AUH

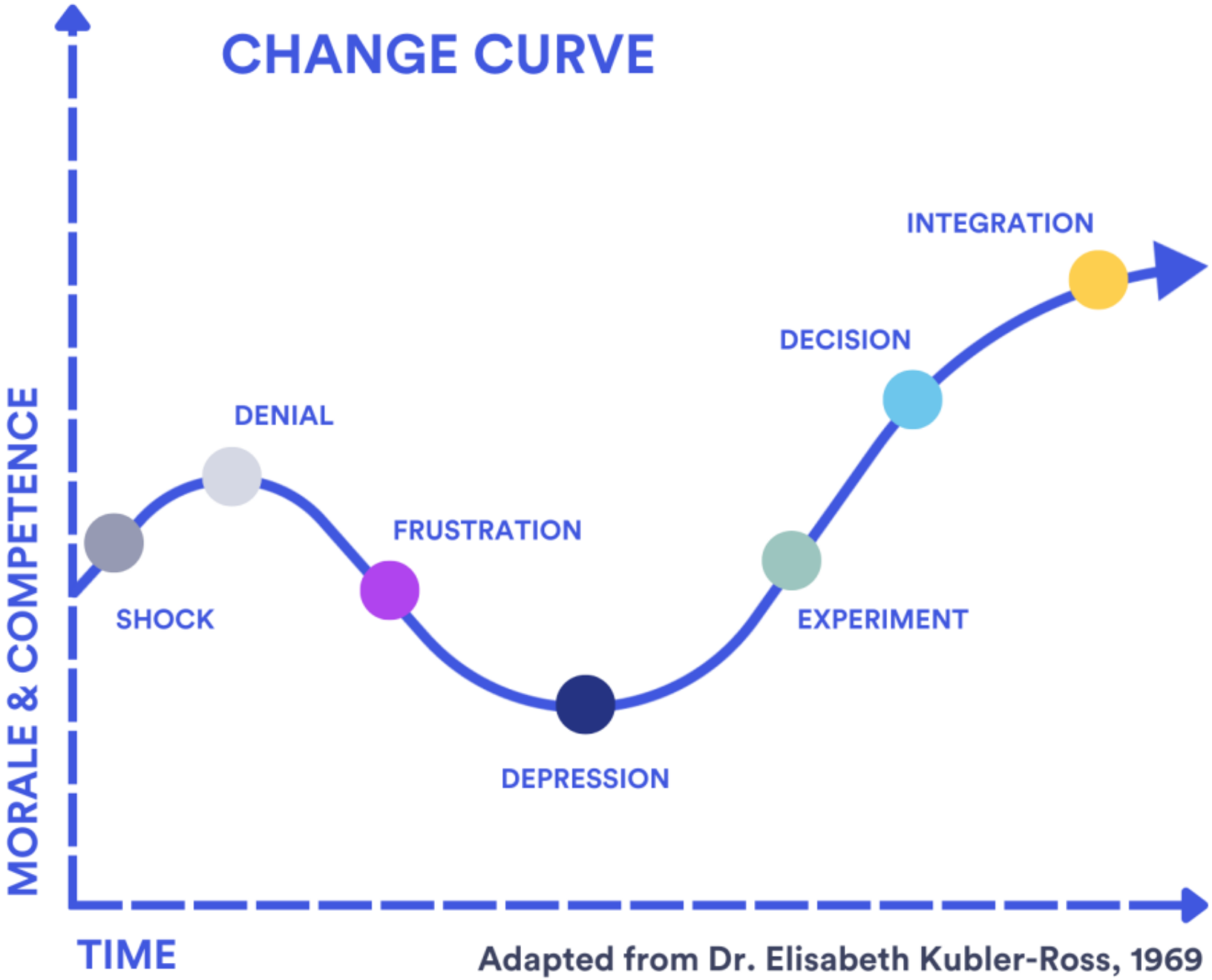
25. Oktober 2024

GENERATIV AI FOR UDDANNELSESLÆGER

Tom Gislev, 2024

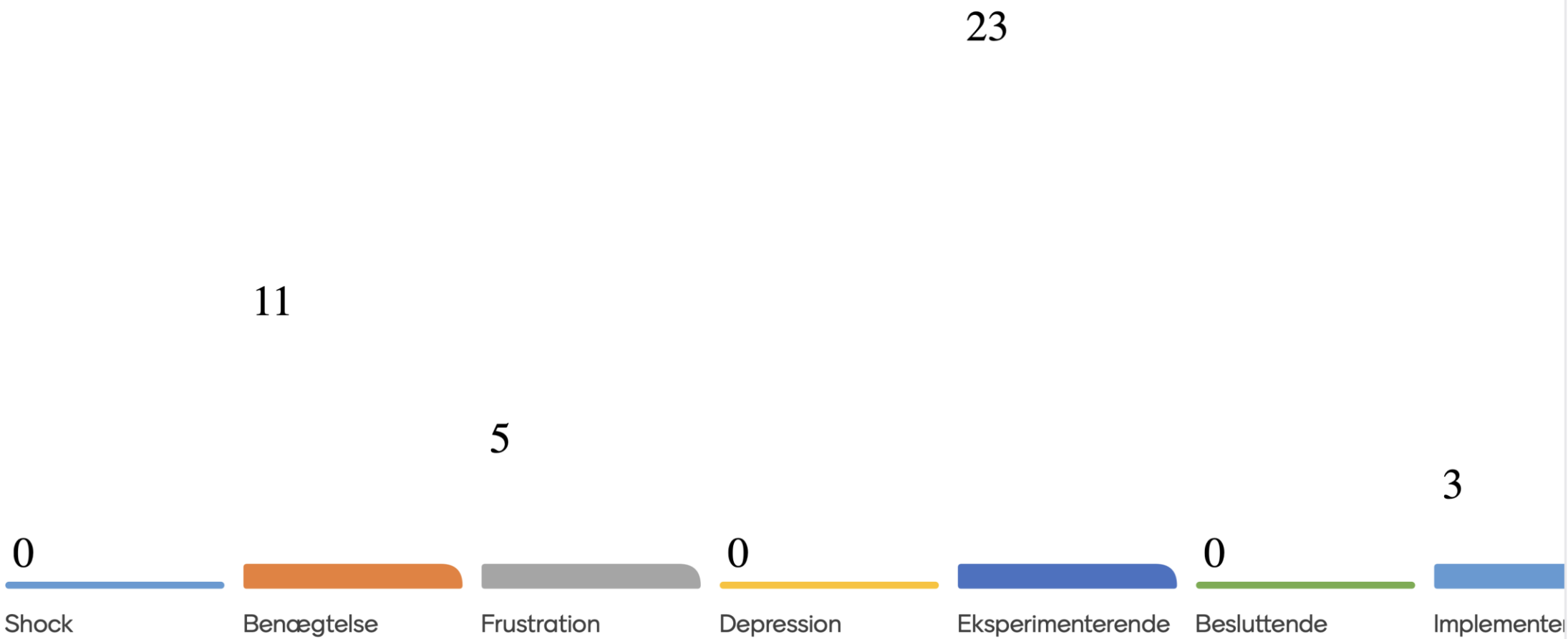
- Introduktion til teknologien bag GenAI.
- Introduktion til prompting
- Pause & check ud.
- Problematikker ved brug af GenAI
- Eksempler fra undervisningen på AU
- Avancerede prompt teknikker

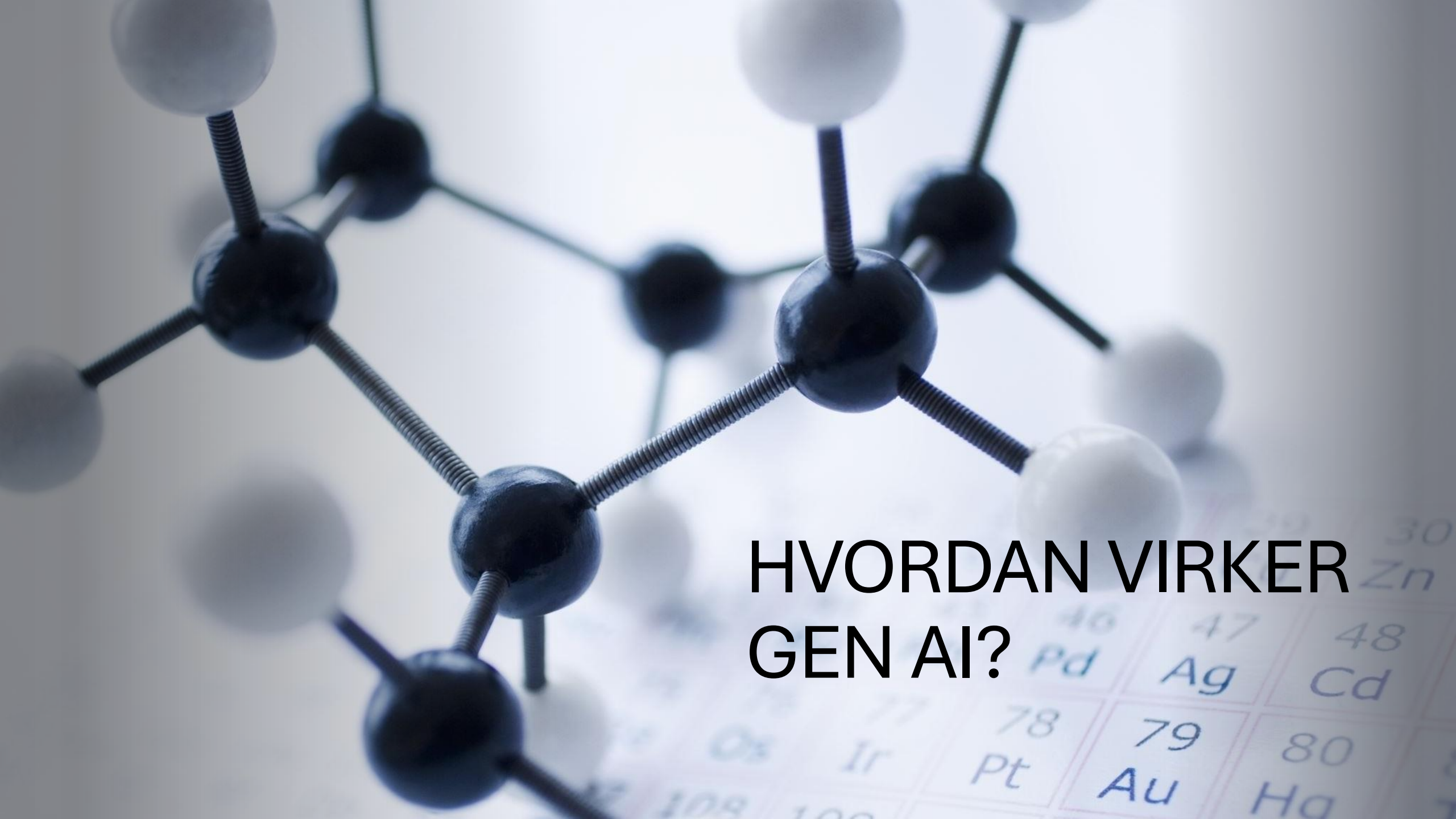
Tom Gislev.
tgislev@au.dk
Tlf: 93508484


CHANGE CURVE
INTEGRATION
DECISION
DENIAL
FRUSTRATION
SHOCK
EXPERIMENT
DEPRESSION
MORALE & COMPETENCE
TIME
Adapted from Dr. Elisabeth Kubler-Ross, 1969

Gå til menti.com | og brug koden 6245 273

Multiple Choice




HVORDAN VIRKER
GEN AI?

GENERATIV AI ER ET PARAPLYBEGREB

Generativ AI er et subfelt inden for kunstig intelligens, der beskæftiger sig med teknologier, der genererer nye data.

STORE SPROGMODELLER

ChatGPT

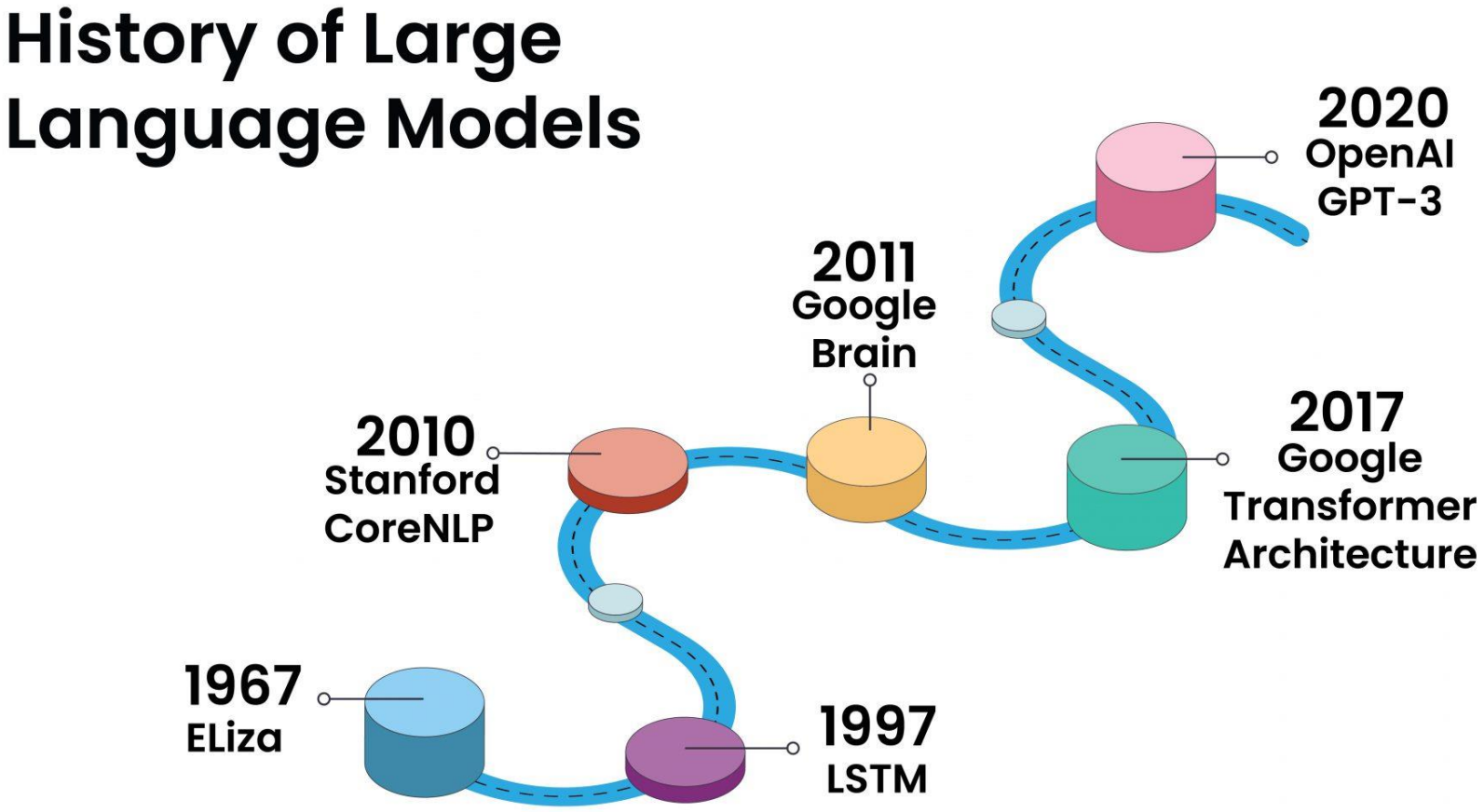
COPILOT

Gemini

MISTRAL AI_

Claude

STORE SPROGMODELLERS HISTORIE



HVAD SKETE DER I NOVEMBER 2022?

SKALA

Ekstremt meget træningsdata.

Korpus på 45 TB

TILGÆNGELIGHED

Enhver kan instruere en computer, ved at bruge klar tale.

CHAT GPT 3.5

ChatGPT 4o

Share

Giv 3 gode råd til en læge, der ønsker at starte på at bruge Generativ AI i sin praksis

Her er tre gode råd til en læge, der ønsker at begynde at bruge Generativ AI i sin praksis:

1. **Start med små, veldefinerede opgaver**: Begynd med at implementere Generativ AI til opgaver, der ikke kræver beslutningstagning på liv og død. Det kan være hjælp til dokumentation, fx udarbejdelse af notater eller breve baseret på input fra samtaler, eller opsummering af store mængder tekst som forskningsartikler eller patienthistorik. Dette kan spare tid og sikre, at du forstår teknologiens potentiale og begrænsninger, før den bruges i mere kritiske situationer.
2. **Prioritér datasikkerhed og patientfortrolighed**: Vær meget opmærksom på, hvordan du beskytter patientoplysninger, når du bruger AI. Tjek altid, om det AI-værktøj, du overvejer at bruge, overholder GDPR og andre relevante databeskyttelseslove, så patienternes fortrolighed bevares. Dette er især vigtigt, når det drejer sig om sundhedsdata, som er ekstra følsomme.
3. **Hold øje med medicinske bias**: Generativ AI kan utilsigtet videreføre eller forstærke bias fra de data, den er trænet på. Derfor er det vigtigt at bruge værktøjerne som et supplement til din egen vurdering og ikke som en erstatning. Brug AI-output med kritisk sans, og vær særligt opmærksom på at evaluere, om de forslag, AI'en genererer, er baseret på relevante og evidensbaserede kilder, specielt når det gælder behandling og diagnose.

INSTRUKTION

GENERERET TEKST

Tom Gislev, 2024

Giv 3 gode råd til en læge, der ønsker at starte på at bruge Generativ AI i sin praksis

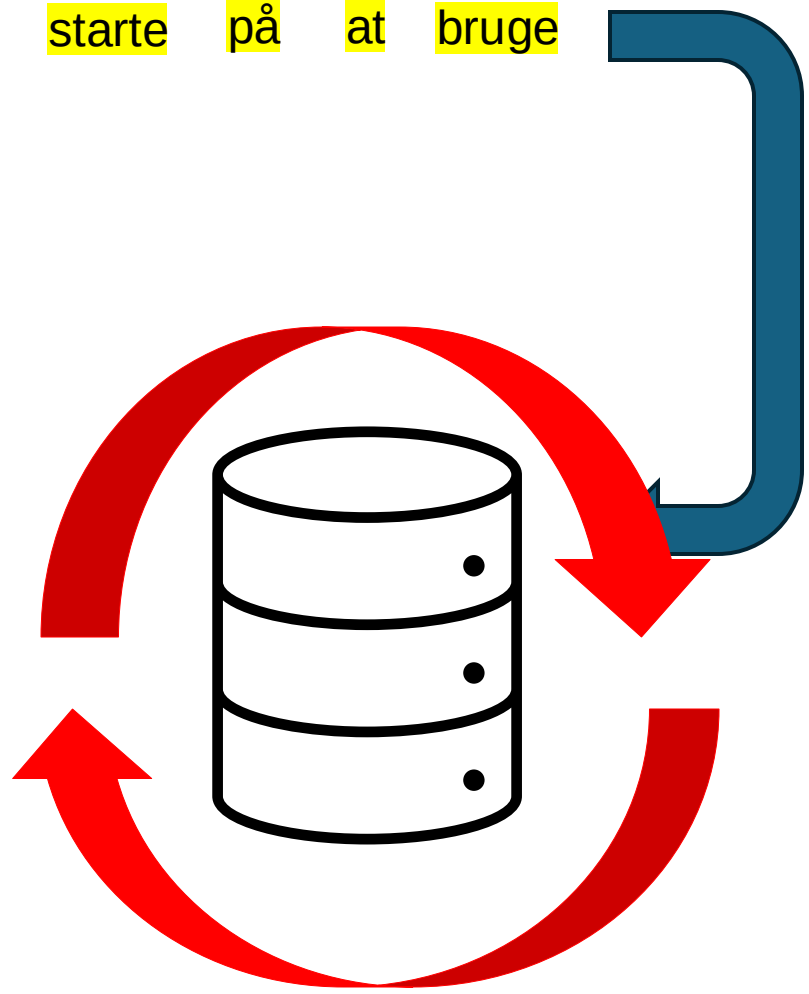
Giv 3 gode råd til en læge , der ønsker at starte på at bruge

Instruktionen i prompten transformeres til tokens.

Tokens fødes ind i modellen

Sprogmodellen beregner en sandsynlighedsfordeling af tokens på baggrund af instruktionen.

Modellen forudser den mest sandsynlige næste token, som herefter ”skrives” i output vinduet.



Her er tre gode råd til en læge, der ønsker at begynde at bruge Generativ AI i sin praksis:

1. **Start med små, veldefinerede opgaver**: Begynd med at implementere Generativ AI til opgaver, der ikke kræver beslutningstagning på liv og død. Det kan være hjælp til dokumentation, fx udarbejdelse af notater eller breve

HVORDAN SER DET UD?

Hvad tror I er det næste ord i den følgende sætning?

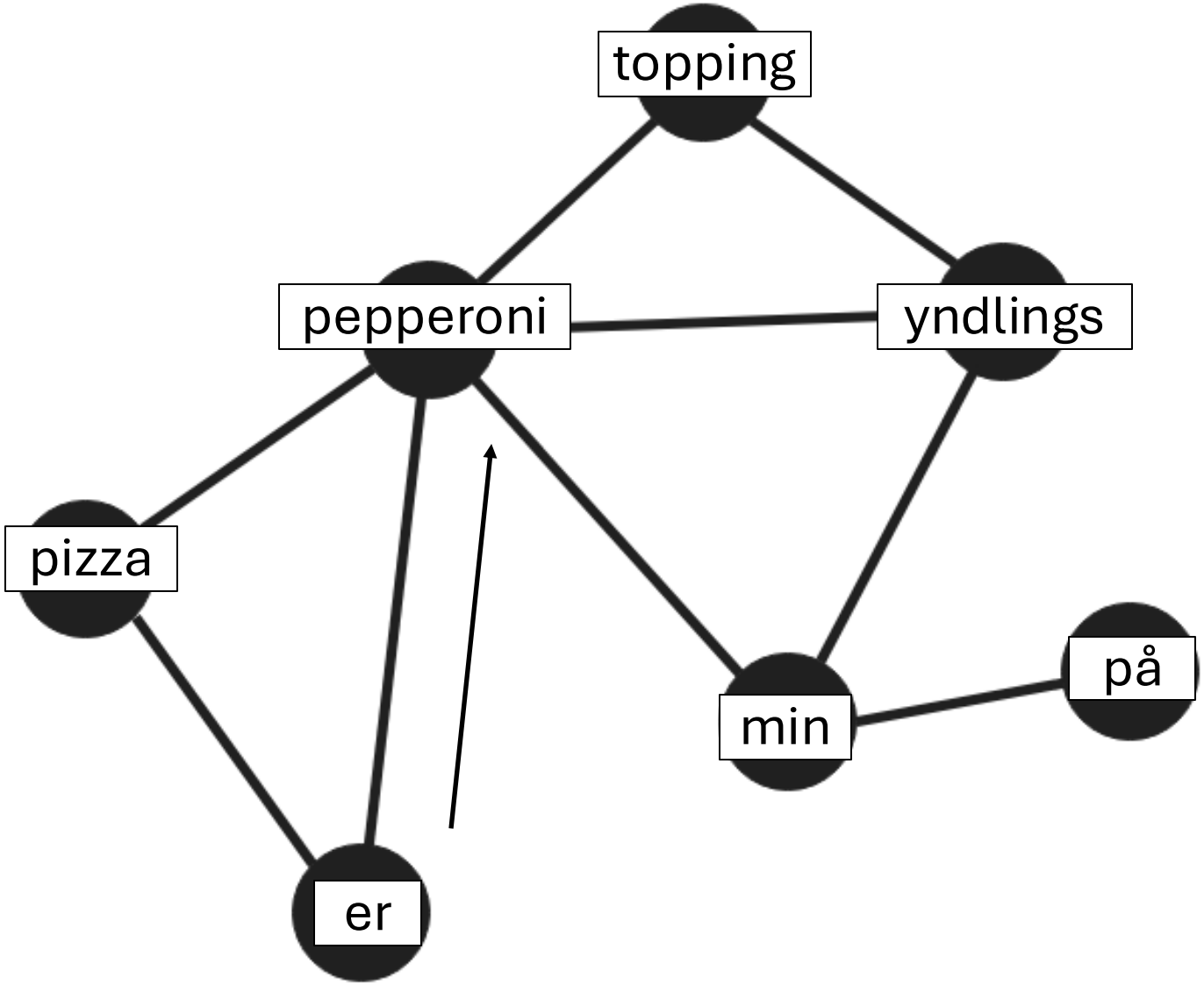
Min yndlings topping på …

Min yndlings topping på **pizza** …

Min yndlings topping på pizza **er** …

Min yndlings topping på pizza er **peperoni**. **skinke**? **ananas**?

MATEMATIKKEN



SPROGMODELLER ER STATISTISKE ALGORITMER

En sprogmodel er en statistisk model af naturligt sprog, der kan generere tilsyneladende sammenhængende og meningsfuld tekst i relation til en given kontekst.

Et program, der er (meget) god til at forudsige det næste ord i en given streng af tekst.

Sprogmodeller er ikke beregnet til at formidle viden, men til at skrive tekst.

For at skabe brugbart og troværdigt indhold med Generativ AI i en akademisk sammenhæng, er det nødvendigt at anvende:

KRITISK TÆNKNING

og

DOMÆNE KENDSKAB

Tom Gislev, 2024

SPØRGSMÅL
&
KOMMENTARER?


PROMPTING 1

OVERORDNEDE ANBEFALINGER

Forvent ikke at første prompt giver en brugbar tekst.

Instruer modellen i at modificere og tilpasse den genererede tekst.

- Vær specifik, tydelig og undgå unødig kompleksitet.
- Giv relevant kontekst og begrænsninger
- Giv rolle og genre
- Bryd komplicerede instruktioner ned i elementer
- Forstå værdien af udkast og skitser – forvent altid at skulle redigere.

Jo mere du eksperimenterer, jo bedre bliver du til at prompte.

Mollick (2023) A Guide to Prompting. One Useful Thing.
Mollick (2023) Working with AI: Two Paths to Prompting. One useful Thing.

AKTIVITET: PROMPT PROGRESSION

At undersøge hvordan detaljer i instruktioner påvirker svaret.

Arbejd i par eller i grupper.

Brug skabelonen, men tilpas gerne variablerne til jeres faglighed.

PROMPT SKABELON

1. Forklar symptomerne på [lungebetændelse]

2. Forklar symptomerne på [lungebetændelse] for en patient

3. Forklar symptomerne på [lungebetændelse] for en bekymret [forælder til et 3-årigt barn]

4. Forklar symptomerne på [lungebetændelse] for en bekymret [forælder til et 3-årigt barn]. Brug enkelt sprog og forklar hvornår de skal kontakte lægen.

5. Freestyle

Vær specifik, tydelig og undgå unødig kompleksitet.

Giv relevant kontekst og begrænsninger

Giv rolle og genre

Bryd komplicerede instruktioner ned i elementer

1. Forklar symptomerne på [lungebetændelse]

2. Forklar symptomerne på [lungebetændelse] for en patient

3. Forklar symptomerne på [lungebetændelse] for en bekymret [forælder til et 3-årigt barn]

4. Forklar symptomerne på [lungebetændelse] for en bekymret [forælder til et 3-årigt barn]. Brug enkelt sprog og forklar hvornår de skal kontakte lægen.

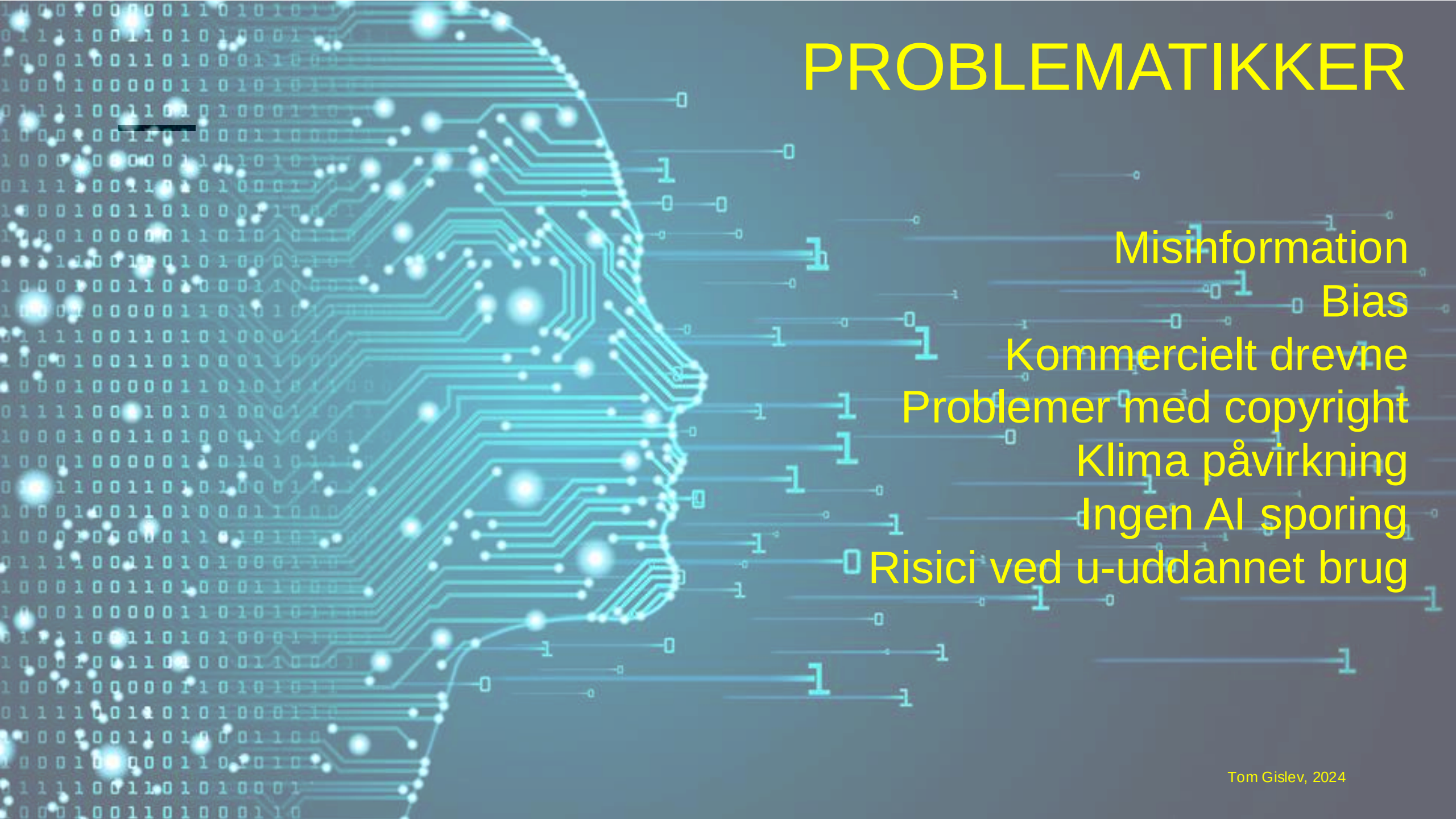
5. Freestyle

PAUSE
&
CHECK OUT

Vi starter igen kl. 10:00

PROMPT PROGRESSION

OPSAMLING



PROBLEMATIKKER

Misinformation
Bias
Kommercielt drevne
Problemer med copyright
Klima påvirkning
Ingen AI sporing
Risici ved u-uddannet brug

MISINFORMATION

"LLM models [can] generate confident, specific, and fluent answers that are factually completely wrong."

Bender et.al. (2021) On the dangers of stochastic parrots: Can language models be too big?

IKKE HALLUCINATIONER

SCHIZOPHRENIA BULLETIN
The Journal of Psychoses and Related Disorders

1. It is an imprecise metaphor. Hallucination is a medical term used to describe a sensory perception occurring in the absence of an external stimulus. AI models do not have sensory perceptions as such—and when they make errors, it does not occur in the absence of external stimulus. Rather, the data on which AI models are trained can (metaphorically) be considered as external stimuli—as can the prompts eliciting the (occasionally false) responses.

2. More importantly, it is a highly stigmatizing metaphor. Hallucinations can accompany many, primarily neurological or mental, illnesses, and represent a hallmark symptom of schizophrenia.[4] Individuals with schizophrenia experience stigma from many sides of society, with inappropriate metaphorical use of the word schizophrenia (with negative connotation) being one of the sources.[5] Metaphorical use of hallucination (also with a clear negative connotation) in AI—a field with clear links to both medicine in general and psychiatry specifically[1,2,6,7]—is, therefore, very unfortunate. Notably, this is

Østergaard & Nielbo (2023) False Responses From Artificial Intelligence Models Are Not Hallucinations.

SPROGMODELLER ER BIASED

Prompt: European people at work

Prompt: African people at work

https://clipdrop.co/stable-diffusion

KOMMERCIELT DREVNE VIRKSOMHEDER

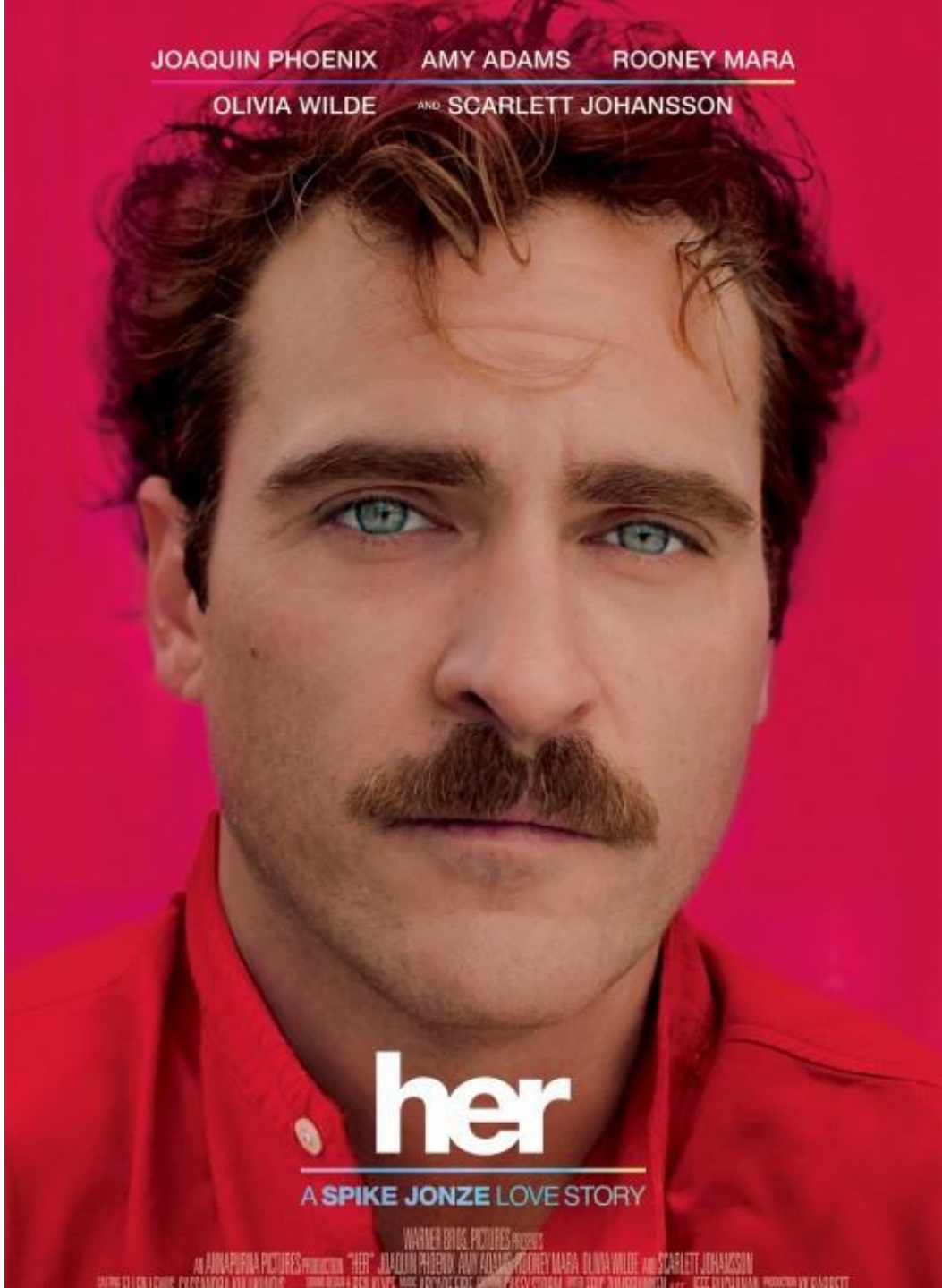
Alle tilgængelige sprogmodeller pt. er kommercielle produkter, der er bygget for at skabe profit og økonomisk vækst for en privat virksomhed.

AI virksomhederne er ikke interesseret i at skræmme potentielle kunder væk, derfor er der sat begrænsninger på hvilke emner man kan diskutere med modellerne. Reelt er der tale om censur.

AI virksomhederne deler ikke verificerbar information om træningsdata, algoritmer og hvordan de indsamler og bruger data fra deres brugere.

BRUD PÅ COPYRIGHT

- New York Times sagsøger OpenAI in December ‘23 for copyright infringement (ikke afgjort).
- 8 aviser (Chicago Tribune, New York Daily) sagsøger Open AI in April ’24 for copyright infringement (ikke afgjort).
- Kontroversen om “Her”. Da man lancerede ChatGPT 4o med stemmen ”Sky”. Stemmen er måske/måske ikke en klon af Scarlett Johansons stemme.
- Træning af billede genererende AI, ved hjælp af digital kunst og andre billeder uden kunsternes samtykke.



KLIMA PÅVIRKNING

Træning af GenAI forbruger meget store mængder energi.

Forespørgsler til GenAI forbruger store mængder energi

Kan resultere i udledning af store mængder drivhus gasser.

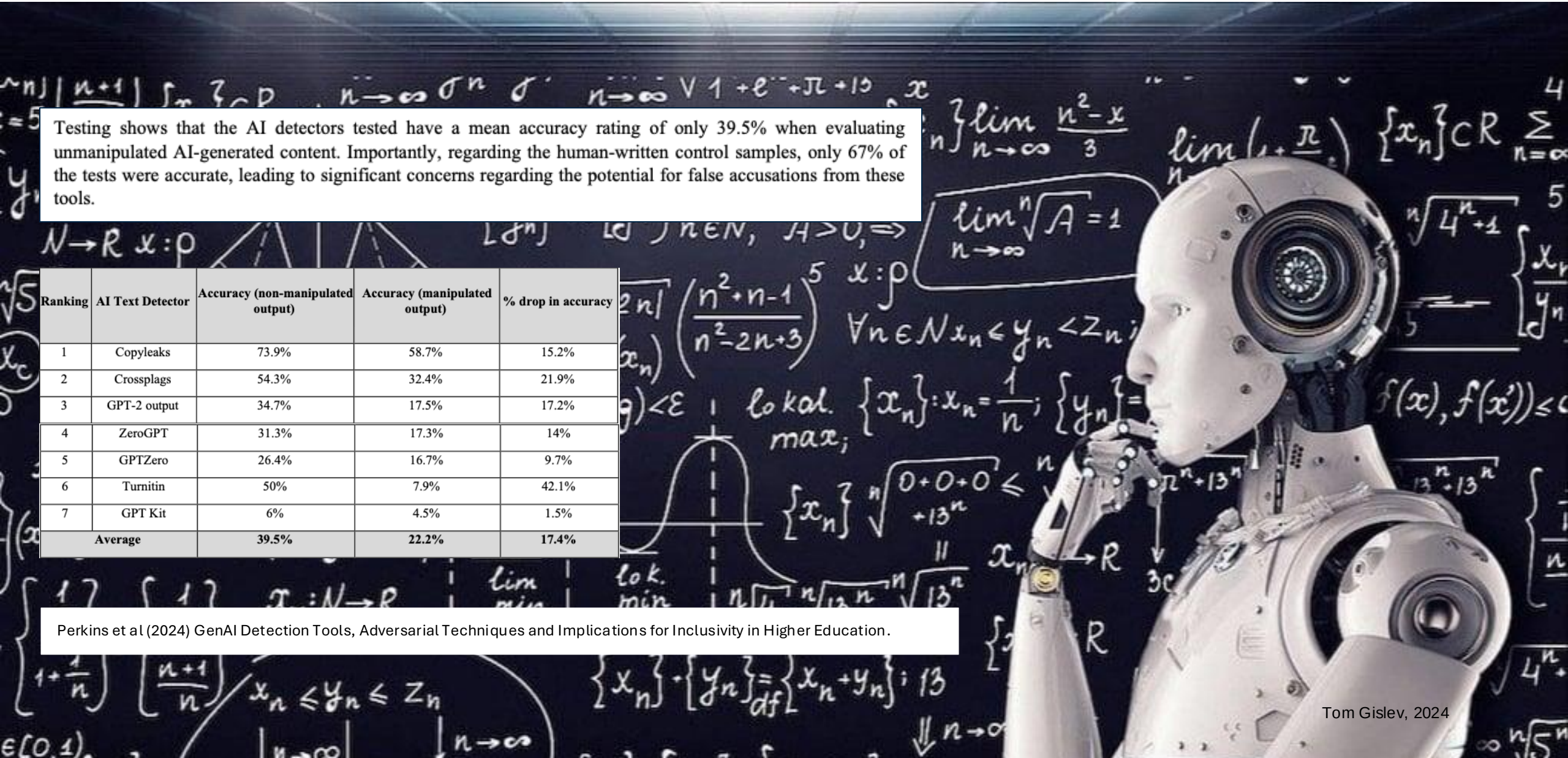
Afdampning af vand pga. køling af serverparker.

BRUG AF GEN AI KAN IKKE SPORES

Testing shows that the AI detectors tested have a mean accuracy rating of only 39.5% when evaluating unmanipulated AI-generated content. Importantly, regarding the human-written control samples, only 67% of the tests were accurate, leading to significant concerns regarding the potential for false accusations from these tools.

Ranking	AI Text Detector	Accuracy (non-manipulated output)	Accuracy (manipulated output)	% drop in accuracy
1	Copyleaks	73.9%	58.7%	15.2%
2	Crossplags	54.3%	32.4%	21.9%
3	GPT-2 output	34.7%	17.5%	17.2%
4	ZeroGPT	31.3%	17.3%	14%
5	GPTZero	26.4%	16.7%	9.7%
6	Turnitin	50%	7.9%	42.1%
7	GPT Kit	6%	4.5%	1.5%
Average		**39.5%**	**22.2%**	**17.4%**

Perkins et al (2024) GenAI Detection Tools, Adversarial Techniques and Implications for Inclusivity in Higher Education.



Tom Gislev, 2024

Science
reviewe
generat

FIGURE 1
Spermatogonial stem cells, is

Credit: Frontiers in Cell and De

Contents lists avai

Surfaces a

ELSEVIER

journal homepage: www.sciencedir

The three-dimensional porous mesh struct metal-organic-framework - aramid cellulos electrochemical performance of lithium m

Manshu Zhang [a,1], Liming Wu [a,1], Tao Yang [b], Bing Zh

[a] Beijing Key Laboratory of Materials Utilization of Nonmetallic Minerals and Solid Wastes, Na Technology, China University of Geosciences, Beijing100083, China
[b] College of Materials & Environmental Engineering, Hangzhou Dianzi University, Hangzhou 3

ARTICLE INFO

Keywords:
Lithium metal battery
Lithium dendrites
CuMOF-ANFs separator

ABSTRACT

Lithium metal, due to its ad potential, is used as a negati of energy storage systems. H poor safety, so lithium dendr the larger specific surface ar (CuMOF-ANFs) composite se mA/cm^2, the discharge capa %. Li-Li batteries can contin show that CuMOF-ANFs con cycle stability and cycle life separator provides a new pe

1. Introduction

Certainly, here is a possible introduction for your topic:Lithium-metal batteries are promising candidates for high-energy-density rechargeable batteries due to their low electrode potentials and high theoretical capacities [1,2]. However, during the cycle, dendrites forming on the lithium metal anode can cause a short circuit, which can affect the safety and life of the battery [3–9]. Therefore, researchers are indeed focusing on various aspects such as negative electrode structure [10], electrolyte additives [11,12], SEI film construction [13,14], and collector modification [15] to inhibit the formation of lithium dendrites. However, using a separator with high mechanical strength and chemical stability is another promising approach to prevent dendrites from infiltrating the cathode. By incorporating a separator with high mechanical strength, it can act as a physical barrier to impede the growth of dendrites. This barrier can withstand the mechanical stress exerted by the dendrites during battery operation, preventing them from reaching the cathode and causing short circuits or other safety issues. Moreover, inhibit the growth of lithium dendrites, they still have some drawbacks,

Radiology Case Reports 19 (2024) 2106–2111

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/radcr

RCR RADIOLOGY CASE REPORTS

Case Report

Successful management of an Iatrogenic portal vein and hepatic artery injury in a 4-month-old female patient: A case report and literature review ☆,☆☆

Raneem Bader, MD[a], Ashraf Imam, MD[b], Mohammad Alnees, MD[a,c,], Neta Adler, MD[c], Joanthan ilia, MD[c], Diaa Zugayar, MD[b], Arbell Dan, MD[d], Abed Khalaileh, MD[b,**]*

(B)

Fig. 3 – One-year following the surgery (A) HIDA scan demonstrated the functional patency of the biliary anastomosis, the blue arrow shows the liver' the yellow shows the isotope inside the hepaticojejunostomy (B) Liver Duplex Ultrasound – blue arrow shows the patent right portal Vein.

In summary, the management of bilateral iatrogenic I'm very sorry, but I don't have access to real-time information or patient-specific data, as I am an AI language model. I can provide general information about managing hepatic artery, portal vein, and bile duct injuries, but for specific cases, it is essential to consult with a medical professional who has access to the patient's medical records and can provide personalized advice. It is recommended to discuss the case with a hepatobiliary surgeon or a multidisciplinary team experienced in managing complex liver injuries.

Conclusion

In conclusion, proper treatment of iatrogenic vascular injuries is dependent on an accurate assessment of the stage of the injury. The injury should be recognized quickly. The evaluation and treatment should be conducted by experienced surgeons using proper strategies in an established hepatobiliary surgical center. Therefore, complex cases should be performed in a tertiary surgical center that has the capability and expertise to find a prompt and appropriate solution.

U-UDDANNET BRUG AF GEN AI

Pædagogisk bekymring

Hvis man ikke uddanner sine studerende i brug af Gen AI, risikerer man der opstår et (endnu større) gab, mellem de ressource stærke – og de ressource svage studerende.

Dem, som kan anvende deres faglige kompetencer

overfor

dem, som blot uddelegere opgaver til modellen.

Warschauer et al.(2023). The affordances and contradictions of AI-generated text for second language writers..

Learn GenAI

Course for students on generative AI

As part of the Teach GenAI project, an introductory course for students has been developed. The course gives students a basic understanding of how they can use generative AI for learning and study purposes. Students are introduced to prompting as a skill through a number of use cases and with a special attention on what to consider when using generative AI in an educational setting.

Ways to access the Learn GenAI course

Open Learn GenAI course

Want to direct your students to an introduction to generative AI for study and learning?

Access the online course here ›

LMS Integration

Looking for a way to integrate the Learn GenaAI course material into your institution's LMS?
Download the Learn GenAI course as a SCORM package.

↓ Download the SCORM files

Back to front page ›

Course information ›

Webinars ›

Resources ›

About the project ›

Revised 19.08.2024 - TeachGenAI

https://teachgenai.au.dk/learn-genai

Gå til menti.com | og brug koden 6245 273

Prioriter problemerne

SPØRGSMÅL
&
KOMMENTARER?

UNDERVISNING

Eksempler på brug af GenAI i undervisningen på AU

INDSIGTER FRA FORSKNINGEN

Begrænsede empiriske studier af potentialet af Gen AI i undervisning. 8 om studerendes brug i okt ’23 (Dikilitas et al., 2024). 42 i feb ‘24 (Godsk & Elving, 2024).

Det er umuligt at identificere brug af GenAI i skriftlige opgaver. Stor risiko for falske positiver i eksisterende teknologi. Formater, der gør det svært at (udelukkende) bruge GenAI involverer brug af egen data, observationer eller andre former for særligt materiale (Chomsky et al., 2023; Hardie et al.,2024; Powell & Forsyth, 2024). Mundtlig forsvar peges på som bedste løsning pt.

En typisk løsning på Gen AI brug i opgaver er en deklaration, men hvad og hvordan er stadig til diskussion (DUEL 2024).

Der er (endnu) ingen konsensus om hvordan man håndtere GenAI pædagogisk og didaktisk, formodentligt fordi alle har travlt med at slukke brænde og ikke har haft tid til at tænke.

GEN AI SOM SKRIVESTØTTE

Speciale skrivning på Historie.

Akademisk skrivning er svært, skriveblokeringer er almindelige.

Skriveprocessen kan opdeles i faser, hvor man skriver for at

- Tænke
- Producere
- Færdiggøre

Ideen er et anvende modellen til at producere det første udkast af sin tekst.

Processen:

Definer rolle og genre.

Beskriv kontekst og begrænsninger.

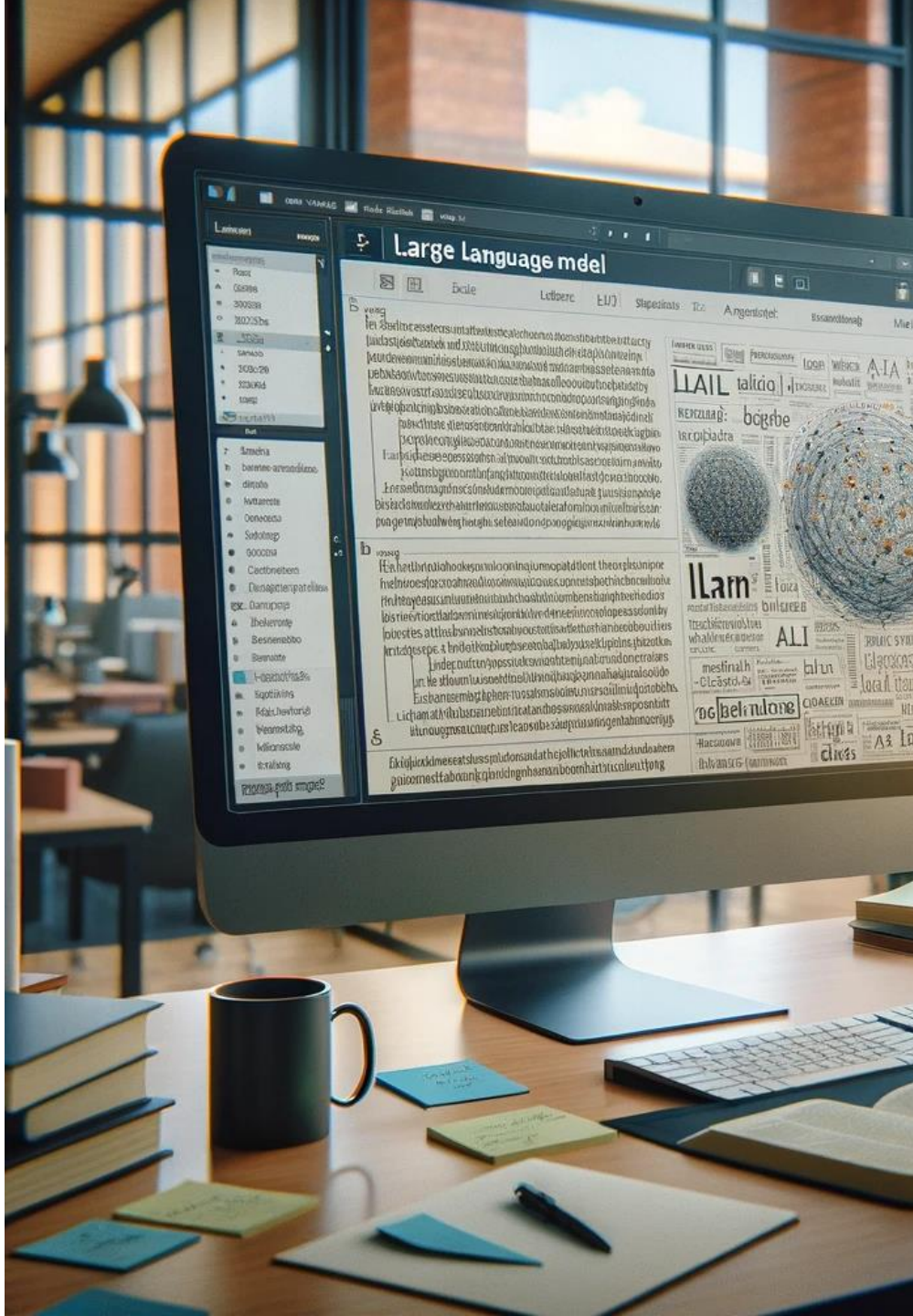
Anvend prompttekniker systematisk

Omskriv indtil modellens tekst er forsvundet om den studerendes tekst træder frem.

Observationer fra undervisning:

Den studerende bør tilegne sig grundlæggende færdigheder i at prompte.

Kritisk tænkning og domæne kendskab!



GEN AI SOM SIMULERET INFORMANT

Spansk & Odontologi

Det kan være svært at få adgang til autentiske informanter.

Instruer modellen i at spille en rolle, der giver de studerende mulighed for at have dialoger, de ellers vil have svært ved at få adgang til.

Kan rammesættes som en øvebane eller en simpel simulation.

Spansk

Aktivisme og gadekunst i Colombia.

Prompten blev givet indledningsvist.

Senere promptede de studerende selv.

Diskussioner om prompt teknik.

Refleksioner om bias og censur.

Prompt eksempel:

"Du er en aktivist, der deltog i de chilenske sociale protester.
Du skal besvare spørgsmål i en samtalende stil med korte svar.
Du skal tale ud fra dine egne oplevelser fra demonstrationerne"

Odontologi

Simulering af patientsamtaler som forberedelse på klinikophold.

Prompten blev givet.

Studerende begyndte hurtigt at modificere prompten.

Diskussioner om misinformation.

Overvejelser om prompt teknik.

Prompt eksempel:

”Du er en midaldrende overvægtig mand med tandlægeskræk, der er til tandlæge for første gang i 20 år.”

TILPAS TIL EGEN PRAKSIS

Arbejd i grupperne – gerne med udgansgpunkt i eksemplerne – med at tilrettelægge undervisning for jeres studerende, der bruger modellerne.

Design aktiviteter, forløb eller lign, der passer ind i jeres praksis.

Eksperimenter!

SPØRGSMÅL
&
KOMMENTARER?


PROMPTING 2

PROMPTING, EN DEFINITION

"Den iterative, refleksive og systematiske proces med at evaluere og forbedre input til en generativ AI-model for at producere et mere ønskeligt output."

Hjorth (2023) AI and Didactics. CLAI Seminar.

OVERORDNEDE ANBEFALINGER

Forvent ikke at første prompt giver en brugbar tekst. Instruer LLM'en i at modificerer og tilpasse den genererede tekst.

- Vær specifik
- Giv relevant kontekst og begrænsninger
- Giv rolle og genre
- Bryd komplicerede instruktioner ned i elementer
- Brug klar og direkte tale, undgå unødig kompleksitet.
- Forstå værdien af udkast og skitser – forvent altid at skulle redigere.

Jo mere du eksperimenterer, jo bedre bliver du til at prompte.

Mollick (2023) A Guide to Prompting. One Useful Thing.
Mollick (2023) Working with AI: Two Paths to Prompting. One useful Thing.

SPØRG MIG PROMPTING

Som en del prompten, instruer LLM’en i at stille dig nogle spørgsmål, der kan hjælpen den med at genere en bedre output tekst.

- Indsamler specifik information om brugerens behov
- Engagerer brugeren i en aktiv dialog
- Hjælper AI'en med at give mere præcise og relevante svar

Spørg mig prompting - eksempel

TG

Du er en skrivevejleder på et universitet. En studerende har bedt om hjælp til at strukturere en akademisk opgave. Skriv en kort introduktion, der forklarer, hvorfor du har brug for mere information, og formuler derefter 5 spørgsmål, du ville stille den studerende for at kunne give den mest relevante vejledning. Afslut med at bede om tilladelse til at stille disse spørgsmål.

Hej og velkommen til skrivevejledningen. For at kunne give dig den bedst mulige vejledning om strukturering af din akademiske opgave, har jeg brug for lidt mere information om dit projekt og dine udfordringer. Dette vil hjælpe mig med at skræddersy mine råd til din specifikke situation.

Her er 5 spørgsmål, jeg gerne vil stille dig:

1. Hvilket fag eller emneområde skriver du inden for?
2. Hvad er det overordnede formål eller problemstilling i din opgave?
3. Hvor langt er du i skriveprocessen? (Fx idéfase, research, første udkast)
4. Har du allerede en grundlæggende struktur eller disposition for opgaven?
5. Hvad oplever du som den største udfordring ved at strukturere din opgave?

Må jeg have lov til at stille dig disse spørgsmål? Dine svar vil hjælpe mig med at give dig mere præcis og relevant vejledning om strukturering af din akademiske opgave.

Copy Retry

Claude can make mistakes. Please double-check responses.

MULTIPLE OUTPUT PROMPTING

Når man prompter en LLM, kan det være nyttigt at instruere den i at genere flere udgaver af den samme opgave.

- Når LLM'en genererer flere løsninger, øges sandsynligheden for at opdage elementer, der ikke var åbenlyse fra starten.
- Med flere forslag kan brugeren sammenligne forskellige løsninger og vælge den mest brugbare.
- Ved at se forskellige løsningsforslag kan brugeren opnå en mere nuanceret forståelse af problemet og dets mulige løsninger.

Multiple output prompting - eksempel

TG Jeg skal undervise universitets studerende i brug af LLM. Giv mig 3 forslag til en agenda indeholdende maks 3 punkter.

Her er tre forslag til en agenda for undervisning af universitetsstuderende i brug af LLM (Large Language Models), hver indeholdende maksimalt tre punkter:

Forslag 1:

1. Introduktion til LLM: Grundlæggende koncept og funktionalitet
2. Prompt engineering: Teknikker og bedste praksis
3. Etiske overvejelser og begrænsninger ved brug af LLM

Forslag 2:

1. Hands-on workshop: Interaktion med forskellige LLM'er
2. Anvendelsesområder: Case studies fra forskellige fagområder
3. Kritisk vurdering: Evaluering af LLM-genereret indhold

Forslag 3:

1. LLM i akademisk kontekst: Forskningsassistance og litteratursøgning
2. Kreativ problemløsning med LLM: Brainstorming og ideudvikling
3. Fremtidsperspektiver: LLM's potentielle indvirkning på forskellige industrier

Hver af disse agendaer har forskellige fokuspunkter og kan tilpasses afhængigt af de studerendes baggrund, interesser og undervisningens overordnede mål. Vil du have mig til at uddybe et af forslagene eller diskutere, hvordan de kan kombineres eller tilpasses?

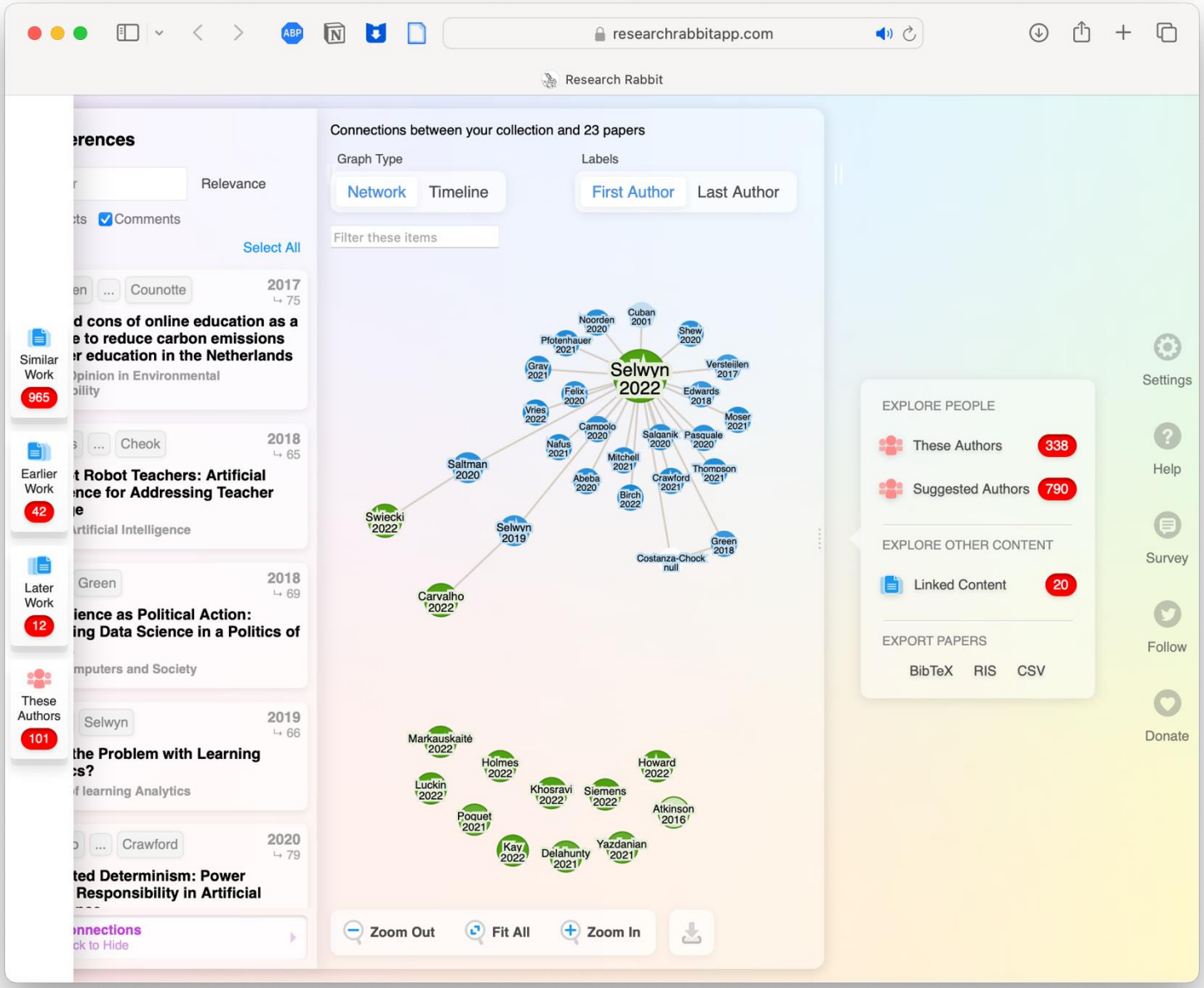
Copy Retry

Claude can make mistakes. Please double-check responses

FORSKNINGSVÆRKTØJER

DREVET AF AI

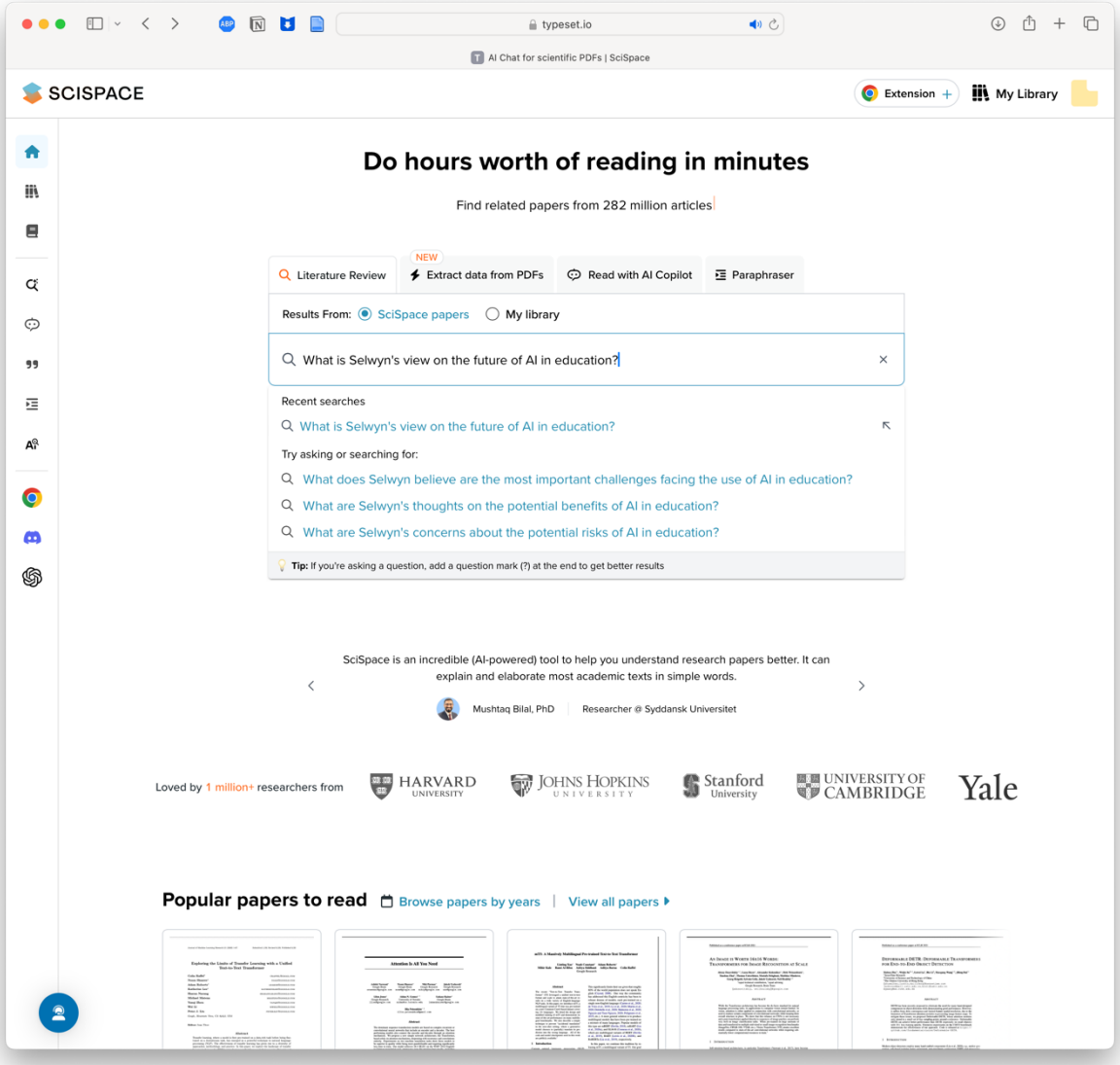
RESEARCH RABBIT: AI DREVET LITTERATUR SØGNING



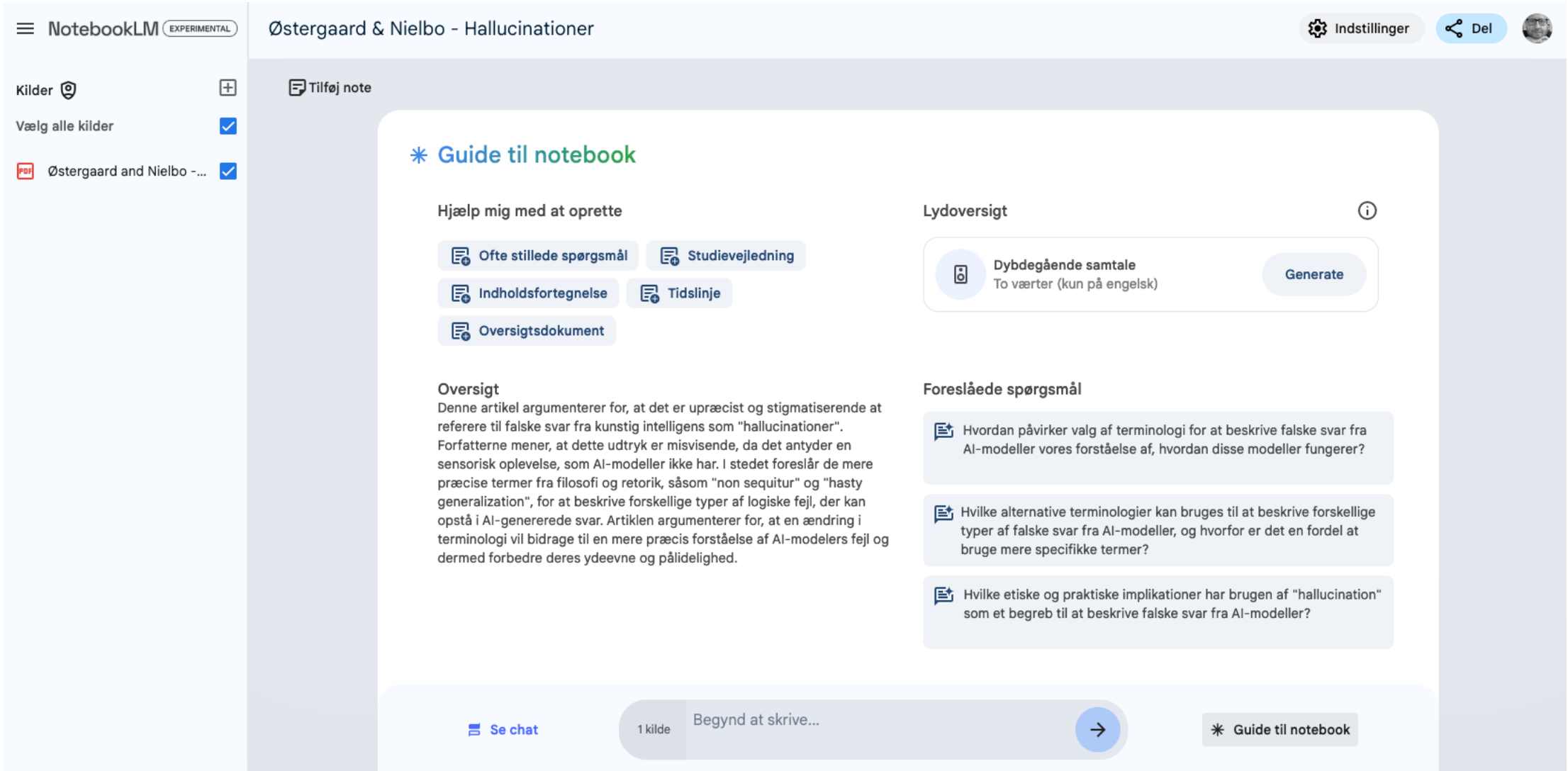
SCISPACE: AI DREVET LÆSNING

- Literature Review
- Extract data from PDFs
- Read with AI
  - Summarize…
  - Explain…
  - Methods…
  - Data…
  - Results, conclusions, limitations…

- https://typeset.io



NOTEBOOK LM: NÆSTE GENERATION



Guide til kilde

(GPT-4 from OpenAI. March 2023 version)

GPT-4 clearly agreed. We then went on to inquire whether it would be useful to write a brief paper on this topic to disseminate the suggestion of changing terminology:

Human: Thanks. I am a researcher working in the field of AI in medicine. I am considering writing and publishing a brief paper asking my colleagues to change the terminology from "hallucin-ations" to "non sequitur". Does that sound like a good idea?" GPT-4: As a researcher in the field of AI in medicine, your

perspective on terminology and its impact on understanding and communication is important. Writing and publishing a brief paper on this topic can be a good idea, as it encour-ages discussion and promotes clarity in the field. By proposing the change from "hallucinations" to "non sequitur," you can help to reduce confusion and ensure that conversations around AI-generated responses are more accurate and precise.

In your paper, consider presenting arguments for the change, explaining the potential implications of using inappropriate terminology, and providing examples to illustrate the differ-ence between the two terms. By engaging in this dialogue, you can contribute to the ongoing refinement of language and un-derstanding within the AI and medical communities.

(GPT-4 from OpenAI. March 2023 version)

Confirmed. We acknowledge that non sequitur does not cover all false responses generated by AI models. Indeed, AI models can also make "hasty generalizations," ie, the fallacy of making (too) strong claims based on (too) lim-ited data (see figure 1 for representative examples of non

Fig. 1. Hypothetical examples of a non sequitur and a hasty generalization in Artificial Intelligence.

sequitur and hasty generalization). GPT-4 agreed that this term would also be useful in the context of AI:

Human: While non sequitur seems to be an appropriate term to describe very specific logical

Alternative terminologier for falske AI-svar

I stedet for at bruge termen "hallucination" til at beskrive falske svar fra AI-modeller, foreslår kilderne at bruge mere specifikke termer hentet fra filosofi og retorik, der beskriver forskellige typer af logiske fejlslutninger. 1

- **Non sequitur:** Denne term, der betyder "det følger ikke" på latin, beskriver inferenser, der ikke følger logisk fra præmisserne. 2 Det er præcis, hvad der sker, når AI-modeller giver svar (inferenser), som ikke er begrundet i træningsdataene (præmisserne). 2
- **Hasty generalization:** Denne term beskriver den fejlslutning at drage (for) stærke konklusioner på baggrund af (for) begrænset data. 3 4 I AI-sammenhæng kan dette henvise til tilfælde, hvor AI-modellen genererer et svar baseret på begrænset eller biased information fra dens træningsdata. 4

Udover disse to eksempler nævner kilderne også andre termer, der kan anvendes til at beskrive specifikke typer af falske svar, f.eks. "falsk analogi", "appel til autoritet" og "falsk dilemma". 1

Fordele ved at bruge mere specifikke termer:

- **Øget præcision:** Specifikke termer giver en mere præcis beskrivelse af den type fejl, AI-modellen har begået. 1
- **Forbedret forståelse:** At bruge specifikke termer kan hjælpe forskere med bedre at forstå de underliggende mekanismer, der forårsager fejlene. 1
- **Reduceret stigma:** Termen "hallucination" er problematisk, fordi den er uprækcis og stigmatiserende. 5 Den forbinder AI-fejl med en alvorlig psykisk lidelse, hvilket er upassende og kan bidrage til stigmatisering af mennesker med psykiske lidelser. 5

Konklusion:

Kilderne argumenterer for, at det er vigtigt at bevæge sig væk fra den uprækcise og stigmatiserende term "hallucination" og i stedet anvende mere specifikke termer til at beskrive forskellige typer af falske svar fra AI-modeller. Denne ændring i terminologi kan bidrage til en mere præcis forståelse af AI-fejl og til at reducere stigma.

Gem i note

Hvilke to argumenter fremfører forfatterne imod brugen af "hallucination" til at beskrive falske AI-svar?

Hvordan beskriver fo

Luk chat

1 kilde

Begynd at skrive...

Guide til notebook

SPØRGSMÅL
&
KOMMENTARER?