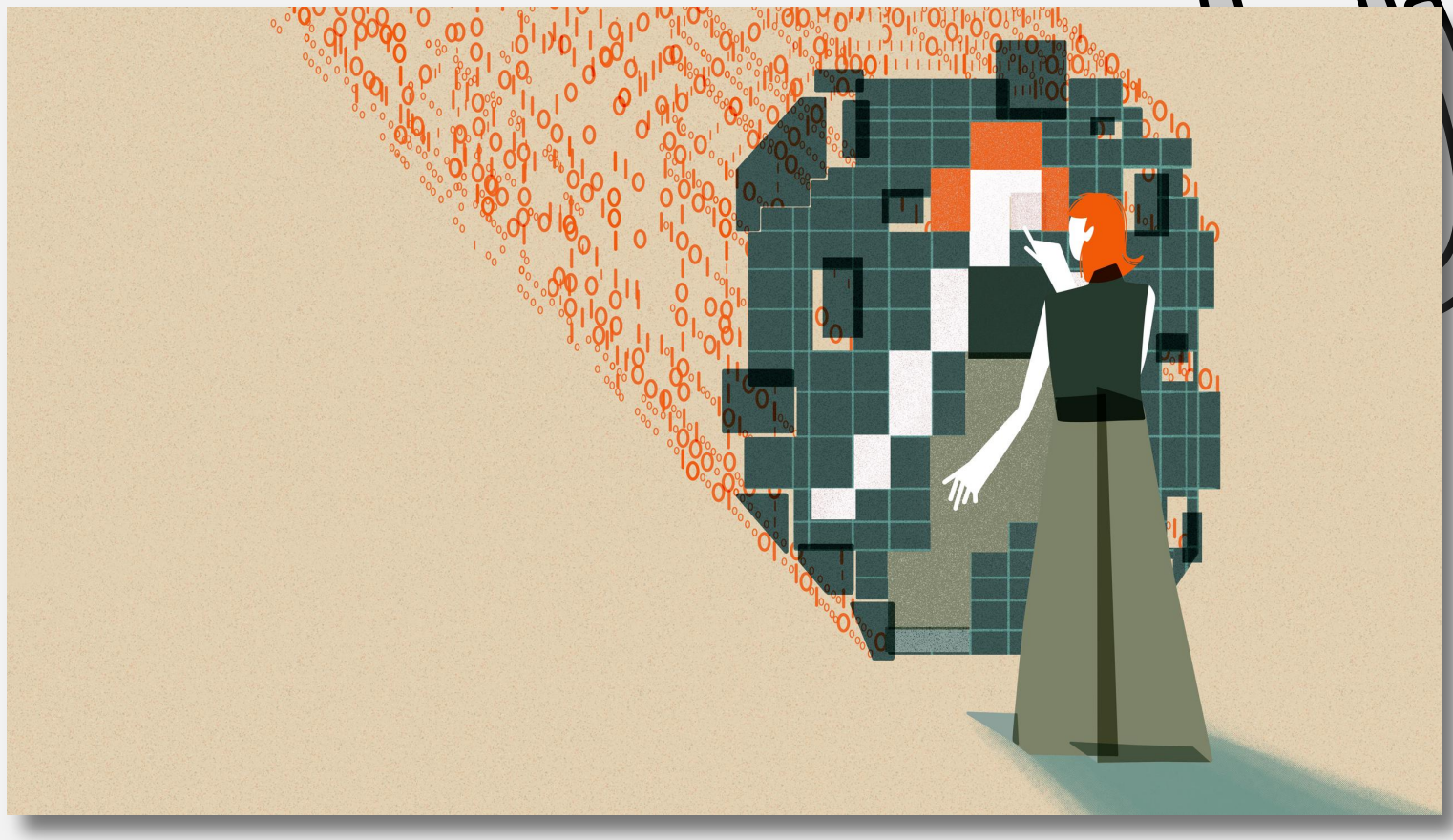
UNIVERSITÄT BIELEFELD
Zentrum für Lehren und Lernen

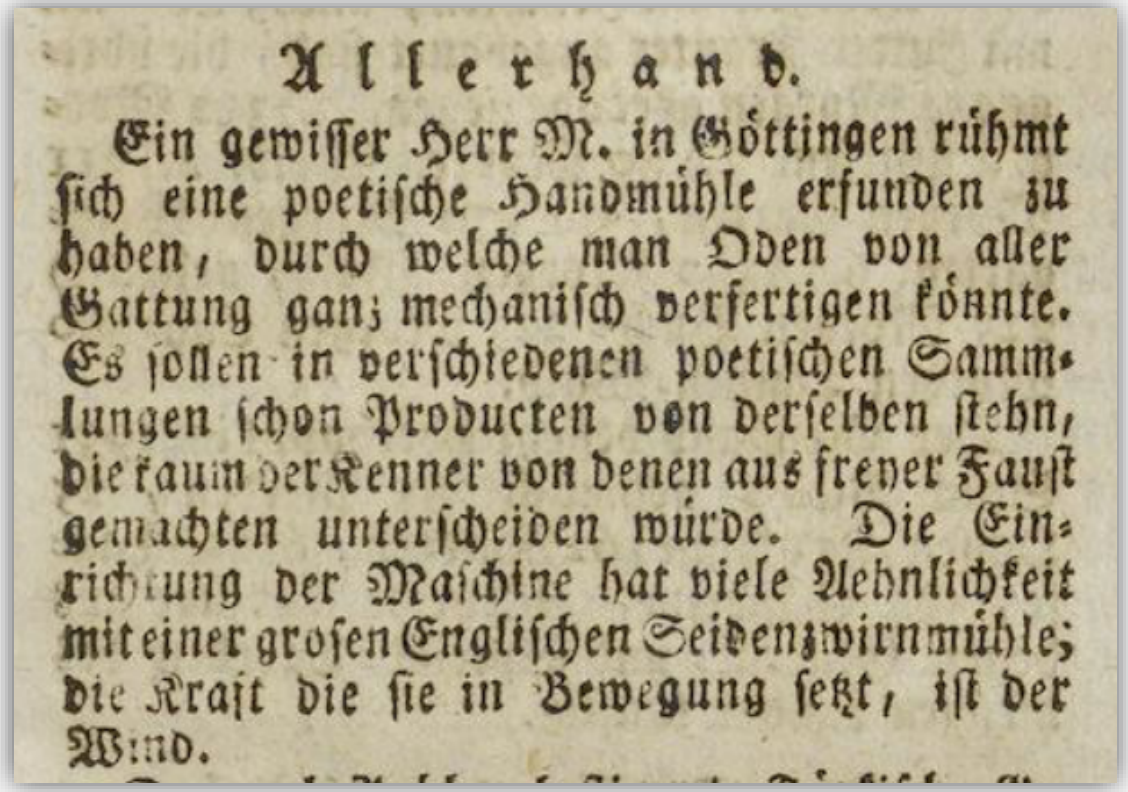
Using gen. AI tools in your PhD



ZLL / HDLE
Benjamin Angerer
July 8, 2026

Prehistory: Big claims

1777



Allerhand.

Ein gewisser Herr M. in Göttingen rühmt sich eine poetische Handmühle erfunden zu haben, durch welche man Oden von aller Gattung ganz mechanisch verfertigen könnte. Es sollen in verschiedenen poetischen Sammlungen schon Producten von derselben stehn, die kaum der Kenner von denen aus freyer Faust gemachten unterscheiden würde. Die Einrichtung der Maschine hat viele Aehnlichkeit mit einer grosen Englischen Seidenzwirnmühle; die Kraft die sie in Bewegung setzt, ist der Wind.

"A certain Mr M. in Göttingen boasts of having invented a poetic hand-mill by means of which odes of all kinds may be produced entirely mechanically. Several collections are said to already contain its products, which even experts can hardly tell from freely written ones. The machine's design bears semblance to a large English silk thread-mill; the force which sets it in motion is the wind."

Hesse-Darmstadt Privileged State Gazette, July 30, 1777

1958

The New York Times

TUESDAY, JULY 8, 1958

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

1959

IMPLICATIONS FOR INDUSTRY

Chairman: THE RT. HON. THE EARL OF HALSBURY, NRDC, London

National Physical Laboratory (1959). *Mechanisation of Thought Processes*. Vol. 1, p. vi, London: Her Majesty's Stationary Office.

UNIVERSITÄT BIELEFELD
Zentrum für Lehren und Lernen

Background

Misunderstandings vs widespread usage

Professor Flunks All His Students After ChatGPT Falsely Claims It Wrote Their Papers

Texas A&M University–Commerce seniors who have already graduated were denied their diplomas because of an instructor who incorrectly used AI software to detect cheating

BY MILES KLEE

MAY 17, 2023

A confused professor handed out 'incomplete' grades to his whole class because he misunderstood how AI chatbots work. ALYH M/ADOBE STOCK

https://www.rollingstone.com/culture/culture-features/texas-am-chatgpt-ai-professor-flunks-students-false-claims-1234736601/

Background

- Productivity gains not necessarily in line with manufacturers' promises

- At universities: Productivity does not equal good science or successful learning

THE WALL STREET JOURNAL.

SUBSCRIBE SIGN IN

LIFESTYLE • WORKPLACE

CEOs Say AI Is Making Work More Efficient. Employees Tell a Different Story.

How much time workers say the technology saves them on the job is vastly different from what executives report

By Lindsay Ellis
Jan. 21, 2026 5:00 am ET

Aa Listen (2 min)

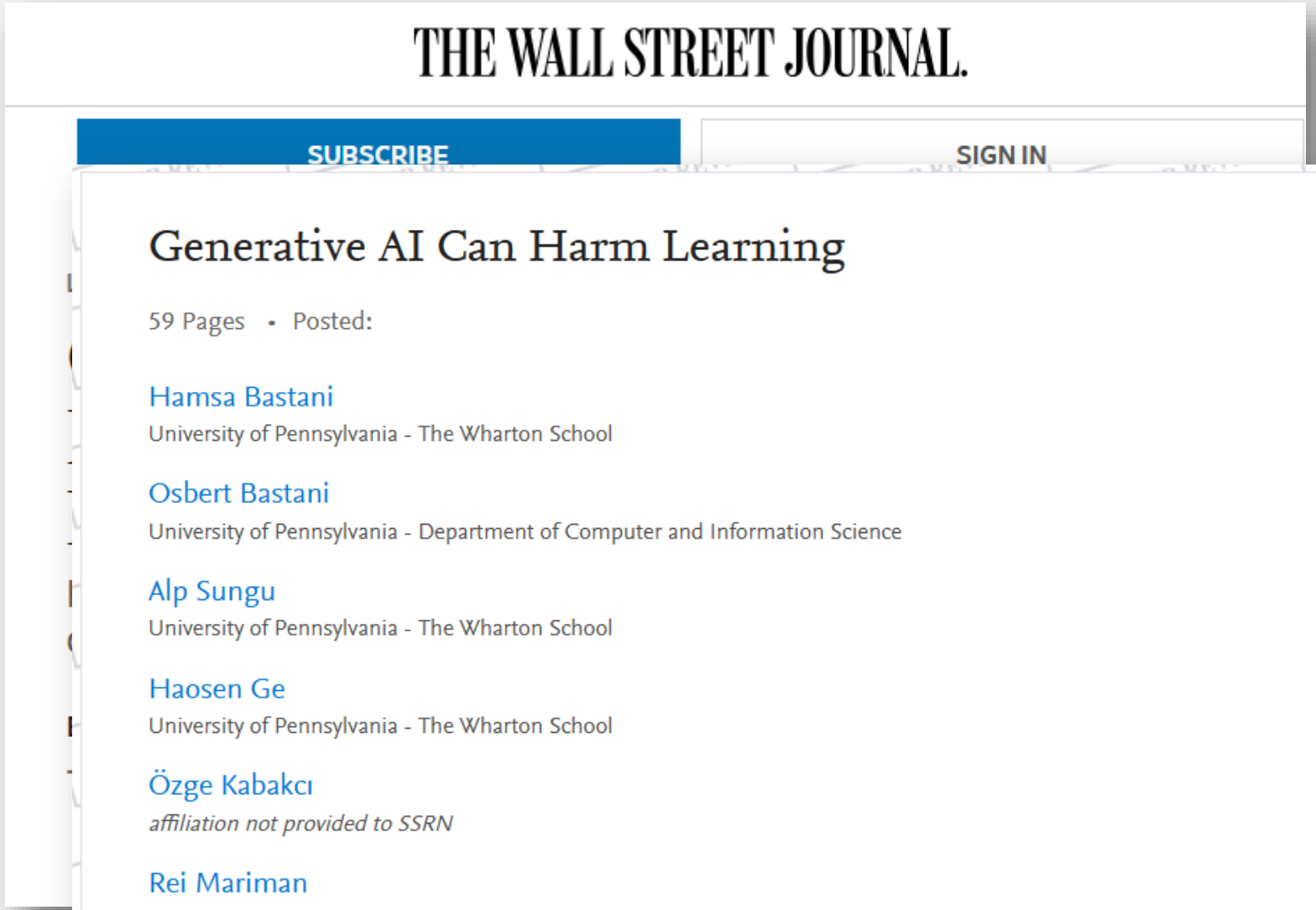
https://www.wsj.com/lifestyle/workplace/ceos-say-ai-is-making-work-more-efficient-employees-tell-a-different-story-6613ce9d
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4895486
https://www.microsoft.com/en-us/research/publication/the-impact-of-generative-ai-on-critical-thinking-self-reported-reductions-in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-workers/

UNIVERSITÄT BIELEFELD
Zentrum für Lehren und Lernen
d Lernen

Background

- Productivity gains not necessarily in line with manufacturers' promises

- At universities: Productivity does not equal good science or successful learning


THE WALL STREET JOURNAL.

SUBSCRIBE SIGN IN

Generative AI Can Harm Learning

59 Pages • Posted:

Hamsa Bastani
University of Pennsylvania - The Wharton School

Osbert Bastani
University of Pennsylvania - Department of Computer and Information Science

Alp Sungu
University of Pennsylvania - The Wharton School

Haosen Ge
University of Pennsylvania - The Wharton School

Özge Kabakcı
affiliation not provided to SSRN

Rei Mariman
Independent

Date Written: July 15, 2024

Abstract

Generative artificial intelligence (AI) is poised to revolutionize how humans work, and has already demonstrated promise in significantly improving human productivity. However, a key remaining question is how generative AI affects *learning*, namely, how humans acquire new skills as they perform tasks. This kind of skill learning is

https://www.wsj.com/lifestyle/workplace/ceos-say-ai-is-making-work-m
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4895486
https://www.microsoft.com/en-us/research/publication/the-impact-of-
in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-

UNIVERSITÄT BIELEFELD
Zentrum für Lehren und Lernen
d Lernen

Background

- Productivity gains not necessarily in line with manufacturers' promises

- At universities: Productivity does not equal good science or successful learning

https://www.wsj.com/lifestyle/workplace/ceos-say-ai-is-making-work-m
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4895486
https://www.microsoft.com/en-us/research/publication/the-impact-of-g
in-cognitive-effort-and-confidence-effects-from-a-survey-of-knowledge-

A vague, ambiguous term

- **AI in general:** Computational system designed to demonstrate skills for the possession of which a human would be attributed intelligence (i.e. also correcting grammar, solving algebra problems, etc.)

- **Moving target:** Once turned into an algorithmic procedure, skills are often deemed not to require “real intelligence” -> only new or unsolved problems remain “AI” in the long run (“AI effect”; McCorduck 2004)

- **Generative AI:** Systems based on current machine learning technologies which take descriptions as input and generate corresponding media (text, pictures, ...)

- **AI >> Generative AI >> large language models (LLMs) >> LLM-based chatbots**

Background

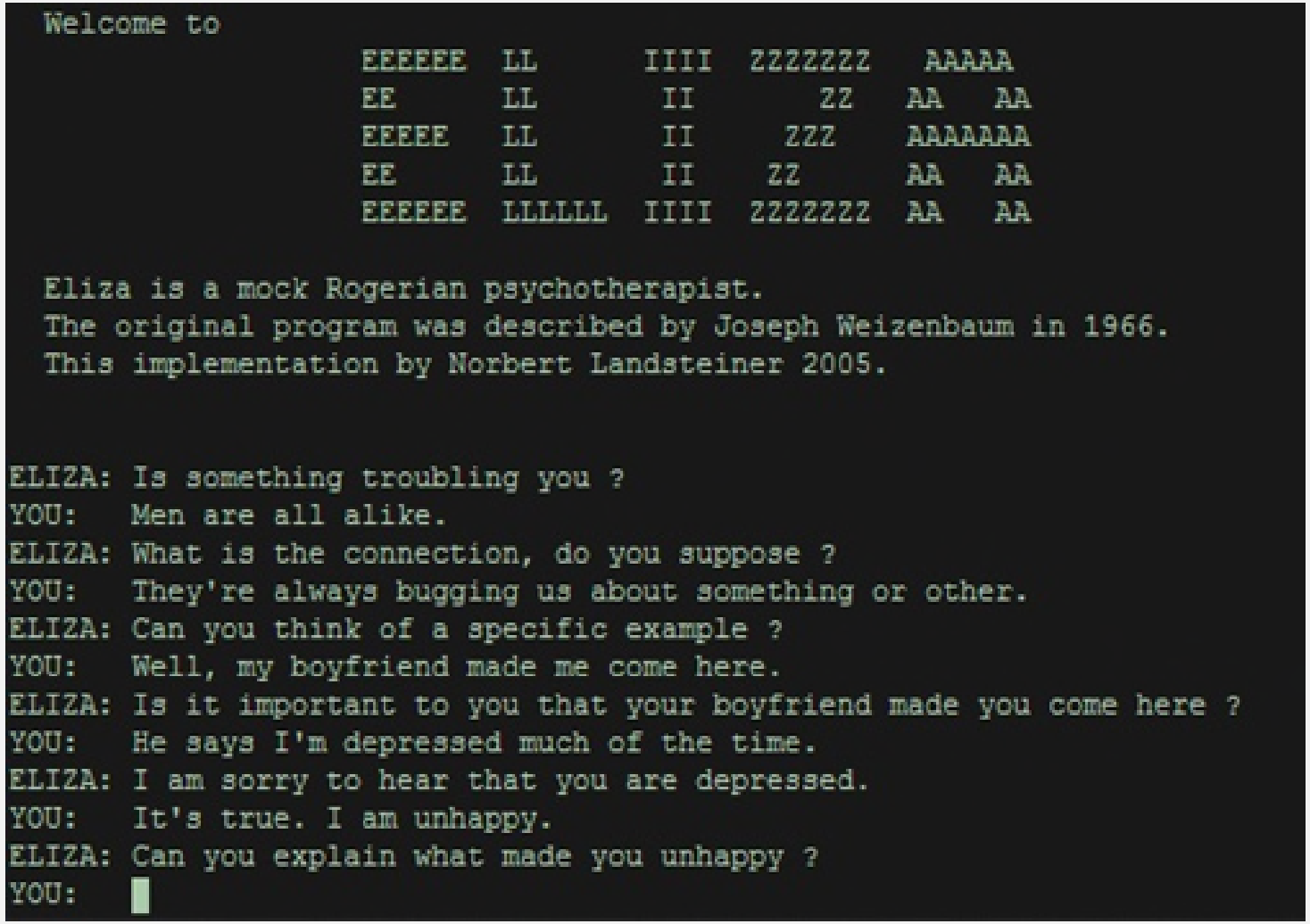
Large language models (LLMs): Statistical models of a large text corpus that can plausibly continue given text inputs. User access to such models is often provided in the shape of dialogue-like chat interfaces.

Models gain this ability through:

1. “Training” on billions of existing pieces of human-written text (web pages, books, articles, …)
2. Further training on bespoke input/output pairs (instruction following / question answering / …)
3. Extensive human feedback regarding (non-)desirable outputs

Background

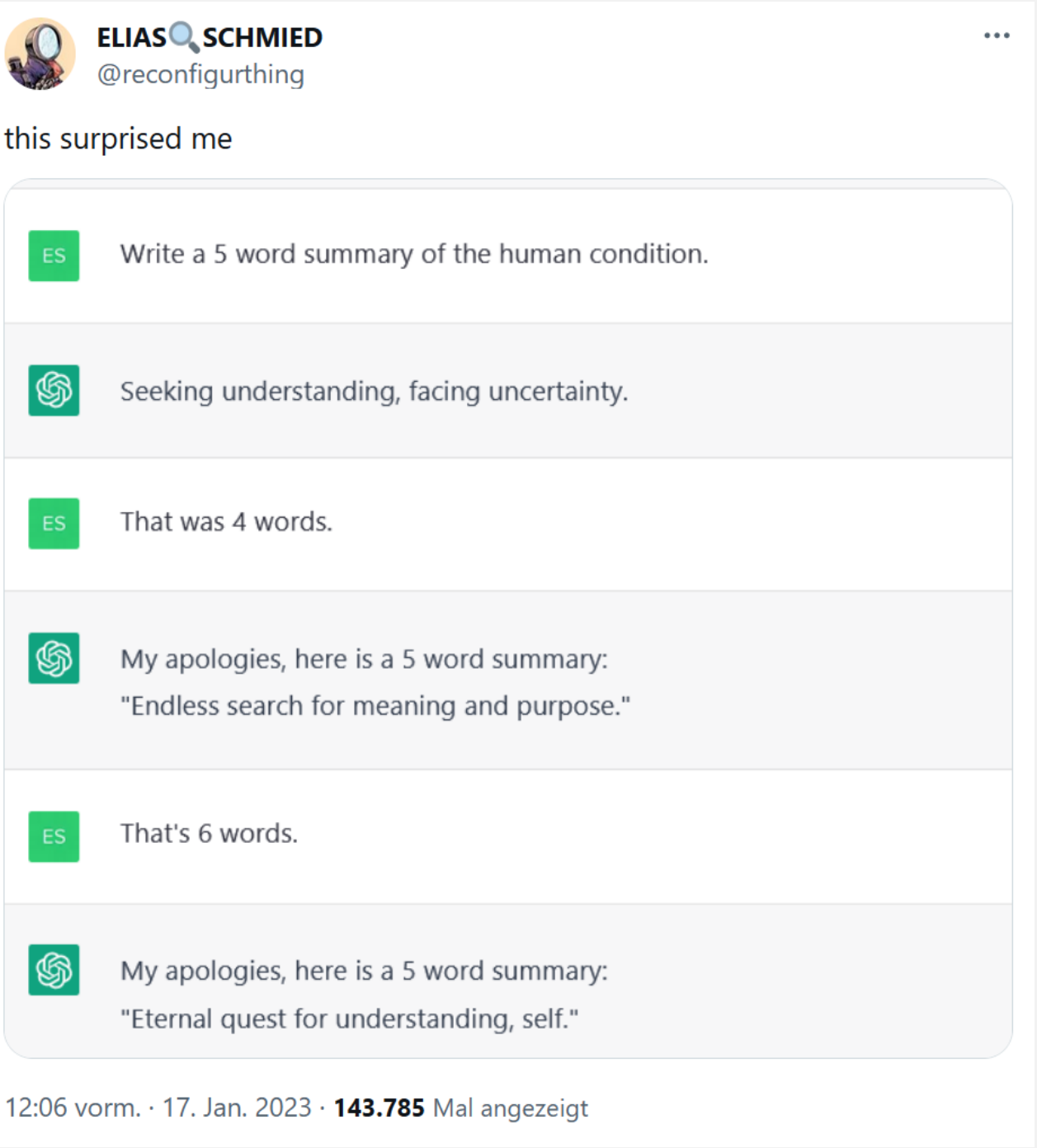
Use of natural language and dialogue creates implicit assumptions of human-like qualities (unconscious inference from outer to inner similarities)



Weizenbaum, Joseph (1966). "ELIZA—a computer program for the study of natural language communication between man and machine". *Communications of the ACM. **9**: 36–45.* doi:10.1145/365153.365168
Dingemanse, Mark (2026). *Interactional foundations for critical AI literacies*. Preprint. https://doi.org/10.5281/zenodo.19560684

Background

Capabilities of language models can be “weird” and don’t always conform to human expectations



https://nitter.net/reconfigurthing/status/1615123364372152321

Background

“Tendency towards frequency”

Frequent statements in training data are frequently reproduced: Stereotypical, toxic, normative statements, but also: large language communities, “big” authors, well-established fields of research, theories, etc.

prompt:
Toys in Iraq
are soldiers with guns

https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/

Background

"Tendency towards frequency"

Even capabilities are dependent on the frequency in which a task occurs in the training data

(E.g., the linear function 9/5x + 32 is the conversion formula for Celsius to Fahrenheit; shifting by 13 letters is a common cipher etc.)

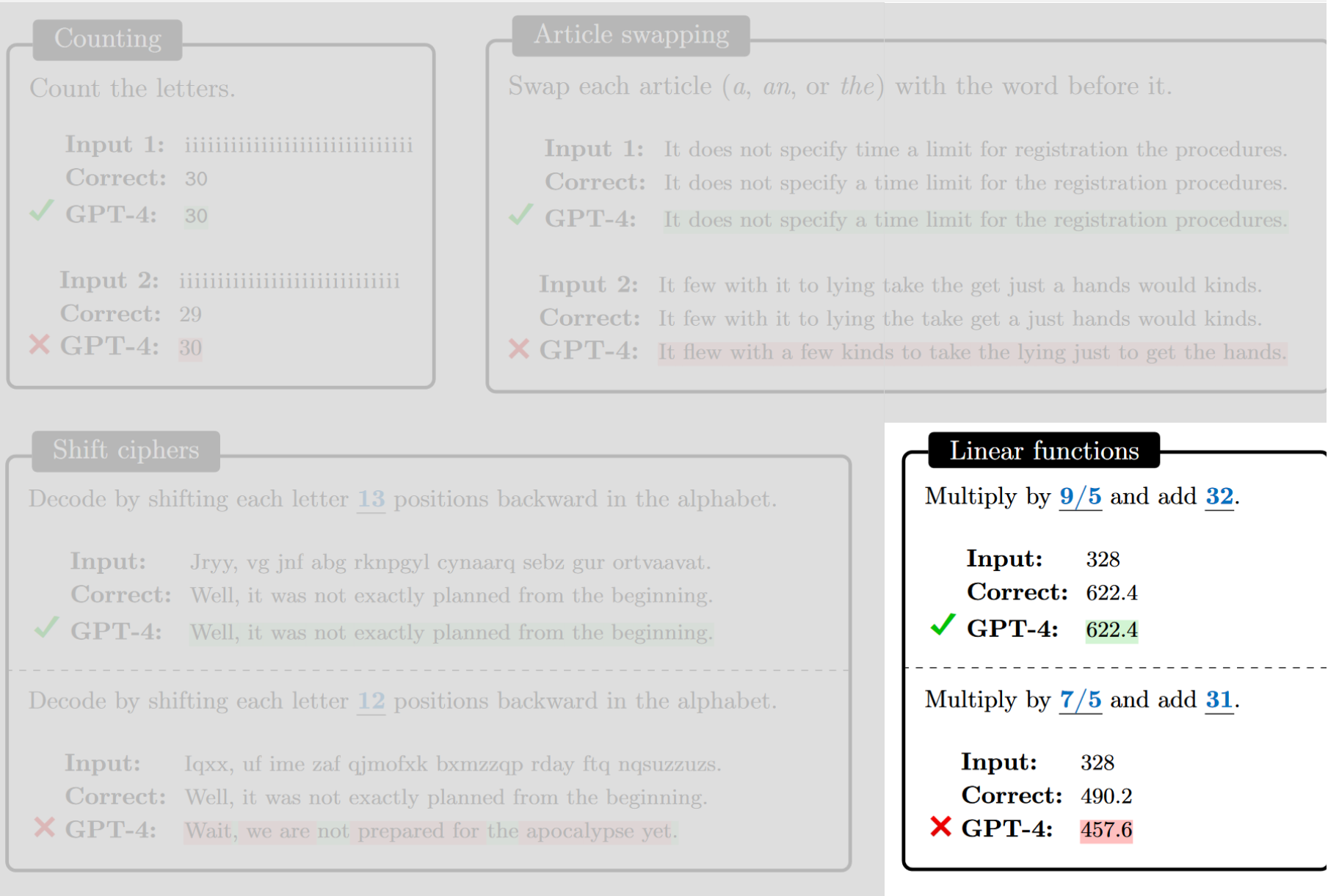
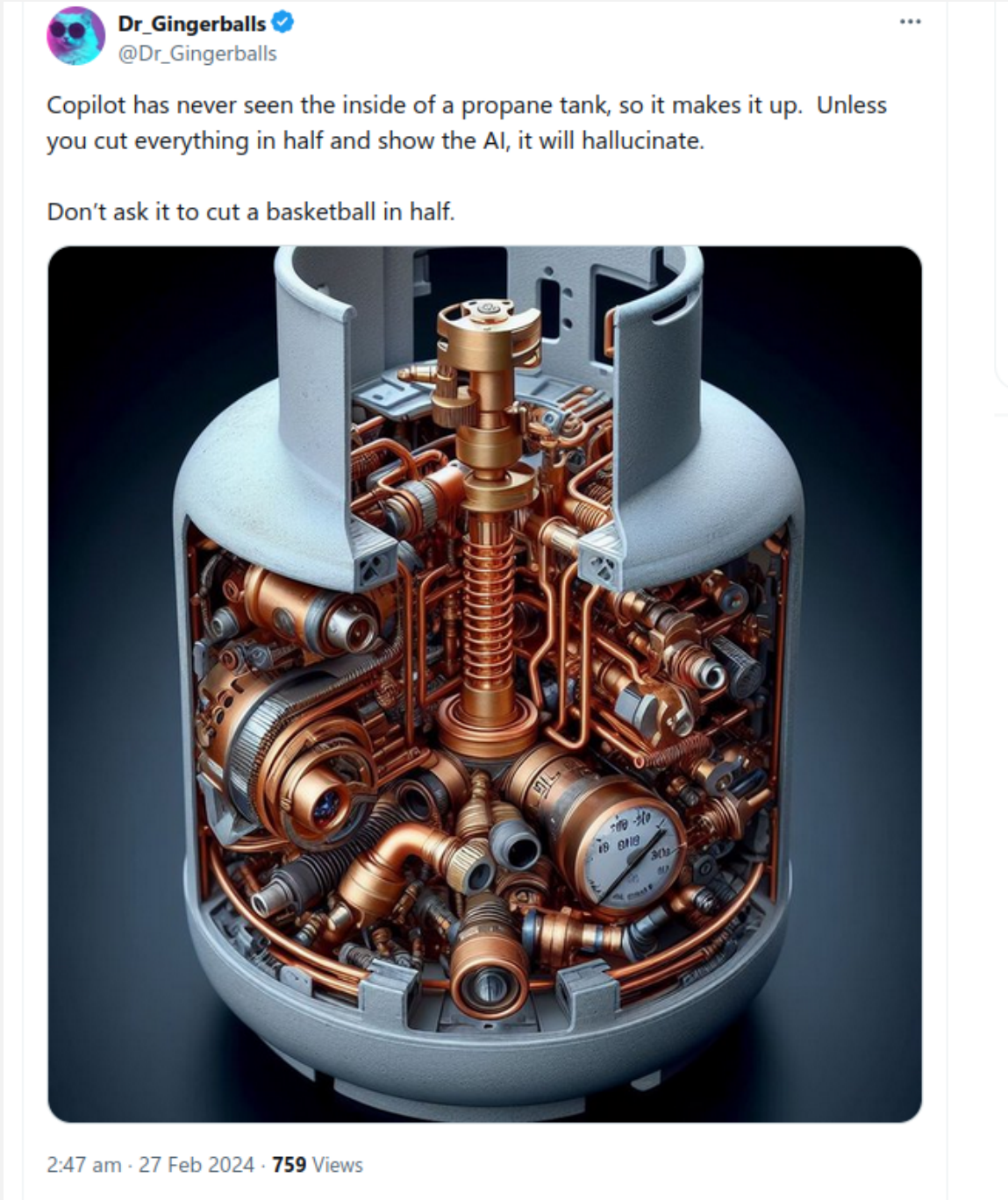


Fig 2 aus van McCoy et al. (2024).
https://www.pnas.org/doi/10.1073/pnas.2322420121

Background

Information that is missing in training data is “filled in” with seemingly plausible output (i.e. with whatever is most closely associatively related)



https://nitter.net/Dr_Gingerballs/status/1762293434038050920

The Interplay of Spatial Representations, Imagery, and Gestures in Spatial Problem-Solving: An Embodied Cognition Perspective

I. Introduction

Spatial problem-solving is a fundamental aspect of human cognition, with applications ranging from navigation to engineering and architecture (Hegarty & Waller, 2004). This essay examines the interplay between spatial representations, imagery, and gestures in spatial problem-solving, drawing on research by Freksa, Tenbrink, Tversky, Kosslyn, Pylyshyn, Goldin-Meadow, Alibali, and Hostetter. We will analyze the concepts, experimental tasks, and empirical studies related to the seminar topic, emphasizing the importance of representational format in human problem-solving. Additionally, we will explore contemporary approaches to spatial imagery that align with embodied cognition to elucidate the relationship between gestures and spatial representations.

II. Spatial Representations

Spatial representations are cognitive structures that enable individuals to encode, store, and manipulate information about spatial layouts (Freksa, 1991). These representations can be mental or external (e.g., maps, diagrams) and are shaped by an individual's experiences and interactions with their environment. Factors such as perspective, scale, and level of detail influence these representations (Tversky, 1991).

Research demonstrates that spatial representations significantly impact spatial problem-solving performance. Tenbrink and Winter (2009) found that participants who used more precise spatial language in describing object locations performed better in subsequent spatial tasks. This suggests that the way spatial information is linguistically encoded affects problem-solving abilities.

Tversky (2011) proposed that the structure of mental representations influences the ease or difficulty of solving spatial problems. Specifically, hierarchically organized mental representations, with clear distinctions between global and local information, are more effective in spatial problem-solving. This hierarchical organization allows for more efficient processing and manipulation of spatial information.

The format of spatial representations also plays a crucial role in problem-solving outcomes. Tenbrink et al. (2011) compared the use of maps versus verbal descriptions in spatial tasks. They found that participants using maps outperformed those using verbal descriptions, highlighting the importance of visually explicit representations in enhancing spatial task performance.

Having established the significance of spatial representations in spatial problem-solving, we can now explore the role of spatial imagery in facilitating the manipulation of these representations.

Angerer, Benjamin
Psychometric study on mental rotation. Says nothing about applications or problem solving

Angerer, Benjamin
A bit grand for a 6 page essay, but students sometimes also promise more than they deliver, so not immediately a red (genAI-)flag

Angerer, Benjamin
This talks about spatial representations, but in terms of a specific symbolic formalism, not about cognitive structures and their encoding, storage, manipulation

Angerer, Benjamin
This sort of says the opposite: spatial mental models (as opposed to images) were shown to be perspective-independent

Angerer, Benjamin
Given the very broad definition of spatial representation above – how could a person even solve a "spatial problem solving" task without the use of spatial representations? So in what sense to they improve performance?

Angerer, Benjamin
No such study is reported in this article

Angerer, Benjamin
Once more, this is a VERY generic claim. The more detailed ideas in the remaining paragraph can NOT be found in Tversky 2011 at all.

Angerer, Benjamin
No such study is reported in this article

III. Spatial Imagery

Spatial imagery refers to the ability to generate and manipulate mental images of objects and their spatial relationships (Kosslyn, 1980). This cognitive process involves using mental representations to manipulate spatial information, including transformations such as mental rotation, translation, reflection, and scaling (Kosslyn, 1980; Shepard & Metzler, 1971; Tversky, 1991).

Research indicates that spatial imagery plays a crucial role in spatial problem-solving abilities. Pylyshyn (2003) argued that mental imagery is essential for manipulating mental representations, enabling individuals to simulate and predict outcomes of various spatial transformations. However, it's important to note that Pylyshyn's perspective challenges the idea of depictive mental imagery, instead proposing that spatial reasoning relies on propositional representations and tacit knowledge about spatial relations.

Contemporary perspectives on spatial imagery align with embodied cognition principles, emphasizing the importance of sensorimotor experiences in shaping spatial imagery (Wilson, 2002). Wilson and Golonka (2013) proposed that spatial imagery is grounded in sensorimotor experiences, with the body playing an active role in its formation. This embodied approach offers a potential explanation for the connection between gestures and spatial representations, suggesting that gestures, as a form of sensorimotor simulation, may facilitate the manipulation of mental images.

Newcombe and Shipley (2015) found that individual differences in spatial imagery proficiency can predict performance in spatial tasks such as mental rotation and navigation. Their research suggests that individuals with higher spatial imagery abilities are more adept at mentally manipulating objects and solving spatial problems requiring such manipulation.

Research on the development of spatial cognition provides valuable insights into how spatial representations and imagery abilities evolve. Newcombe and Huttenlocher (2003) propose that spatial development progresses through several stages, from early sensorimotor experiences to more abstract representational abilities. For instance, young children initially rely heavily on egocentric spatial representations, gradually developing the ability to use allocentric frames of reference as they mature.

A meta-analysis conducted by Uttal et al. (2013) of spatial training studies revealed that spatial skills are malleable and can be enhanced through targeted interventions across all age groups. This malleability indicates that the dynamic nature of spatial representations extends beyond immediate problem-solving contexts to longer-term developmental trajectories. The authors also discovered that different types of spatial skills (e.g., mental rotation, spatial visualization) demonstrate varying developmental patterns and responsiveness to training, emphasizing the multifaceted nature of spatial cognition.D

These developmental perspectives highlight the importance of considering how the interplay between spatial representations, imagery, and gestures may change over time, potentially influencing an individual's approach to spatial problem-solving at different life stages.

Angerer, Benjamin
Quite generic definition; reference is an older book that I don't have at hand to check. But Kosslyn 1994 defined clear differences between spatial and visual imagery that would need better explanation here, and are e.g. mentioned in an easily accessible and readable review of the topic (Sima et al. 2013)

Angerer, Benjamin
Not exactly wrong, but a strange way of putting it and not really the point of this article

Angerer, Benjamin
Strange jump from the discussion of Pylyshyn with no link provided (pasted paragraph?)

Angerer, Benjamin
No mention of spatial imagery in this article (mental imagery in general though, also (non-imagined) spatial tasks)

Angerer, Benjamin
Imagery (of any kind) is not mentioned in this article at all.

Angerer, Benjamin
Theory and review chapter, not a study

Angerer, Benjamin
Distinct stages are a Piagetian idea not embraced by N&H. No time to look up their exact claims (this is a lengthy book).

First two pages from an essay generated in multiple steps. In red: Citations from sources that exist, but don't contain the statements for which they are being cited.

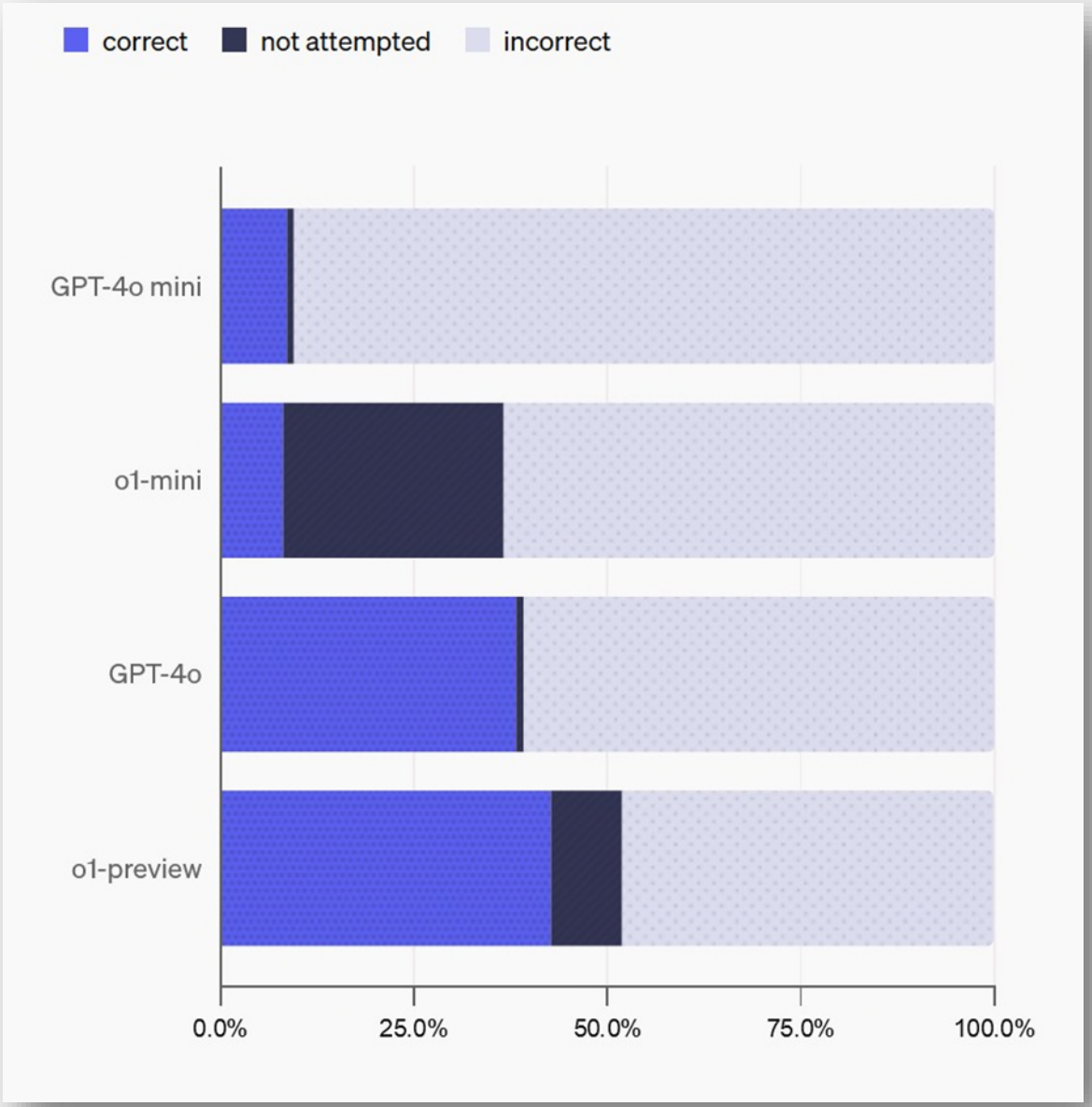
Inaccurate or false outputs

Sources/causes:

- Indirect
  - Missing training data
  - Statistical biases
- Direct
  - Training data contain falsehoods
  - (further effects in data modelling and inference)

Inaccurate or false outputs

- Still a problem



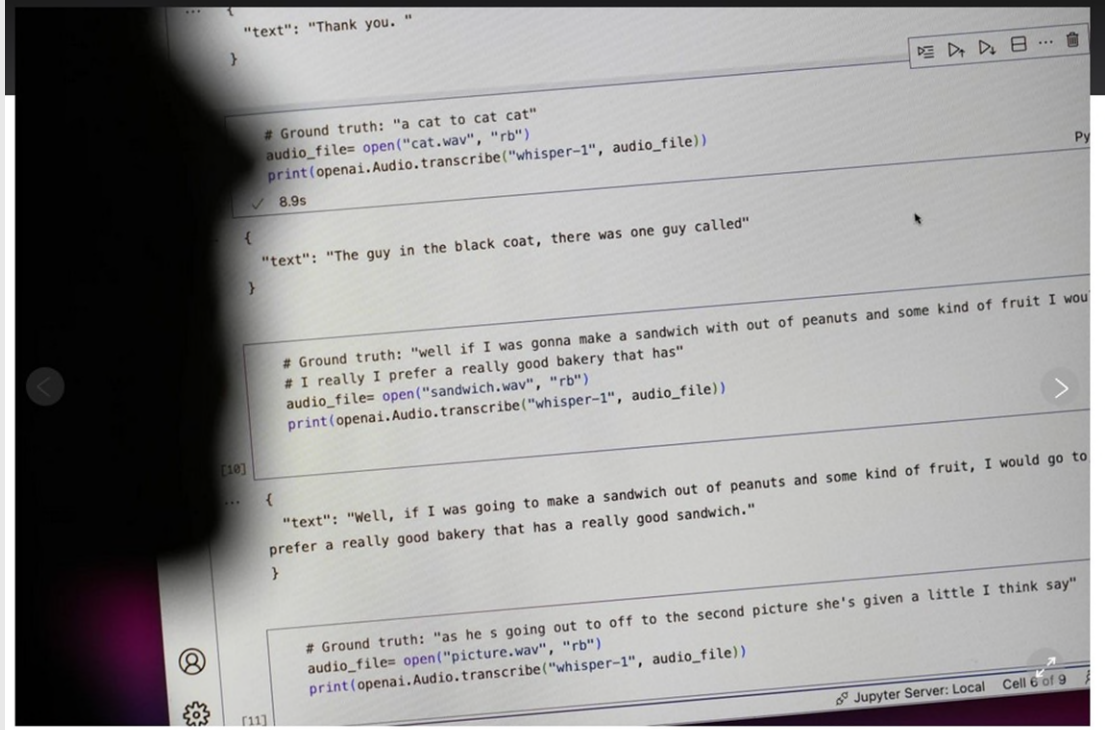
https://openai.com/index/introducing-simpleqa/
https://apnews.com/article/ai-artificial-intelligence-health-business-90020cdf5fa16c79ca2e5b6c4c9bbb14
https://www.bbc.co.uk/mediacentre/2025/new-ebu-research-ai-assistants-news-content

Inaccurate or false outputs

- Still a problem

TECHNOLOGY

Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said



1 of 6 | Assistant professor of information science Allison Koenecke, an author of a recent study that found hallucinations in a speech-to-text transcription tool, works in her office at Cornell University in Ithaca, N.Y., Friday, Feb. 2, 2024. The text preceded by "#Ground truth" shows what was actually said while the sentences preceded by ""text"" was how t... Read More

BY GARANCE BURKE AND HILKE SCHELLMANN
Updated 8:22 PM MEZ, October 26, 2024

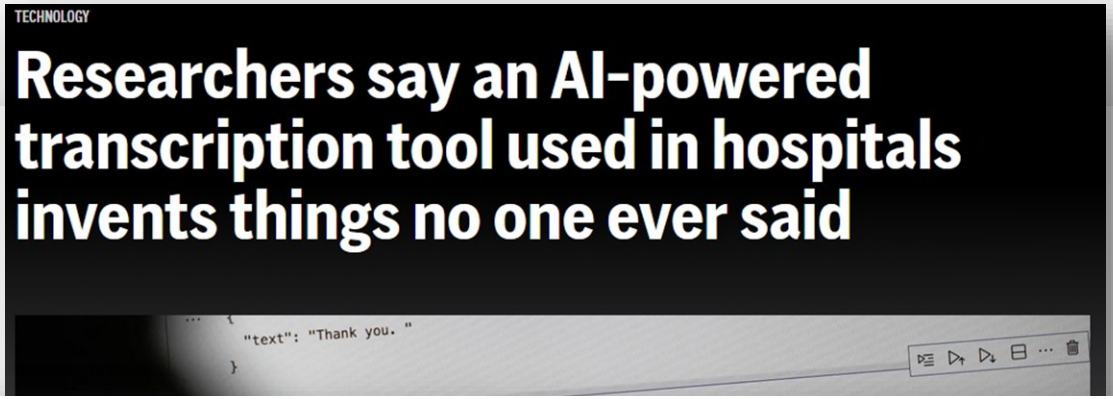
SAN FRANCISCO (AP) — Tech behemoth OpenAI has touted its artificial intelligence-powered transcription tool Whisper as having near "human level robustness and accuracy."

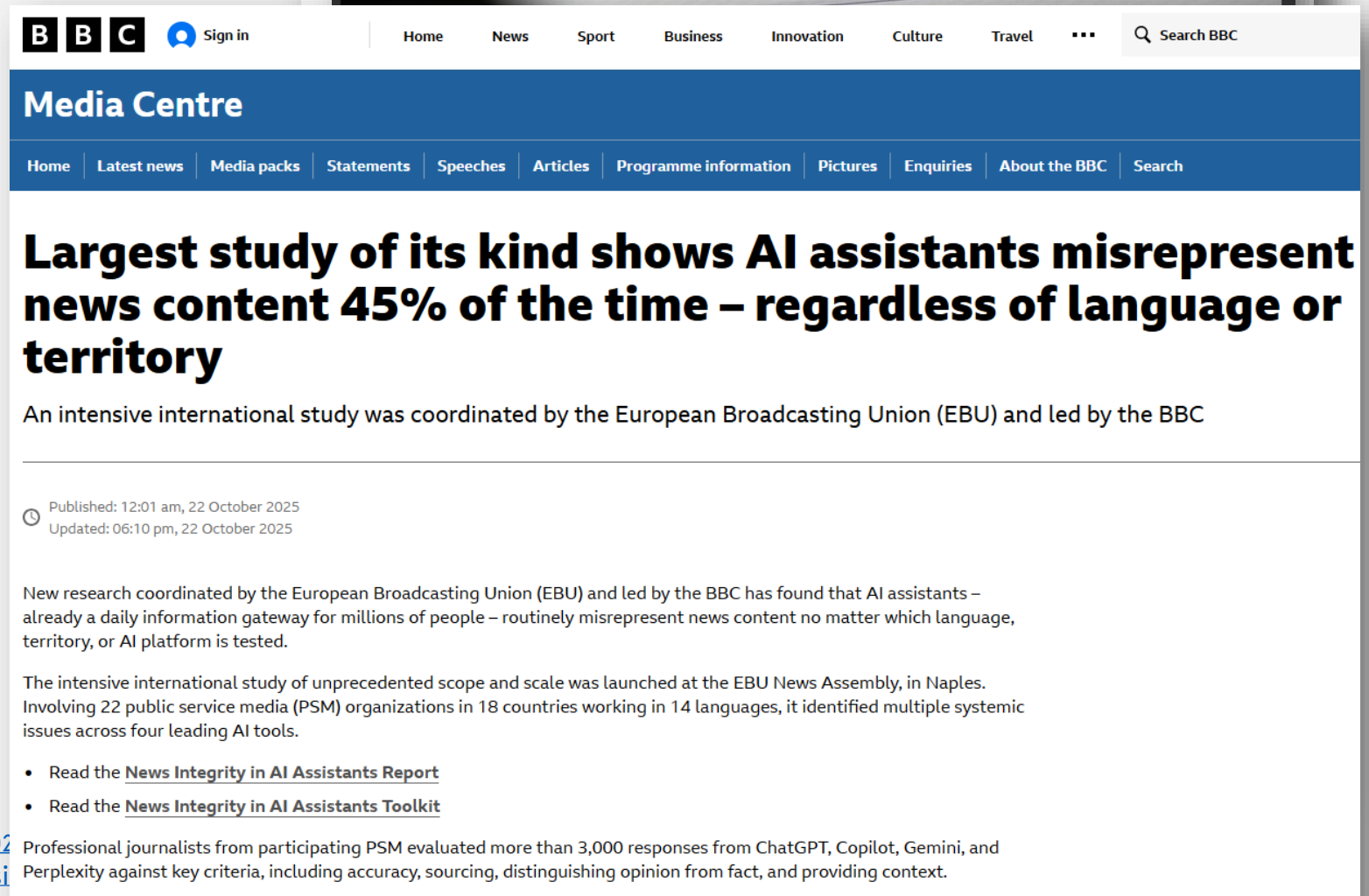
But Whisper has a major flaw: It is prone to making up chunks of text or even entire sentences, according to interviews with more than a dozen software engineers,

https://openai.com/index/introducing-simpleqa/
https://apnews.com/article/ai-artificial-intelligence-health-business-90020cdf5fa16c79ca2e5b6c4c9bb
https://www.bbc.co.uk/mediacentre/2025/new-ebu-research-ai-assistants-news-content

Inaccurate or false outputs

- Still a problem

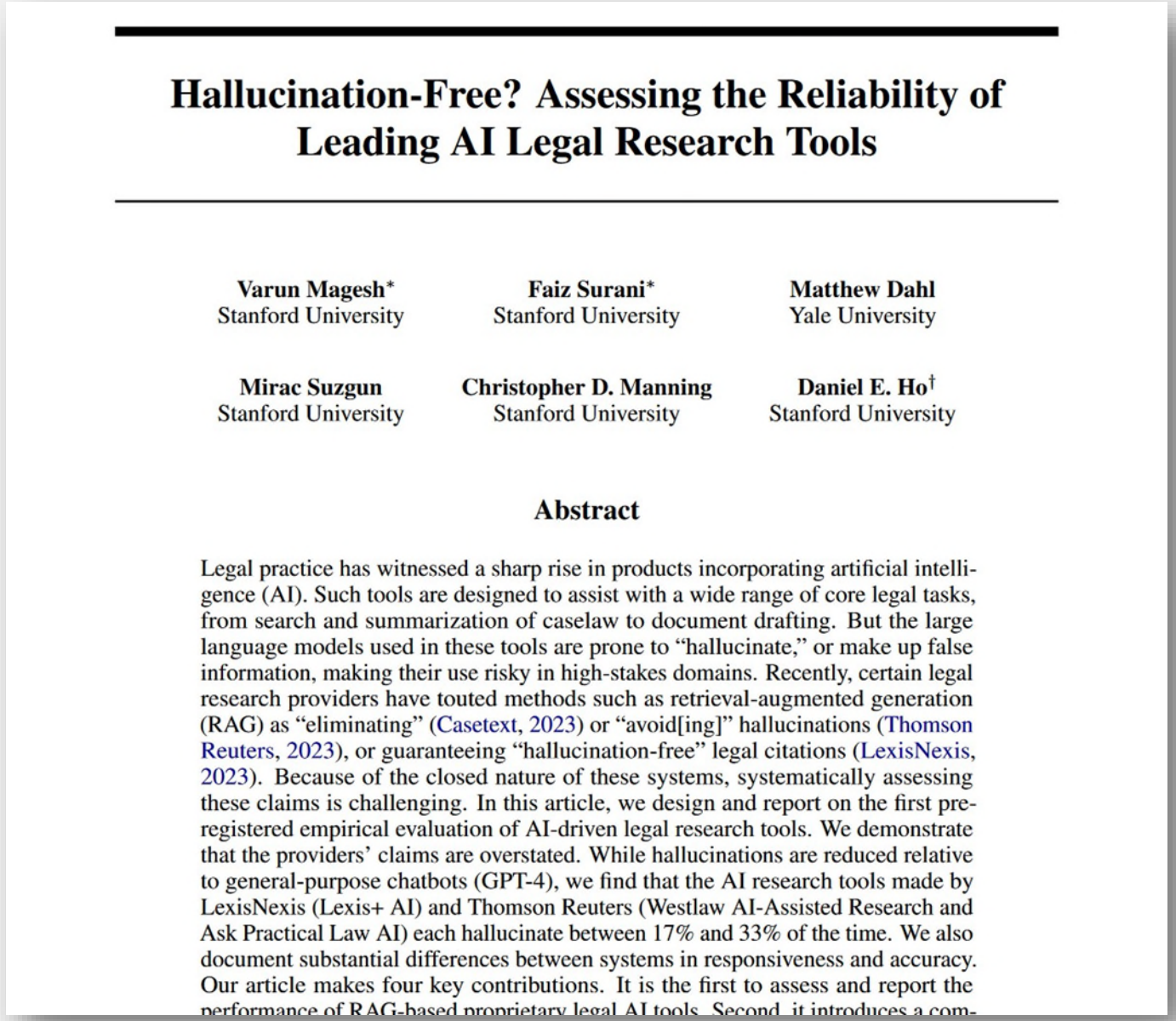




https://openai.com/index/introducing-simpleqa/
https://apnews.com/article/ai-artificial-intelligence-health-business-9002
https://www.bbc.co.uk/mediacentre/2025/new-ebu-research-ai-assi

Inaccurate or false outputs

- Still a problem
- Suggested solutions (such as field-specific RAG) or SAG just lower the percentage of inaccuracies (e.g. Stanford study about “legal AI tools - GPT4: 58-82% vs. Legal AI tools: 17-33%)



Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools

Varun Magesh* Stanford University
Faiz Surani* Stanford University
Matthew Dahl Yale University
Mirac Suzgun Stanford University
Christopher D. Manning Stanford University
Daniel E. Ho† Stanford University

Abstract

Legal practice has witnessed a sharp rise in products incorporating artificial intelligence (AI). Such tools are designed to assist with a wide range of core legal tasks, from search and summarization of caselaw to document drafting. But the large language models used in these tools are prone to “hallucinate,” or make up false information, making their use risky in high-stakes domains. Recently, certain legal research providers have touted methods such as retrieval-augmented generation (RAG) as “eliminating” (Casetext, 2023) or “avoid[ing]” hallucinations (Thomson Reuters, 2023), or guaranteeing “hallucination-free” legal citations (LexisNexis, 2023). Because of the closed nature of these systems, systematically assessing these claims is challenging. In this article, we design and report on the first preregistered empirical evaluation of AI-driven legal research tools. We demonstrate that the providers’ claims are overstated. While hallucinations are reduced relative to general-purpose chatbots (GPT-4), we find that the AI research tools made by LexisNexis (Lexis+ AI) and Thomson Reuters (Westlaw AI-Assisted Research and Ask Practical Law AI) each hallucinate between 17% and 33% of the time. We also document substantial differences between systems in responsiveness and accuracy. Our article makes four key contributions. It is the first to assess and report the performance of RAG-based proprietary legal AI tools. Second, it introduces a com-

https://www.damiencharlotin.com/hallucinations/
https://doi.org/10.48550/arXiv.2503.05777
https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries

Inaccurate or false outputs

- Still a problem
- Suggested solutions (such as field-specific RAG) or SAG just lower the percentage of inaccuracies (e.g. Stanford study about “legal AI tools - GPT4: 58-82% vs. Legal AI tools: 17-33%)

Medical Hallucination in Foundation Models and Their Impact on Healthcare

Yubin Kim[‡1], Hyewon Jeong[§1], Shan Chen[2], Shuyue Stella Li[3], Mingyu Lu[§3], Kumail Alhamoud[1], Jimin Mun[4], Cristina Grau[1], Minseok Jung[1], Rodrigo Gameiro[1], Lizhou Fan[2], Eugene Park[1], Tristan Lin[8], Joonsik Yoon[§5], Wonjin Yoon[2], Maarten Sap[4], Yulia Tsvetkov[3], Paul Liang[1], Xuhai Xu[7], Xin Liu[6], Daniel McDuff[6], Hyeonhoon Lee[5], Hae Won Park[1], Samir Tulebaev[§2], Cynthia Breazeal[1]

[1] Massachusetts Institute of Technology [2] Harvard Medical School [3] University of Washington [4] Carnegie Mellon University [5] Seoul National University Hospital [6] Google [7] Columbia University [8] Johns Hopkins University.

Abstract

Foundation Models that are capable of processing and generating multi-modal data have transformed AI’s role in medicine. However, a key limitation of their reliability is *hallucination*, where inaccurate or fabricated information can impact clinical decisions and patient safety. We define *medical hallucination* as any instance in which a model generates misleading medical content. This paper examines the unique characteristics, causes, and implications of medical hallucinations, with a particular focus on how these errors manifest themselves in real-world clinical scenarios. Our contributions include (1) a taxonomy for understanding and addressing medical hallucinations, (2) benchmarking models using medical hallucination dataset and physician-annotated LLM responses to real medical cases, providing direct insight into the clinical impact of hallucinations, and (3) a multi-national clinician survey on their experiences with medical hallucinations. Our results reveal that inference techniques such as Chain-of-Thought (CoT) and Search Augmented Generation can effectively reduce hallucination rates. However, despite these improvements, non-trivial levels of hallucination persist. These findings underscore the ethical and practical imperative for robust detection and mitigation strategies, establishing a foundation for regulatory policies that prioritize patient safety and maintain clinical integrity as AI becomes more integrated into healthcare. The feedback from clinicians highlights the urgent need for not only technical advances but also for clearer ethical and regulatory guidelines to ensure patient safety. A repository organizing the paper resources, summaries, and additional information is available at https://github.com/mitmedialab/medical_hallucination.

‡Corresponding Author: ybkim95@mit.edu
§Authors with MD degrees have contributed to the clinical expertise of this work.

https://www.damiencharlotin.com/hallucinations/
https://doi.org/10.48550/arXiv.2503.05777
https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries

Inaccurate or false outputs

- Still a problem
- Suggested solutions (such as field-specif
  RAG) or SAG just lower the percentage o
  inaccuracies (e.g. Stanford study about '
  AI tools - GPT4: 58-82% vs. Legal AI tools:
  33%)

Medical Hallucination in Foundation Models and Their Impact on Healthcare

Yubin Kim[‡1], Hyewon Jeong[§1], Shan Chen[2], Shuyue Stella Li[3], Mingyu Lu[§3], Kumail Alhamoud[1], Jimin Mun[4], Cristina Grau[1], Minseok Jung[1], Rodrigo Gameiro[1], Lizhou Fan[2], Eugene Park[1], Tristan Lin[8], Joonsik Yoon[§5], Wonjin Yoon[2], Maarten Sap[4], Yulia Tsvetkov[3], Paul Liang[1], Xuhai Xu[7], Xin Liu[6], Daniel McDuff[6], Hyeonhoon Lee[5], Hae Won Park[1], Samir Tulebaev[§2], Cynthia Breazeal[1]

AI Hallucination Cases

This database tracks legal *decisions*[1] in cases where generative AI produced hallucinated content – typically fake citations, but also other types of AI-generated arguments. It does not track the (necessarily wider) universe of all fake citations or use of AI in court filings.

While seeking to be exhaustive (**1730** cases identified so far), it is a work in progress and will expand as new examples emerge. This database has been featured in news media, and indeed in several decisions dealing with hallucinated material.[2]

If you have any questions about the database, a FAQ is available here.
And if you know of a case that should be included, feel free to contact me.[3]

Click to Download CSV

Last updated: 7 July 2026

Based on this database, I have developed an automated reference checker that also detects hallucinations: PelAIkan. Check the Reports in the database for examples, and reach out to me for a demo.

For weekly takes on cases like these, and what they mean for legal practice, subscribe to Artificial Authority.

Search cases, courts, tools, t

I want graphs

State

Argentina (8)

Case	Court / Jurisdiction	Date ▼	Party Using AI	AI Tool	Nature of Hallucination	Outcome / Sanction	Monetary Penalty
Pooja Ramesh Singh v. Jammu and Kashmir Bank Ltd. & Anr.	Supreme Court (India)	2 July 2026	Judge	Implied	Fabricated Case Law (3) False Quotes Case Law (3)	NCLT and NCLAT judgments set aside	—
Patterson v.	CA California	2	Pro Se	Unidentified	Fabricated	Monetary	500 USD

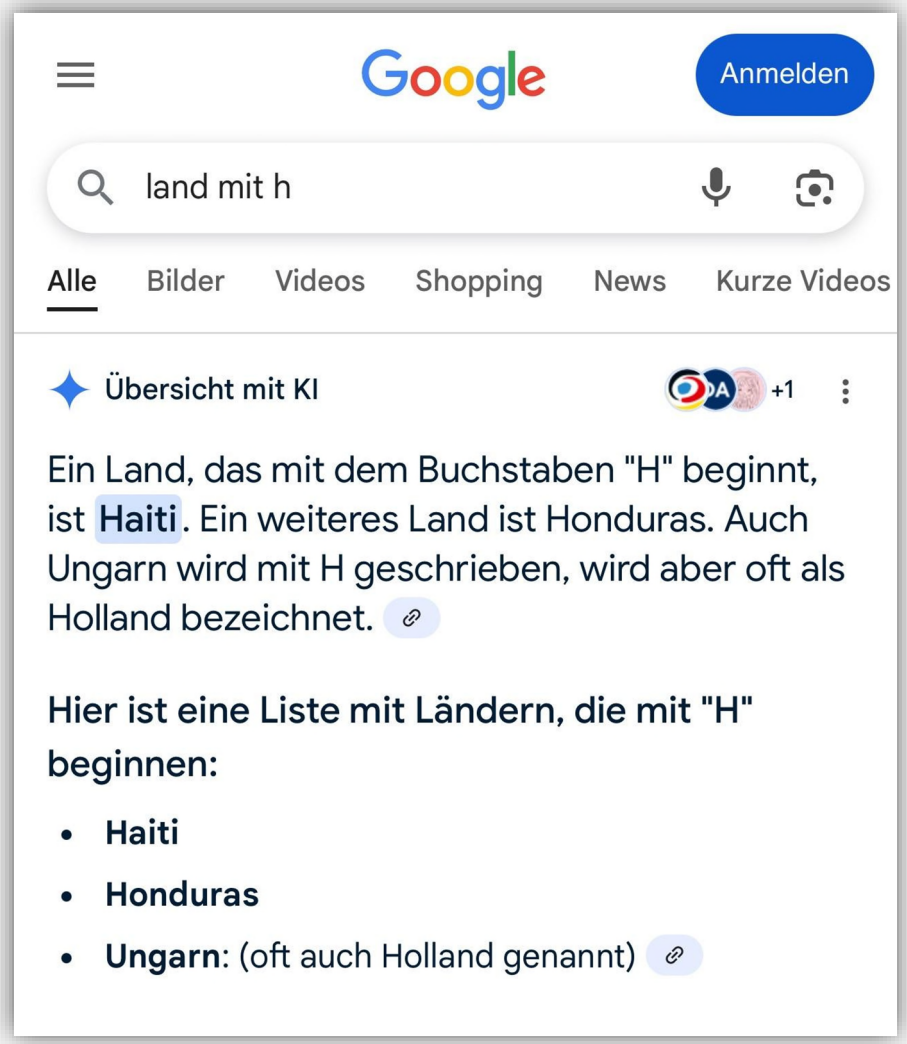
https://www.damiencharlotin.com/hallucinations/
https://doi.org/10.48550/arXiv.2503.05777
https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-o

Things to keep in mind

- LLMs generate plausible, but not necessarily true outputs
- Different fields, languages, methods, authors are covered differently well
- Problem of bias still remains
- Potential violation of copyright and privacy laws
- Main recommendations:

(1) Users always need to possess solid background knowledge about the topic they want to prompt an LLM for (no “oracle” usage)

(2) Outputs need to be verified by humans -> good scientific practice



In situations where these recommendations can’t be followed (or negate any advantage from using gen. AI in the first place) we recommend to just not use gen. AI tools!

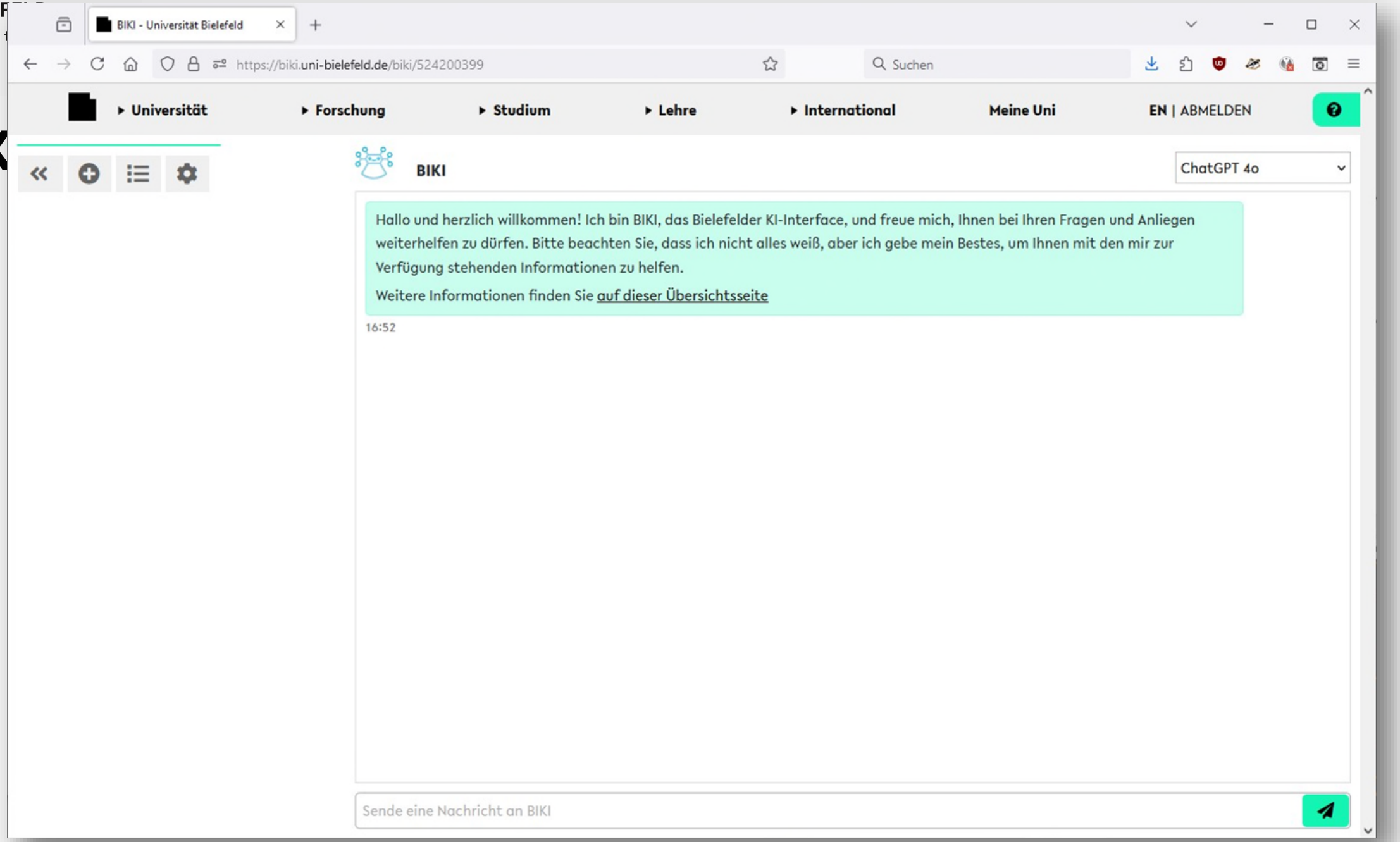
UNIVERSITÄT BIELEFELD
Zentrum für Lehren und Lernen

BIKI

UNIVERSITÄT BIELEFELD

Zentrum f

BIK

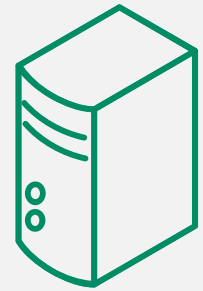


UNIVERSITÄT BIELEFELD
Zentrum für Lehren und Lernen

BIKI

Direct use

Users create accounts with providers, transfer their data to them directly; pay usage fees themselves

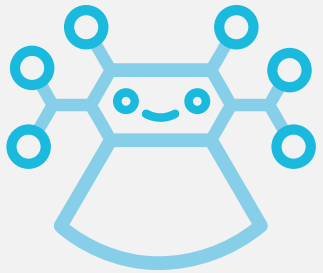


users

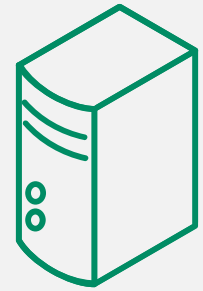
externally hosted gen. AI systems (e.g. OpenAI)

Use through BIKI

Users transfer data to university server (BITS)



BIKI transfers data to external provider



BIKI (web interface on BITS servers)

Possible applications

- Engaging with generative AI as phenomenon of interest
- Tasks for which there is sufficient background knowledge and generating + verifying has tangible advantages over doing it yourself
  - Steps with results that can easily be discarded (e.g. generating follow-up questions, variations on a theme)
  - Steps with easily verifiable results (e.g. paraphrases and/or summaries of one's own writing)

...

Responsibility for accuracy, legality etc. remains with authors

Gen. AI in studying and teaching: **Self-study course (German)**

- https://moodle.uni-bielefeld.de/course/view.php?id=7663
- Target groups: Lecturers and students; open for self-enrollment
- Time frame: about 90 min
- No technical course, but aims to teach conceptual, educational and legal bas
- Learning objectives:
  - Basic understanding of underlying technologies and relevant regulations
  - Usage of BIKI
  - Ability to assess the use of gen. AI systems in the context of one's own tasks

Gen. AI in studying and teaching: **Didactic guidelines (German)**

- https://www.uni-bielefeld.de/lehre/digitale-lehre/ki-tools/2025-09-Didaktische-Handreichung-genKI.pdf
- Similar range of topics as Moodle course
- Aims to collate relevant information in a single PDF

UNIVERSITÄT BIELEFELD
Zentrum für Lehren und Lernen

Thank you for your attention!