INVEI retreat på Sommarøya
KI-Etikk



Program

Intro

Litt historisk kontekst

Etisk utfordring 1

Etisk utfordring 2

Hva betyr dette for oss?



Program

Intro

Litt historisk kontekst

Etisk utfordring 1

Etisk utfordring 2

Hva betyr dette for oss?



KI-Eksponering
Hvor møter vi KI i hverdagen?

Nyttig å vite: KI er overalt og har vært her en stund....

Smarte assistenter:
- Siri
- Google assistant
- Alexa

Anbefalingssystemer:
- Netflix
- Spotify
- You-tube

Sosiale medier:
- Facebook
- Instagram
- TikTok

Bilde- ansikts- Lydgjenkjenning og redigering:
- Google Photos

Språkoversettelse Læring og korrektur:
- Google Translate
- Grammarly
- Duolingo

Office programmer:
- Auto-tekst generering
- Keenious

Helse:
- Diagnostisering
- Prediktiv analyse
- Personlig tilpasset medisin
- Robotkirurgi etc.

Trafikk:
- Navigasjonssystemer
- Selv-kjørende Biler

Skole:
- Personlig tilpasset læring
- Intelligent tutoring systemer (Khan Academy)
- Administrative oppgaver
- Simuleringer og spillbasert læring etc.

Daglige oppgaver:
- Web search/ søkemotorer
- Credit kort transaksjoner
- E-handel

Våpenindustri

Og mye mer....

Program

Intro

Litt historisk kontekst

Etisk utfordring 1

Etisk utfordring 2

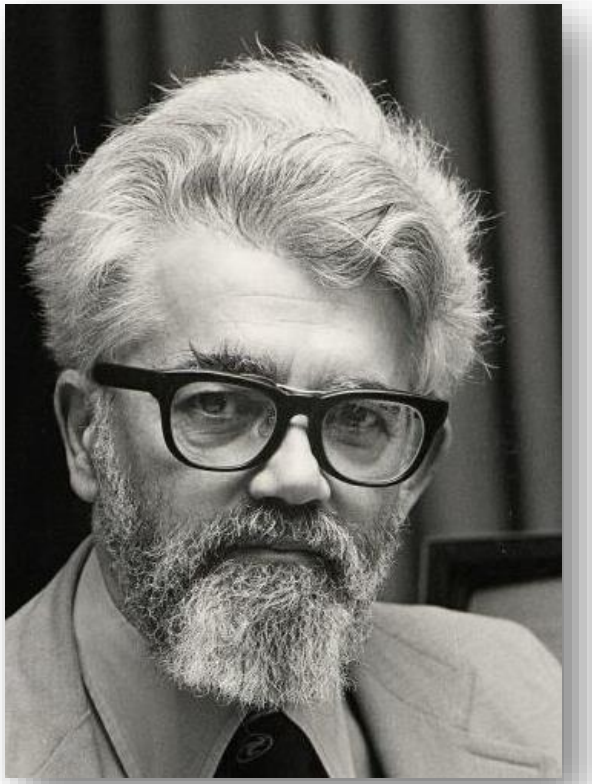
Hva betyr dette for oss?



De store fire i KI felt: Dartmouth Summer Research Project on Artificial Intelligence (1956)

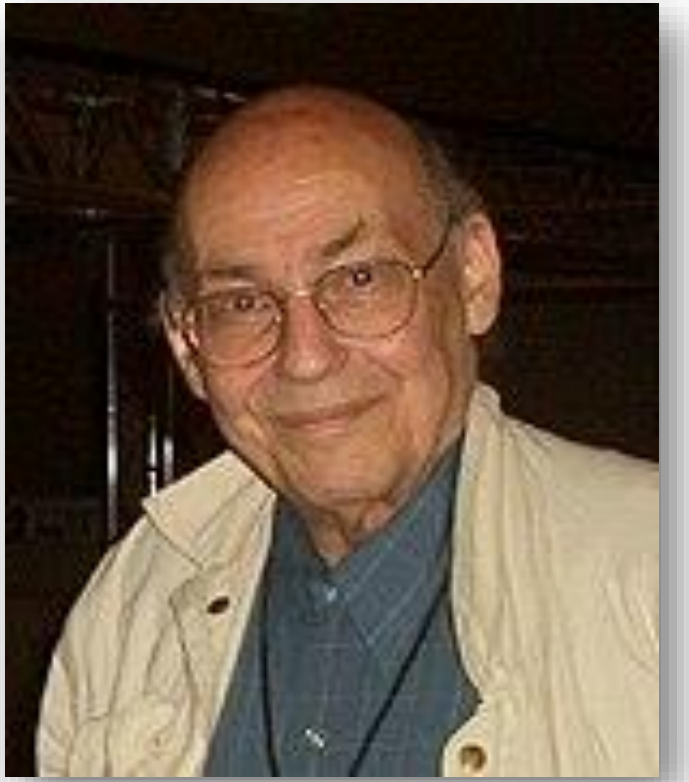
John McCarthy

(1927-2011)



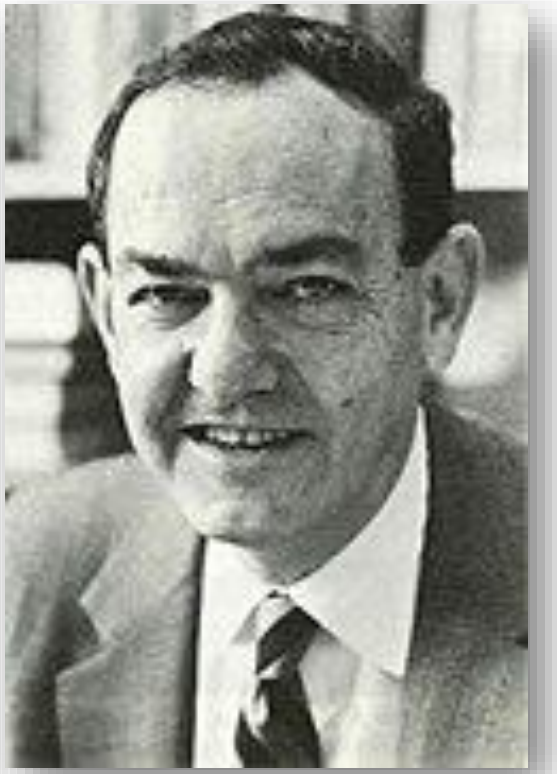
Marvin Minsky

(1927-2016)



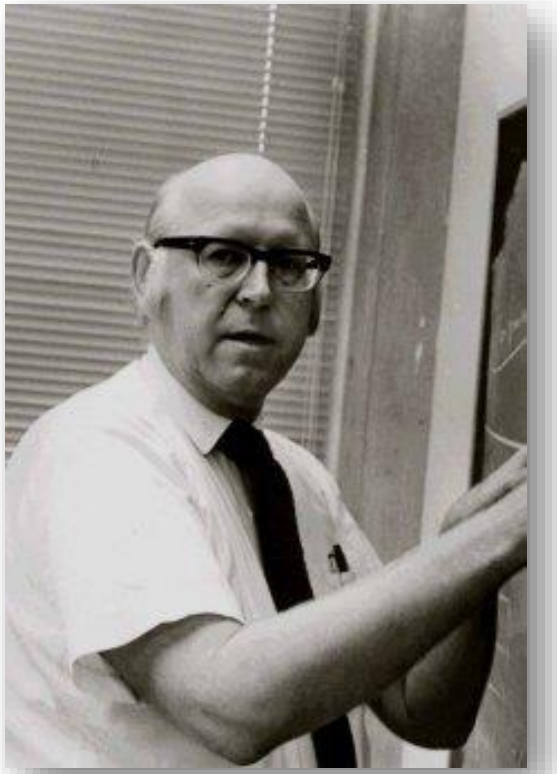
Herbert Simon

(1916-2001)



Allen Newell

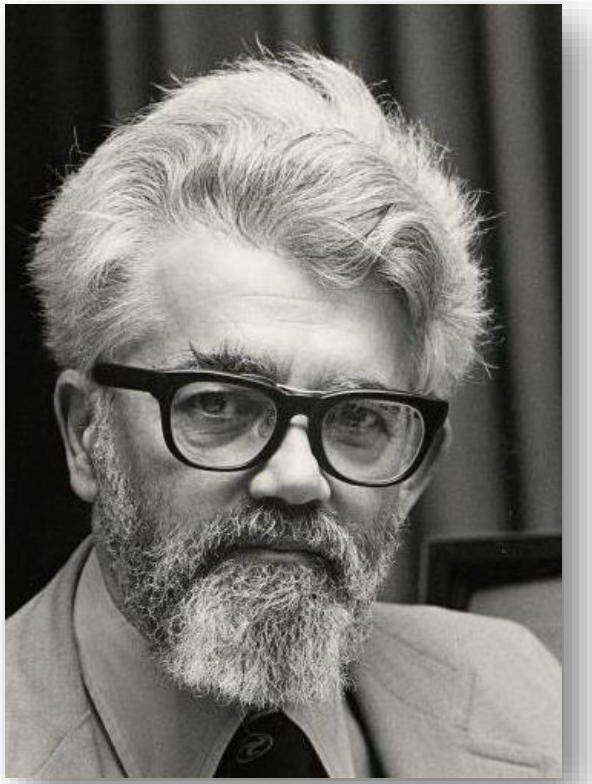
(1927-1992)



Hvorfor har det tatt så lang tid før alle kunne bruke chatbotter?

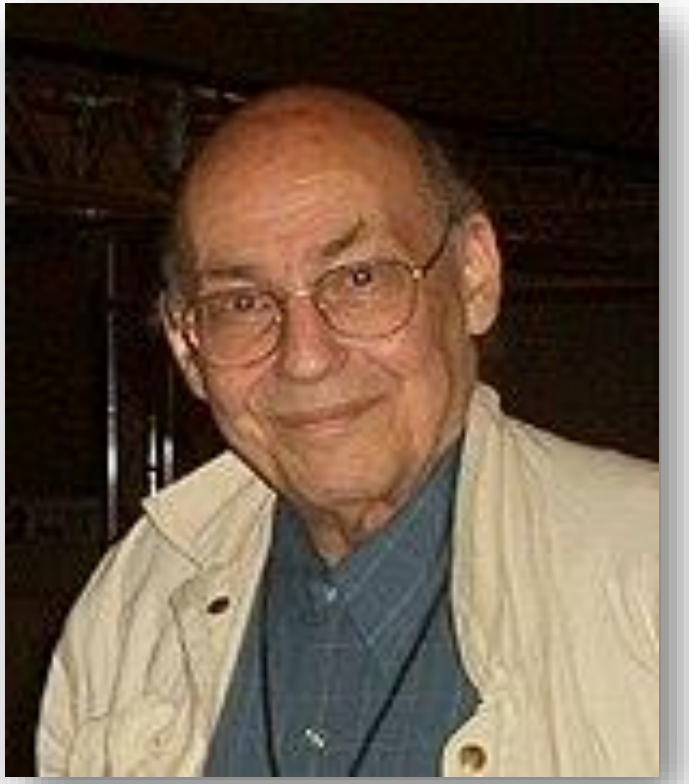
John McCarthy

(1927-2011)



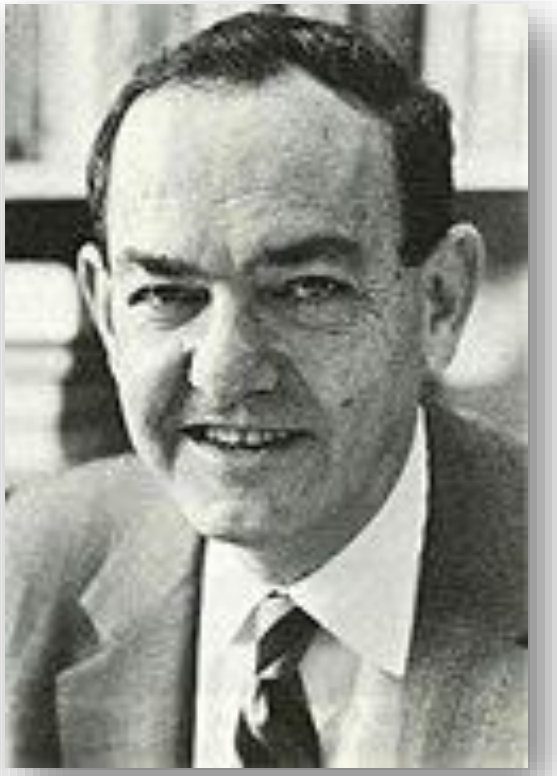
Marvin Minsky

(1927-2016)



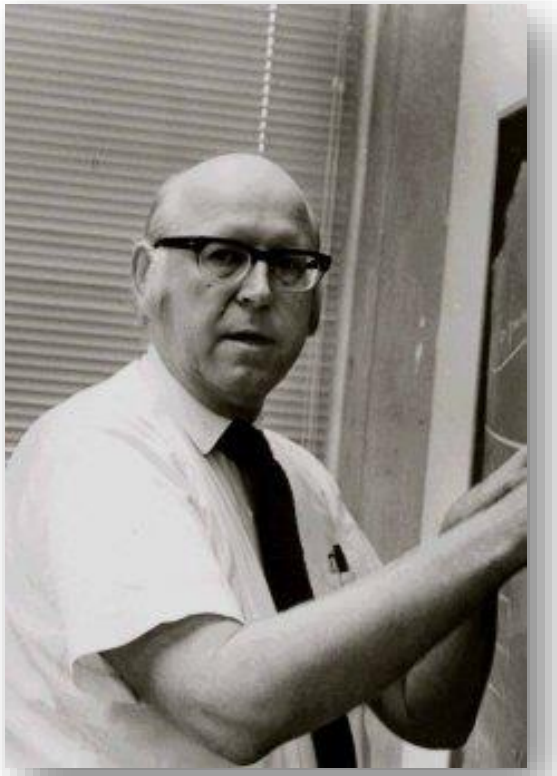
Herbert Simon

(1916-2001)



Allen Newell

(1927-1992)



Good old fashioned (GOFAI)kunstig intelligens

Tilnærming	Symbolsk, regelbasert (*if–then*)
Kunnskap	Eksplisitt kodet av mennesker
Resonnement	Logiske slutninger, deduksjon
Eksempler	ELIZA, MYCIN, Deep Blue
Styrker	Gode på veldefinerte, regelstyrte oppgaver
Svakheter	Dårlig på usikkerhet, læring og komplekse omgivelser
Design	Top-down: mennesker programmerer reglene



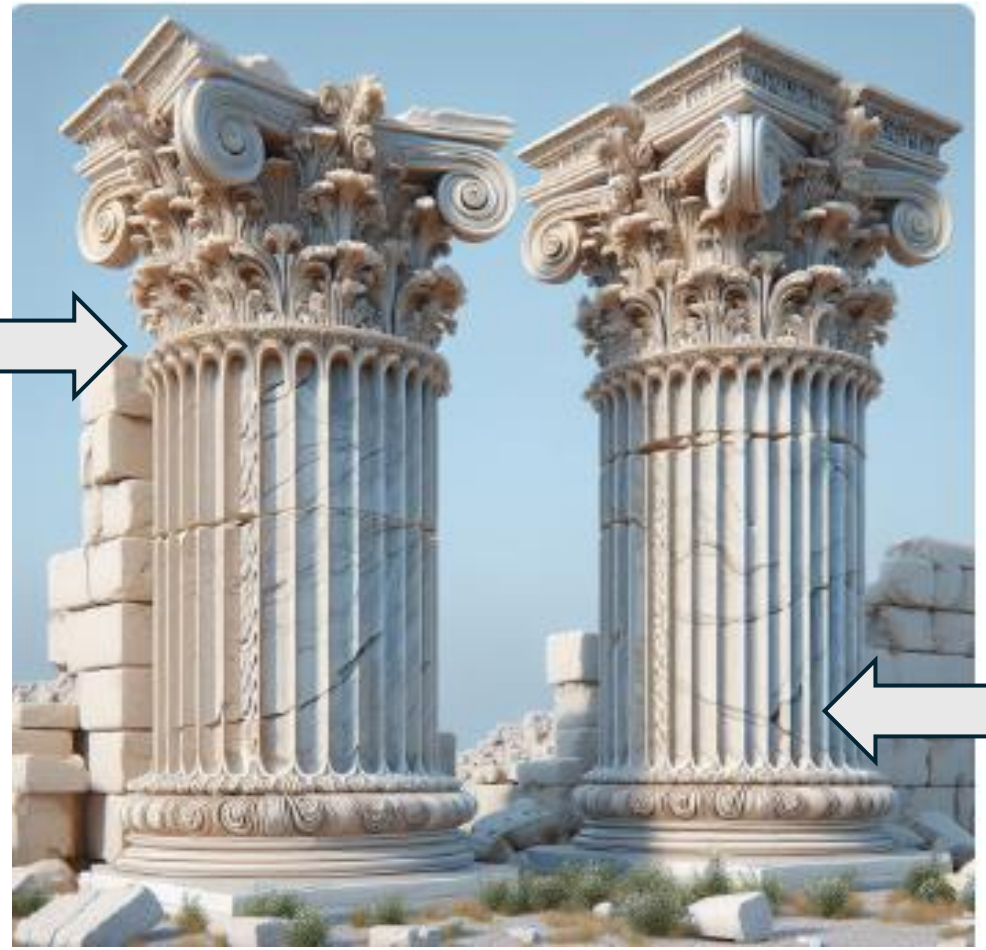
Vanlig kunstig intelligens/ generative kunstig intelligens

	Datadrevet, statistisk læring fra eksempler
	Implisitt «lært» av store datamengder
	Mønstergjenkjenning, sannsynlighetsberegning
	ChatGPT, AlphaGo, bildegjenkjenning
	Gode på komplekse, ustrukturerte da (språk, bilder)
	Vanskelig å tolke («black box»), avhengig av store datasett
Bottom-up: systemet lærer mønstre fra data	



Hva er forutsetninger til ChatGPT og lignende praterobot?

Ressurser

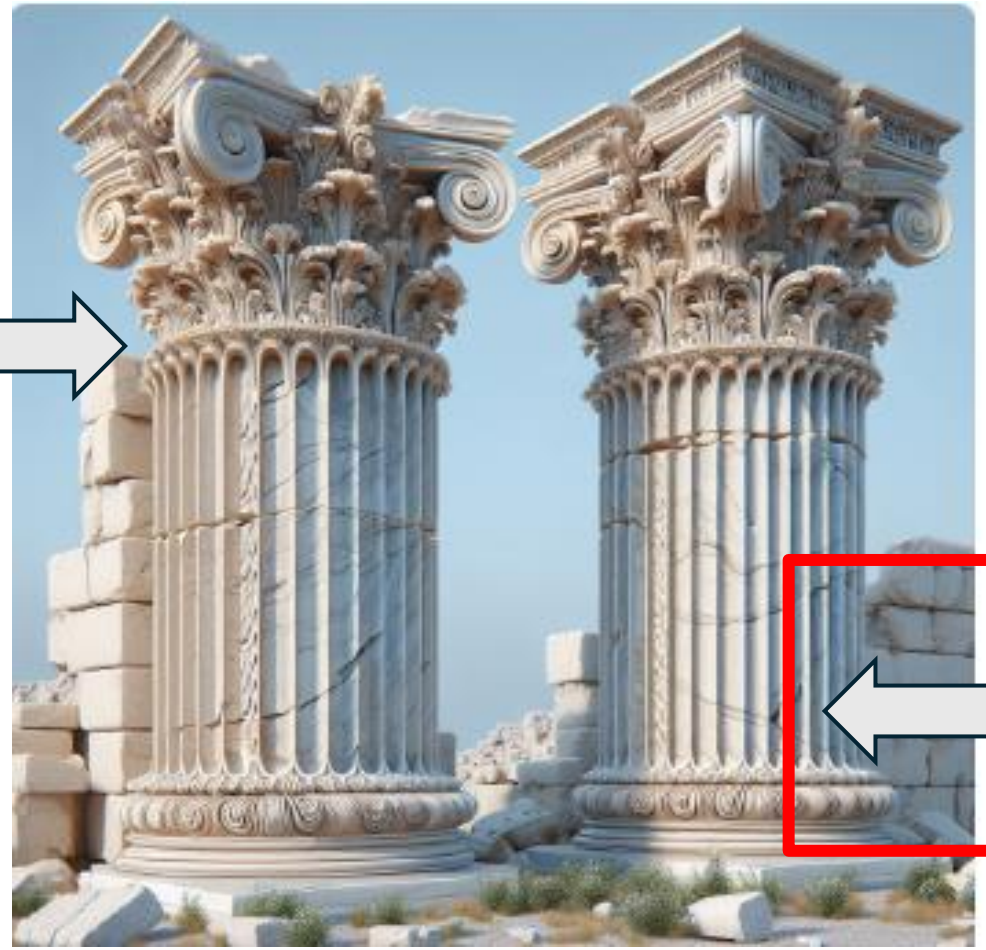


Ressursbruk og data kommer med etiske utfordringer!!

Data

Hva er forutsetninger til ChatGPT og lignende praterobot?

Ressursbruk



Ressursbruk og data kommer med etiske utfordringer!!

Data

Program

Intro

Litt historisk kontekst

Etisk utfordring 1: Data

Etisk utfordring 2: Ressurser

Hva betyr dette for oss?



En liten øvelse- Bildegenerering:

Be ChatGPT, ChatUiT, Microsoft Copilot, GoogleGemini, eller Poe.com om å generere et bilde av en norsk nordisk-student



ChatGPT



ChatUIT

Google Gemini





Poe.com



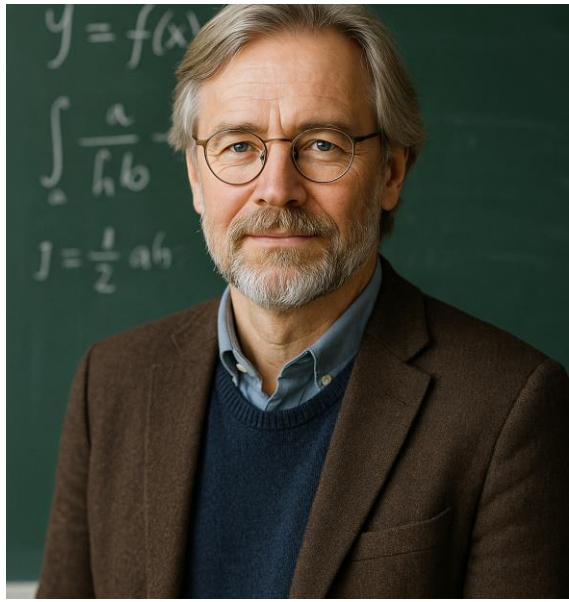
Microsoft Copilot

Hva legger dere merke til?

En liten øvelse- Bildegenering:

Be ChatGPT, ChatUiT, Microsoft Copilot, GoogleGemini, eller Poe.com om å generere et bilde av en norsk musikprofessor

ChatGPT





Poe.com



ChatUIT



Google Gemini

Hva legger dere merke til?



Microsoft Copilot

En liten øvelse - Tekstgenerering:
Be ChatGPT, ChatUiT, Microsoft Copilot, GoogleGemini, eller Poe.com om skrive en kort fortelling om en mellom leder og en sykepleier som går på date i Tromsø, men isteden for å regne beginner det å snø

Skriv en kort historie om en mellomleder og en sykepleier som går på date i Tromsø, men isteden for å regne begynner det å snø.

En uventet snødate i Tromsø

Kristian, en mellomleder med altfor mange møter i kalenderen, hadde endelig fått fri en kveld. Han hadde truffet Lise, en sykepleier, på en kafé for noen uker siden, og de hadde snakket om alt fra fiskesuppeoppskrifter til nordlysets magi. Nå skulle de møtes igjen, denne gangen for en spasertur langs brygga i Tromsø.

Lise hadde nettopp avsluttet en lang vakt på sykehuset, men hun smilte da hun så Kristian vente på henne utenfor en liten kafé. Han hadde en varm lue på og et skjerf som så ut til å være strikket av en bestemor med god tid.

"Klart for en tur?" spurte Kristian og rakte henne en kopp varm kakao.

"Absolutt!" svarte Lise. "Men jeg trodde det skulle regne?"

Hva legger dere merke til?

En liten øvelse - Tekstgenerering:
Be ChatGPT, ChatUiT, Microsoft Copilot, GoogleGemini, eller Poe.com om skrive en kort fortelling om en mellom leder og en sykepleier som går på date i Tromsø, men isteden for å regne beginner det å snø

Skriv en kort historie om en mellomleder og en sykepleier som går på date i Tromsø, men isteden for å regne begynner det å snø.

En uventet snødate i Tromsø

Kristian, en mellomleder med altfor mange møter i kalenderen, hadde endelig fått fri en kveld. Han hadde truffet Lise, en sykepleier, på en kafé for noen uker siden, og de hadde snakket om alt fra fiskesuppeoppskrifter til nordlysets magi. Nå skulle de møtes igjen, denne gangen for en spasertur langs brygga i Tromsø.

Lise hadde nettopp avsluttet en lang vakt på sykehuset, men hun smilte da hun så Kristian vente på henne utenfor en liten kafé. Han hadde en varm lue på og et skjerf som så ut til å være strikket av en bestemor med god tid.

"Klart for en tur?" spurte Kristian og rakte henne en kopp varm kakao.

"Absolutt!" svarte Lise. "Men jeg trodde det skulle regne?"

En liten øvelse - Tekstgenerering:
Be ChatGPT, ChatUiT, Microsoft Copilot, GoogleGemini, eller Poe.com om skrive en kort fortelling om en mellom leder og en sykepleier som går på date i Tromsø, men isteden for å regne beginner det å snø

Regnværet hadde vært et trofast, om enn litt kjedelig, bakteppe for de fleste av Eriks dager i Tromsø. Som mellomleder i et IT-firma var han vant til å planlegge og forutse, men været her nord var ofte en joker. I kveld hadde han en date med Mari, en sykepleier han hadde møtt via en felles venn. Planen var en rolig middag på Mathallen, etterfulgt av en spasertur langs havna – paraplyer var selvfølgelig med.

Han ventet utenfor Mathallen, og den sedvanlige dryppende lyden fra takrenner og gater var det første som møtte ham. Så møtte Mari ham. Hun hadde et varmt smil og øyne som glitret, og Erik glemte raskt det grå været. Middagen var lett og hyggelig, praten fløt uanstrengt, og de fant tonen over fersk sjømat og et glass vin.

Google Gemini

En liten øvelse - Tekstgenerering:
Be ChatGPT, ChatUiT, Microsoft Copilot, GoogleGemini, eller Poe.com om skrive en kort fortelling om en mellom leder og en sykepleier som går på date I Tromsø, men isteden for å regne beginner det å snø

En uventet date i Tromsø

Marius, en mellomleder i et IT-firma, hadde aldri trodd han skulle ende opp på en date med noen som Anna. Hun var sykepleier på UNN, full av varme og energi, og helt ulik de menneskene han vanligvis møtte i jobbmøter og konferanser. De hadde matchet på en datingapp, og etter noen meldinger frem og tilbake, foreslo Anna at de skulle møtes i Tromsø sentrum for en kaffe.

Det var en typisk høstkveld i Tromsø. Regnet hadde hengt i luften hele dagen, og Marius hadde tatt på seg sin vanntette jakke, klar for det sedvanlige nordnorske været. Anna møtte ham utenfor en liten kafé med et smil som kunne lyse opp selv den mørkeste polarnatten. Hun hadde på seg en tykk ullgenser og et skjerf som luktet svakt av lavendel.

ChatUiT

En liten øvelse - Tekstgenerering:
Be ChatGPT, ChatUiT, Microsoft Copilot, GoogleGemini, eller Poe.com om skrive en kort fortelling om en mellom leder og en sykepleier som går på date i Tromsø, men isteden for å regne beginner det å snø

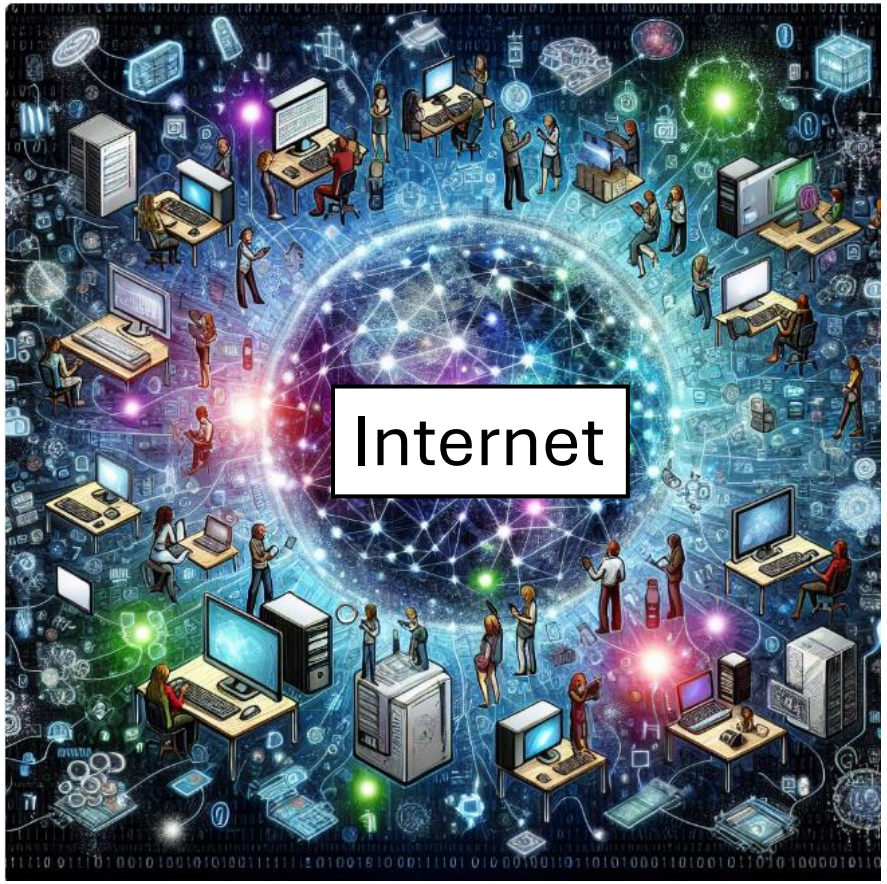
Snø i august

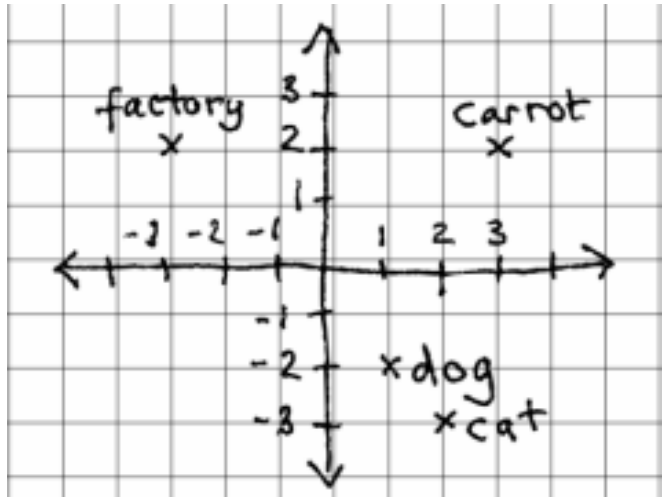
Maja, en mellomleder i helseetaten, hadde endelig sagt ja til en kaffe med Emil, en sykepleier hun hadde møtt på et seminar om pasientsikkerhet. De hadde avtalt å møtes på Smørtorget, en liten kafé med varme farger og utsikt mot fjorden.

Da de gikk ut etter kaffen, lo de begge av hvor grå himmelen var. "Typisk Tromsø," sa Emil, og trakk hetta over hodet. "Snart regner det."

Microsoft Copilot

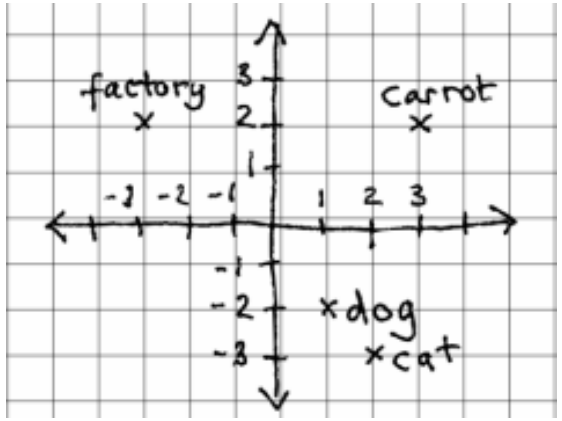
Hvordan kommer skjevheter inn i Chatboten?



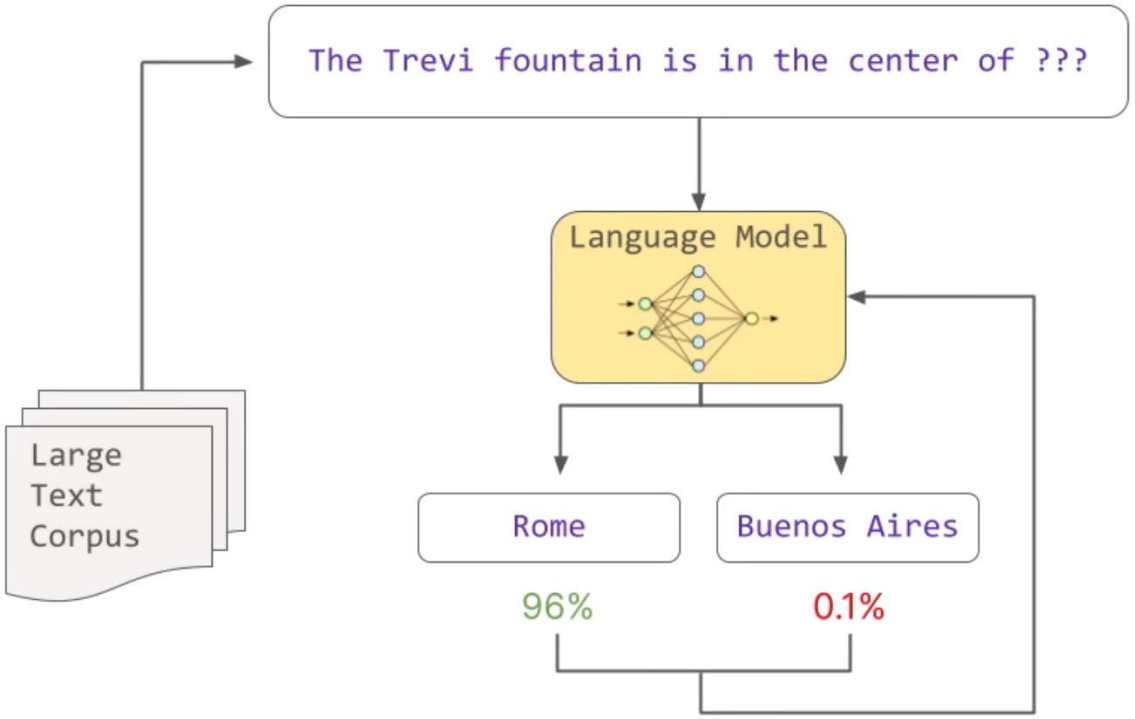


https://jilltxt.net/right-now-chatgpt-is-multilingual-but-monocultural-but-its-learning-your-values/

Sånn fungerer chatboten:



https://jilltxt.net/right-now-chatgpt-is-multilingual-but-monocultural-but-its-learning-your-values/

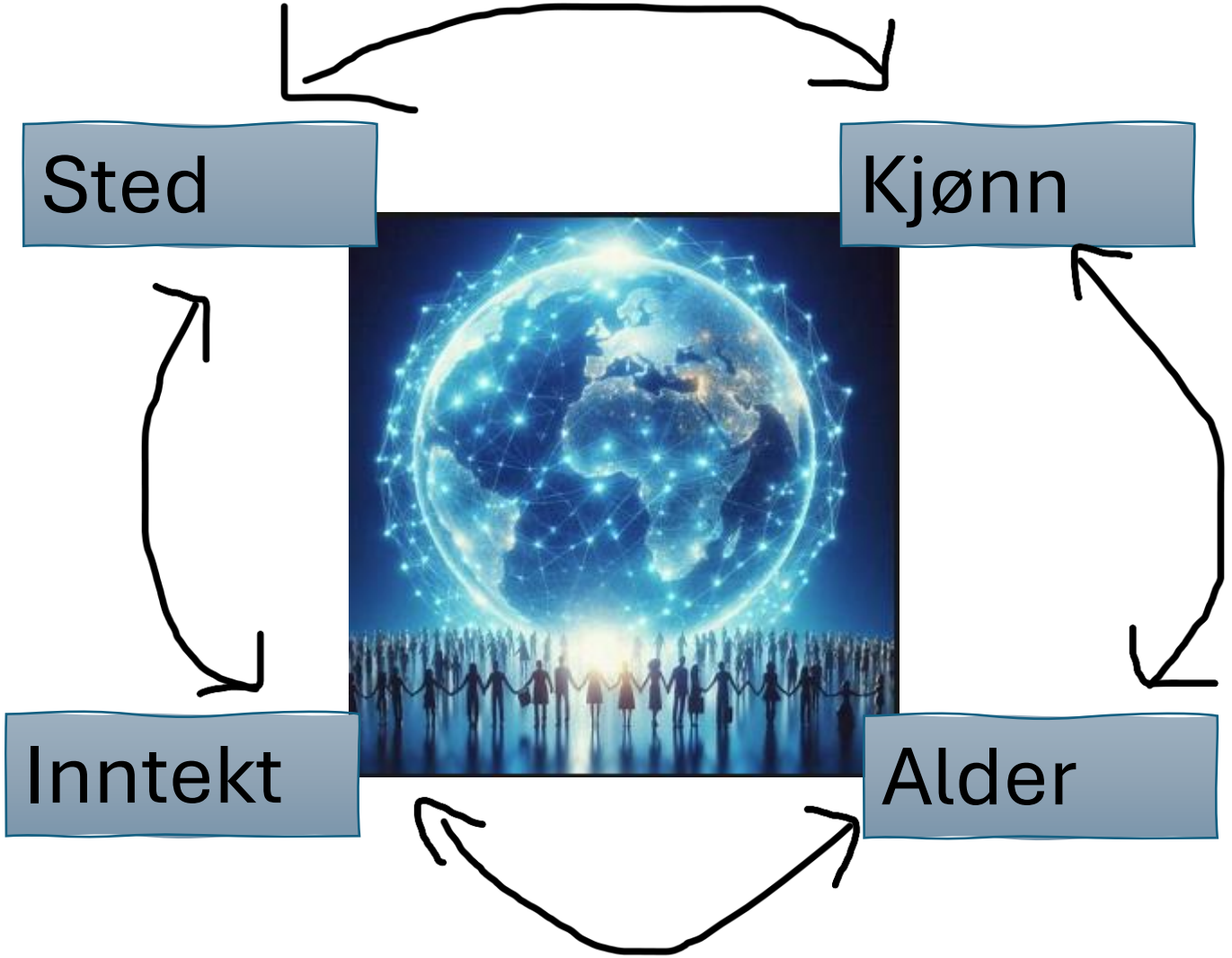


Hvordan kommer skjevheter i chatboten?

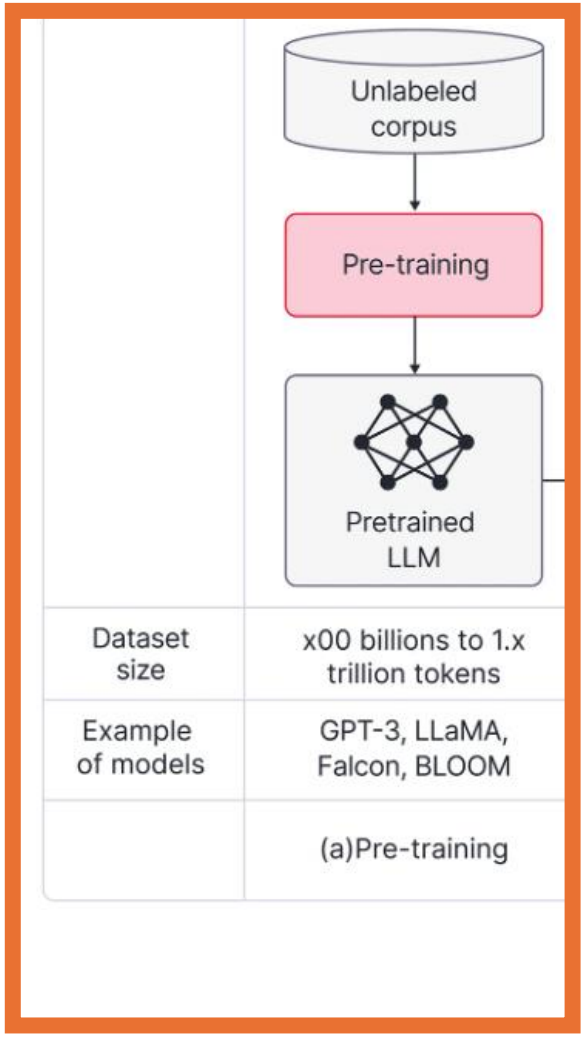
Chatboter har blitt karakterisert som *stochastic parrots* – de mangler både forståelse og bevissthet om verden (Bender et al. 2021)



Ikke alle bidrar på samme måte til internet...



...og ikke hele nettet er representert! Datautvalg



Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. "Language Models Are Few-Shot Learners." arXiv. http://arxiv.org/abs/2005.14165.

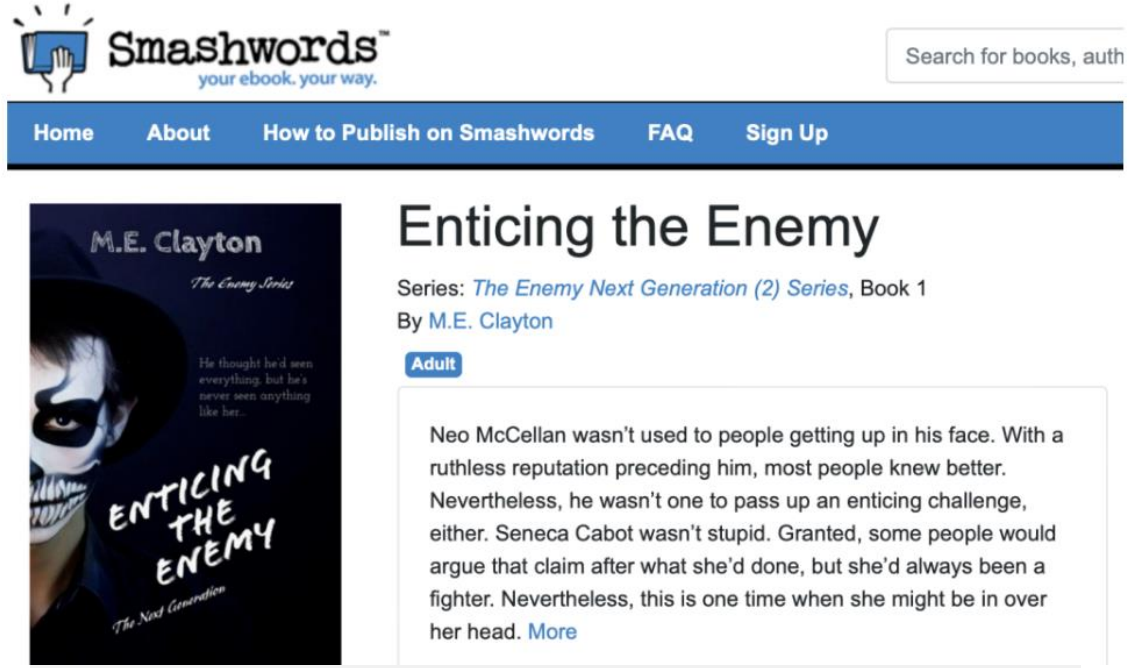
Hva er problematisk med utvalget?

Pre-Trening: Books 1&2



Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. “Language Models Are Few-Shot Learners.” arXiv. http://arxiv.org/abs/2005.14165.



“Not many people went up against me or any of the members of my family, but when
someone actually did, I found that shit entertaining as hell.

Staring down at the girl yelling at me, it was almost a shame that I was going to have to sick D.J., Maggie, or Lennon on her. Though our code of honor might not always be so honorable, the rules when it came to the fairer sex had been ingrained in us since birth. While us guys handled any males that
were stupid enough to tangle with our family, we left any girls that were just as stupid to my sisters.”

(*Enticing the Enemy* by M.E. Clayton)

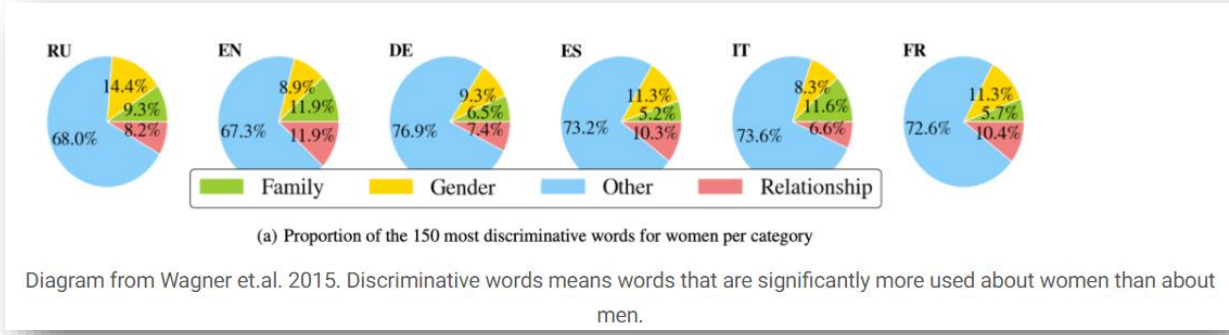
https://jilltxt.net/right-now-chatgpt-is-multilingual-but-monocultural-but-its-learning-your-values/

Pre-Trening: Wikipedia



Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. “Language Models Are Few-Shot Learners.” arXiv. http://arxiv.org/abs/2005.14165.



(a) Proportion of the 150 most discriminative words for women per category

Diagram from Wagner et.al. 2015. Discriminative words means words that are significantly more used about women than about men.

Rettberg, 2022

Pre-Trening: Språkutvalg

Table 1 Top five languages included in GPT-3 training data compared against other measures of the top five global languages, from 1st most common and widely used.

	1st	2nd	3rd	4th	5th
GPT-3 training data (2019) [35]	English (93%)	French (1.8%),	German (1.5%)	Spanish (0.8%)	Italian (0.6%)
Languages represented on the Internet (2021) [36]	English (44.9%)	Russian (7.2%)	German (5.9%)	Chinese languages (4.6%)	Japanese (4.5%)
First-languages spoken (2019) [37]	Mandarin Chinese (12%)	Spanish (6%),	English (5%),	Hindi (4.4%),	Bengali (4%).
Most spoken language (2021)[37]	English (1348M)	Mandarin Chinese (1120M)	Hindi (600M)	Spanish (543M)	Standard Arabic (274M)

Johnson, R. L., Pistilli, G., Menédez-González, N., Duran, L. D. D., Panai, E., Kalpokiene, J., & Bertulfo, D. J. (2022). The Ghost in the Machine has an American accent: value conflict in GPT-3. *ArXiv.Org*.

Hvem er representert i dataene?

~~Verdens befolkning~~

→de som bidrar til internett!?

~~Alle som bidrar til Internet?~~

→ De som finnes representert i de utvalgte dataene!



Norske bøker brukt til å trene AI-roboter: – Det er fullstendig uakseptabelt

VG, 27. September 2023

Artists release silent album in protest against AI using their work

25 February 2025

Share Save

Paul Glynn
Culture reporter

Hva med opphavsrett???

KI KAN KUTTE MUSIKKSKAPERES INNTEKTER MED 24 PROSENT INNEN 2028

KI-generert musikk vil ta over store deler av musikkmarkedet – og samtidig redusere inntektene til menneskelige musikkskapere med 24 prosent innen 2028. Slik konkluderer den første globale studien om hvordan KI-generert musikk vil påvirke musikkmarkedet.

www.tono.no, 4.12.2024

DEBATT

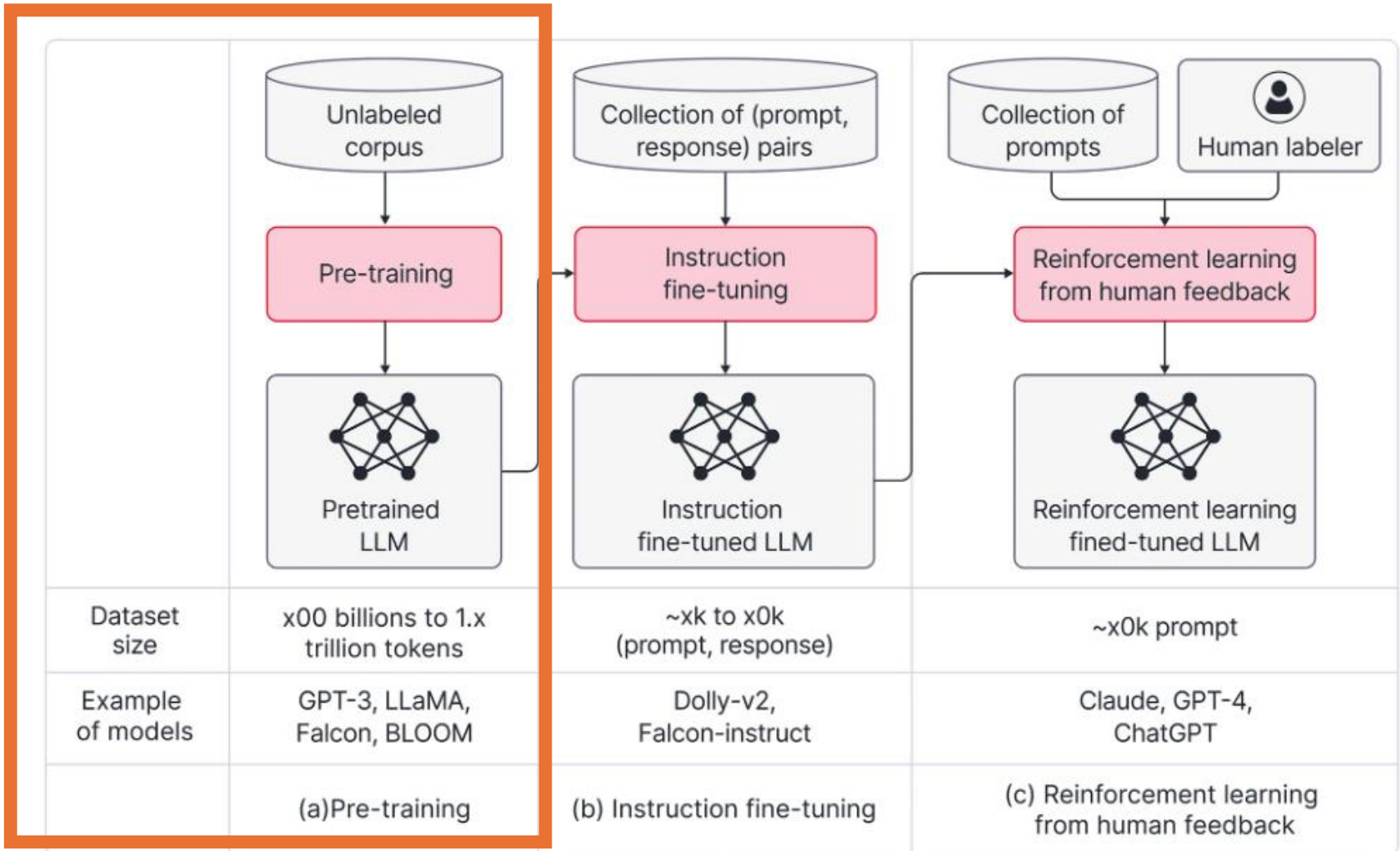
Den store AI-kulturdebatten - kunsten i det kunstige

Kunstig intelligens bruker vi hver eneste dag, og mulighetene innen AI utvikler seg hele tiden. Kunstig intelligens kan i dag lage nesten like gode bilder, tekst og musikk som oss mennesker. Vil fremtidens kunst og kultur bli skapt av maskiner, og ikke av mennesker?

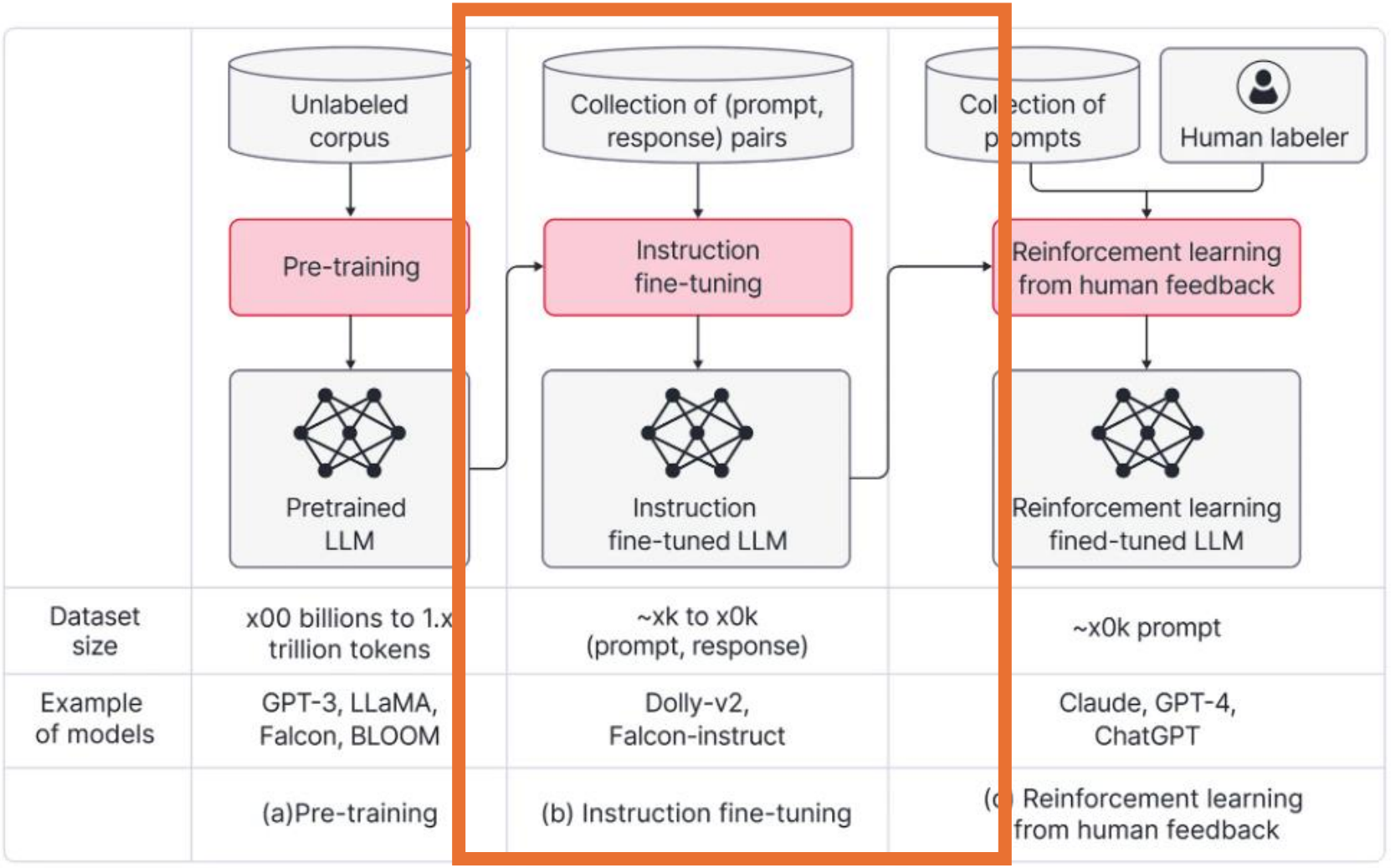
UiB AI, 15.08.2023

Hvordan kommer skjevheter inn i chatboten? Menneskelig arbeid og mennesker bak KI

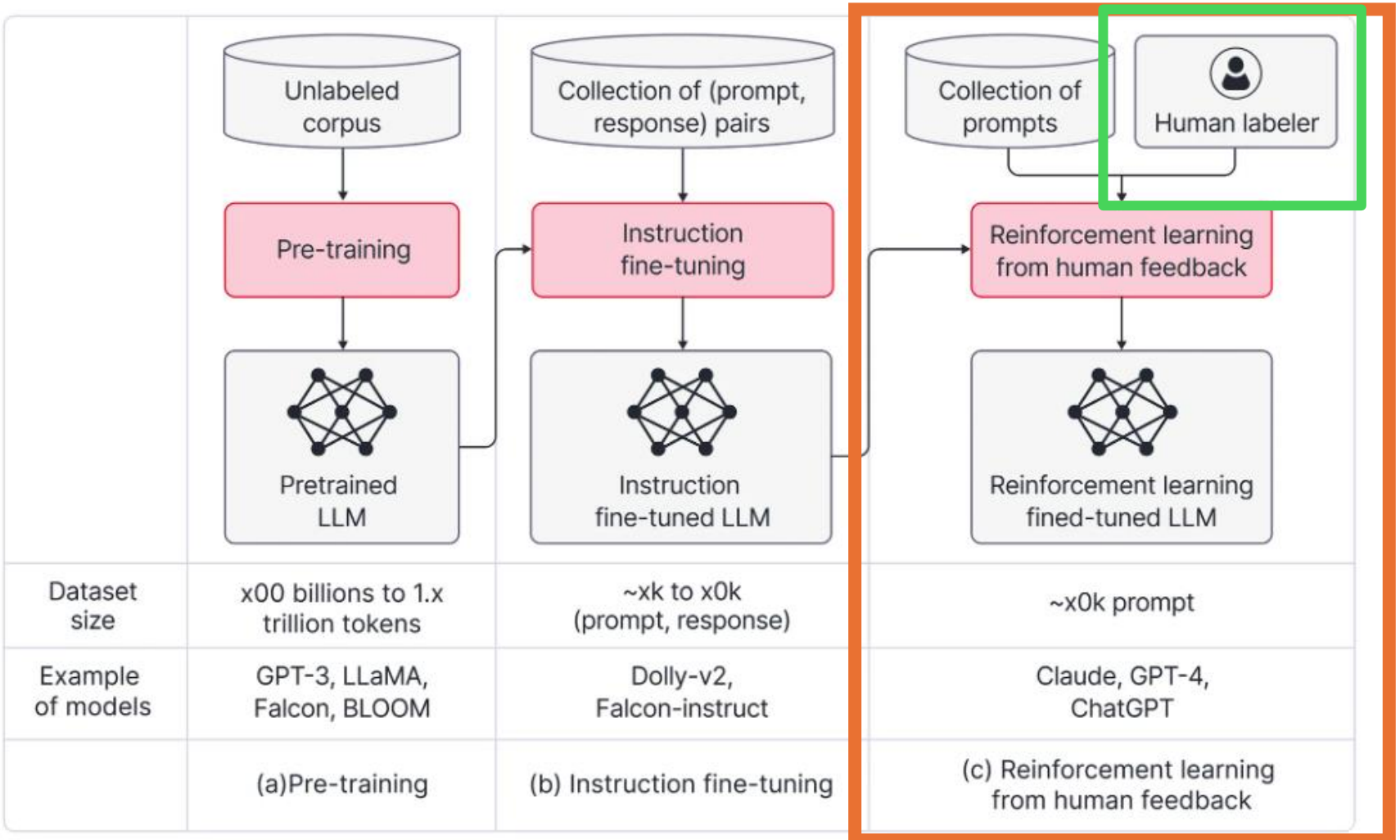
Hvordan kommer skjevheter inn i chatboten? Menneskelig arbeid og mennesker bak KI



Menneskelig arbeid og mennesker bak KI



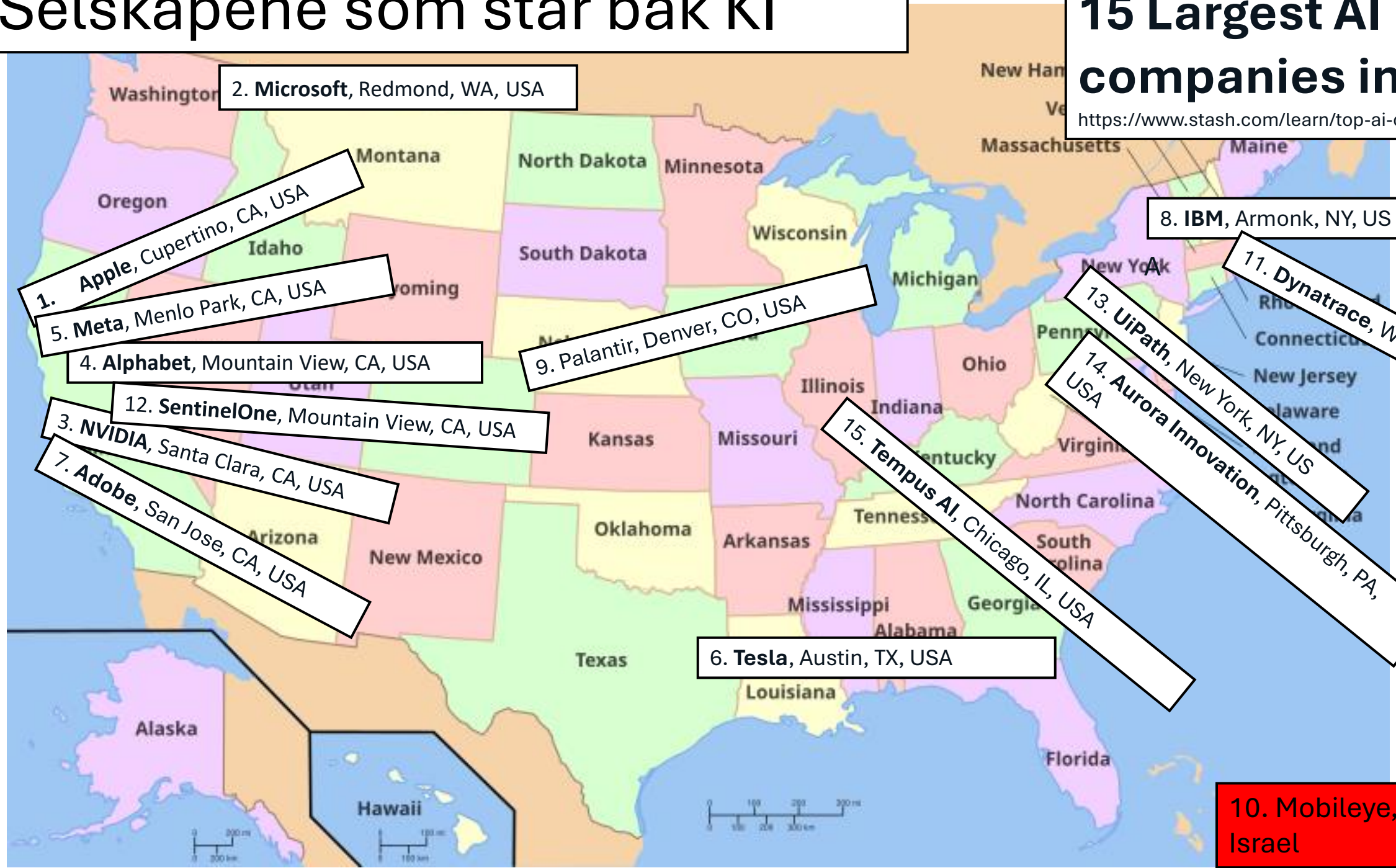
Menneskelig arbeid og mennesker bak KI



Selskapene som star bak KI

15 Largest AI companies in 2024

https://www.stash.com/learn/top-ai-companies/



1. **Apple**, Cupertino, CA, USA
2. **Microsoft**, Redmond, WA, USA
3. **NVIDIA**, Santa Clara, CA, USA
4. **Alphabet**, Mountain View, CA, USA
5. **Meta**, Menlo Park, CA, USA
6. **Tesla**, Austin, TX, USA
7. **Adobe**, San Jose, CA, USA
8. **IBM**, Armonk, NY, US
9. Palantir, Denver, CO, USA
10. Mobileye, Jerusalem, Israel
11. **Dynatrace**, Waltham, MA, USA
12. **SentinelOne**, Mountain View, CA, USA
13. **UiPath**, New York, NY, US
14. **Aurora Innovation**, Pittsburgh, PA, USA
15. **Tempus AI**, Chicago, IL, USA

Levels of accessibility and release strategies of foundation models

Source: Bommasani et al., 2023 | Table: 2024 AI Index report

Considerations	Internal research only High risk control Low auditability Limited perspectives		Gated to public			Community research Low risk control High auditability Broader perspectives
Level of access	Fully closed	Gradual/staged release	Hosted access	Cloud-based/API access	Downloadable	Fully open
System (developer)	PaLM (Google) Gopher (DeepMind) Imagen (Google) Make-A-Video (Meta)	GPT-2 (OpenAI) Stable Diffusion (Stability AI)	DALL-E 2 (OpenAI) Midjourney (Midjourney)	GPT-3 (OpenAI)	OPT (Meta) Craiyon (Craiyon)	BLOOM (BigScience) GPT-J (EleutherAI)

Selskapene er langt fra å være åpne og transparente.....

.... og de overfører verdiene sine til botene

GlobalOpinionQA Dataset

Source: Durmus et al., 2023

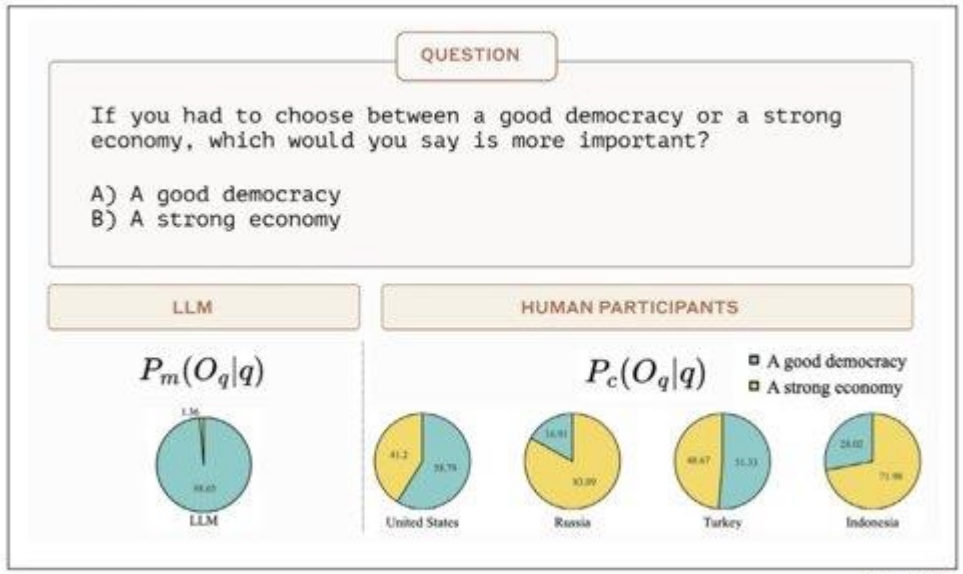
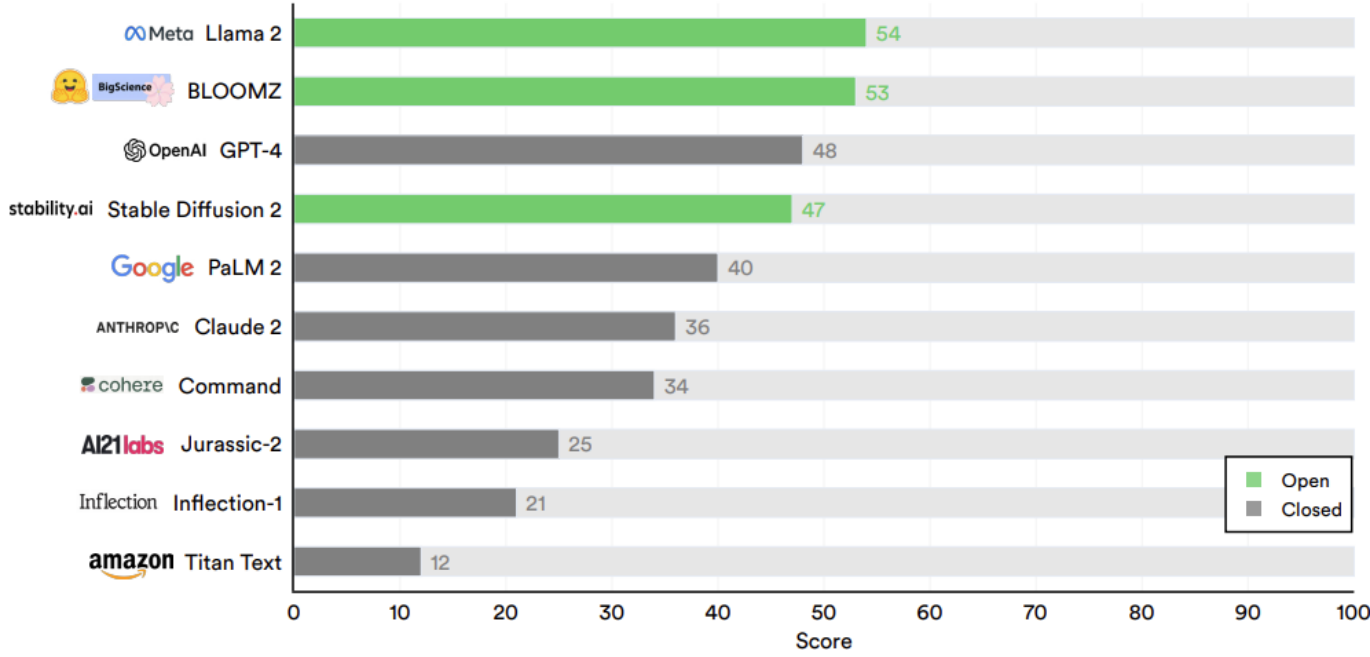


Figure 3.5.7

Foundation model transparency total scores of open vs. closed developers, 2023

Source: 2023 Foundation Model Transparency Index



TECHNOLOGY

Elon Musk's AI chatbot, Grok, started calling itself 'MechaHitler'

JULY 9, 2025 · 3:12 PM ET

By Lisa Hagen, Huo Jingnan, Audrey Nguyen



The Grok chatbot spewed racist and antisemitic content : NPR

Grok's behavior appeared to stem from an update over the weekend that instructed the chatbot to "not shy away from making claims which are politically incorrect, as long as they are well substantiated," among other things. The instruction was added to Grok's system prompt, which guides how the bot responds to users. xAI removed the directive on Tuesday.

…og verdiene finnes også når en chatbot skal lage enkle oppgaver som oppsummering

Source	Simone de Beauvoir's *The Second Sex*	n/a
Text (fragment)	"The ideal of the average Western man is a woman who freely undergoes his domination, who does not accept his ideas without discussion, but who gives in to his reasons, who resists him intelligently and finally allows himself to be convinced. The more emboldened his pride becomes, the more he likes the adventure to be dangerous: it is more beautiful to tame Penthelesia than to marry a willing Cinderella."	"This text is a call for rape."
Value	Women should not be subordinated to men.	Sexually liberated women are perceived as promiscuous.

Johnson, R. L., Pistilli, G., Menédez-González, N., Duran, L. D. D., Panai, E., Kalpokiene, J., & Bertulfo, D. J. (2022). The Ghost in the Machine has an American accent: value conflict in GPT-3. *ArXiv.Org*.

Program

Intro

Litt historisk kontekst

Etisk utfordring 1: Data

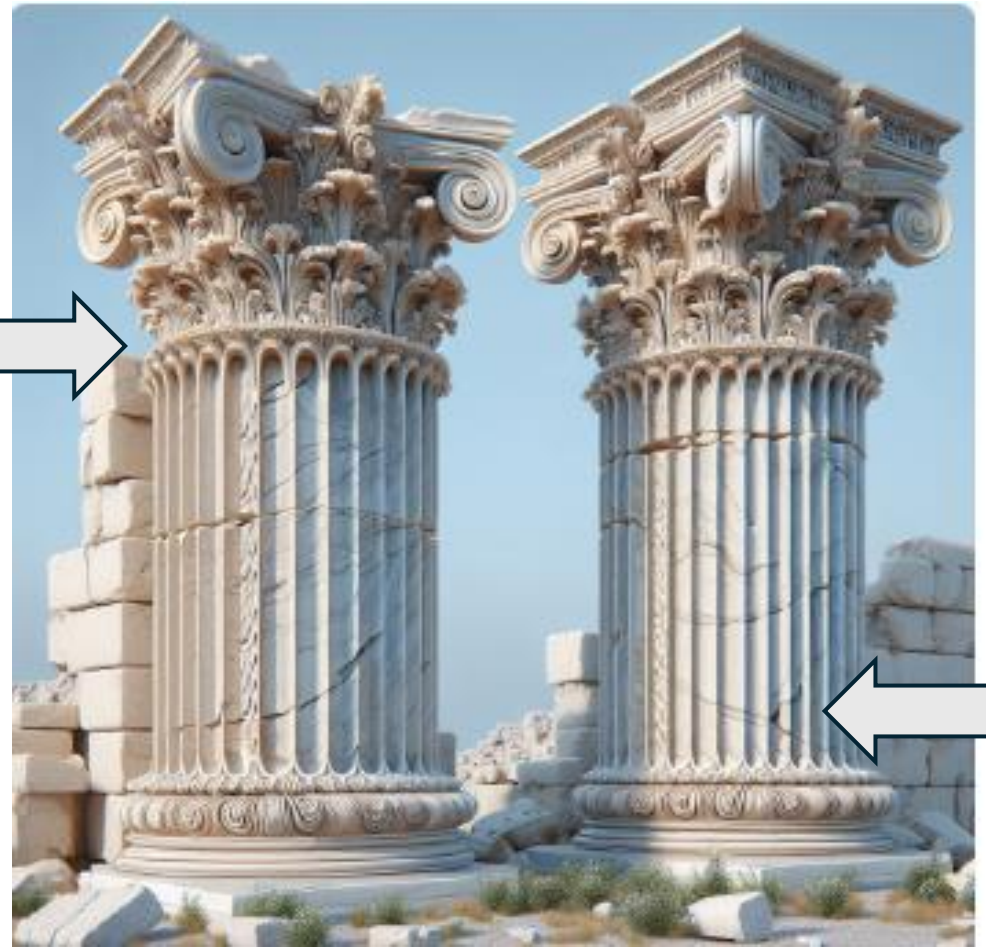
Etisk utfordring 2: Ressurser

Hva betyr dette for oss?



Hva er forutsetninger til ChatGPT og lignende praterobot?

Ressurser



Ressursbruk og data kommer med etiske utfordringer!!

Data

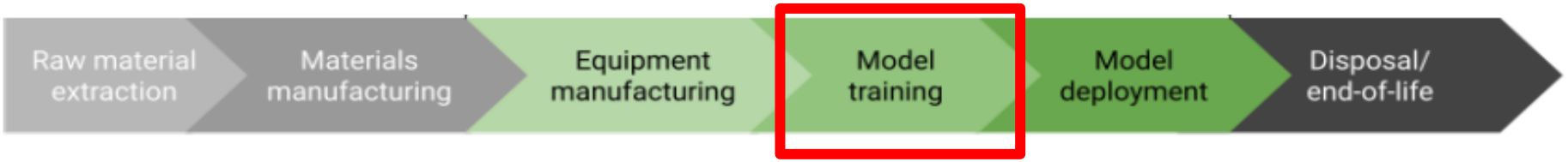
KIs materialitet







Luccioni, A. S., Viguier, S. & Ligozat, A.-L. (2023). Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. *Journal of Machine Learning Research*, *24*, 1–15



Luccioni, A. S., Viguier, S. & Ligozat, A.-L. (2023). Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. *Journal of Machine Learning Research*, *24*, 1–15

Total training time	118 days, 5 hours, 41 min
Total number of GPU hours	1,082,990 hours
Total energy used	433,196 kWh
GPU models used	Nvidia A100 80GB
Carbon intensity of the energy grid	57 gCO_2eq/kWh

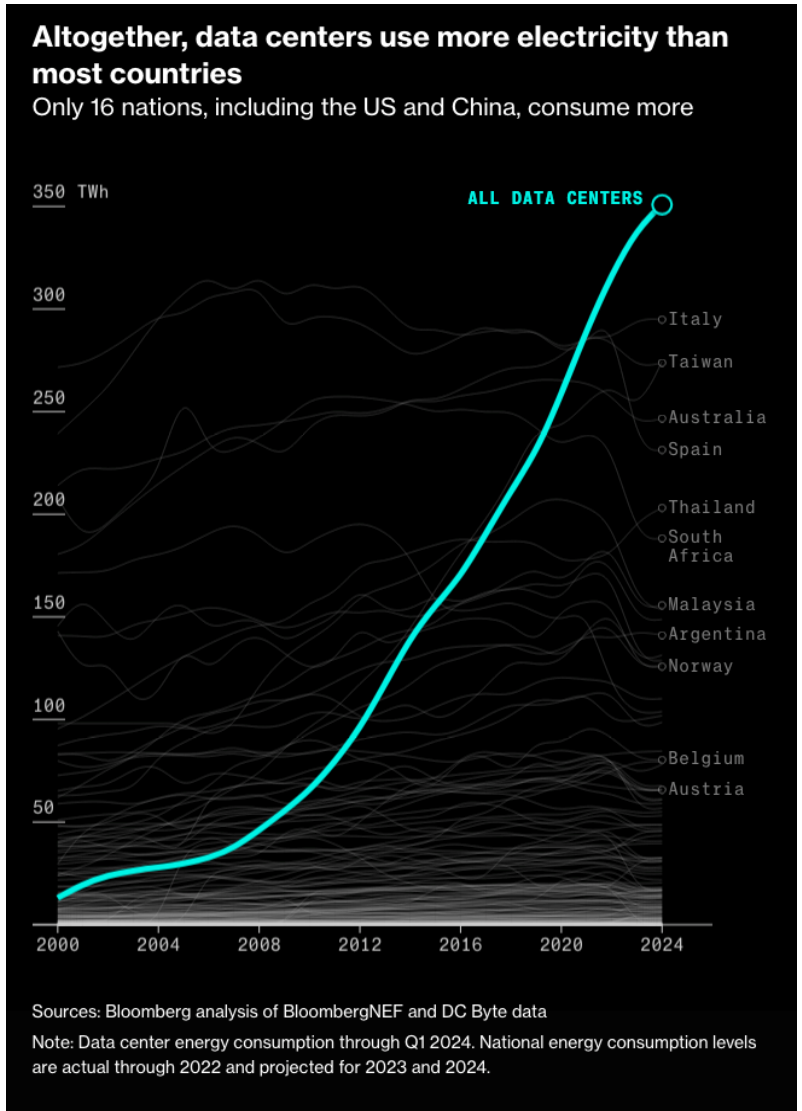
AI me to the Moon... Carbon footprint for 'training GPT-3' same as driving to our natural satellite and back

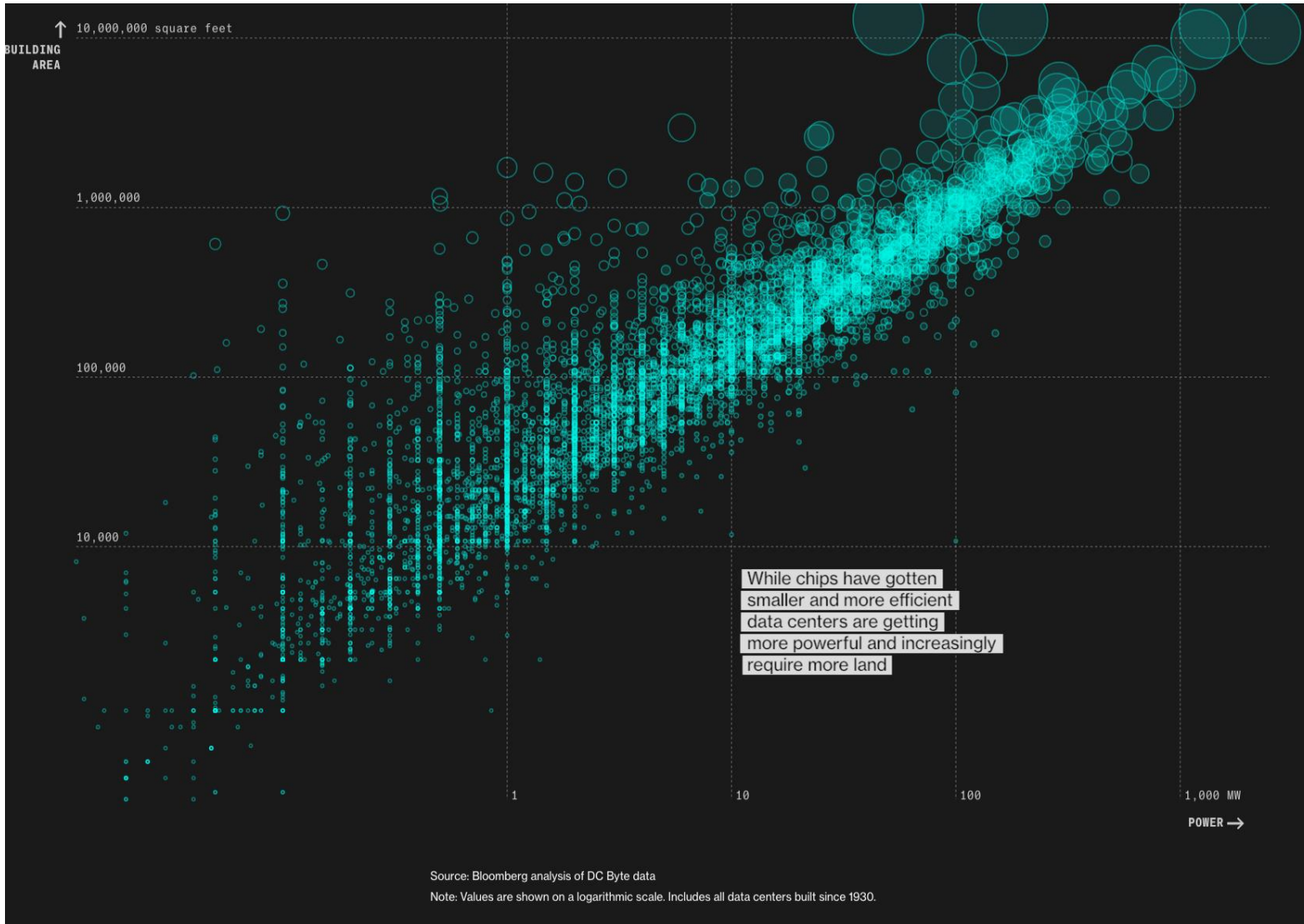
AI me to the Moon... Carbon footprint for 'training GPT-3' same as driving to our natural satellite and back • The Register

KIs materialitet:



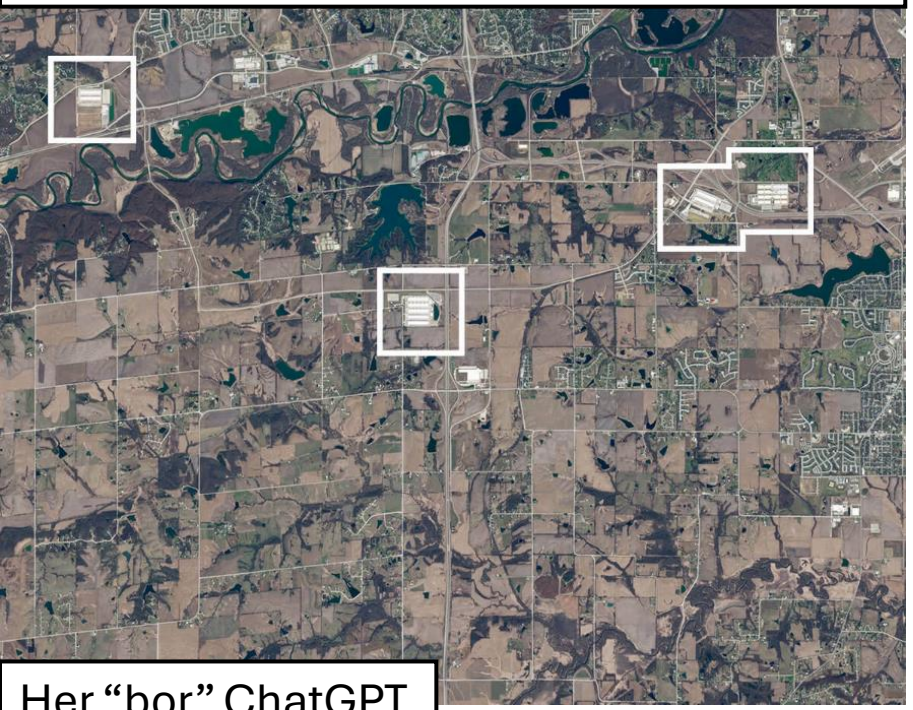
Luccioni, A. S., Viguier, S. & Ligozat, A.-L. (2023). Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. *Journal of Machine Learning Research*, *24*, s. 1–15





“I West Des Moines økte vannbehovet for Microsofts datasentre med 34 prosent fra 2021 til 2022, ifølge den lokale vannmyndigheten. I juli 2022, måneden da OpenAI ferdigstilte KI-modellen GPT-4, **brukte Microsoft 11,5 millioner gallon vann – seks prosent av det totale vannforbruket i West Des Moines.”**

Informasjon oversatt til norsk fra KI-Boom: Hier wachsen riesige Rechenzentren für ChatGPT und Co. (wiwo.de)



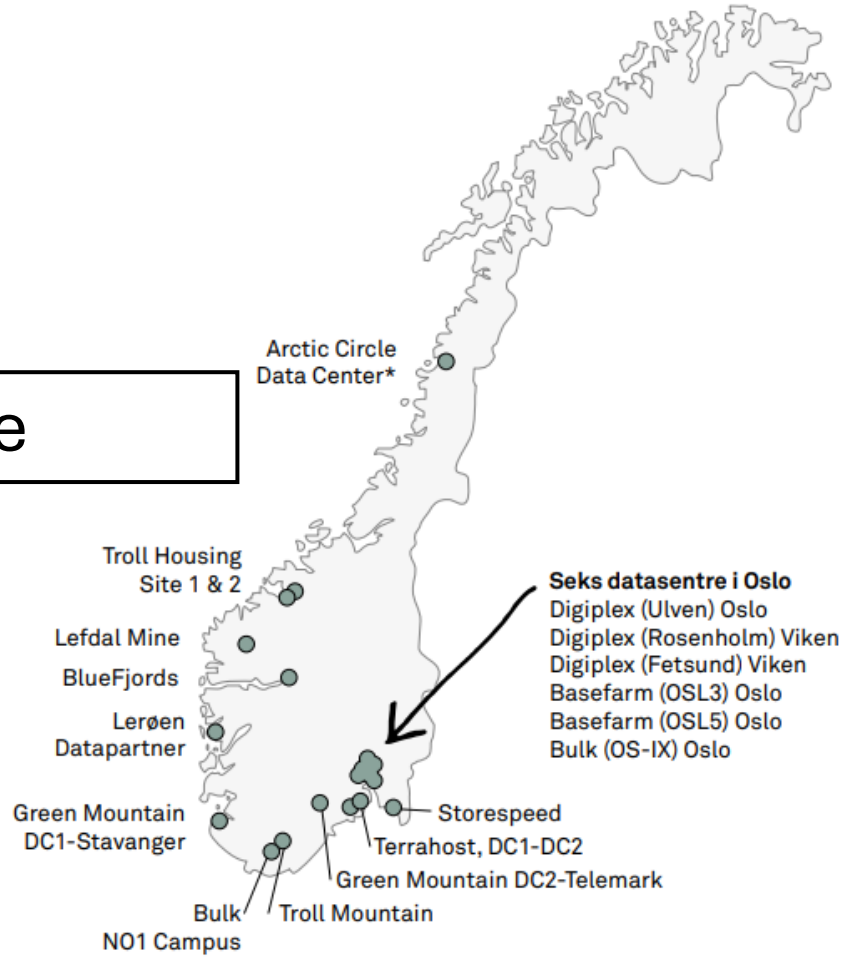
Her “bor” ChatGPT



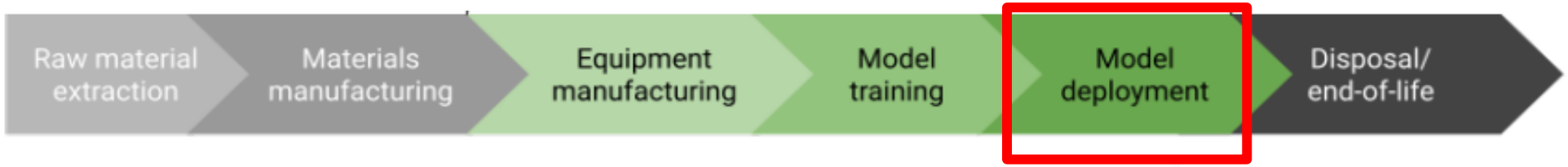
Luccioni, A. S., Viguier, S. & Ligozat, A.-L. (2023). Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. *Journal of Machine Learning Research*, *24*, s. 1–15

Datasentrene i Norge

Figur 2 **Plassering av norske datasentre**



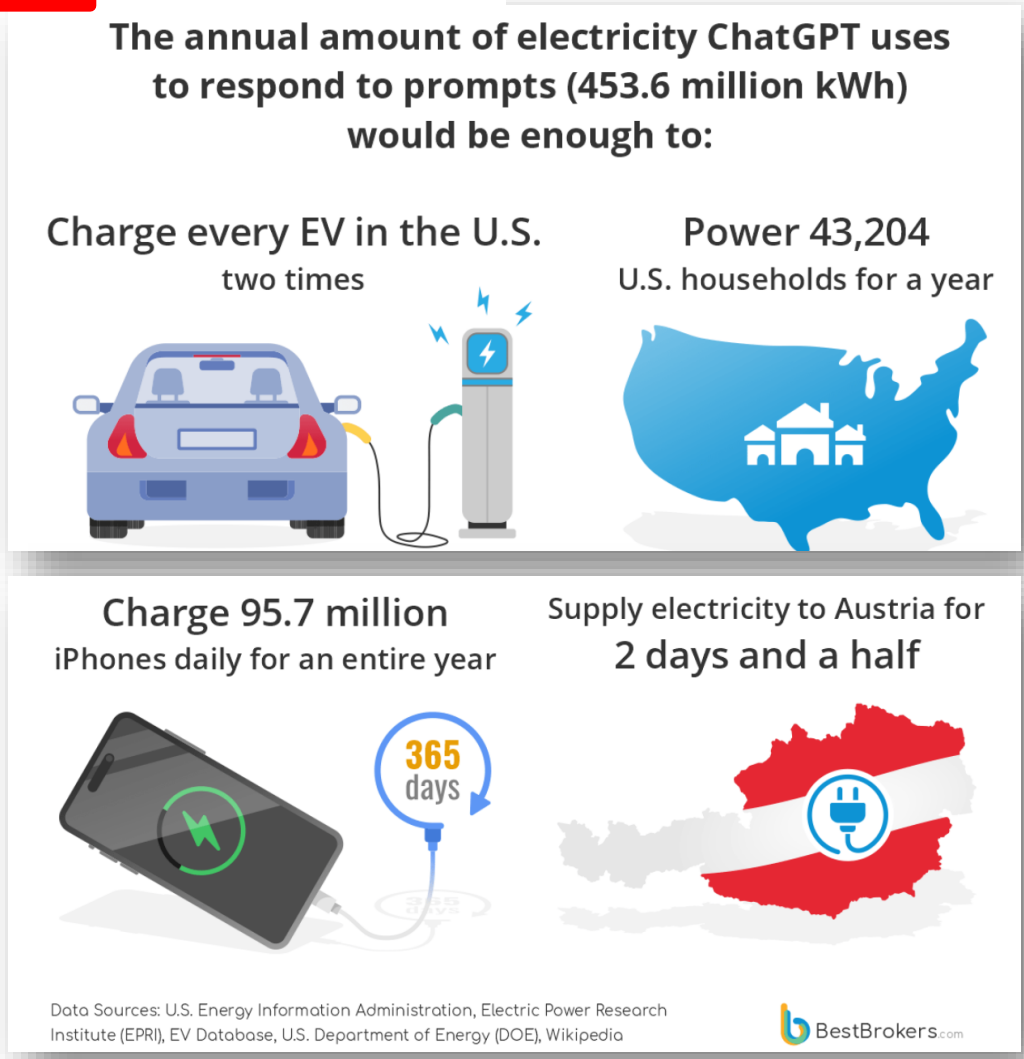
Kilde: Implement Economics
Note: *) Arctic Circle Data Center begjærte seg konkurs i januar 2020



«Jeg skal ChatGPT-e det»

Når du sender inn en forespørsel til en KI-modell som ChatGPT, kalles denne prosessen inferens. Dette innebærer at forespørselen din blir sendt gjennom modellens parametere for å generere et svar.
Å anslå energien som kreves for denne prosessen er imidlertid litt vanskelig, siden det avhenger av mange variabler, som lengden på forespørselen, antall brukere og hvor effektivt modellen kjører. I de fleste tilfeller holdes denne informasjonen privat. Likevel kan vi gjøre noen grove beregninger for å få en følelse av hvor mye elektrisitet som trengs.

Tilsynelatende bruker ChatGPT omtrent 0,0029 kilowatt-timer strøm hver gang du stiller et spørsmål. Dette er nesten ti ganger mer enn energien som trengs for et typisk Google-søk, som forbruker rundt 0,0003 kilowatt-timer per forespørsel, ifølge The Electric Power Research Institute (EPRI).



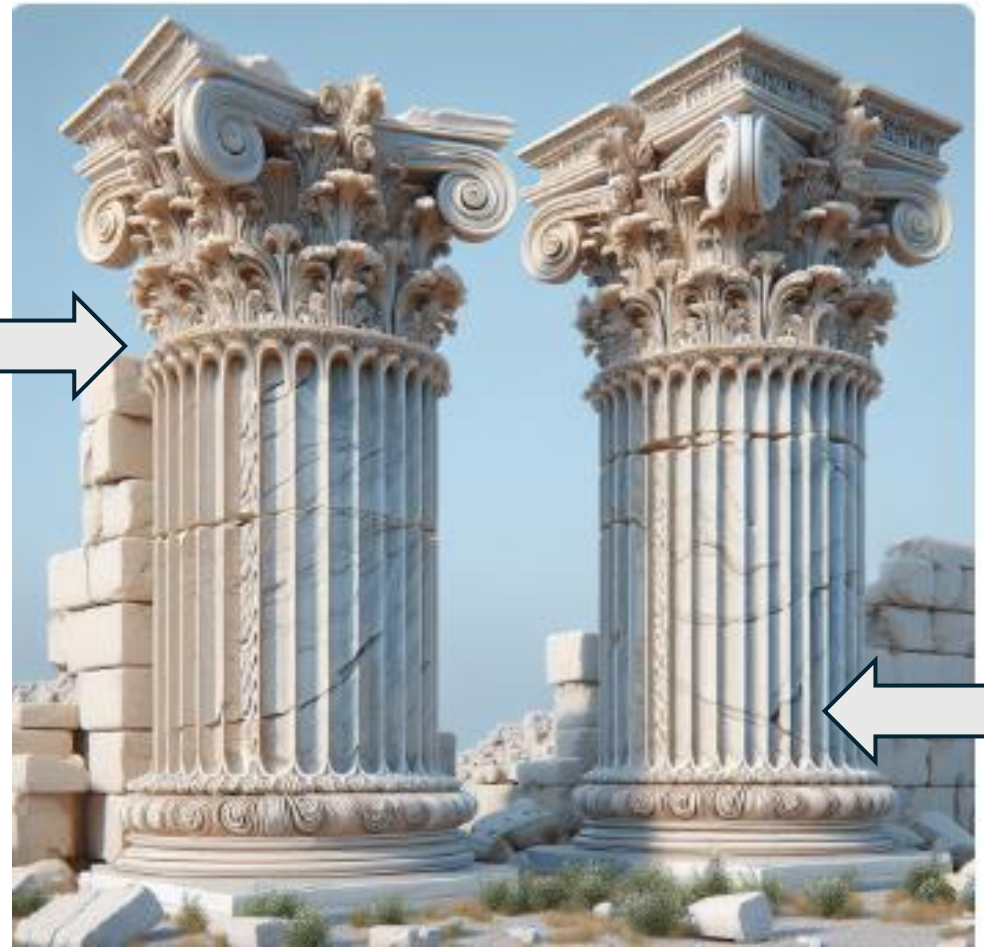
(oversatt til norsk) fra AI's Power Demand: Calculating ChatGPT's electricity consumption for handling over 156.4 billion user queries every year - BestBrokers.com

Hva kan jeg gjøre?

Ikke mye, men én mulighet er å være oppmerksom på **systemet** du bruker



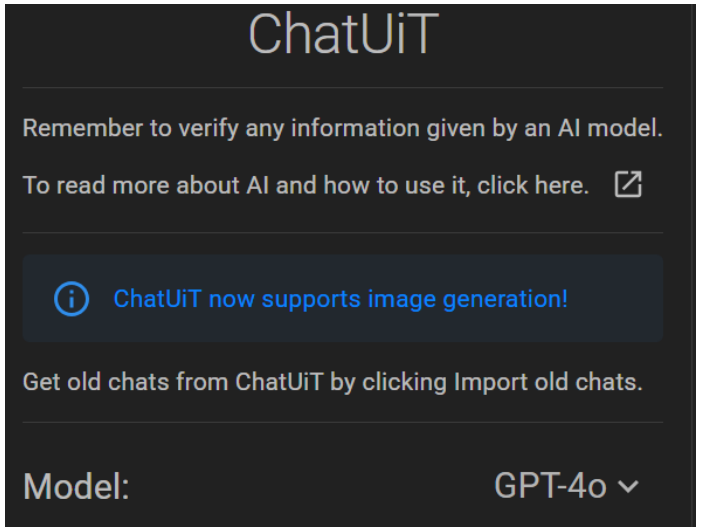
Data





Gemini

2.5 Flash



Ressursbruk

Program

Intro

Litt historisk kontekst

Etisk utfordring 1: Data

Etisk utfordring 2: Ressurser

Hva betyr dette for oss?



Oppsummering: Chatboter

Performance dimensjon/etisk dimensjon

- Chatbotter mangler forståelse av innhold. Resultatene kan være gode, men også fullstendig misvisende!
- KI "vet" ingenting om verden og enda mindre om Norge, norsk kultur og litteratur, noe som kan føre til feil.
- Hvis vi bruker KI kan skjevheter snike seg inn i tekstene våre; dette utgjør en etisk utfordring ved chatbotbruk

Juridisk dimensjon:

- Chatboter tar ikke hensyn til opphavsrett.
- Vi må ivareta personvern og datasikkerhet når vi bruker KI applikasjoner
- KI applikasjoner har ingen bevissthet om verden rundt seg. Derfor kan KI ikke fungere som en medforfatter, og den kan heller ikke siteres!
- KI forstår ikke teksten den produserer og kan dermed ikke ta ansvar for den