



Foto: DALL-E (via ChatGPT 4o), CC0 1.0

AI & hallucinationer – om misinformation, validitet og reliabilitet

Generativ AI er en avanceret form for kunstig intelligens, der er trænet og udviklet på store mængder af data fra internettet, og som ved hjælp af machinelearning kan skabe noget nyt. Men da den gætter sig til næste ord i en sætning – ord for ord for ord – kan der være tale om mange fejl, eller hallucinationer. Det rejser vigtige metodiske spørgsmål: Hvor valid og pålidelig er den viden, AI producerer? Kan man bruge AI som kilde i samfundsfagligt arbejde? Og hvis ja, hvordan sikrer man sig så, at informationen er troværdig og anvendelig?

Mens kunstig intelligens har eksisteret i årtier, kom generativ kunstig intelligens (generativ AI – også kaldet sprogmodeller, LLMs) til i 2018 og blev udbredt og verdenskendt med ChatGPT, en tjeneste fra det amerikanske firma OpenAI i 2022.

Generativ AI kan vha. et prompt (en kommando) udføre research- og skriveopgaver, opsummere og strukturere tekster og føre samtaler med mennesker både i tekst og lyd. Det er tjenester som ChatGPT, CoPilot, Mistral, Claude, Gemini, DeepSeek og Perplexity. De er alle designet på en måde, så man får et svar, som var det fra et menneske.

Generativ AI er en sandsynlighedsmaskine. Den gætter på det næste ord i en sætning – ord for ord for ord – ud fra, hvad flertallet af mennesker gør og skriver. Er der 75 % chance for, at det næste ord er 'hende' og 76 % chance for, at det er 'ham', så bliver det et ham. Teknologien har fået sine evner ved at se på, hvad mennesker har skrevet igennem mange, mange år. Men da den gætter på det næste ord, er den ofte fyldt med fejl – den hallucinerer, som det hedder. Flere undersøgelser peger på, at der i cirka 20 procent af alle prompt er faktuelle fejl i svarene.

Reflektér individuelt og tag stilling

Vurder kritisk graden af validitet og reliabilitet af viden genereret med AI

- Kan man bruge AI til at lave empiriske undersøgelser – hvorfor/hvorfor ikke?
- Skal man faktatjekke alt, hvad der kommer ud af generativ AI?
- Vil du gerne vide, når noget er menneskeskabt og AI-genereret? Hvorfor?
- Hvad har mest værdi for dig: menneskeskabt eller AI-skabt?

Diskutér i grupper

- Er det nemmere at faktatjekke, hvis man har en vis viden om emnet?
- Hvordan faktatjekker man?
- Skal man markere tekst (og fotos), som er genereret med AI, så mennesker kan se forskel på, hvad der er ægte og hvad der er kunstigt?
- Hvordan kan man undersøge, om AI gengiver politisk eller ideologisk bias?