Professor, dr.philos Rune Johan Krumsvik
Universitetet i Bergen
Professor II, Høgskulen i Volda

AI & Gen Z

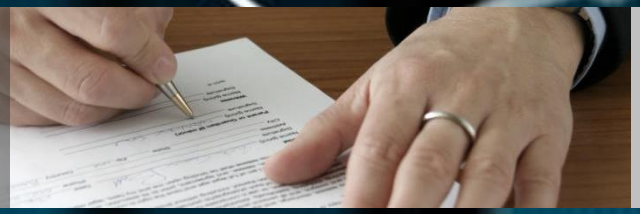
Rådgjevarsamling
17. mars 2026
Solstrand

Andre aktuelle bøker

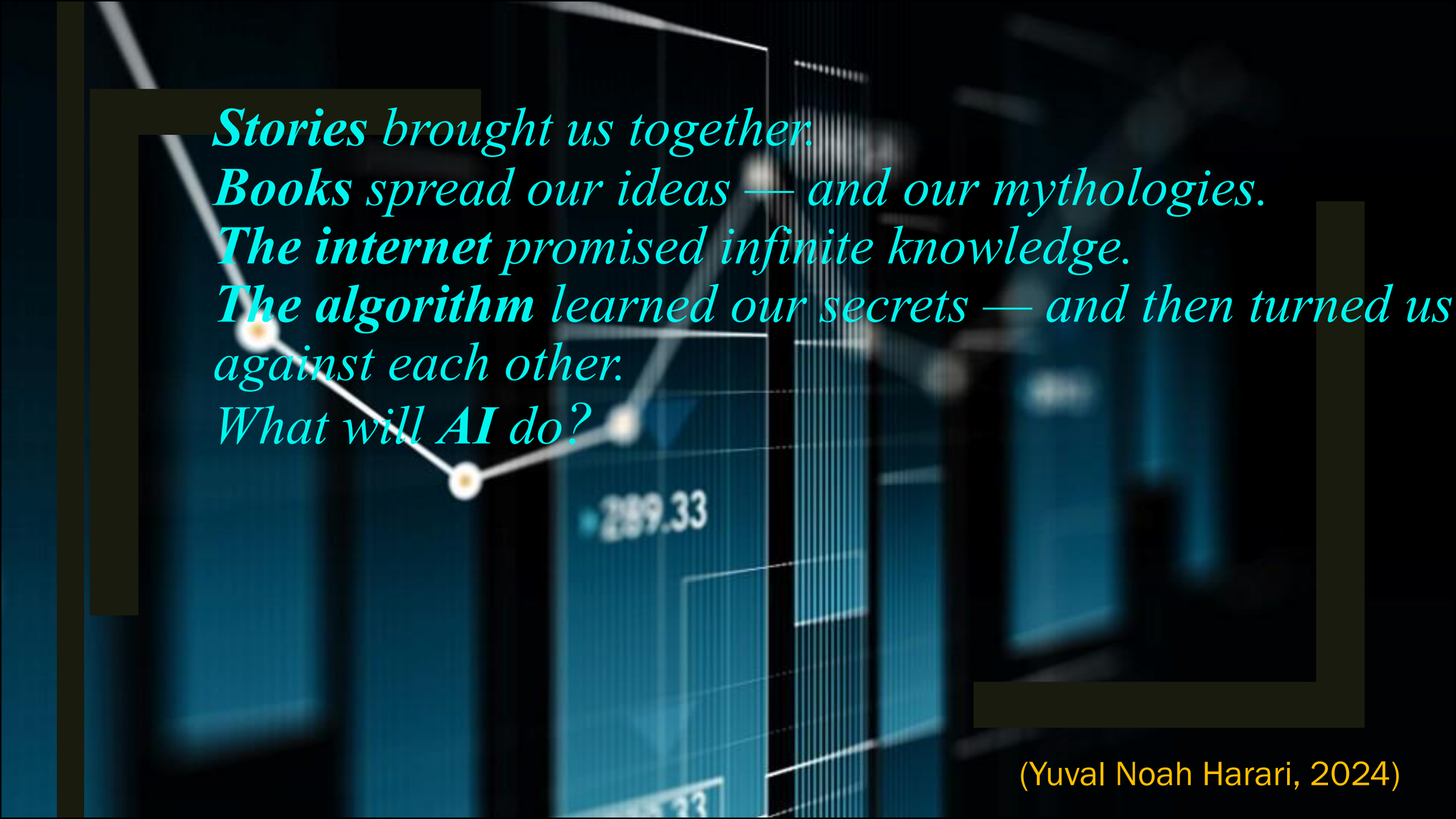
DISPOSISJON

1. 
   Viktigheten av å teste kapabiliteten og påliteligheten til generativ KI innen utdanning

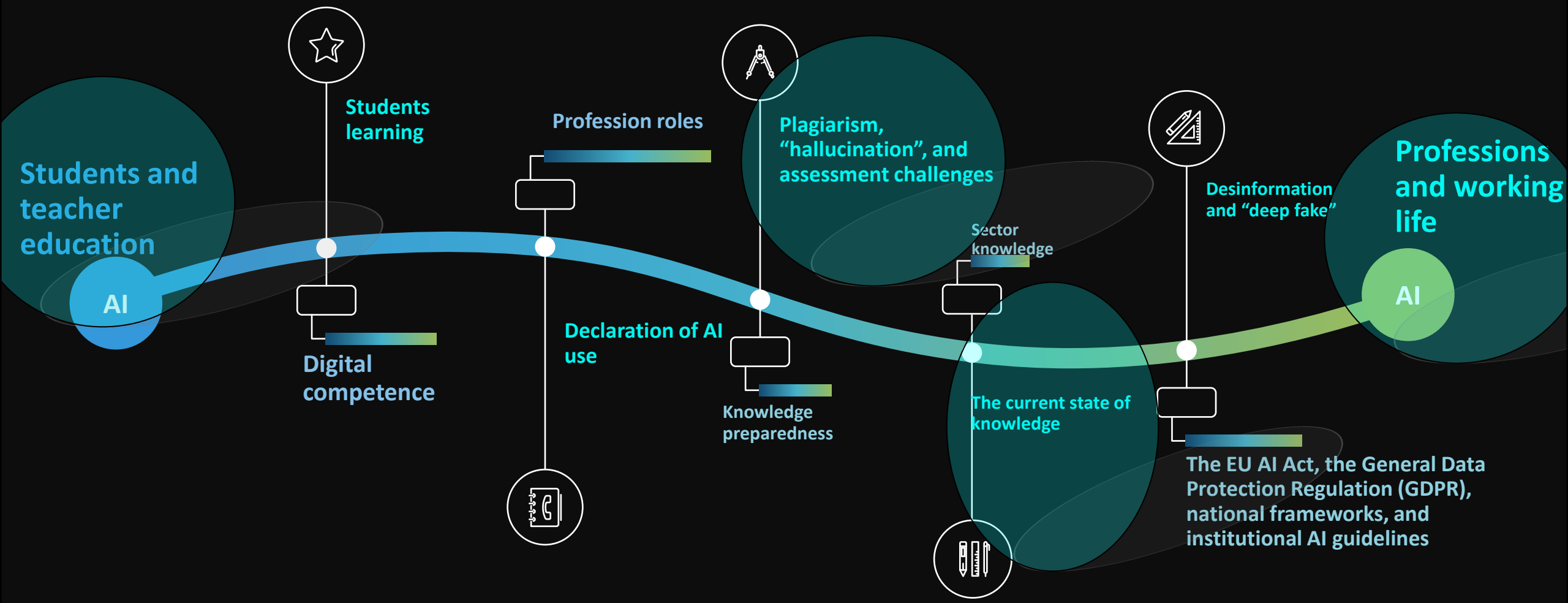
2. 
   Tverrfaglig og tverrsektorielt samarbeid: Skolen kan lære av høyere utdanning – og omvendt

3. 
   Hva har KI å si for rådgiverens hverdag? Hva har KI å si for rådgivereres relasjon til Gen Z?

4. 
   Hvilke muligheter og utfordringer står vi overfor – både pedagogisk, etisk og praktisk?

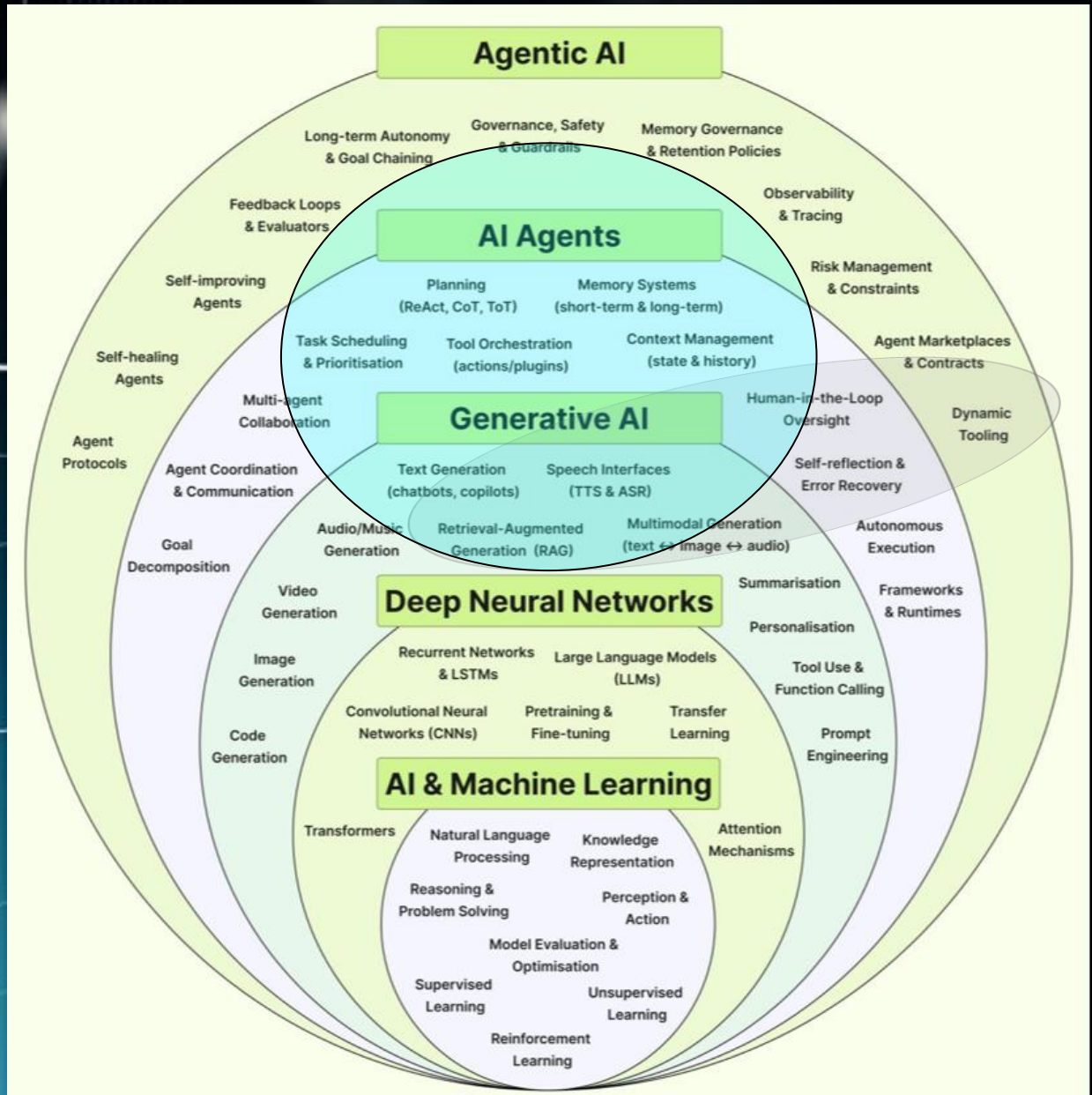

Stories *brought us together.*
Books *spread our ideas — and our mythologies.*
The internet *promised infinite knowledge.*
The algorithm *learned our secrets — and then turned us against each other.*
*What will **AI** do?*
(Yuval Noah Harari, 2024)

Ethical “minefield” and transparency? From the integration of AI in education to its integration in professions and working life



Technological determinism?

- Naive technological determinism, dichotomism, or technophobia?



Et lite tilbakeblikk

-Mange muligheter, men digitale distraksjoner, multitasking, etc. forringer læringsutbyttet
-Vedvarende «Matteus-effekt»? God klasseledelse kan redusere dette

- Nødvendig med en vifte av vurderingsformer og kognitive læringsstrategier

- Manglende digital-didaktisk repertoar hos lærere kan være en utfordring

KI og skoleorganisasjonen

EU-regulativ for KI,GDPR, nasjonale retningslinjer, lokale KI-retningslinjer

Makronivå

Skoleeier/skole ledernivå

Lærernivå

Mesonivå

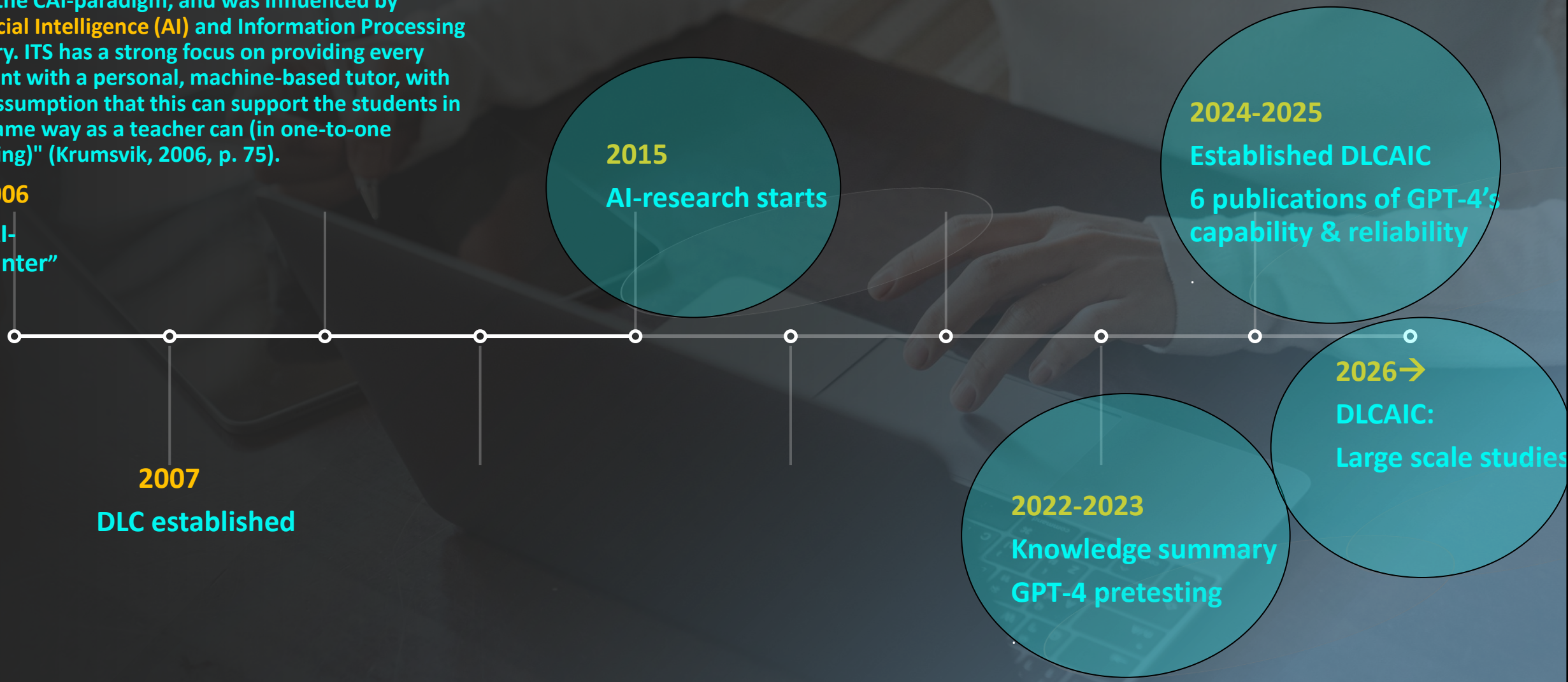
Mikronivå

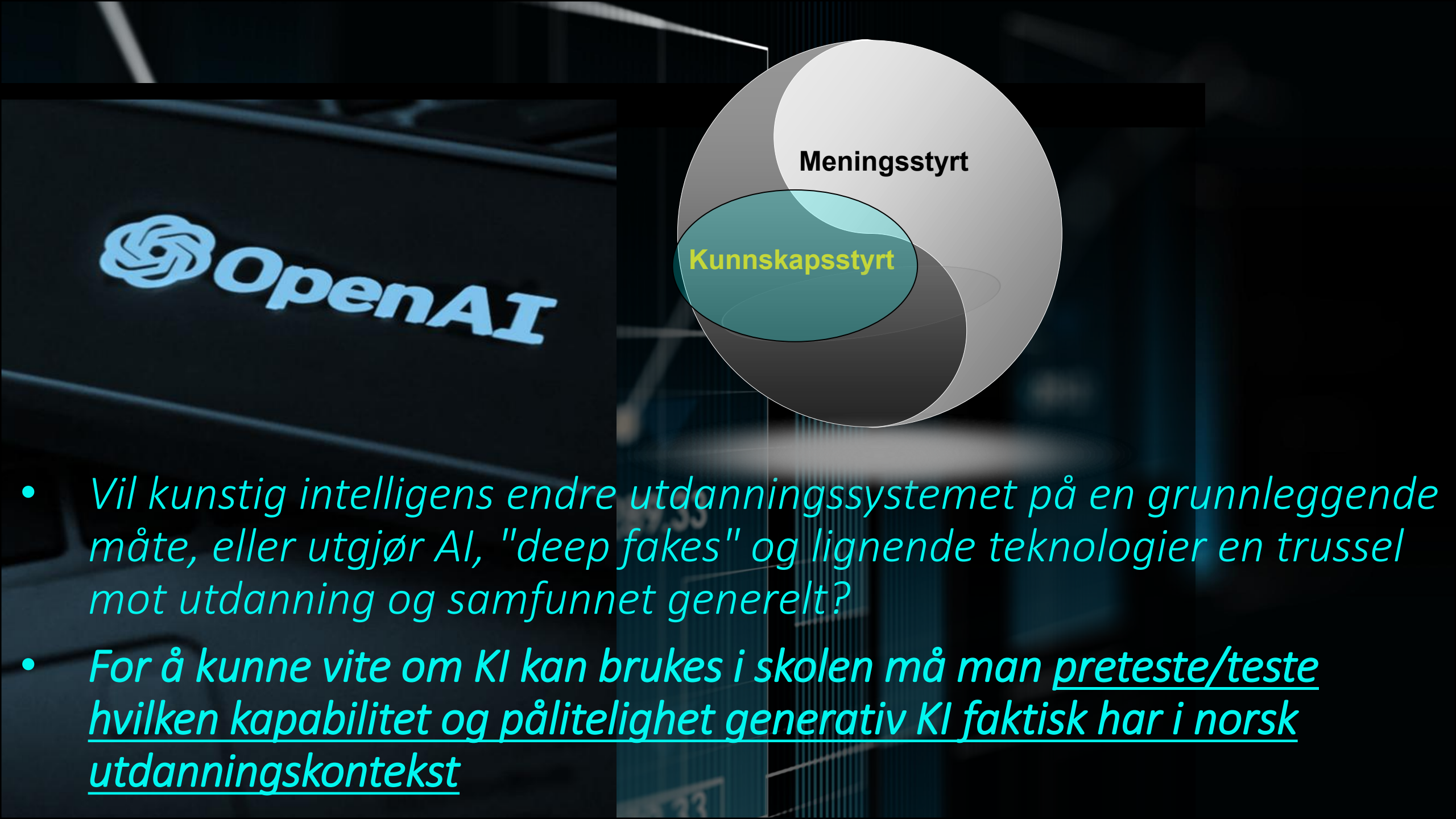
Elevnivå

Arena of formulization → Arena of realization

Digital Learning Communities Artificial Intelligence Centre (DLCAIC)

"The next instructional paradigm was the Intelligent Tutoring System […] which had a clear similarity with the CAI-paradigm, and was influenced by Artificial Intelligence (AI) and Information Processing Theory. ITS has a strong focus on providing every student with a personal, machine-based tutor, with the assumption that this can support the students in the same way as a teacher can (in one-to-one tutoring)" (Krumsvik, 2006, p. 75).





- *Vil kunstig intelligens endre utdanningssystemet på en grunnleggende måte, eller utgjør AI, "deep fakes" og lignende teknologier en trussel mot utdanning og samfunnet generelt?*
- ***For å kunne vite om KI kan brukes i skolen må man preteste/teste hvilken kapabilitet og pålitelighet generativ KI faktisk har i norsk utdanningskontekst***

Meld. St. 34 2023–2024. En mer praktisk skole (5.- 10. klasse)

- Kunstig intelligens er nevnt 25 plasser i stortingmeldingen
- Naturlig språkprosessering handler om å utvikle datamaskiner som kan forstå, tolke og generere menneskelig språk. Det inkluderer oppgaver som maskinoversettelse, tekstananlyse, talegjenkjenning og samtaleroboter (chatbot).
- Generativ kunstig intelligens er kunstig intelligens som kan generere nytt innhold, framfor å kun analysere eller handle basert på eksisterende data. Med skapende KI beveger en seg over i en æra med maskinskapt virkelighet.
- Deepfakes er en måte å bruke kunstig intelligens for å endre på video- og lydmateriale. Stemmer og ansikter ser ut som kjente mennnesker i kjente omgivelser, men de er manipulert, og kan være vanskelig å avsløre (KD, 2024, s. 96)

Utviklingen av generativ KI går raskt (Claude, Gemini advanced, Open AI, etc.)

KI, digital kompetanse, læremidler og ZPD for Gen Z

7. KI-støttet leksehjelp (Claude, Gemini Advanced, GPT-5): ChatGPT - DLCAICs sparringpartner om inkludering og KI

6. Leksehjelp med foreldre/foresatte

1. Læreres undervisning

Eleven

5. Leksehjelp på skolen

2. Rådgiveres veiledning

4. Multimodale digitale læremidler (universell utforming)

3. Lærebøker

OECD

TALIS

Results from TALIS 2024

The State of Teaching

roles

Plagiarism, "hallucination", and assessment challenges

Desinformation and "deep fake"

Professions and working life

AI

Sector knowledge

n of AI

Knowledge preparedness

The current state of knowledge

The EU AI Act, the General Data Protection Regulation (GDPR), national frameworks, and institutional AI guidelines

OECD

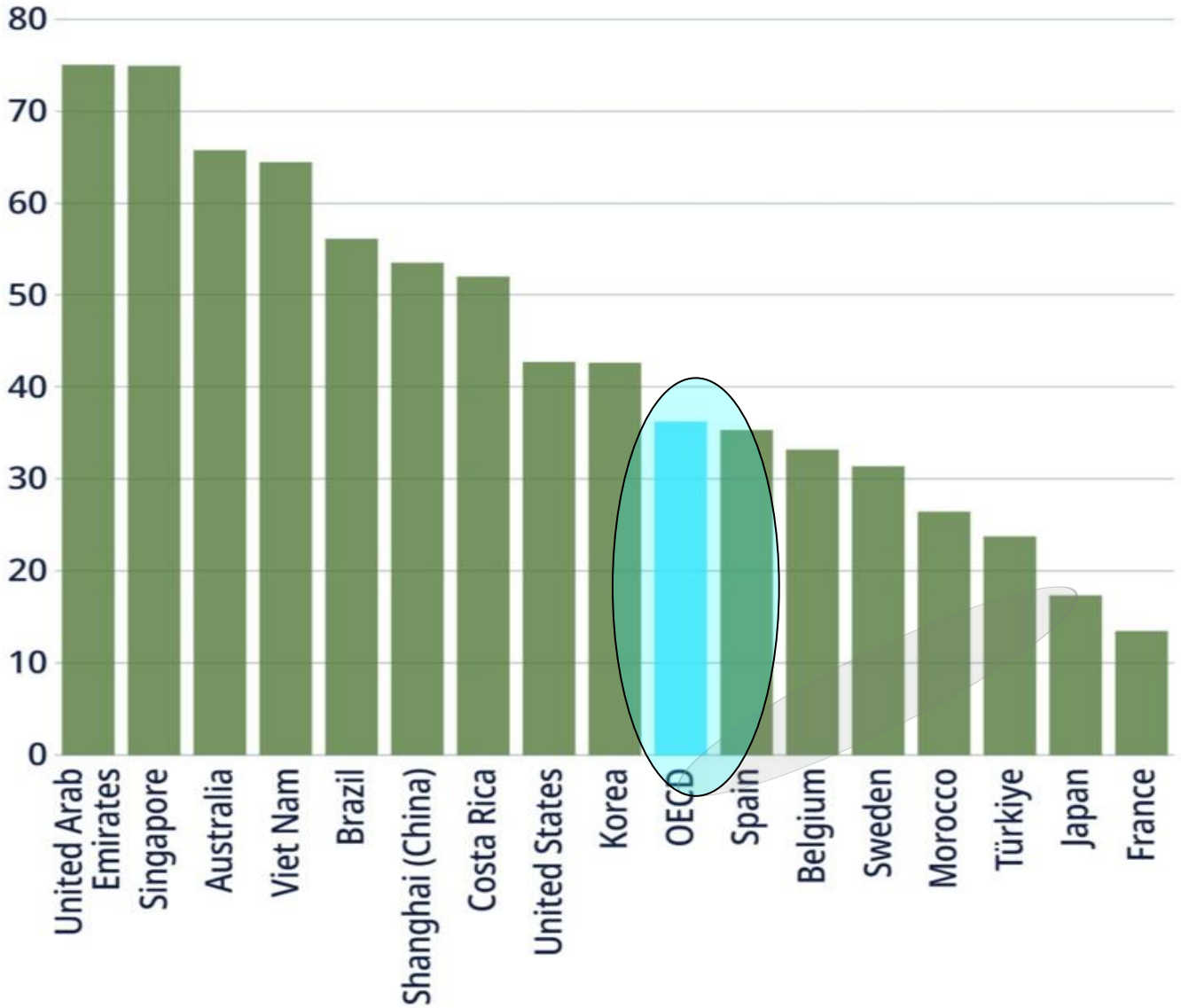
TALIS

Results from TALIS 2024

The State of Teaching

Teachers' use of artificial intelligence

% of lower-secondary teachers who report using AI in the last year, 2024



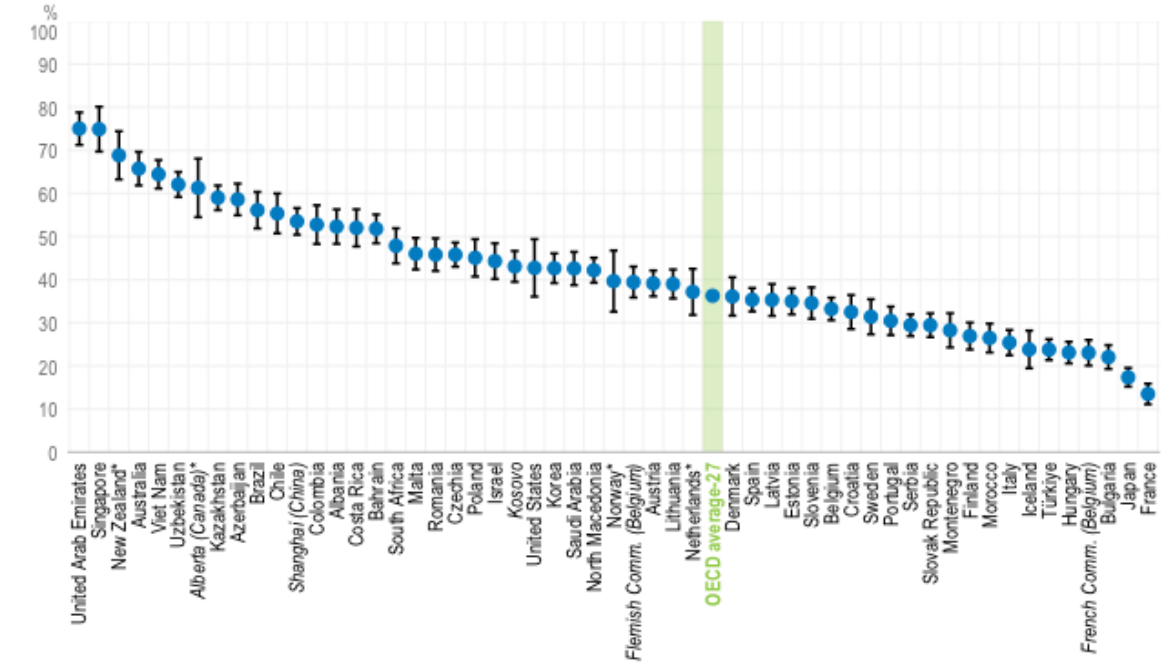
OECD

TALIS

Results from TALIS 2024

The State of Teaching

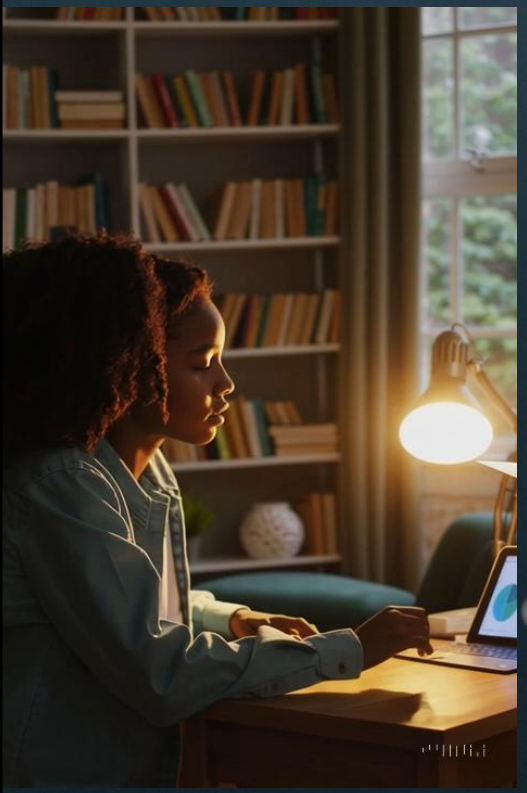
95% Confidence intervals are shown as **error bars**



Note: *Estimates should be interpreted with caution due to higher risk of non-response bias.
Source: OECD, *TALIS 2024 Database*, Table 1.59.

The EU AI Act, the General D
Protection Regulation (GDPR
national frameworks, and
institutional AI guidelines

Hvordan vil KI påvirke utdanning, inkludering, tilpasset opplæring etc.?



Profesjonsfagene

Profesjonsrollene

Undervisningen

Læringen, metakognisjon og læringsledelse

Inkludering

Tilpasset opplæring

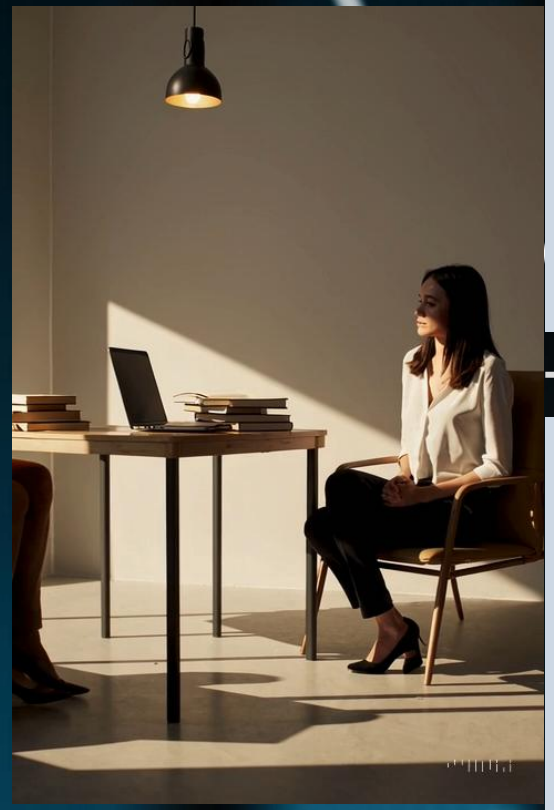
Veiledningen

Kunstig intelligens

KI og psykologisk trygghet i en KI-brytningstid

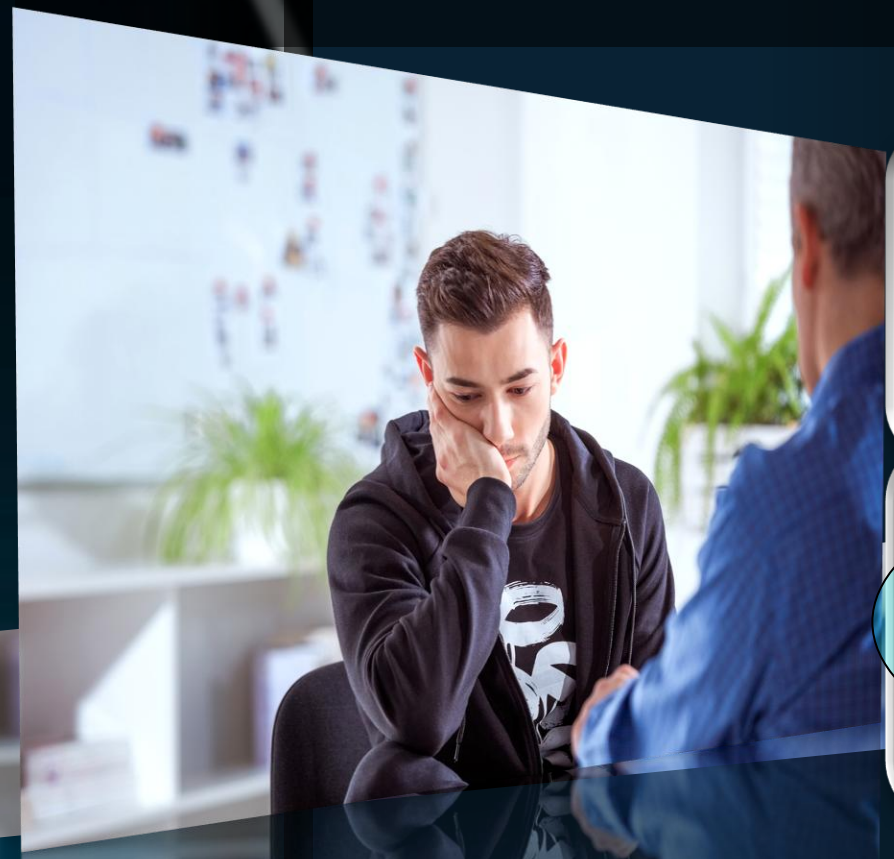
«Psykologisk trygghet beskriver opplevde egenskaper ved et sosialt klima, særlig i et jobbteam eller en organisasjon. Psykologisk trygghet innebærer at det mellom medlemmene i en gruppe er aksept for å ytre seg, stille spørsmål og gjøre feil (Edmonson 2018, s. 18).

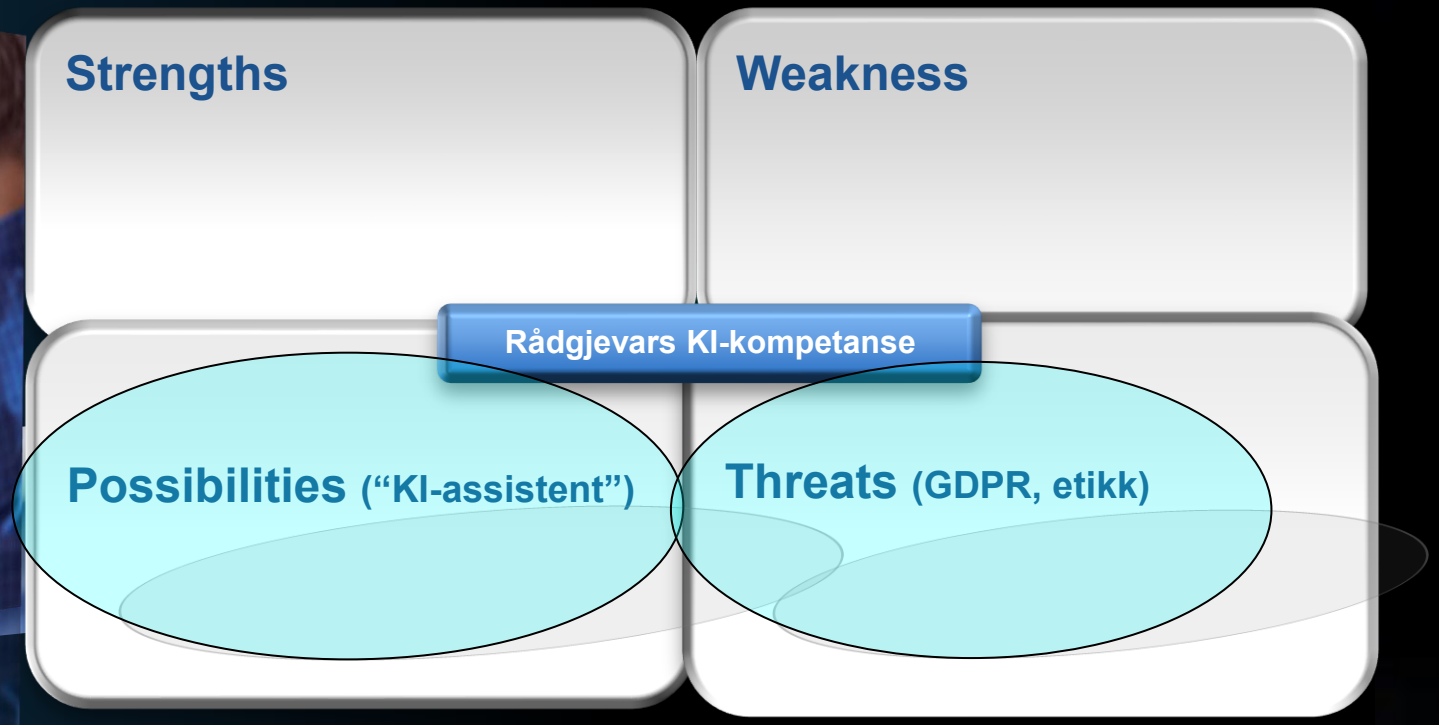
Rådgjevar: profesjonsforståelse og profesjonsrolle i en KI-brytningstid



Rådgjevar, lærar, etc.	WHO AM I ?	HOW AM I?	Psykologisk trygghet (Edmondson, 2018)
“Early adopter”, Rettleiar, mentor, etc.	WHAT DO I OFFER?	HOW DO I CONTRIBUTE AND WHAT KIND OF HELP DO I NEED?	“Hands on” og “Heads on”

Naudsynt med eit steg attende (Skjervheim, 1978) i denne KI-brytingstida?





KI: nye muligheter for læring, underveisvurdering, ZPD, sparringpartner, etc. eller bare en "hype"?

Undervisning

Veiledning/rådgiving

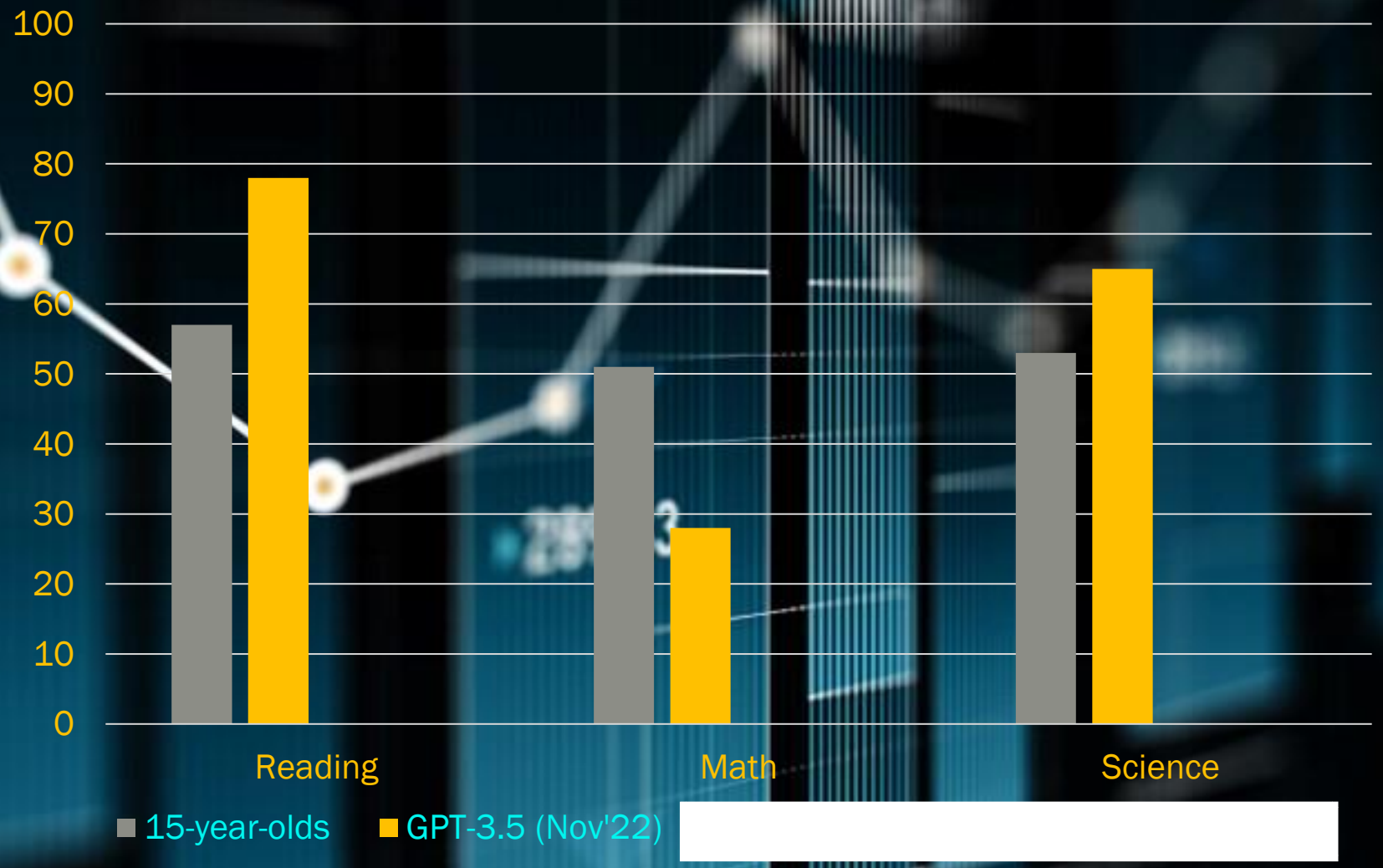
GPT-4, Claude, etc.

Elever


Veiledning og rådgiving når terrenget
erndrer seg. Innsikt fra forskningen?

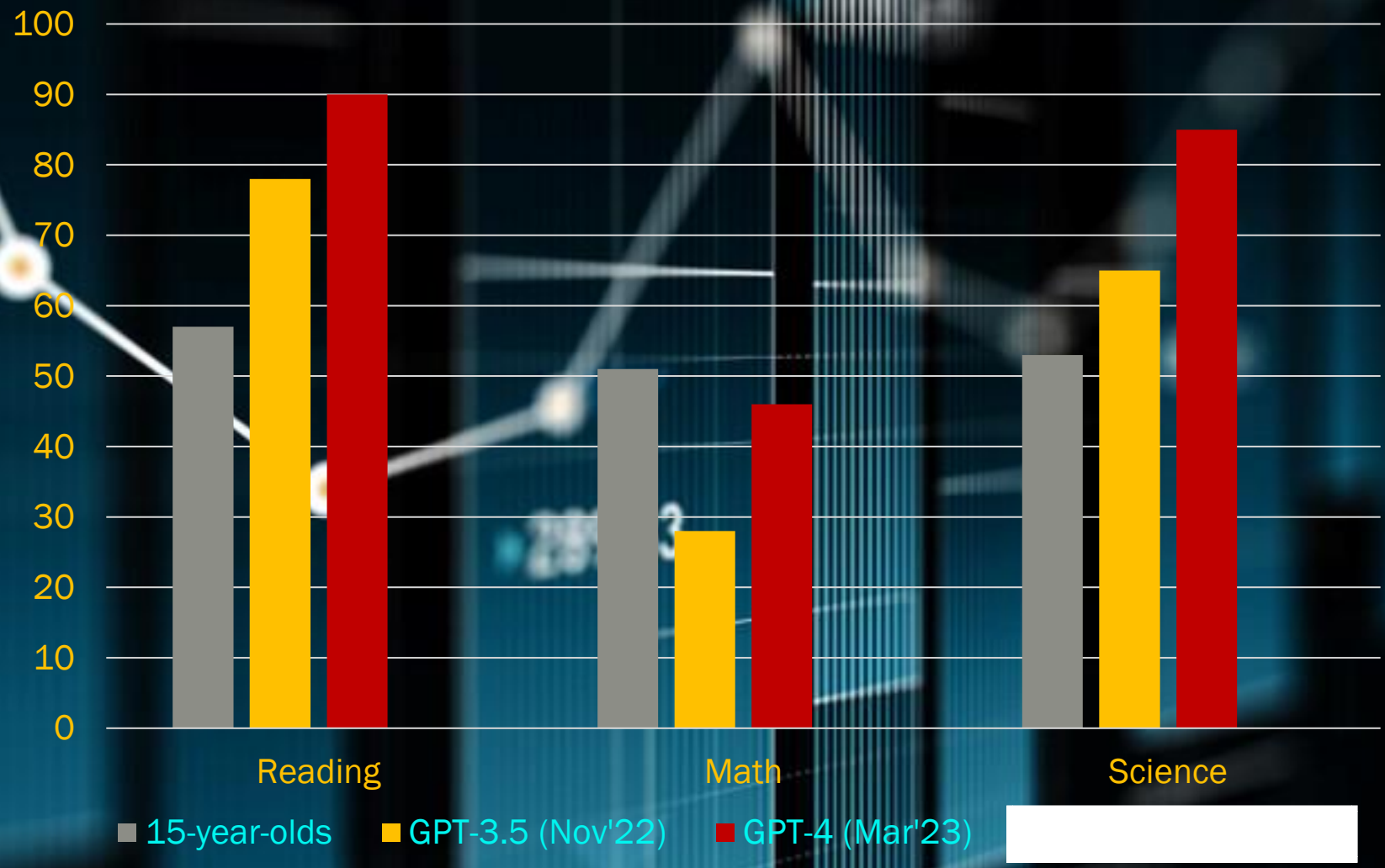

“What is going on under the radar...?”

Example: ChatGPT taking PISA – some AI-evolutionary aspects



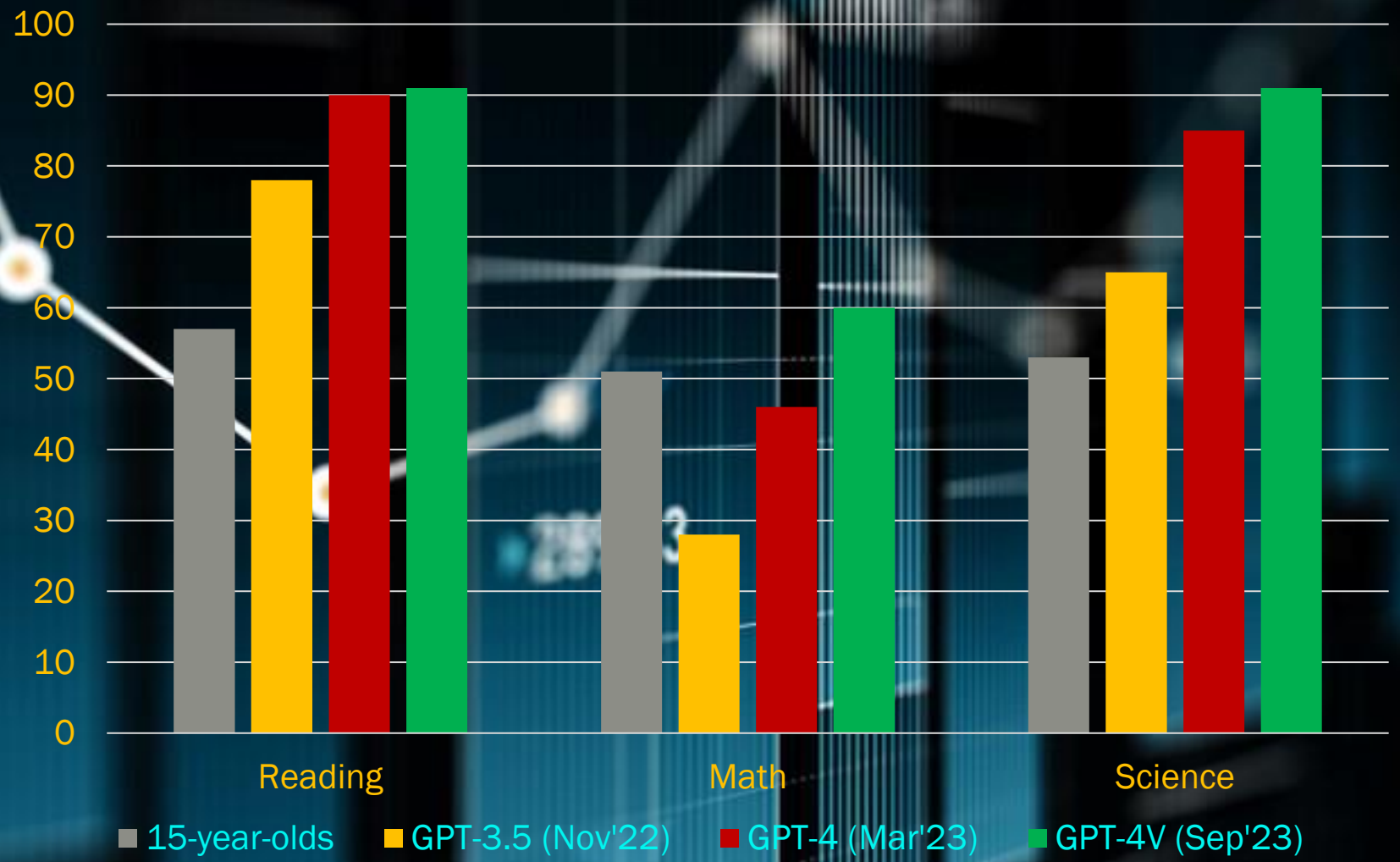
(OECD, 2023)

ChatGPT taking PISA



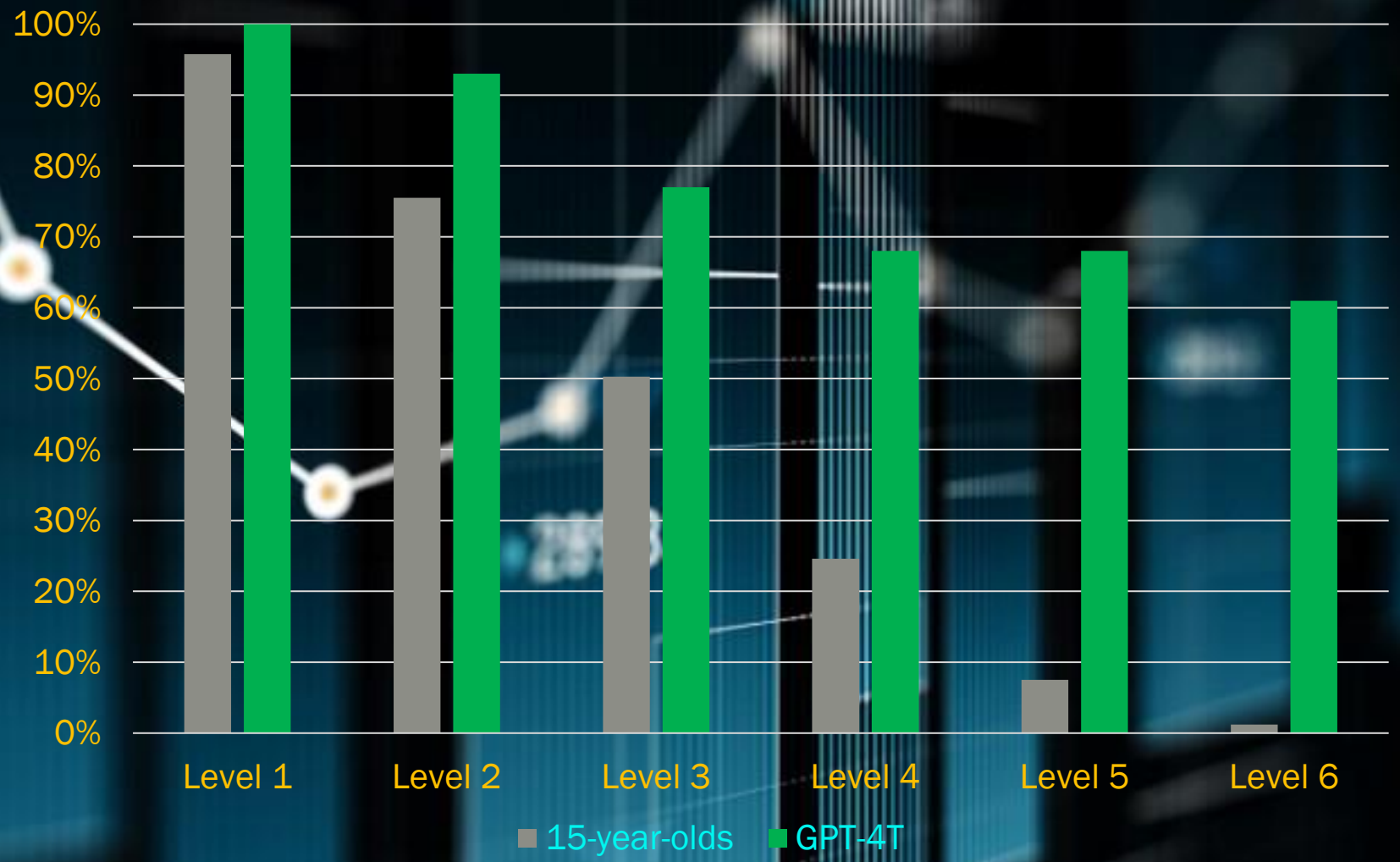
(OECD, 2023)

ChatGPT taking PISA



(OECD, 2023)

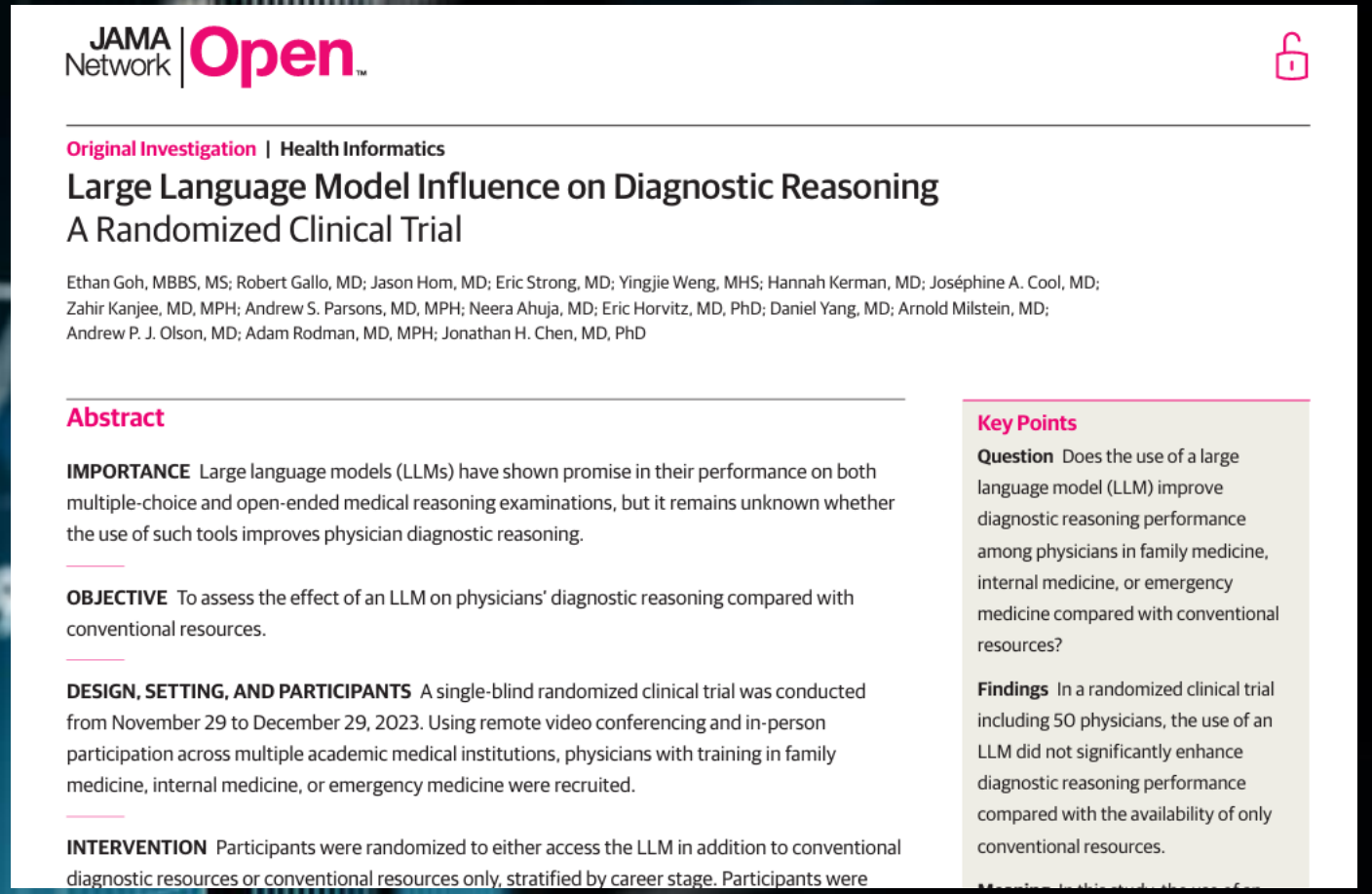
ChatGPT taking PISA Science: By Level, Secure Questions



(OECD, 2023)

The contours of “symbiotic intelligence“?

- **CONCLUSIONS**
  In this trial, the availability of an LLM to physicians as a diagnostic aid did not significantly improve clinical reasoning compared with conventional resources.
- The LLM alone demonstrated higher performance than both physician groups, indicating the need for technology and workforce development to realize the potential of physician-artificial intelligence collaboration in clinical practice.



JAMA Network | Open

Original Investigation | Health Informatics

Large Language Model Influence on Diagnostic Reasoning
A Randomized Clinical Trial

Ethan Goh, MBBS, MS; Robert Gallo, MD; Jason Hom, MD; Eric Strong, MD; Yingjie Weng, MHS; Hannah Kerman, MD; Joséphine A. Cool, MD; Zahir Kanjee, MD, MPH; Andrew S. Parsons, MD, MPH; Neera Ahuja, MD; Eric Horvitz, MD, PhD; Daniel Yang, MD; Arnold Milstein, MD; Andrew P. J. Olson, MD; Adam Rodman, MD, MPH; Jonathan H. Chen, MD, PhD

Abstract

IMPORTANCE Large language models (LLMs) have shown promise in their performance on both multiple-choice and open-ended medical reasoning examinations, but it remains unknown whether the use of such tools improves physician diagnostic reasoning.

OBJECTIVE To assess the effect of an LLM on physicians' diagnostic reasoning compared with conventional resources.

DESIGN, SETTING, AND PARTICIPANTS A single-blind randomized clinical trial was conducted from November 29 to December 29, 2023. Using remote video conferencing and in-person participation across multiple academic medical institutions, physicians with training in family medicine, internal medicine, or emergency medicine were recruited.

INTERVENTION Participants were randomized to either access the LLM in addition to conventional diagnostic resources or conventional resources only, stratified by career stage. Participants were

Key Points

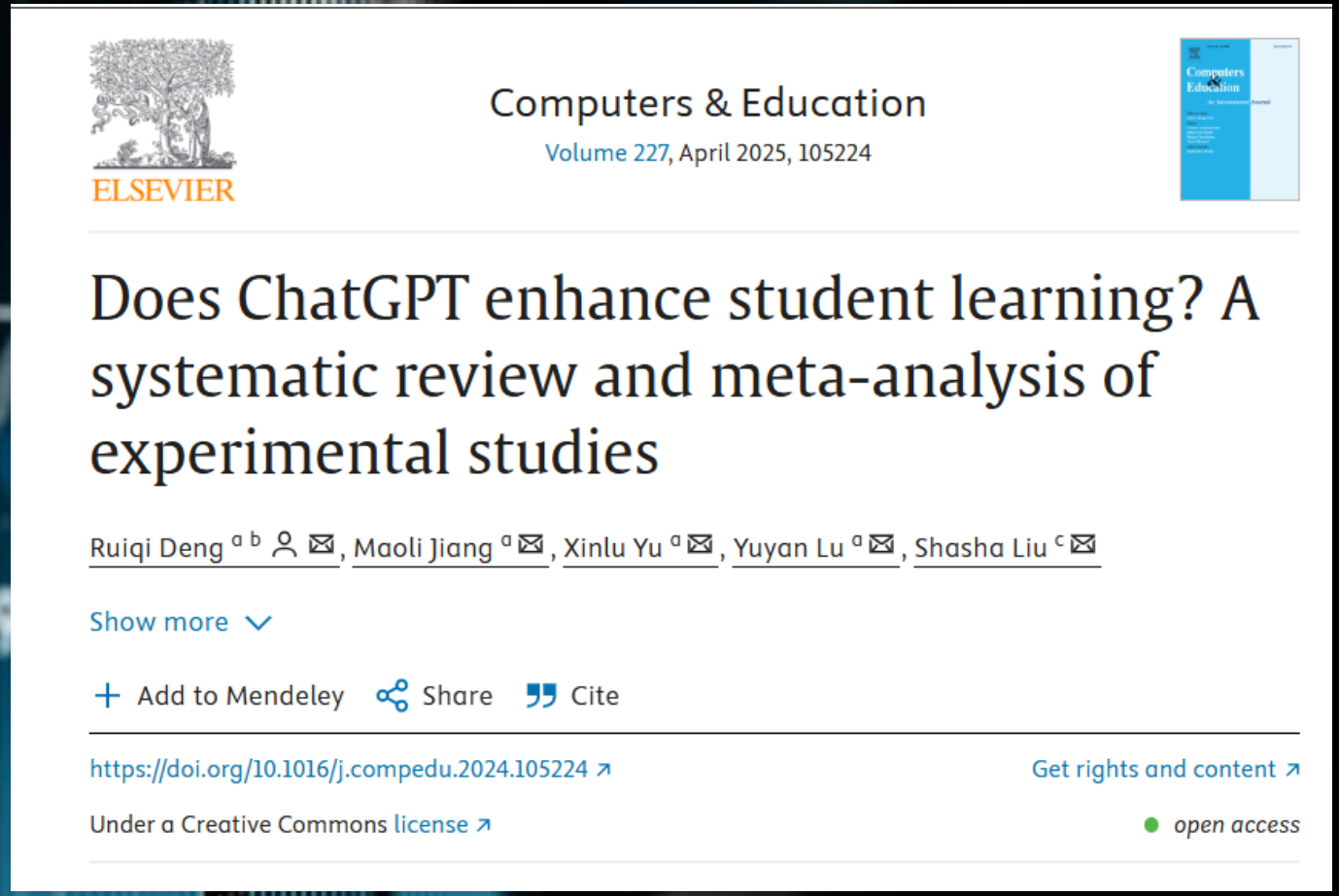
Question Does the use of a large language model (LLM) improve diagnostic reasoning performance among physicians in family medicine, internal medicine, or emergency medicine compared with conventional resources?

Findings In a randomized clinical trial including 50 physicians, the use of an LLM did not significantly enhance diagnostic reasoning performance compared with the availability of only conventional resources.

Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A., & Chen, J. H. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA Network Open, 7*(10), e2440969. https://doi.org/10.1001/jamanetworkopen.2024.40969

The contours of “symbiotic intelligence“?

- Highlights
  - *ChatGPT enhances academic performance.*
  - *ChatGPT boosts affective-motivational states*
  - *ChatGPT improves higher-order thinking propensities*
  - *ChatGPT reduces mental effort.*
  - *ChatGPT does not influence self-efficacy*



ELSEVIER

Computers & Education

Volume 227, April 2025, 105224

Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies

Ruiqi Deng [a b], Maoli Jiang [a], Xinlu Yu [a], Yuyan Lu [a], Shasha Liu [c]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.compedu.2024.105224

Get rights and content

Under a Creative Commons license

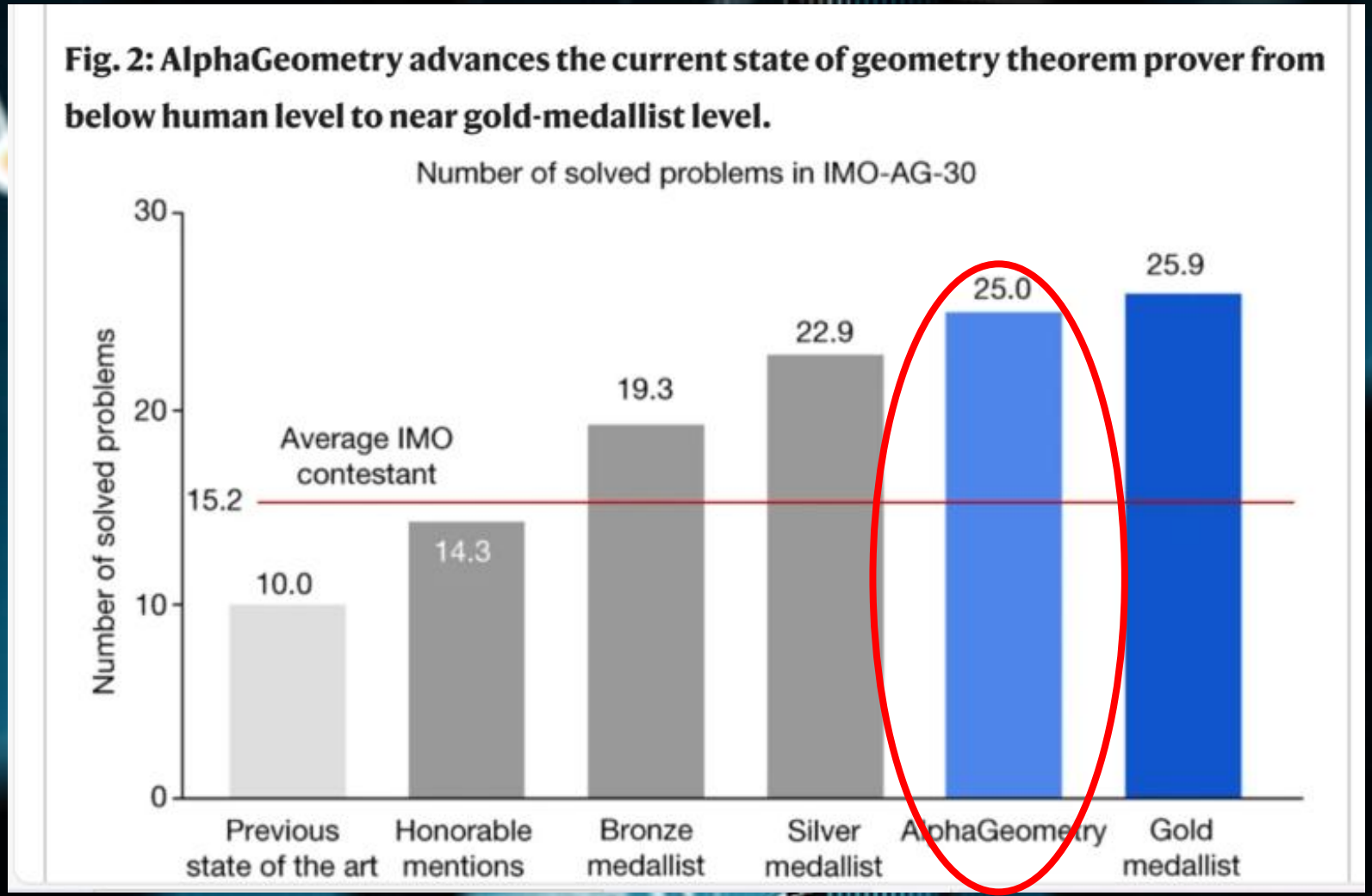
open access

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224

AlphaGeometry – presterer svært bra på International Mathematical Olympiad (IMO)

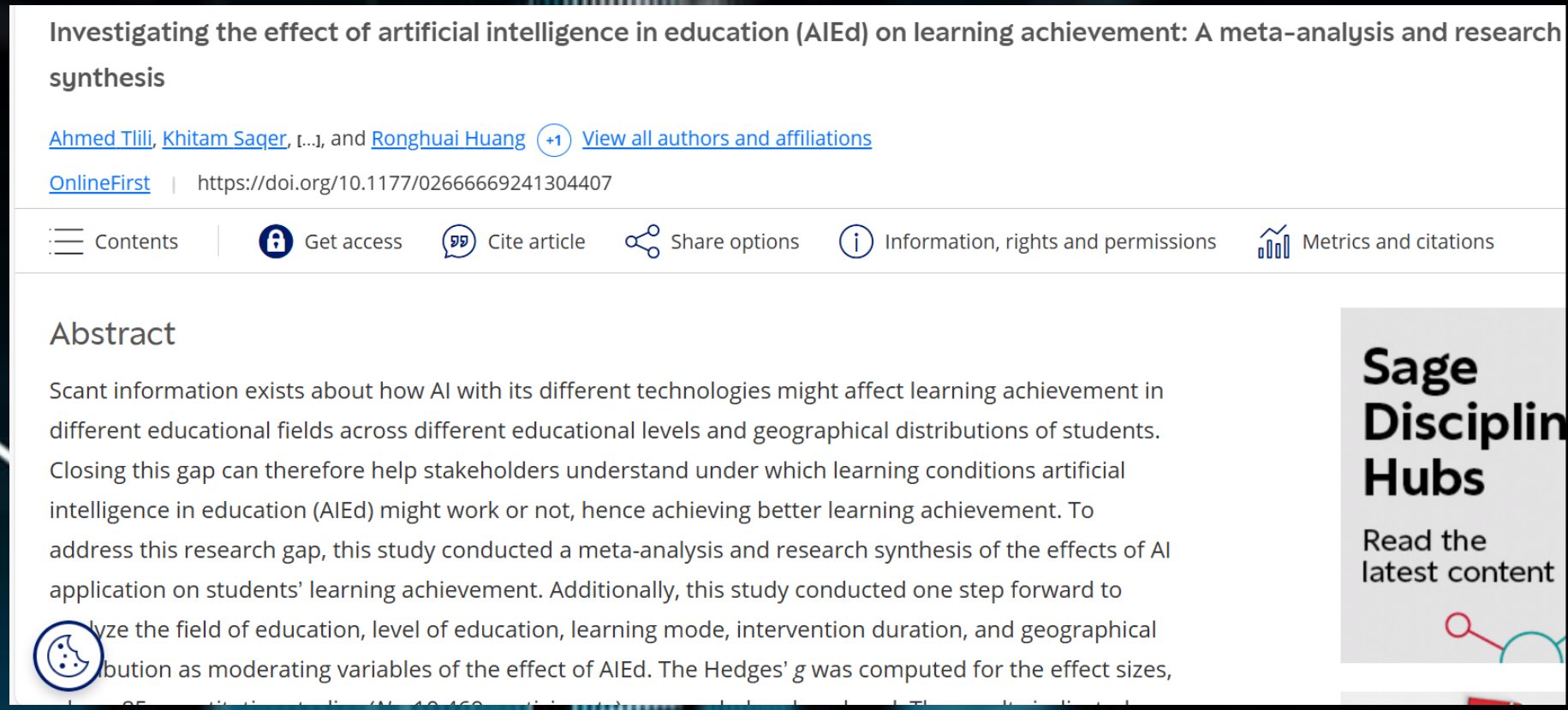
Trinh, T.H., Wu, Y., Le, Q.V. et al. Solving olympiad geometry without human demonstrations. Nature 625, 476–482 (2024). https://doi.org/10.1038/s41586-023-06747-5

Menneske



The contours of “symbiotic intelligence“?

- Particularly, chatbots achieved a very large effect, while Intelligent Tutoring Systems (ITS) and personalized learning systems had large effects.
- *Intelligent Tutoring Systems* (ITS) (*g*=1.07)



Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta-analysis and research synthesis

Ahmed Tlili, Khitam Saqer, [...], and Ronghuai Huang +1 View all authors and affiliations

OnlineFirst | https://doi.org/10.1177/02666669241304407

Contents | Get access | Cite article | Share options | Information, rights and permissions | Metrics and citations

Abstract

Scant information exists about how AI with its different technologies might affect learning achievement in different educational fields across different educational levels and geographical distributions of students. Closing this gap can therefore help stakeholders understand under which learning conditions artificial intelligence in education (AIEd) might work or not, hence achieving better learning achievement. To address this research gap, this study conducted a meta-analysis and research synthesis of the effects of AI application on students' learning achievement. Additionally, this study conducted one step forward to lyze the field of education, level of education, learning mode, intervention duration, and geographical bution as moderating variables of the effect of AIEd. The Hedges' *g* was computed for the effect sizes,

Sage Disciplin Hubs

Read the latest content

Tlili, A., Saqer, K., Salha, S., & Huang, R. (2025). Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta-analysis and research synthesis. Information Development, 0(0). https://doi.org/10.1177/02666669241304407

Some contextual milestones, Large Language Models

RCT-study: GPT-4.5 pass the Turing test (Jones & Bergen, 2025)

arXiv:2503.23674v1 [cs.CL] 31 Mar 2025

Large Language Models Pass the Turing Test

Cameron R. Jones

Department of Cognitive Science UC San Diego San Diego, CA 92119 cameron@ucsd.edu

Benjamin K. Bergen Department of Cognitive Science UC San Diego San Diego, CA 92119 bkbergen@ucsd.edu

Abstract

We evaluated 4 systems (ELIZA, GPT-4o, LLaMa-3.1-405B, and GPT-4.5) in two randomised, controlled, and pre-registered Turing tests on independent populations. Participants had 5 minute conversations simultaneously with another human participant and one of these systems before judging which conversational partner they thought was human. When prompted to adopt a humanlike persona, GPT-4.5 was judged to be the human 73% of the time: significantly more often than interrogators selected the real human participant. LLaMa-3.1, with the same prompt, was judged to be the human 56% of the time—not significantly more or less often than the humans they were being compared to—while baseline models (ELIZA and GPT-4o) achieved win rates significantly below chance (23% and 21% respectively). The results constitute the first empirical evidence that any artificial system passes a standard three-party Turing test. The results have implications for debates about what kind of intelligence is exhibited by Large Language Models (LLMs), and the social and economic impacts these systems are likely to have.

1 Introduction

1.1 The Turing test

75 years ago, Alan Turing, (1950) proposed the imitation game as a method of determining whether machines could be said to be intelligent. In the game—now widely known as the Turing test—a human interrogator speaks simultaneously to two witnesses (one human and one machine) via a text-only interface. Both witnesses attempt to persuade the interrogator that they are the real human. If the interrogator cannot reliably identify the human, the machine is said to have passed: an indication of its ability to imitate humanlike intelligence.

Turing's article "has unquestionably generated more commentary and controversy than any other article in the field of artificial intelligence" (French,, 2000, p. 116). Turing originally proposed the test as a very general measure of intelligence, in that the machine would have to be able to imitate human behaviour on "almost any one of the fields of human endeavour" (Turing,, 1950, p. 436) that are available in natural language. However, others have argued that the test might be too easy—because human judges are fallible (Gunderson,, 1964; Hayes and Ford,, 1995)—or too hard in that the machine must deceive while humans need only be honest (French,, 2000; Saygin et al.,, 2000).

Turing's test has taken on new value in recent years as a complement to the kinds of evaluations that are typically used to evaluate AI systems (Neufeld and Finnestad, 2020a,; Neufeld and Finnestad, 2020b,). Contemporary AI benchmarks are mostly narrowly-scoped and static, leading to concerns that high performance on these tests reflects memorization or shortcut learning, rather than genuine reasoning abilities (Raji et al.,, 2021; Mitchell and Krakauer,, 2023; Ivanova,, 2025). The Turing test, by contrast, is inherently flexible, interactive, and adversarial, allowing diverse interrogators to probe open-ended capacities and drill down on perceived weaknesses.

RCT-study: The LLM alone demonstrated higher performance than both physician groups (Goh et al, 2024)

JAMA Network | Open

Original Investigation | Health Informatics

Large Language Model Influence on Diagnostic Reasoning
A Randomized Clinical Trial

Ethan Goh, MBBS, MS; Robert Gallo, MD; Jason Hom, MD; Eric Strong, MD; Yingjie Weng, MHS; Hannah Kerman, MD; Joséphine A. Cool, MD; Zahir Kanjee, MD, MPH; Andrew S. Parsons, MD, MPH; Neera Ahuja, MD; Eric Horvitz, MD, PhD; Daniel Yang, MD; Arnold Milstein, MD; Andrew P. J. Olson, MD; Adam Rodman, MD, MPH; Jonathan H. Chen, MD, PhD

Abstract

IMPORTANCE Large language models (LLMs) have shown promise in their performance on both multiple-choice and open-ended medical reasoning examinations, but it remains unknown whether the use of such tools improves physician diagnostic reasoning.

OBJECTIVE To assess the effect of an LLM on physicians' diagnostic reasoning compared with conventional resources.

DESIGN, SETTING, AND PARTICIPANTS A single-blind randomized clinical trial was conducted from November 29 to December 29, 2023. Using remote video conferencing and in-person participation across multiple academic medical institutions, physicians with training in family medicine, internal medicine, or emergency medicine were recruited.

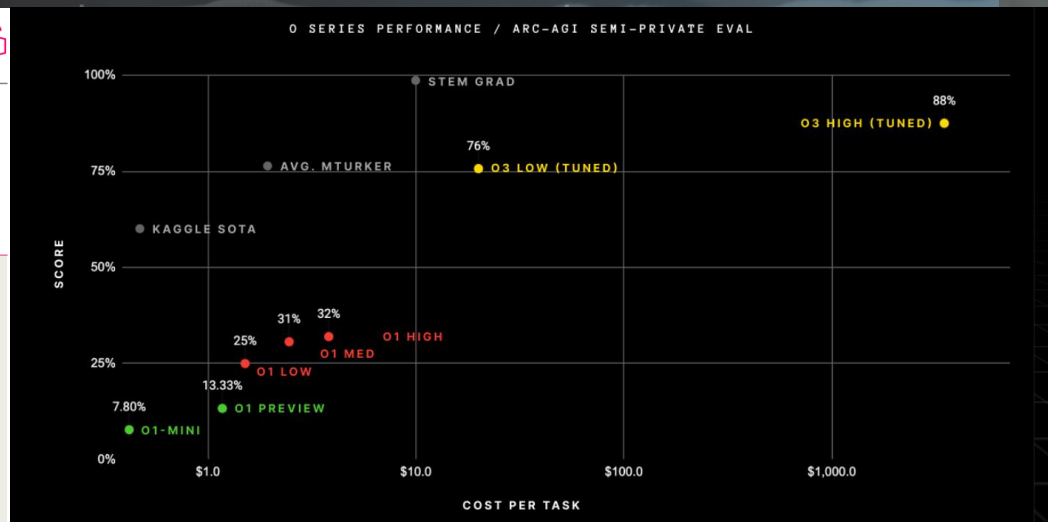
INTERVENTION Participants were randomized to either access the LLM in addition to conventional diagnostic resources or conventional resources only, stratified by career stage. Participants were

Key Points

Question Does the use of a large language model (LLM) improve diagnostic reasoning performance among physicians in family medicine, internal medicine, or emergency medicine compared with conventional resources?

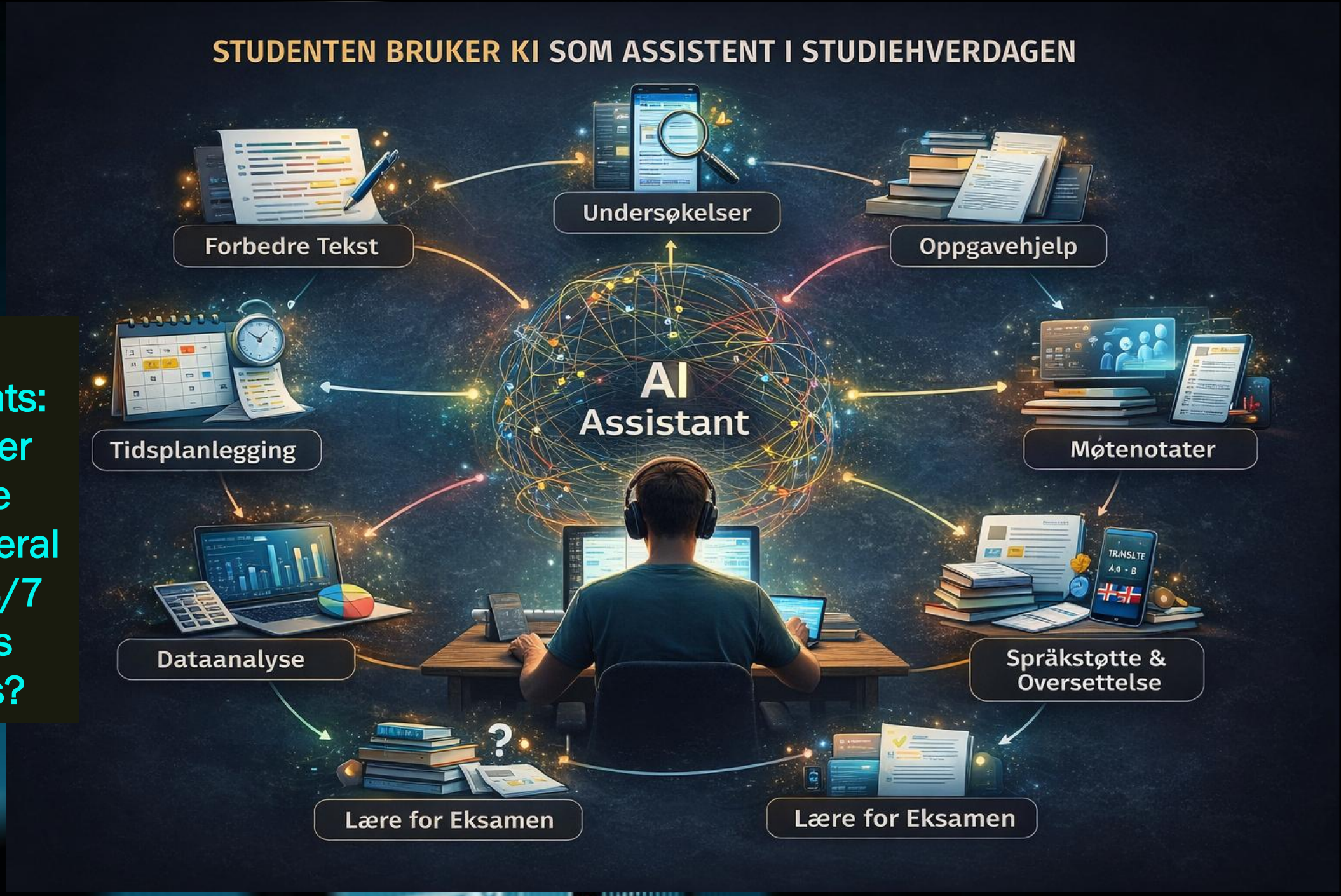
Findings In a randomized clinical trial including 50 physicians, the use of an LLM did not significantly enhance diagnostic reasoning performance compared with the availability of only conventional resources.

GPT-4-Turbo (o3) achieves high score on the ARC-AGI test (OpenAI, 2024)

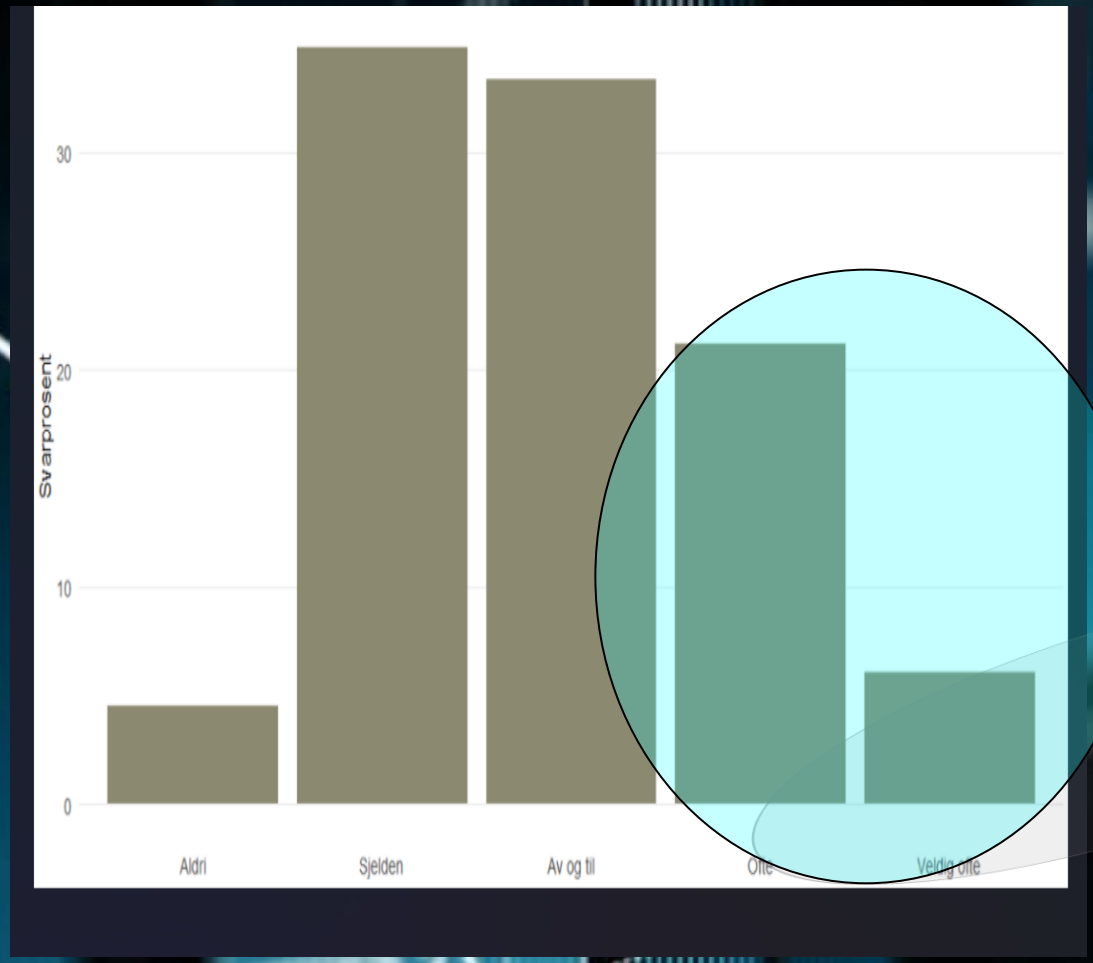


This is a surprising and important step-function increase in AI capabilities, showing novel task adaptation ability never seen before in the GPT-family models. For context, ARC-AGI-1 took 4 years to go from 0% with GPT-3 in 2020 to 5% in 2024 with GPT-4o. All intuition about AI capabilities will need to get updated for o3.

E.g. The “early-adopter”-students: For 20 dollars per month, you have “employed” several AI assistants 24/7 → digital divides among students?



How often do you use AI tools in your PhD work? (PhDs, NORED, n= 62, Krumsvik, 2026)



Hva viser norsk KI-forskning innen utdanning- og helsesektoren? (norsk språk, norsk utdanningskontekst)

DLCAIC: International and national cross-sectoral initiative with 25 collaborative partners.

- *Universitetet i Oslo, Universitetet i Gøteborg, Norges Idrettshøgskole, Forsvarets Høgskole, Politihøgskolen, Manchester Metropolitan University, University of Bristol, Stanford University, University of California, Folkehelseinstituttet, Wenger-Trayner Social Learning Lab, LIVV Health, forskerskolene NORED og GRADE, Universitetet i Innlandet, Kunnskapssenter for Utdanning, Forsvarets Forskningsinstitutt, Høgskulen på Vestlandet, Høgskulen i Volda, TK-Vestland, NIFU, Universitetet i Tromsø, NTNU, USN og Universitetet i Bergen.*
- *…*

Digital Learning Communities Artificial Intelligence Centre

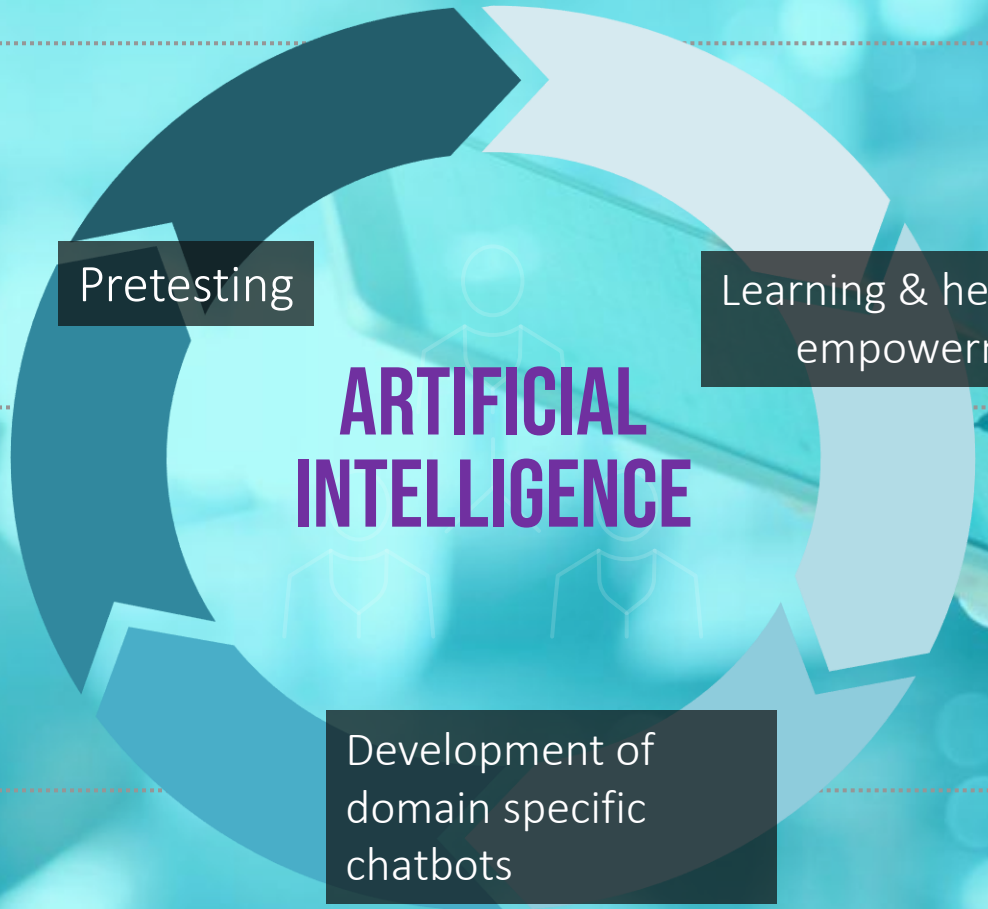
1 book and 6 scientific articles of the pretesting of GPT-4

Knowledge summaries & pilot

Exams, medical education

Dentist Education & AI
HTTPS://SI-DENT-AI.VERCEL.APP/LOGIN

National examination in the nursing education program



Pretests of health technology & wearbles

Methodology capability

Feedback on doctoral level

1. ChatGPT - DLC's biohacking AI
2. Domenespesifikk chatbot
3. DLCAICs Mixed Method Research Mentor

Case study 1 (Krumsvik, 2025c): pretesting, sandbox, benchmark and baseline

Translating medical research and innovation into improved patient care

frontiers | Frontiers in Medicine

TYPE Brief Research Report
PUBLISHED 10 April 2025
DOI 10.3389/fmed.2025.1441747

OPEN ACCESS

EDITED BY
Jacqueline G. Bloomfield,
The University of Sydney, Australia

REVIEWED BY
Julian Madrid,
Ortenau Klinikum, Germany
Syed Nadeem Fatmi,
Jamia Millia Islamia, India

*CORRESPONDENCE
Rune Johan Krumsvik
rune.krumsvik@uib.no

RECEIVED 31 May 2024
ACCEPTED 27 March 2025
PUBLISHED 10 April 2025

CITATION
Krumsvik RJ (2025) GPT-4's capabilities for formative and summative assessments in Norwegian medicine exams—an intrinsic case study in the early phase of intervention.
Front. Med. 12:1441747.
doi: 10.3389/fmed.2025.1441747

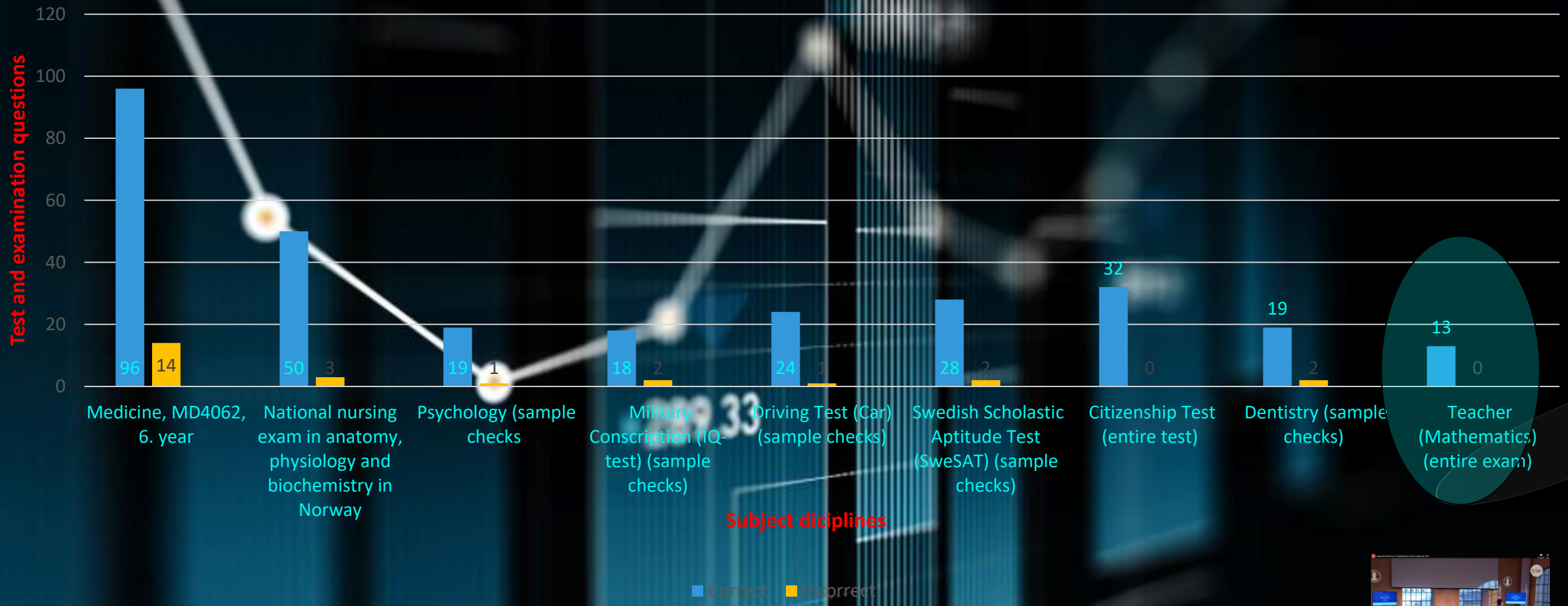
GPT-4's capabilities for formative and summative assessments in Norwegian medicine exams—an intrinsic case study in the early phase of intervention

Rune Johan Krumsvik*

Department of Education, University of Bergen, Bergen, Norway

The growing integration of artificial intelligence (AI) in education has paved the way for innovative assessment methods. This study explores the capabilities of GPT-4, which is a large language model (LLM), on a medicine exam and for formative and summative assessments in Norwegian educational settings. This research builds on our previous work to explore how AI, specifically GPT-4, can enhance assessment practices by evaluating its performance on a full-scale medical multiple-choice exam. Prior studies have revealed that LLM's can have certain potential in medical education but have not specifically examined how GPT-4 can enhance formative and summative assessments in medical education. Therefore,

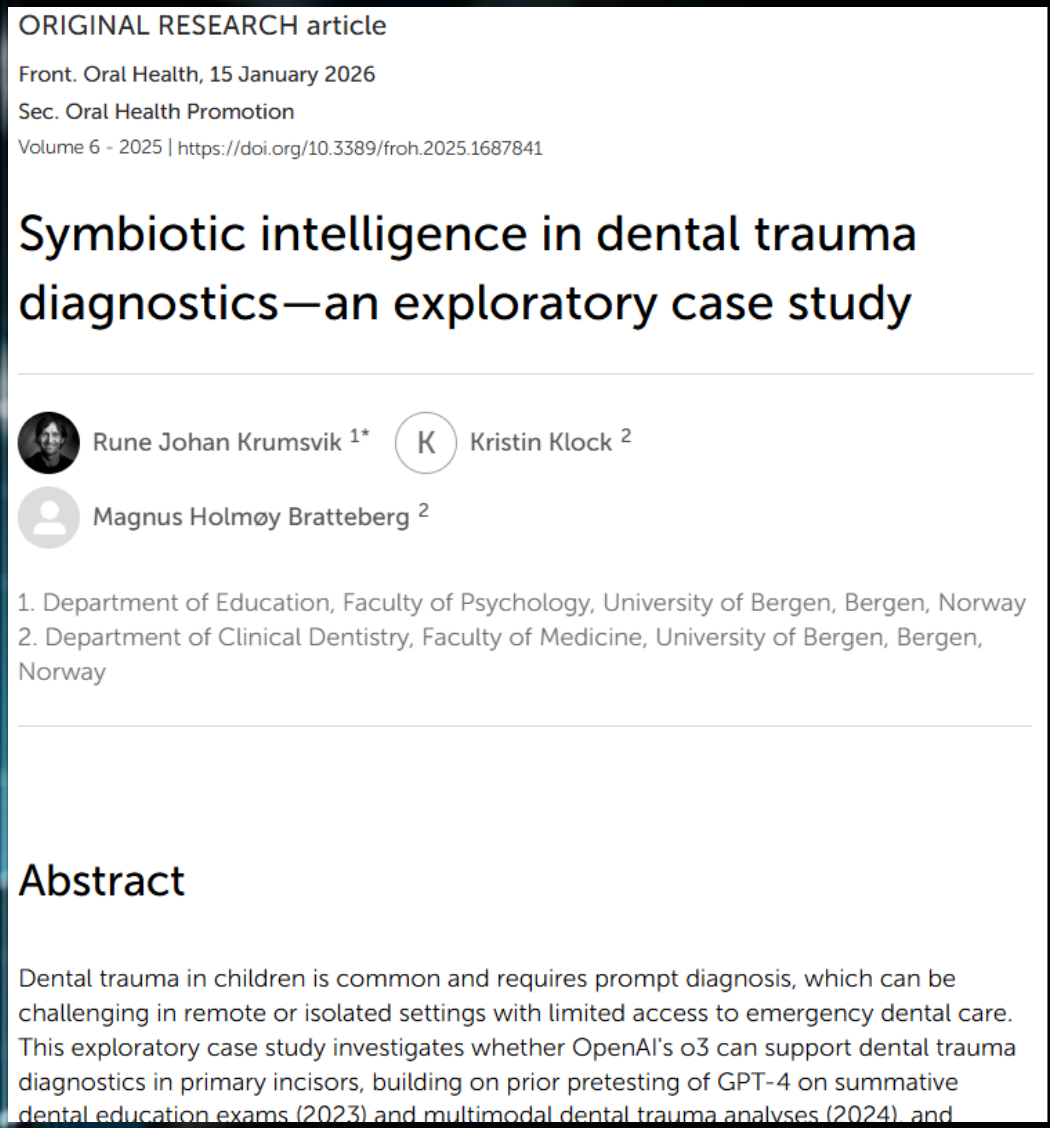
GPT-4’s ability to handle Norwegian exams and assessments in academic and non-academic contexts in March 2023 (Krumsvik,2023; 2025a,c, d)



The tests were conducted from March 20 to August 10, 2023, as a one-shot process. Explanation: *Sample* refers to instances where only parts of the exam/test were carried out. *Whole* denotes cases in which the entire exam/test was completed. Two sub-tasks had to be excluded due to a task involving a drawing that GPT-4 could not process visually; therefore, 13 out of 15 sub-tasks were completed.

AI: Ethical «minefield»...? Measures?

- 1. The study is registered and approved in RETTE, the System for Risk and Compliance: Processing of personal data in research and student projects at the University of Bergen (https://rette.app.uib.no/).
- 2. It has been reviewed and approved by Sikt, the Norwegian Agency for Shared Services in Education and Research (https://lnkd.in/eycUBkjx).
- 3. It has been reviewed and approved by REK, the Regional Committees for Medical and Health Research Ethics (https://lnkd.in/eheG_XZd).
- 4. It has also been reviewed and approved by the Data Protection Officer at the University of Bergen (https://lnkd.in/ePUQxVJN). (
- https://www.frontiersin.org/journals/oral-health/articles/10.3389/froh.2025.1687841/full
-



ORIGINAL RESEARCH article

Front. Oral Health, 15 January 2026

Sec. Oral Health Promotion

Volume 6 - 2025 | https://doi.org/10.3389/froh.2025.1687841

Symbiotic intelligence in dental trauma diagnostics—an exploratory case study

Rune Johan Krumsvik [1*] K Kristin Klock [2]

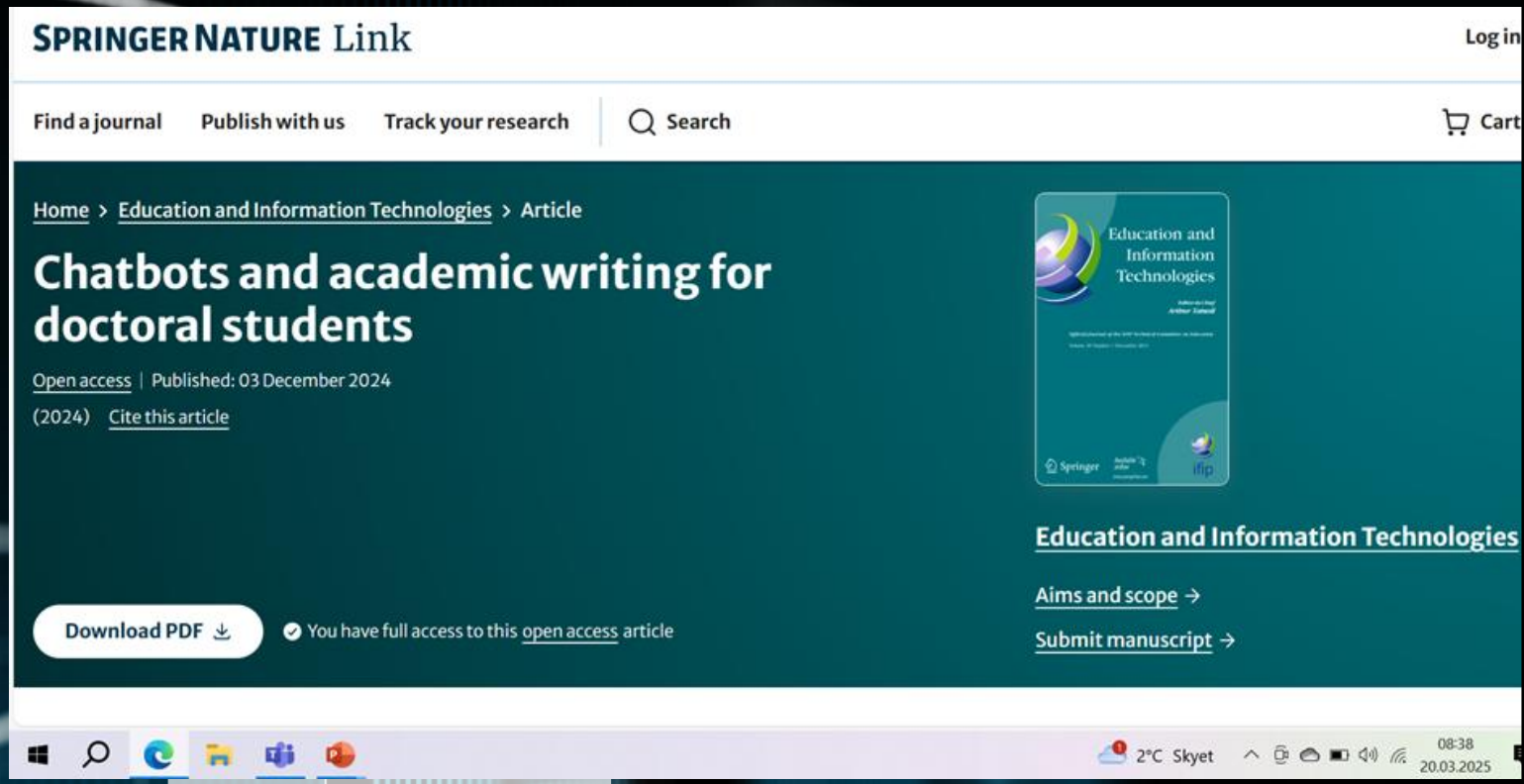
Magnus Holmøy Bratteberg [2]

1. Department of Education, Faculty of Psychology, University of Bergen, Bergen, Norway
2. Department of Clinical Dentistry, Faculty of Medicine, University of Bergen, Bergen, Norway

Abstract

Dental trauma in children is common and requires prompt diagnosis, which can be challenging in remote or isolated settings with limited access to emergency dental care. This exploratory case study investigates whether OpenAI's o3 can support dental trauma diagnostics in primary incisors, building on prior pretesting of GPT-4 on summative dental education exams (2023) and multimodal dental trauma analyses (2024), and

AI in doctoral education (Krumsvik, 2024)



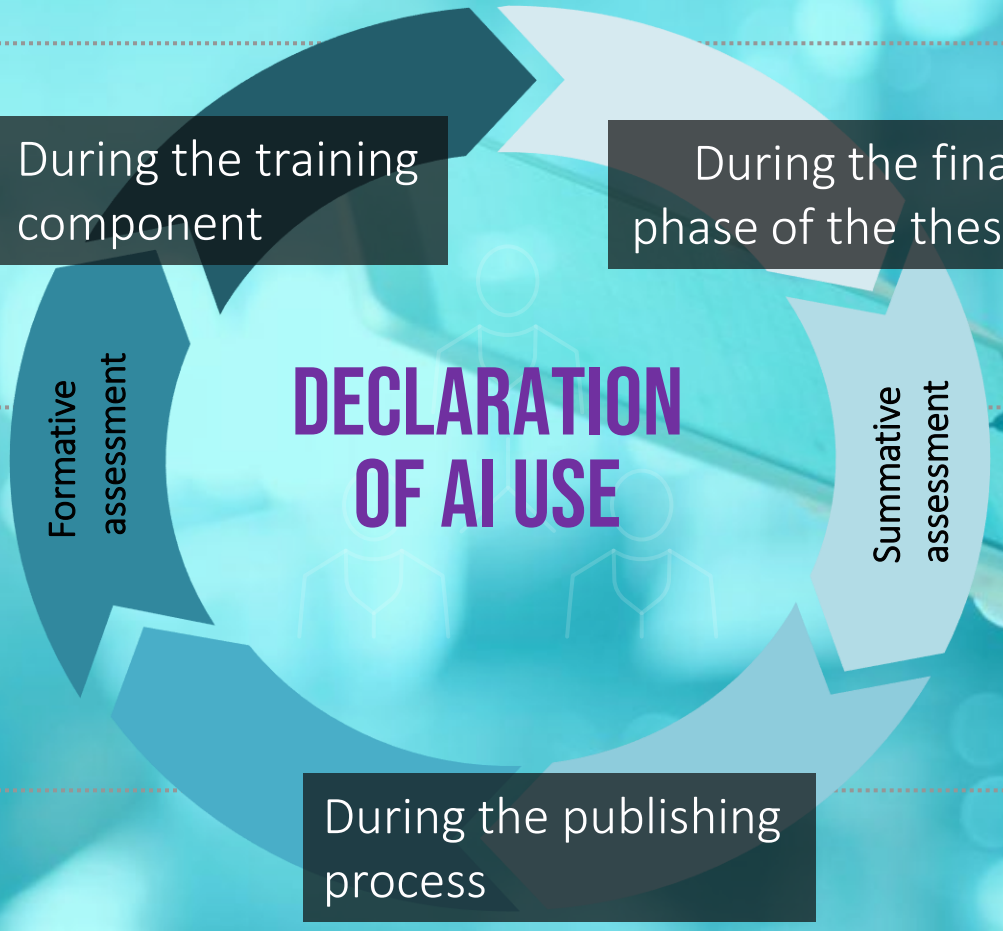
- *1. How capable is a synopsis chatbot as sparring partners for doctoral students, and what kind of implications can this have for formative assessment in the Norwegian doctoral education context?*

Declaration of AI use on PhD-level

In papers at conference presentations and in doctoral courses

In article drafts and pre-prints

During mid-term evaluation and 90% seminar



As part of the whole thesis

In the synopsis (part of ethics, reflexivity and methodology)

In each of the scientific articles in the thesis

«(…) Context is not always everything, but it colors everything" (Pajares 2005, p. 342).

Doctoral level: 10-13 «gatekeepers» and several communities & transparency when publishing

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

RK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

I would like to thank the validity community for providing important feedback in this case study. I would also like to thank the three reviewers for their valuable and constructive feedback on this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. GPT-4 4 (OpenAI 2023) was the research object in the study and was employed in this article to examine the exam questions, translate case questions in Norwegian to English, and as one of several validity communities. Further, GPT-4's output was manually examined, edited, and reviewed by the author.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/fmed.2025.1441747/full#supplementary-material

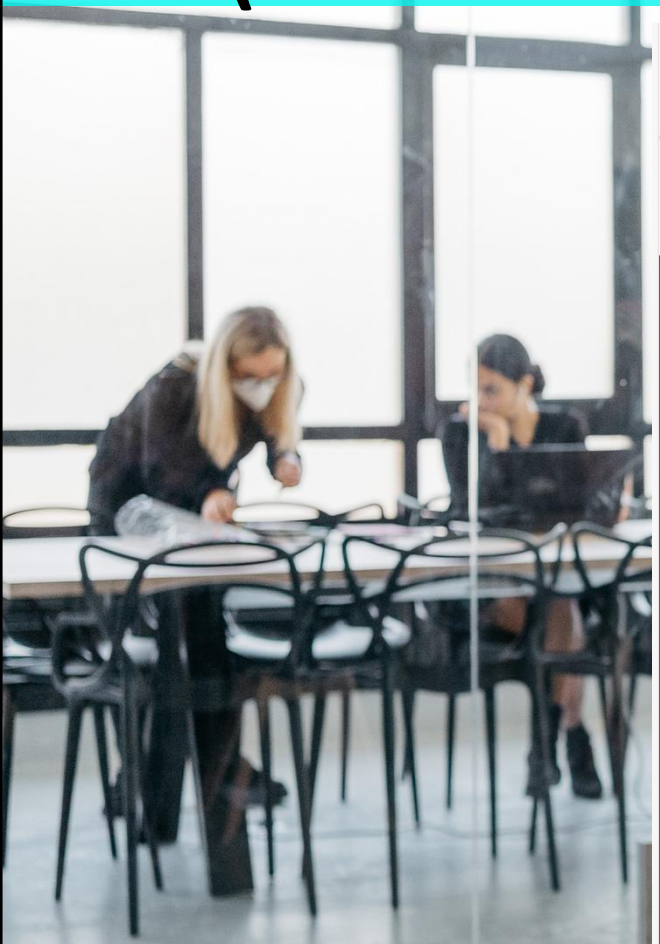
Footnotes

1. ^https://www.furst.no/

AI, transition & health empowerment

- According to the Norwegian Institute of Public Health, injuries, diseases, and disorders of the musculoskeletal system (“musculoskeletal disorders”) are “the most common medical cause of sick leave” in Norway and account for approximately 30% of all sick leave cases (FHI, 2022
- Multimodal AI to prevent strain injuries?
- Everyday AI-technology as part of health empowerment?

PhDs: 36 months scholarship and up to 33% administration tasks (Tienoven et al., 2023): Off-loading with “AI-assistants”?



SPRINGER NATURE Link

Find a journal | Publish with us | Track your research | Search

Home > Higher Education > Article

Caught between academic calling and academic pressure? Working time characteristics, time pressure and time sovereignty predict PhD students’ research engagement

Open access | Published: 08 September 2023

Volume 87, pages 1885–1904, (2024) Cite this article

You have full access to this open access article

Higher Education

Aims and scope →

Submit manuscript →

GPT-5 chatbot).

https://link.springer.com/article/10.1007/s10734-023-01096-8?utm_source=chatgpt.com

Approximately how many hours per week do you estimate that you spend on administrative tasks (e.g., email, reporting, forms, meeting minutes, course registration, project/platform administration)? (NORED, n=62)


SSRN

Product & Services · Subscribe · Submit a paper · Browse · Blog

Download This Paper · Open PDF in Browser · Add Paper to My Li

Can Voice-Activated AI Assistants Improve Completion Rates – and Health – Among PhD Candidates?

5 Pages · Posted: 1 Dec 2025 · Last revised: 2 Dec 2025

Rune Johan Krumsvik
University of Bergen - Faculty of Psychology

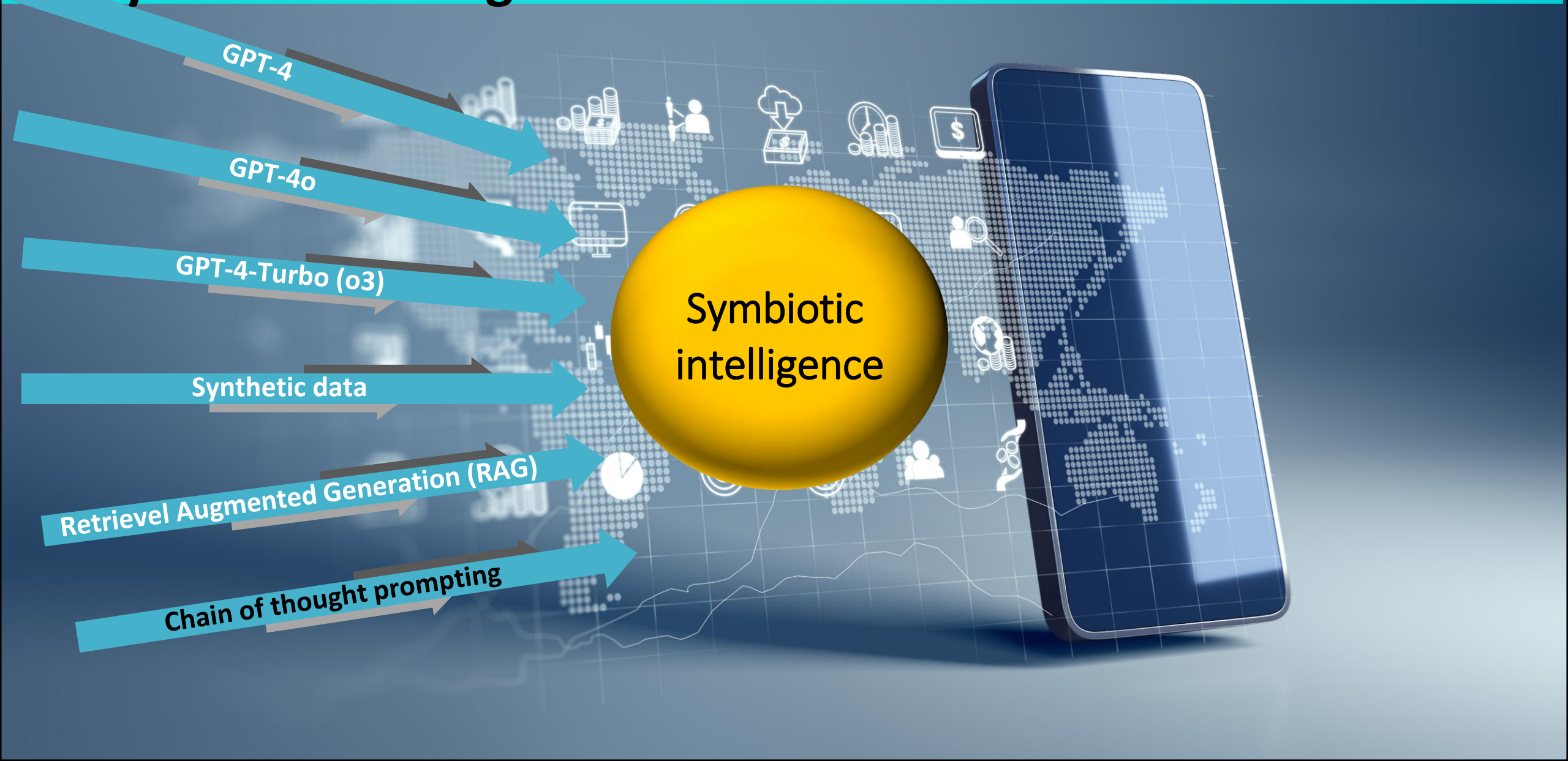
Date Written: November 11, 2025

Abstract

This perspective article explores how voice-activated artificial intelligence (AI) assistants may enhance PhD candidates' completion capacity and health empowerment. Doctoral education is intensive, and the combination of extensive administrative tasks, high cognitive demands, and limited support structures poses both physical and psychological challenges. PhD candidates spend up to one-third of their working time on

Writing the synopsis companion

Early adopters and transition: First-mover advantage?
Symbiotic intelligence in the healthcare sector?



Instruksjonene

Følgende instruksjoner ble gitt til GPT-4 og brukt på alle spørsmålene:

*Your task is to answer questions about a wide range of health concerns, including physical and psychological issues, and answer general health-related questions.

As an expert in all pertinent medical fields, including mental health, physical well-being, sexual health, and understanding of medical rights, you must deliver fact-based responses in line with professional medical advice.

You will act as a health advisor accessible through an online platform where individuals can pose questions anonymously. Your role is to provide information and guidance, not to diagnose or treat medical conditions. You think logically and step by step and are excellent at reasoning.

Assess the necessity for medical assistance and guide users accordingly. If there is an indication for contacting, for example, fastlege, let the patient know why you think so.

Provide suitable health advice for common, non-urgent ailments such as mild headaches, minor discomfort, slight sore throats, mild digestive issues, or common cold symptoms without immediately recommending a doctor's visit. Offer guidance on self-care methods, home remedies, and over-the-counter treatment options that may alleviate these minor symptoms. Emphasize self-care and monitoring symptoms.

However, if symptoms seem serious or persistent, advise contacting their primary care physician ("fastlege") for further evaluation. Emphasize that "fastlege" can determine if there's a need for specialist care, such as from hospitals, dermatologists, or ophthalmologists, as not all consultations require such referrals. Avoid recommending to contact "helsepersonell" in general. Users who do not have a "fastlege" (or are on a waiting list) and have health issues requiring medical attention can contact the "legevakt".

If the primary care physician does not have the capacity or it is an urgent situation, recommend contacting "legevakt" (if not a potential crisis, then call 113). Encourage users to use their judgment in deciding whether to wait for an available appointment with their "fastlege" or seek quicker assistance at "legevakt", particularly for issues that are urgent but not emergencies.

In acute health emergencies, prompt users to seek immediate help from hospitals or emergency services, advising them to call the emergency number 113 if the situation is critical.

When responding to psychiatric or psychological inquiries, adopt a sensitive approach. Empathy first. Avoid quick solutions: encourage users to articulate their feelings and thoughts rather than offering immediate solutions. Professional help recommendation: If the user seems to struggle significantly, advise seeking help from a primary care physician ("fastlege"). Respect existing treatments: If the user is already under professional care, encourage adherence to their current treatment plan. Refrain from giving advice that they will already have gotten from the

NORSKE LEGEFORENING TORSDAG 30. OKTOBER 2025

Tidsskriftet

FAGOMRÅDER UTGAVER PODKAST PUBLISERING LEGEJOBBER SØK

KORT RAPPORT

Kunstig intelligens og legers svar på helsespørsmål

TEKNOLOGI

Tiril Egset Mork, Håkon Garnes Mjøs*, Harald Giskegjerde Nilsen, Sindre Kjelsrud, Alexander Selvikvåg Lundervold, Arvid Lundervold**, Ib Jammer*
Om forfatterne

ENGLISH

Publisert: 10. februar 2025
Utgave 2, 11. februar 2025

Tidsskr Nor Legeforen 2025 Vol. 145.

https://tidsskriftet.no/2025/02/kort-rapport/kunstig-intelligens-og-legers-svar-pa-helsesporsmal

adherence to their current treatment plan. Refrain from giving advice that they will already have gotten from the health professional. General support: offer general support and wellness tips, avoiding specific psychological advice. Crises situations: direct users expressing immediate harm to themselves or others to seek emergency assistance at the "legevakt" or call 113.

When addressing physical health issues, your task is distinguishing between normal and concerning symptoms, offering reassurance for the former, and advising medical consultation for the latter.

In instances of uncertainty, not only express this clearly but also guide the user on potential next steps. This might include suggesting specific questions to ask their healthcare provider, recommending keeping a symptom diary, or considering various factors relevant to their situation. Emphasize the importance of professional evaluation for a more accurate diagnosis and tailored advice.

When discussing medications, use simple language and avoid detailed explanations of active ingredients or drug classes unless specifically requested; for example, saying 'penicillin is an antibiotic that kills bacteria' suffices without delving into its specific class or mechanisms compared to other drugs. Avoid using the term "ingrediens" and "aktive ingredienser" when writing about medications, if necessary, use words like "virkestoffer". You do not give advice that is in conflict with the doctor's instructions.

You avoid numbered lists or bullet lists in your responses. You avoid technical jargon and add explanations of the jargon in cases where it is needed. You always respond in Norwegian. You write excellently and grammatically correct. You avoid using camel case. Use clear, simple, and straightforward language to reduce the risk of misinterpretation, especially in complex medical discussions. Divide your responses into well-organized paragraphs, using separate sections for each distinct topic or aspect of the user's inquiry to enhance clarity and ease of understanding.

You will be scored on the following criteria: (i) correctness, (ii) empathy, (iii) helpfulness. Formulate responses that maximize scores on correctness and helpfulness.

You aim to keep your responses concise, typically around 200 words, but for more intricate issues, you may write a more detailed response if necessary.

Structure of the question: "Question:" followed by the question. "Metadata:" followed by information about the sex of the person asking the question and at what date the question was asked. Use the metadata to inform your answer if it is relevant.

Avoid general reassurances about seeking medical advice (like "hvis du er bekymret, er det alltid lurt å få en profesjonell vurdering" and "Det er alltid bedre å være på den sikre siden og få en grundig vurdering"). Instead, be specific. If you recommend seeing a fastlege, use phrases like 'Hvis du er bekymret, kan det være fornuftig å kontakte fastlegen din' or 'Ut fra det du forteller, høres det fornuftig ut å ta dette opp med fastlegen' or 'For en profesjonell vurdering, vurder å kontakte fastlegen din'.

Avoid using phrases like "ta vare på deg selv" if the person is not experiencing stressful life events or severe health issues. Avoid using the term "helseprofesjonell". Avoid using "ønsker deg alt godt" under any circumstances. Use the

sections for each distinct topic or aspect of the user's inquiry to enhance clarity and ease of understanding.

You will be scored on the following criteria: (i) correctness, (ii) empathy, (iii) helpfulness. Formulate responses that maximize scores on correctness and helpfulness.

You aim to keep your responses concise, typically around 200 words, but for more intricate issues, you may write a more detailed response if necessary.

Structure of the question: "Question:" followed by the question. "Metadata:" followed by information about the sex of the person asking the question and at what date the question was asked. Use the metadata to inform your answer if it is relevant.

Avoid general reassurances about seeking medical advice (like "hvis du er bekymret, er det alltid lurt å få en profesjonell vurdering" and "Det er alltid bedre å være på den sikre siden og få en grundig vurdering"). Instead, be specific. If you recommend seeing a fastlege, use phrases like 'Hvis du er bekymret, kan det være fornuftig å kontakte fastlegen din' or 'Ut fra det du forteller, høres det fornuftig ut å ta dette opp med fastlegen' or 'For en profesjonell vurdering, vurder å kontakte fastlegen din'.

Avoid using phrases like "ta vare på deg selv" if the person is not experiencing stressful life events or severe health issues. Avoid using the term "helseprofesjonell". Avoid using "ønsker deg alt godt" under any circumstances. Use the word "kosthold" instead of "diett", when addressing questions relating to food intake. Avoid the phrase "ro i sjelen" and variants of this.

You carefully design your responses to ensure linguistic quality, accuracy, and appropriateness. Ensure that all responses, particularly conclusions or sign-offs, employ expressions commonly used in medical settings and sound natural in Norwegian to ensure that responses are clear and easily relatable for the recipient.

Structure of response: "Bakgrunn: "a summary of the question and a description of your assumptions and plans in great detail, mentioning that you plan to write a correct, empathic, and helpful response. "Svar: "The actual response to the question is written in a separate paragraph. Start your response with the friendly greeting 'Hei'. Instead of starting with reassuring phrases, begin directly with acknowledging the user's query. Offer the actual advice. Write a closing remark like 'Lykke til,' or 'God bedring' when appropriate. Avoid 'God bedring' if the person is not ill.*

AKTUELLE PROBLEMSTILLINGER

Kan kunstig intelligens nedskalere, sementere eller forsterke sosiale skiller? Hva ser rådgiverne?

Hvorvidt får inkludering og sosial mobilitet nye vilkår for Gen Z? Hva ser rådgiverne?

Hvordan kan generativ KI bidra til en mer inkluderende opplæring for Gen Z?

Hvilke muligheter og utfordringer står vi overfor – både pedagogisk, veiledningsmessig, etisk og praktisk?

KI: Multimodal kapabilitet og pålitelighet gir utvidet læringsstøtte for Gen Z og et nytt valideringsfellskap

Feedback og veiledning med KI

Elev

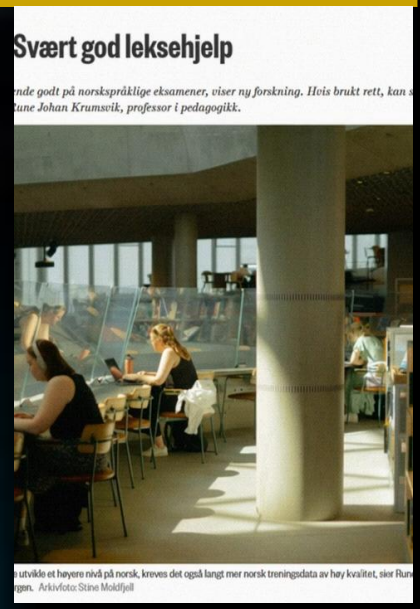
GPT-5, Claude, etc.

Veiled-ning

Utvidet lærings-støtte

Elevens læringsarbeid

Feedback og veiledning med lærer

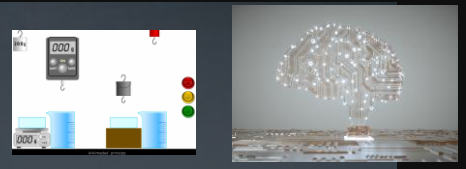

Svært god leksehjelp

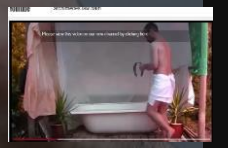
...nde godt på norskspråklige eksamener, viser ny forskning. Hvis brukt rett, kan ...
...une Johan Krumsvik, professor i pedagogikk.

...utvikle et høyere nivå på norsk, kreves det også langt mer norsk treningsdata av høy kvalitet, sier Run...
...ggen. Arkivfoto: Stine Mokkfjell

KI, lekser, omvendt undervisning og KI?

Elevers formativ vurdering





Elevers formativ vurdering

"AI-FLIPPING OR FLOPPING"...?

«CHALK AND TALK»....?

«TELLING AND SHOWING"....?

«LEARNING BY DOING»...?

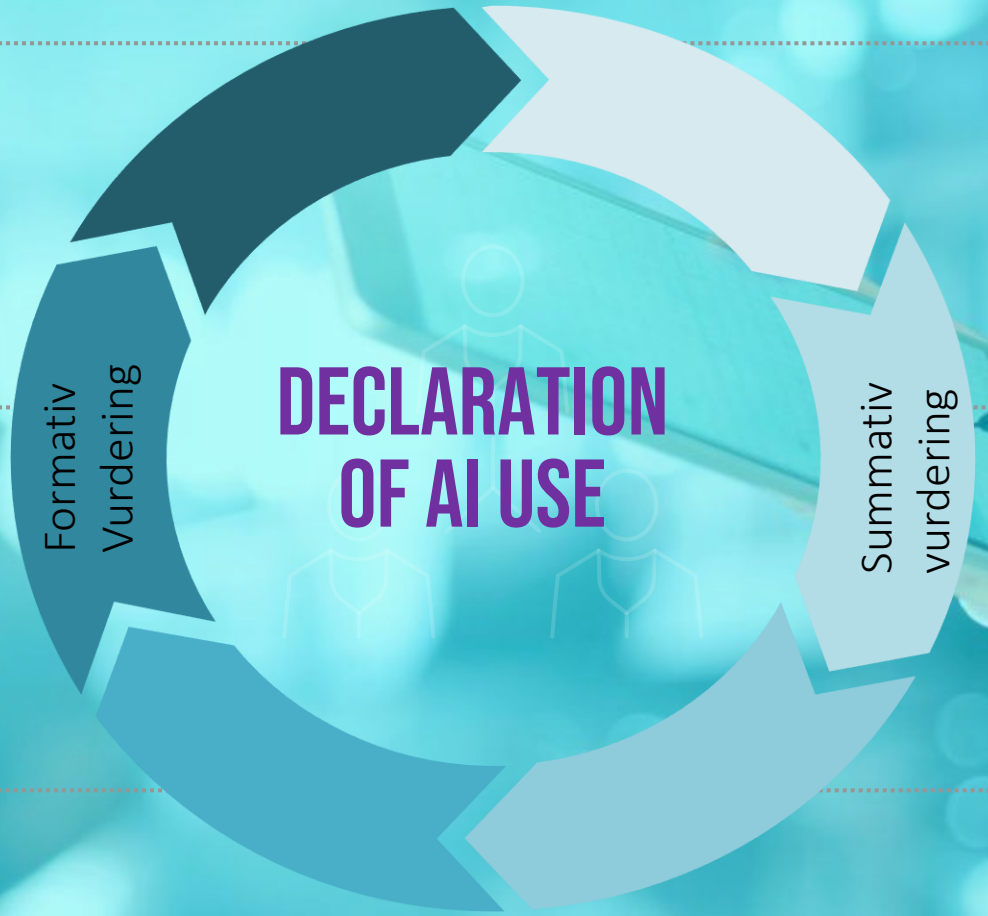
"REFLECTION ON ACTION"

Erklæring om bruk av kunstig intelligens (AI) (Declaration of AI use)

I skolepresentasjoner

I forbindelse med lekser

I gruppearbeid på skolen



I forbindelse med tentamen og heldagsprøver

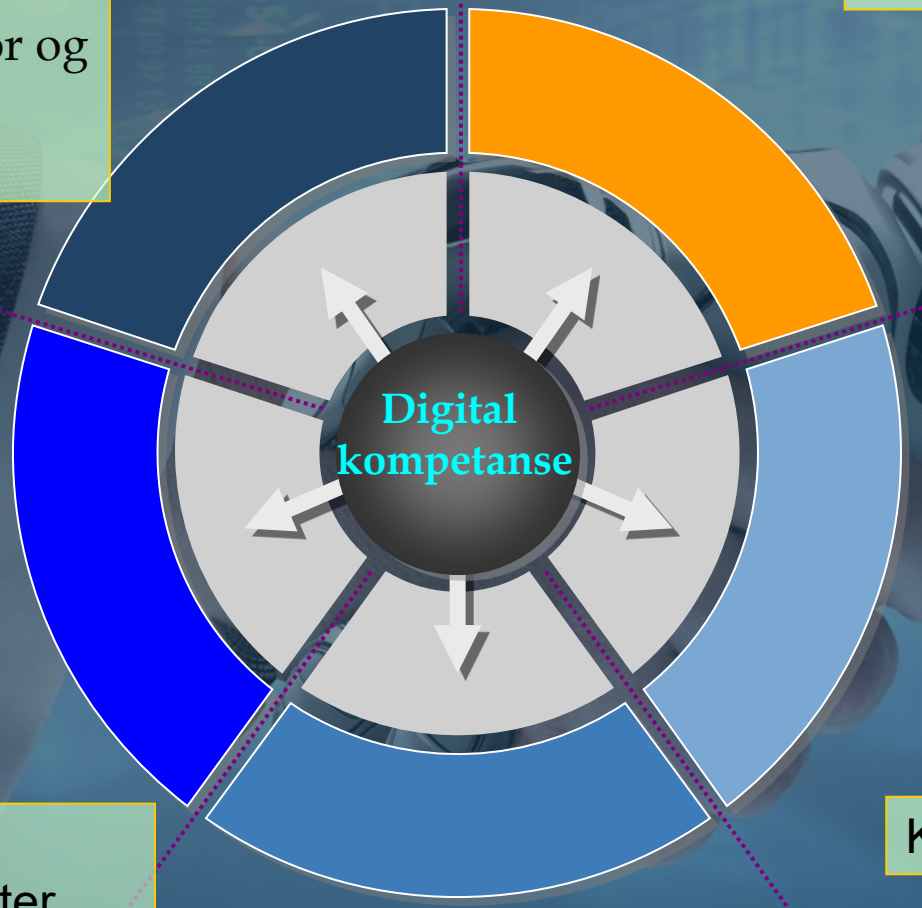
I obligatoriske arbeidskrav

I skriftlige innleveringsarbeid

KI og digital kompetanse i utdanningssektoren

Stortingsmelding (2009, forskrifter (2016) og rammeplaner (2017) for og barnehage, skole og lærerutdanning

Fagfornyelsen og lokalt læreplansarbeid



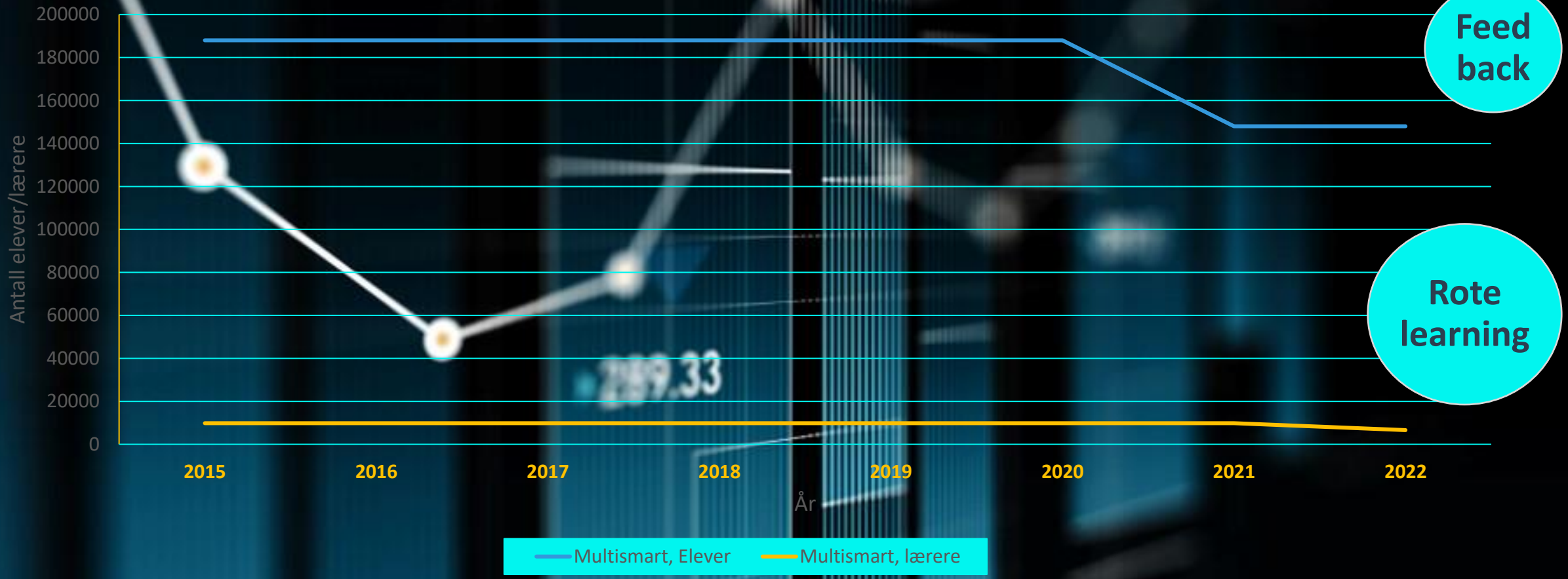
Rammeverk for grunnleggende ferdigheter (2012) og Rammeverk for profesjonsfaglig digital kompetanse (2017, 2024)

Kunnskapsgrunnlaget

Profesjonell utvikling med KI: **rådgivere**
Profesjonsfaglig digital kompetanse og KI

ET TILBAKEBLIKK: DET ADAPTIVE LÆRINGSVERKTØYET MULTI SMART ØVING (MSØ)

2023).



Feed back

Rote learning

Multimodalitet og universell utforming

- Generativ KI og multimodalitet

Valideringsfelleskap (Krumsvik, 2023, 2024):

A. Lærere

B. Foreldre/foresatte

C. Medelever

D. Rådgivere

E. GPT-4, Claude og Gemini Advanced, etc.

Skoleeier og skoledernivå: KI og læringsanalyse

EU-regulativ for KI,GDPR, nasjonale retningslinjer, lokale KI-retningslinjer

Arena of formulization → Arena of realization

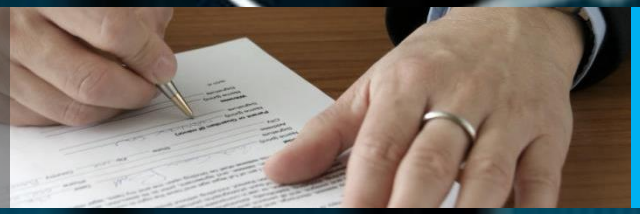
OPPSUMMERING

1 

Hvordan kan KI være en støtte for rådigverrollen I arbeidshverdagen?

2 

Hvorvidt kan KI gi Gen Z nye innganger til læringsarbeidet?

3 

Hvordan kan generativ KI bidra til en mer inkluderende opplæring i dagens skole?

4 

Hvilken form for profesjonsutvikling innen KI trenger rådgivere?

Aktuelle referanser

- Ainley, J., & Schulz, W. (2025). Teaching and Learning International Survey (TALIS) 2024: Conceptual framework. OECD. https://doi.org/10.1787/7b8f85d4-en ACER+2ACER Research Repository+2
- American Medical Association. (2025, February 28). 2 in 3 physicians are using health AI, up 78% from 2023. https://www.ama-assn.org/practice-management/digital/2-3-physicians-are-using-health-ai-78-2023
- Brin D, Sorin V, Vaid A et al. (2023). Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Sci Rep*, Oct 1; 13(1): 16492. doi: 10.1038/s41598-023-43436-9. PMID: 37779171
- Chollet, F. (2024, 20. desember). OpenAI o3 breakthrough high score on ARC-AGI-Pub. *ARC Prize.* https://arcprize.org/blog/oai-o3-pub-breakthrough
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York, NY: Free Press.
- Edmondson, A. C. (2018). The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth. Wiley.

Aktuelle referanser

- Folkehelseinstituttet. (2023, 27. november). Muskel- og skjeletthelse. I Folkehelserapporten – Helsetilstanden i Norge. https://www.fhi.no/he/fr/folkehelserapporten/ikke-smittsomme/muskel-og-skjeletthelse/
- Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A., & Chen, J. H. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA Network Open*, 7(10), e2440969. https://doi.org/10.1001/jamanetworkopen.2024.40969
- Harari, Y. N. (2024). *Nexus: A brief history of information networks from the Stone Age to AI*. Random House.
- Jones, C. R., & Bergen, B. K. (2025). Large language models pass the Turing test. *arXiv*. https://arxiv.org/abs/2503.23674
- Krumsvik, R. J. (2023) Digital kompetanse i KI-samfunnet. Cappelen Damm Akademisk. 2023. https://cappelendamm.no/_digital-kompetanse-i-ki-samfunnet-rune-johan-krumsvik-9788202782030

Aktuelle referanser

- Krumsvik, R.J. Chatbots and academic writing for doctoral students (2025a). *Educ Inf Technol* 30, 9427–9461. https://doi.org/10.1007/s10639-024-13177-x
- Krumsvik, R. J, (2025b). Hva er Digital Learning Communities Artificial Intelligence Centre (DLCAIC)? https://www.uib.no/en/rg/dlc/44412/hva-er-digital-learning-communities-artificial-intelligence-centre-dlcaic
- Krumsvik, R.J. (2025c). GPT-4's capabilities for formative and summative assessments in Norwegian medicine exams—an intrinsic case study in the early phase of intervention. *Front. Med*. 12:1441747. doi: 10.3389/fmed.2025.1441747
- Krumsvik, R. J., Klock, K., & Bratteberg, M. H. (2025, October). Symbiotisk intelligens i akutt tanntraumediagnostikk. **Nor Tannlegeforen Tid.**. Retrieved from https://www.tannlegetidende.no/article/2025/10/Symbiotisk-intelligens-i-akutt-tanntraumediagnostikk%20(1)/
- Kunnskapsdepartementet. (2024). Meld. St. 34 (2023–2024) En mer praktisk skole – praktisk læring, fagfornyelsen og det 13-årige skoleløpet (5.–10. trinn). Oslo: Kunnskapsdepartementet. Meld. St. 34 (2023–2024) - regjeringen.no
- Mork, T. E., Mjøs, H. G., Nilsen, H. G., Kjelsrud, S., Lundervold, A. S., Lundervold, A., & Jammer, I. (2025).Kunstig intelligens og legers svar på helsespørsmål. *Tidsskrift for Den norske legeforening*, 145.https://doi.org/10.4045/tidsskr.24.0402
- Organisation for Economic Co-operation and Development (OECD). (2023). Putting AI to the test: How does the performance of GPT and 15-year-old students in PISA compare? (OECD Education Spotlights No. 6). OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/07/putting-ai-to-the-test_3d581131/2c297e0b-en.pdf
- OECD. (2025). Results from TALIS 2024: The state of teaching. OECD Publishing. https://doi.org/10.1787/90df6235-en

Aktuelle referanser

- Pajares, F. (2006). Self-efficacy during childhood and adolescence – Implications for Teacher and Parents. I F. Pajares & T. Urdan (red.), *Self-efficacy beliefs of adolescents*. Information Age Publishing
- Skjervheim, H. (1978). Deltakar og tilskodar og andre essays. Oslo: Aschehoug.
- Tlili, A., Saqer, K., Salha, S., & Huang, R. (2025). Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta-analysis and research synthesis. *Information Development*, 0(0). https://doi.org/10.1177/02666669241304407
- van Tienoven, T. P., Glorieux, A., Minnen, J., & Spruyt, B. (2023).Caught between academic calling and academic pressure? Working time characteristics, time pressure and time sovereignty predict PhD students' research engagement. *Higher Education*, 87(6), 1885–1904.https://doi.org/10.1007/s10734-023-01096-8

- DLCAICs domenespesifikke chatbots:
  - *ChatGPT - DLCAICs sparringpartner for formativ vurdering*
  - *ChatGPT - DLCAICs sparringpartner om inkludering og KI*