys og andre bemærkelsesværdige ting såsom for mange fingre
- **Kig efter et vandmærke** fra AI-udbydere
- Læg mærke til, hvad billedet viser: Er det logisk? Giver det mening?
- Ved brug af detektionsværktøjer, **så vælg et billede med høj opløsning eller en tidlig version af billedet frem for et, der er blevet delt flere gange.**