

Så upptäcker och förstår du AI-genererat och digitalt manipulerat innehåll

Tillgängliga AI-modeller som är lätta att använda kan hjälpa folk att hitta information och skapa innehåll men de kan också **öka de risker som des- och misinformation utgör** för öppna samhällen och demokratiska diskurser. Det är viktigt att se till att våra gemensamma informationsmiljöer inte fylls av AI-genererad och digitalt manipulerad des- och misinformation.

En del av lösningen är ny teknologi såsom initiativ för att identifiera innehållets ursprung samt spårningsprogram. **Men teknologiska lösningar är långt ifrån perfekta och vi behöver oberoende faktagranskare för att ge samhället gemensamma verifierade uppgifter.**

Här är en sammanfattning om vad professionella oberoende faktagranskare gör för att identifiera och granska AI-genererad mis- och desinformation och vad du kan lära dig från dem.

Mängden AI-genererat innehåll ökar

Idag utgör AI-genererad mis- och desinformation en liten andel av påståenden som professionella oberoende faktagranskare undersöker. Digitalt manipulerat innehåll förekommer oftare i faktagranskarnas arbete.

Men en intern enkät till EFCSN-medlemmar visar att **de flesta faktagranskarna är överens om att AI-genererat och digitalt manipulerat innehåll kommer att bli allt viktigare** i framtiden. [Aktuella exempel](#) i samband med EU-valet stöder prognosen.

OBSERVERA: *Digitalt manipulerat* syftar på allt innehåll som modifierats betydligt för att manipulera eller förändra de ursprungliga budskapen, inklusive ändringar som gjorts med AI-verktyg. Detta omfattar inte förändringar som gjorts för klarhet eller kvalitet.

AI-genererat syftar på allt innehåll som skapats med ett AI-system.



Teknologin utvecklas snabbt men vi kan inte enbart förlita oss på den.

AI-experter och professionella faktagranskare är överens om att **enbart AI-detekteringsverktyg inte är tillräckligt för att upptäcka eller granska AI-genererat eller digitalt manipulerat innehåll.**

Experter rekommenderar att man bekantar sig med AI- och detekteringsverktyg innan man använder detektorer. Genom att kombinera kunskap om hur modeller tränas med statistik kan faktagranskare uppskatta verktygens styrkor och svagheter samt beräkna hur sannolik deras framgång är. **Verktyg kan trots allt vara en bra utgångspunkt.**

Initiativ för att identifiera innehållets ursprung
exempelvis C2PA Specifications kan hjälpa att verifiera källan och historien för medieinnehållet men vattenmärken och verifiering är inte felfria.

Hur AI-desinformation påverkar människor

“Varje gång man reagerar, tillåt mig att säga, med magkänslan, förbigår man övervägandet.”

– Christine Dugoin*

PSYKOLOGI: Påverkansoperationer är ofta skapade så att de utnyttjar psykologiska fördomar.

Att förstå dina egna och din publiks fördomar kan hjälpa med att bekämpa desinformation.

MÅL: Varför använder en illasinnad aktör AI för att skapa eller sprida desinformation? Vilken är den avsedda effekten i den verkliga världen?

- Utöka sin räckvidd i ett annat land eller gemenskap?
- Slippa detektion eller överbelasta faktagranskare genom att skapa flera varianter av liknande påståenden?
- Påverka tankar eller övertygelser genom att bygga upp trovärdighet med nätverk av falska konton?

* Christine Dugoin forskar informationspåverkan vid La Sorbonne.

Granskning kräver en mångfacetterad strategi och nyanserad förståelse

Om detekteringsverktyg inte fungerar, vad kan man då göra? Det är lika viktigt att förstå sammanhanget av påståendet som dess innehåll. Professionella faktagranskare är experter på olika metoder som kan användas för att granska påståenden. Här är några tips.

“Detekteringsverktyg kommer aldrig att fungera till 100% – jag förväntar mig inte att de någonsin kommer att göra det.”
– Henk van Ess**



TÄNK PÅ KÄLLAN: Kan du verifiera källans identitet? Vad talar de om och vad delar de? Vem läser deras innehåll? Vilken påverkan kan innehållet ha på läsarna?



UPPSKATTA TROVÄRDIGHETEN: Verifiera informationen på ett oberoende sätt genom trovärdiga källor såsom experter som har konkreta erfarenheter inom området. Är det som beskrivs vettigt baserat på vad du vet?



Använd **FORENSISKA MEDIATEKNIKER** för att komplettera traditionell granskande journalistik och dokumentär forskning. Dessa inkluderar bland annat: dataskrapning, geolokalisering, biometrisk identifiering, mönsteranalys.



LÄR & ANPASSA DIG: Skaparna av AI-genererad mis- och desinformation utvecklas hela tiden. Anpassa dina strategier till förändringarna.

DELA DITT ARBETE

Experter rekommenderar att man presenterar en transparent analys och länkar till källorna tillsammans med det granskade påståendet. Detta hjälper läsarna att följa granskningen och förstå en nyanserad historia. I vissa fall är granskningen viktigare än frågan om innehållet är AI-genererat.

** Henk van Ess är en expert på OSINT- och faktagranskningstekniker.

EFCSN Tips för att granska AI-genererat och digitalt förändrat innehåll

Nedan finns det tips som kan antyda att innehåll är AI-genererat eller digitalt förändrat. Tillsammans med de övriga tipsen i denna guide (kontext, granskningstekniker och detektionsverktyg) kan det hjälpa dig att förstå sanningen bakom det du ser.

Text

- Har ofta (men inte alltid) **bättre grammatik** än en människoskapad text.
- Använder sannolikt **för formellt eller strukturerat språk**, särskilt i kontexten av sociala medier.
- **För många adverb eller adjektiv.**
- Brist på mänskliga känslor, humor, sarkasm och idiomatiska uttryck.
- **Specifika detaljer** (namn, datum, platser) eller originella idéer kan saknas.
- Det viktigaste: är uppgifter i texten korrekta?

Video

- Använd inte ett detektionsverktyg för AI-genererade bilder på stillbilder från en video.
- Titta på **ansiktsuttryck och rörelser**, såsom blinkandet, och om munrörelser passar audion.
- Kan ha **skarpa övergångar eller snitt.**

Audio

- **Jämför audion med ett autentiskt exempel** med verktyg som kan upptäcka skillnader i tals- och andningsmönster, intonation...
- Undvik exempel av låg kvalitet med statistiskt eller bakgrundsljud när du använder detekteringsverktyg.
- Kan ha **onaturliga eller mekaniska talmönster**, en brist på pauser eller naturlig andning.

Bilder

- Leta efter områden med **onaturliga detaljer**: perfekt hud, suddiga bakgrunder, onaturlig skönhet eller ljus och konstigheter såsom ytterligare fingrar.
- **Leta efter vattenmärken** från vanliga bildgeneratorer.
- Observera detaljer i vad beskrivs: är det logiskt? Är det lämpligt?
- **Välj en högupplöst eller tidigt uppladdad version av bilden hellre en version som delats flera gånger** för detekteringsverktyg.