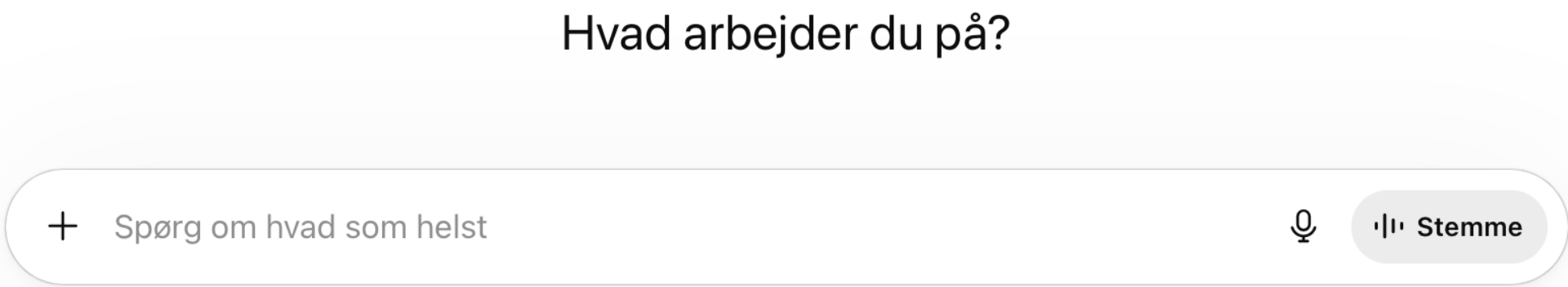
Hvordan kan man anvende AI på en etisk forsvarlig måde?

BRUG AF AI I JORD- OG GRUNDVANDSARBEJDE, 20. MAJ 2026

ANNA JESPERSEN, KØBENHAVNS UNIVERSITET

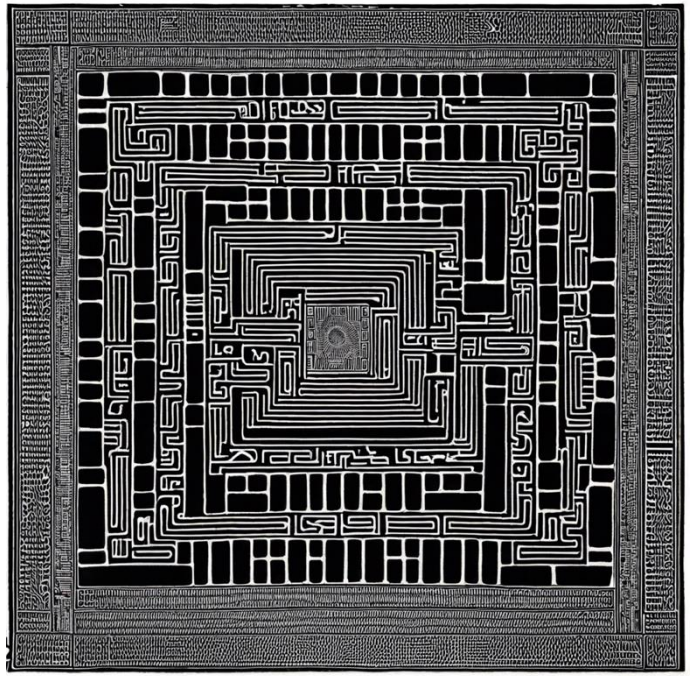
Hvorfor dette oplæg?

- LLMer er relativt nye, ugennemsigtige, og uprøvede på flere parametre
- AI-modeller kan levere en række produkter, som mennesker også kan levere (bliver vi udfaset?)
- LLMer kommunikeres med på brugerens naturlige sprog og kan levere naturligt sprog tilbage



Brugerflade, ChatGPT

Hallucinationer



Figur 1: Output fra DALL-E på prompts'ne "Create an image of a text" + "try again" (2024)

Hvor meget hallucinerer AI?

- OpenAI sagde i august 2025 at de har opnået “significant advances in reducing hallucinations” (OpenAI 2025)
- Men LLMers opbygning gør, at hallucinationer er svære at komme af med
- “[E]ven the best-performing AI tools still generate false information at a non-zero baseline rate, regardless of how they are used” (Shao 2026)
- Performance er ujævn og varierer med kontekst og opgave

Hvorfor er det vigtigt for jer?

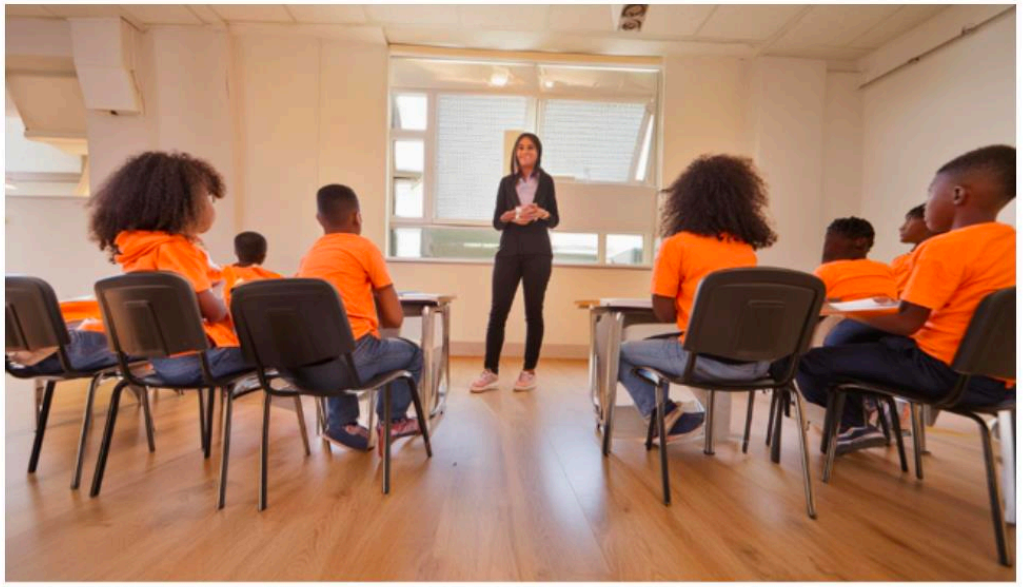
- Ukritisk brug kan lede til upræcise resultater og fejlbehæftede produkter → det er vigtigt at validere AI output (så godt man kan)

<u>Usynlighed og synlighed</u>

- Vi bruger det til informationssøgning (f.eks. 46% af amerikanere; IPSOS 2025) men vi ved ofte ikke, at det er AI, vi bruger (Maese 2025)
- Øget viden om hallucinationer i offentligheden → Misinformation og informationsusikkerhed → Aftagere kan være usikre på AI-genereret materiale

Bias: York et al. 2024

Figur 2: Forsøg på at generere en sort kvindelig underviser og sorte elever (DALL-E)



Figur 3: Forsøg på at generere en sort kvindelig underviser og sorte elever (FireFly)

Bias i tekst – et eksempel

Peters & Chin-Yee (2025): 10 AI-modellers opsummeringer af 4900 videnskabelige artikler

- De fleste LLMer overdrev mere end 50% af artiklernes resultater (26%–73%)
- AI overgeneraliserede 4.85 gange så ofte som mennesker
- Nyere modeller (f.eks. GPT-4.5) blev udkonkurreret af ældre modeller (f.eks. GPT-4o)

Glazing – kunden har altid ret

- AI-modeller har sykofantiske tendenser
  - De roser brugerens intelligens, originalitet, osv.
  - De giver brugeren ret
  - Kritik nedtones
- I et studie af feedback på fejlbehæftede matematiske argumenter gav GPT-5 sykofantisk feedback 29% af tiden (Pertov et al. 2025).

→ Feedback på sprog og indhold kan ikke altid tages for gode varer

Hvordan påvirker AI *dig*?

Ejerskab og datasikkerhed

Brug af andres produkter: AI-firmaer scraper data fra internettet

Ejerskab over egne data? AI-modeller høster brugerdata, der kan opbevares af AI-firmaer og i værste fald hackes og tilgås af mange

- Eksempel: "private" samtaler med ChatGPT bruges som bevismateriale i retten i USA
- Eksempel: Steffens problemer med copyright

Altså: Man skal tænke over, hvilke data, man vil dele med sin AI-model (og dermed AI-firmaerne)

Ejerskab i bredere forstand

Du har prompted ChatGPT til at skrive en rapport på baggrund af dine data.

Mit spørgsmål er nu: **Hvem har skrevet rapporten?**

Den er ikke nem!

LLMer destabiliserer vores forståelse af en teksts forfatter (Coeckelberg & Gunkel 2025)

Hvem har skrevet din AI-tekst?

Er det tekst uden et menneske bag?

Er det tekst med mange mennesker bag?

Hvor meget skal du involveres før du er forfatteren?

Ejerskab: hvorfor er det vigtigt?

Who's in charge? Ved du, hvordan dine resultater er fremkommet?

Du hæfter for eget arbejde!

Kan din aftager have tillid til dig?

Kan AI-modeller *ændre* dig?

Studier peger på, at brugen af AI over længere tid kan føre til:

- *Automation bias* (at man stoler overdrevet på AI) og *automation complacency* (at man ukritisk overlader arbejdet til AI; Cummings 2004; Spatola 2024)
- Besvær med at huske og forstå informationer (Gerlich 2025)
- Besvær med at tage beslutninger på egen hånd (Pearson et al. 2026)
- Ændringer i kognitive evner som f.eks. kritisk og analytisk tænkning (Zhai et al. 2024)
- Permanente ændringer i hjernens opbygning? (Kosmyna et al., under review)

MEN · Effekten afhænger af brug, og · generativ AI er ny teknologi, og vi har ikke det komplette billede

Sidst men ikke mindst...

The big one: miljøpåvirkning

Antropomorfisme – hvordan tænker du på din AI?

Humanistens holdning

- Tjek efter, om der er hallucinationer eller tegn på bias. Det er dig, der hæfter.
- Tænk over, hvad du deler med AI-modellen og hvad eventuelle AI-agenter har adgang til (f.eks. adgang til din mail eller computer)
- Vær gennemsigtig – ved tvivl, angiv brugen af AI for aftager.

Og mere generelt:

- Brug det når det er nødvendigt eller nyttigt, ikke når du kan løse opgaven selv.
- Faglighed og *trustworthiness* kommer til at have større værdi i fremtiden

Kilder (1/2)

- Coeckelberg, M. & D. J. Gunkel (2025) *Communicative AI: A Critical Introduction to Large Language Models*. Cambridge: Polity Press.
- Cummings, M. L. (2004) Automation bias in intelligent time critical decision support systems. In: *Collection of technical papers – AIAA 1st intelligent systems technical conference*, 2, 557–562; https://doi.org/10.2514/6.2004-6313
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006
- IPSOS. (2025). *Did you know? As more people engage with AI tools, concerns persist despite technology's recognized role in enabling progress*. Ipsos. https://www.ipsos.com/sites/default/files/ct/news/documents/2025-01/ipsos-essentials-infographic-january-2025.pdf
- Maese, E. (2025, January 15). *Americans use AI in everyday products without realizing it*. Gallup. https://news.gallup.com/poll/654905/americans-everyday-products-without-realizing.aspx
- OpenAI. (2025, August 7). *Introducing GPT-5 for developers*. https://openai.com/index/introducing-gpt-5-for-developers/

Kilder (2/2)

- Pearson, J. E. D. Itiel, E. Jayes, G.-R. Whordley, G. Mason & S. Nightingale. (2026). Examining human reliance on artificial intelligence in decision making. In: Nature: Scientific Reports, 16 (5346). 1–12. https://doi.org/10.1038/s41598-026-34983-y
- Peters, U. & B. Chin-Yee (2025) Generalisation bias in large language model summarization of scientific research. In: R. Soc. Open Sci, 12(4): 241776. https://doi.org/10.1098/rsos.241776. .
- Shao, A. (2025). New sources of inaccuracy? A conceptual framework for studying AI hallucinations. *Harvard Kennedy School (HKS) Misinformation Review*. https://doi.org/10.37016/mr-2020-182
- Spatola, N. (2024) The efficiency-accountability tradeoff in AI integration: Effects on human performance and over-reliance. In: *Computers in Human Behavior: Artificial Humans*, 2, 2. 100099. https://doi.org/10.1016/j.chbah.2024.100099.
- York, E. J., E. Brumberger & L. V. A. Harris (2024). Prompting Bias: assessing representation and accuracy in AI-generated images. In: The 42nd ACM International Conference on Design of Communication (SIGDOC '24), October 20–22, 2024, Fairfax, US. https://doi.org/10.1145/3641237.3691658.
- Zhai, C., Wibowo, S. & Li, L.D. (2024) The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn. Environ.* 11: 28. https://doi.org/10.1186/s40561-024-00316-7

Bias – ekstraslide

Xu et al 2026: Sammenligning af jobansøgninger

- Test af de syv mest brugte AI-modeller
  (GPT-4o, GPT4o-mini, GPT-4o-turbo, LAMA 3.3-70B, Qwen 2.5-72B, DeepSeek-V3, Mistral-7B).
  - Alle AI-modeller vurderede AI-skrevne ansøgninger til at være bedre (f.eks. GPT-4o: 97.6%. LLaMA-3.3-70B: 96.3%, DeepSeek-V3 hit 95.5%)
  - Vurderinger fra mennesker viste ikke denne skævvridning
  - AI-modellerne kunne genkende egne output og gav konsekvent sig selv en højere vurdering
- Altså: AI-modeller udviser en stærk grad af bias mod visse typer af sprog