AI is a *Gungfly*: Why We Need New Mental Models for LLMs

Kevin J. Hunt
School of Art and Design
Nottingham Trent University
kevin.hunt@ntu.ac.uk

Sam Nemeth
Poppe and Partners
Amsterdam, Netherlands
sam.nemeth@gmail.com

Lars Erik Holmquist
School of Art and Design
Nottingham Trent University
lars.erik.holmquist@gmail.com

Large Language Models (LLMs) are now used for a large area of applications. However, everyday users need better mental models to understand and navigate these Artificial Intelligence (AI) systems. In HCI, mental models are historically shaped by deterministic computer systems delivering consistent outputs. LLMs overturn this principle, delivering what are essentially probabilistic outputs. This can be a strength, but the variability of outputs and notably the untruths can cause misalignment between deterministic user expectations and probabilistic system functionality. Because the inner workings of AI systems are hard to determine, we used synthetic ethnography to observe user inputs and system outputs. Using a synthetic probe, we observed variable iterative outputs from a simple calendar-based question. Based upon this, we suggest users accept the data landscape of LLMs is not reliable terra firma, but more like a gungfly – a Swedish term for a place where the ground is unstable and changeable.

Keywords : Artificial Intelligence. Large Language Models. Mental Models. Synthetic Ethnography.

1. INTRODUCTION

Since OpenAI launched ChatGPT, on 30 November 2022, natural language Artificial Intelligence (AI) systems have exploded in popularity. Engaging with Large Language Models (LLMs) is now commonplace in areas such as writing, coding and data retrieval. OpenAI's study of ChatGPT usage, from its launch to July 2025, showed 'Practical Guidance,' 'Seeking Information' and 'Writing' account for nearly 80% of interactions [4]. LLMs have also become integrated with search engines, expanding their sociotechnical role within everyday systems [12, 28, 30]. Although tech companies issue disclaimers about reliability, vast numbers of people augment their knowledge and skills using AI systems on a daily basis. LLMs are presented as go-to sources for information and guidance through sociotechnical integration, system positionality and media hype [21].

Before AI became popularised, everyday user expectations of computing were primarily shaped by deterministic interactions. Modern computing is based upon systems or algorithms producing results that are reliable and consistent, appropriate to each input, assuming a predictable sequence of states. Conventional computing is therefore defined by fundamentally deterministic principles, whereas LLMs are essentially probabilistic. This represents a substantial change in everyday HCI interactions, which we would argue is not clearly communicated to everyday users. Producing probabilistic outputs is a partial strength of LLMs that appears to be integral to their functionality. Probabilistic computing enables LLM replication of language and adaptability in outputs. However, while the capacity to communicate through natural language (through pattern making rather than understanding) encourages human engagement and suggests trust, it does not ensure accuracy or truth in outputs.

Research indicates that hallucinations – convincing outputs that are inaccurate, misleading or nonsensical – may be systemically endemic. Hallucinations appear to be an outcome of latent space data mapping and processing within neural networks that cannot entirely be engineered away [17, 18, 19, 24]. While we support efforts to minimise or eradicate hallucinations by the makers of the models, our interest is in the paradox that they are considered on the one hand problematic and on the other hand can have potential benefits such as an increase in the perceived creativity of the output. Concerns about hallucinations infer underlying user expectations (mental models) that successful LLM functionality requires outputs to be consistently reliable. Expectations of reliability, shaped by the legacy of conventional computing, therefore continues to influence the mental models of everyday users.

© *Hunt, Nemeth and Holmquist.*
Published by BCS Learning and Development Ltd.
Proceedings of BCS HCI 2026, UK

Mental models are the cognitive processes through which we make sense of real-world interactions, whereby past and ongoing experiences shape present or future expectations [6, 15, 20]. When LLMs hallucinate, such as including inaccurate elements within written outputs, it can become a source of frustration for users who identify the hallucinated content and also a source of consternation for an unaware user who later is confronted with hallucinated content. Misleading outputs may have personal, professional or even societal impacts in real world contexts.

Our paper argues that everyday user mental models for LLMs need to adapt to the probabilistic nature of these AI systems, and that better mental models will help users navigate the kinds of data landscapes being produced through AI. We used 'synthetic ethnography' [9] to 'probe' [8] the behaviour of a large AI powered search engine, based upon a simple calendar-based calculation. Our probe was designed around an everyday enquiry with an objectively mathematical answer. Observing the probabilistic variability within AI system outputs, where the same or similar input frequently produces a different output, provides insight into user experience and a glimpse of latent space data processing within a neural network [9].

To improve mental models, we propose thinking about this issue metaphorically. If everyday users think of conventional computing as terra firma, indicating solid ground and reliability, then the data landscape produced by LLMs is more like a gungfly[1] – the Swedish term for an inconsistent terrain of floating islands, which might unpredictably change underfoot.[2] The term 'gungfly' is not intended to convey negativity, it is about raising awareness for users to be cautious when exploring a new kind of landscape that is changeable and unfamiliar.



Figure 1: *An image of a 'gungfly' at Holmtjärnen, Sweden. Credit: Wikipedia contributor SiberianJay, licensed under the Creative Commons Attribution-Share Alike 4.0 International license.*

[1] C.f. Gungfly. https://sv.wikipedia.org/wiki/Gungfly

[2] The word is composed of the Swedish words 'gunga', as in rocking, and 'fly', as in disappear or escape. Although superficially similar, English terms such as 'quagmire' or 'swamp' indicate getting stuck or bogged down, which is different to the more subtle meaning of 'gungfly'.

2. RELATED WORK

2.1 Everyday User Mental Models

In 1943, Kenneth Craik conceptualised 'internal models' to explain how cognitive brain function enables human thought to 'model' external reality [6], leading cognitive science and psychology to identify the role of 'mental models' within human understanding of real-world interactions [6, 15]. Mental models inform individual human reasoning, decision making and aspects of behavior [16]; they are changeable subjective constructs shaped by beliefs, values and experiences [20].

In design, mental models become popularised by Donald Norman [20] who distinguished between the designer's 'conceptual model' and the user's 'mental model,' which is shaped by user interaction with a designed system [17]. If misalignment occurs between system function and how that system's functionality is presented (the 'image system') the user will develop the wrong mental model for interaction [17]. Mental models are integral to everyday activities, from simple to complex tasks, including interactions with AI models and related media [23].

Mental model research suggests users adopt human-type social norms with LLMs, compared to more abrupt commands for conventional computing [25]. This aligns with aspects of the 'media equation,' which suggests individual user interactions with media are shaped by social and natural responses that 'come from people, not the media themselves' [22]. As Emily Bender et al. argue in their paper about Language Models (LMs) as 'stochastic parrots,' the meaning produced through an interaction relies upon coherence being constructed by the user, existing 'in the eye of the beholder' [1].

Human-to-human linguistic communication is a jointly constructed activity, informed by narrative intent and real-world experiences [1]. If the person generating language is unknown to us, 'we build a partial [mental] model of who they are and what common ground we think they share with us, and use this in interpreting their words' [1]. However, the natural language text generated by LLMs 'is not grounded in communicative intent, any model of the world, or any model of the reader's state of mind' [1]. This has implications for mental models when engaging with LLMs because the user may be assuming communicative intent, without understanding the AI system outputs as probabilistic.

Everyday users seem to be caught in a complex scenario, where their mental models are shaped by factors including expectations of deterministic

reliability, presumptions of communicative intent, and the tendency to bring social and natural responses to human-computer interactions. Studies indicate that users tend to anthropomorphise (project human-like qualities onto) LLMs, which can include falsely projecting consciousness onto these systems [7, 21, 27]. Such misunderstandings place a sense of urgency upon mental model realignment. Users who regard LLMs as digital friends, companions, assistants or oracles of knowledge – beliefs that are often supported by AI hype [21] – can put misplaced levels of trust into LLM outputs, allowing factors like sycophantic or personalised interactions to cloud objective judgement [5, 27].

Contemporary mental model research includes surveys relating to conversational agents [3] and the role of mental models in emerging fields like Human-Centered Explainable AI (HCXAI) [14, 10, 28, 29]. However, mental models remain underexplored as a way of improving everyday user understanding of LLMs. Factors such as inference temperature, variance, cost functions and gradient descent provide important context for variable outputs across LLMs, including actions that can minimise randomness but not eradicate hallucinations [17]. However, everyday users are not typically aware of detailed technical insights. More immediate influences come from personal interactions with LLMs in everyday environments, media hype and mainstream representations of AI capability, and misplaced projections of human social meaning onto probabilistic outputs.

2.2 Synthetic Ethnography

AI models learn to represent vast datasets in condensed forms through high-dimensional vectors that help construct latent space [8]. Understanding how LLMs organise and process data is very challenging because AI systems are typically black boxes and the data is complex. Synthetic ethnography acknowledges these limitations and provides a methodology – informed by phenomenology (the observation of experiences) [26] – that observes human inputs and AI system outputs to better understand how humans experience AI systems [8].

Synthetic ethnography was initially developed for qualitative studies facilitating insights into algorithmic processes, training datasets and the latent space of Generative AI systems [9, 13]. It has since been adapted to study hallucinations as 'untruths' in LLMs [13]. Our approach uses a 'synthetic probe' [8] to understand user experience of Google Gemini as a multimodal LLM integrated with Google Search. A synthetic probe refers to a purposefully designed object, or in our case a designed series of interactions, used to probe 'a model's latent space to provoke the generation of outputs which can be analyzed iteratively and comparatively' [8].

3. PROBING AN AI-ENABLED SEARCH ENGINE

Everyday users are now turning to AI systems such as OpenAI's ChatGPT and Google's Gemini to get answers to factual questions, reinforced by the integration of AI responses in popular search engines and AI-based 'answer engines' such as Perplexity.[3] To understand the user experience of AI-enhanced search, we probed the dominating search engine, Google Search. We started each search in a new private window to avoid earlier searches or user profiles influencing the results.

Our probe was designed around a practical problem that many users might encounter: calculating the number of working days within a given period of calendar time. Although this can vary based upon national holidays, the answer can be verified objectively as a mathematical calculation.

The initial query we used was:

> how many working days are there between 8 of august 2025 and the 12th of December 2025

A correct answer would be 90 (in the country of the authors, where there would be one holiday day in the period), or 91 (in a country with no holidays in the period).

Our first query was performed on 17 December 2025. The first result to be presented to the user was a so-called AI overview, which appeared before any search results:

> There are approximately 70 working days between August 8, 2025, and December 12, 2025, assuming a standard Monday-Friday work week and excluding UK bank holidays, though the exact count depends on which specific public holidays (like Christmas bank holidays) fall within that period and if they're observed, but typically around 70-75 days is expected for this span.
>
> [Followed by a detailed breakdown resulting in a different answer, 85 days]

Although it acknowledges the possibility of UK bank holidays, both figures given by the AI summary, 70 and 85 days, would be incorrect in any country. Running the query again, on 18 December, the AI summary gave the following, substantially different, and more correct, answer:

> There are 91 working days between August 8, 2025, and December 12, 2025, assuming a

[3] https://www.perplexity.ai

> standard five-day work week (Monday to Friday) and not including public holidays.
>
> [Followed by a detailed breakdown resulting in the same answer]

This answer does not consider regional holidays, which is correctly acknowledged in the answer. We modified the search slightly and ran it again:

> how many working days are there between 8 of august 2025 and the 12th of december 2025 in the uk
>
> There are 85 working days between August 8, 2025, and December 12, 2025, in the UK (specifically England and Wales), assuming Monday to Friday are working days and the start and end dates are included. This number varies slightly in Scotland and Northern Ireland due to different bank holidays.

A detailed breakdown followed, before another answer was given, this time correct:

> [Continued from above]
>
> Answer: There are 90 working days between August 8, 2025, and December 12, 2025, in the UK, for all regions including England, Wales, Scotland, and Northern Ireland, assuming a standard Monday to Friday working week and including the start and end dates as potential working days.

Finally, we tried the same query but for a different country:

> how many working days are there between 8 of august 2025 and the 12th of december 2025 in Sweden
>
> The number of working days between August 8, 2025, and December 12, 2025, in Sweden, inclusive of both start and end dates (if they are working days), is 95 working days.
>
> [A detailed breakdown followed, resulting in the same answer, 95 days]

This time the answer of 95 days is not only incorrect but also impossible, even accounting for the regional holidays, because there are only 91 possible working days (i.e. weekdays) in the time period.

Taken together, this selective sample shows that even the results from a leading AI-assisted search engine can be inconsistent, in this case giving anything from 70 to 95 days as the answer. Even when arriving at the correct answer for the UK (90 days) this was preceded by an incorrect answer (85 days), making it impossible for the user to identify which one is valid. Uncertainty was only indicated once, in the term ‘approximately 70 working days.’ In all other cases the answers are stated as fact, even when they differ within the same response.

The experience meeting the user is like being upon shifting ground, akin to navigating a gungfly. Responses change from one query to the next, without any way to ascertain if they are correct or not. To make a reasonable judgement requires a different user mental model to the one used for a traditional search engine (conventional computing), which assumes a (deterministic) factually correct and consistent answer to the same query each time.

4. DISCUSSION: THE GUNGFLY METAPHOR

Our paper identifies a different way of thinking about everyday user experiences of LLMs. We are highlighting a problem – the misalignment of mental models – and speculatively proposing realignment through user consideration of LLMs being like a gungfly. This metaphor has been chosen as a memorable provocation, based upon the comparison between terra firma and a gungfly landscape of floating islands (pictured above). It provides an intuitive way of conceptualizing the changed grounds of HCI when moving between the determinism of conventional computing and the probabilistic computational processing of LLMs.

LLMs are not fundamentally designed to provide solid ground and consistency. Instead, they produce variable probabilistic outputs tailored to the likelihood of coherent meaning, within the context of any given interaction with an individual user. The grounds upon which an everyday user and LLM interact always includes the potential for variable outputs, manifested in the way LLMs construct a natural language formation for each output. When randomness in outputs is minimised within a LLM there is still no guarantee its outputs will be factually accurate, even if they are more consistent.

LLMs pattern making capabilities enable them to produce high levels of coherence using natural language, but this does not necessarily correlate with reliable transfer of information. The widespread integration of LLMs within public-facing systems, such as customer services and search engines, infers trustworthiness for data processing, data management, and data retrieval [21, 28, 30]. Yet LLM system functionality may not be reliably or consistently aligned with an underlying factual record of information [1, 13].

This represents a gap between everyday user mental models for LLM interactions and system functionality. AI hype and media reporting tends to suggest that ongoing public engagement with AI systems, alongside ongoing scaling of models, is on a progressive trajectory of constant improvement [21]. But AI systems are new forms of data processing and how well they will perform over time remains unknowable. What we do know is that user mental models can adapt based upon experience, perception and beliefs.

We propose using a visual and linguistic metaphor to reshape mental models and argue it is appropriate because of the way LLM functionality is tied to linguistic structure. As already stated, the natural language produced by a LLM is based upon probability without direct connection to real world experience. As Bender et al. suggest, LLMs produce language structures that are unmoored from direct knowledge of the world, which is the basis of human language and communication [1]. The concept of LLMs being like a gungfly similarly refers to a form of floating landscape that has no direct roots in the earth. LLMs provide an ungrounded system of information transfer through sophisticated pattern making, based upon data sets and training focused primarily upon language structure.

One way of thinking about LLMs is as 'a new kind of cultural and social technology,' [11] whereby AI systems allow humans to access and take advantage of vast amounts of information that other humans have accumulated (including knowledge from real world experiences). This information has been processed as huge data sets in machine-readable form [11]. LLMs have the capability to selectively make probabilistically identified segments of this knowledge available to any user, while also enabling those vast data sets of information to be 'reorganized, transformed, and restructured in distinctive ways.' [11]

However, the opportunity LLMs offer of human engagement with data processing on previously unknown levels is counterbalanced by the fact these systems only process data patterns for interpretation; they do not intrinsically 'understand' the data patterns they are producing. Importantly, these systems are trained to find a pattern if they are prompted to do so, even when the pattern is formulated by producing outputs beyond the edges of their training data [13]. Pattern recognition in LLMs includes the potential to share, reveal, combine, distort and fabricate data, among other outcomes, without awareness of which of these qualities is operational or where one process may have seamlessly blended with another. Hallucinations, producing untruths, for example, are typically integrated within a coherent presentation of information, where some content is reliable and some is fabrication [13].

Realigning mental models around the concept of AI being like a gungfly is not intended to convey negativity. Gungfly has a different meaning to the English concept of a quagmire or bog. The gungfly concept is intended to evoke caution and curiosity in the user, aligned with the awareness of someone wishing to explore a new kind of terrain that is unfamiliar and that they do not fully understand. Gungfly infers a (data) landscape that can change quickly and where a pathway that may appear solid and reliable might instead be unstable and unreliable. The user should be self-aware of the shifting grounds they are traversing, and this awareness should enhance their ability to navigate and explore without getting into trouble.

Everyday user mental models that consider AI systems to be like a gungfly, notably LLMs, will be better prepared for hallucinations and other probabilistic outputs that are disconnected from factual or reliable information. It is a more realistic way of thinking about LLMs, based upon their actual functionality, rather than considering them as oracles of knowledge or digital assistants. It places emphasis upon the user to take responsibility for appropriately interpreting information.

Importantly, realigning mental models might also support and encourage new or different ways for users to interact with LLMs. The significant potential within these systems as social and cultural technologies is, in part, related to how they might help humans identify new patterns of knowledge or different ways of understanding data sets that have not previously been recognised. Such insights are most likely to occur by exploring and probing LLMs to see what unexpected, creative, or unpredictable ways they have of organizing data, and then experimenting with and assessing the value of those insights for meaningful real-world application.

5. CONCLUSION

Our paper is partly a provocation and partly proposes a new way of thinking about human interaction with LLMs, framed around mental models as an under-valued area in HCI research of AI systems. We are identifying a problem for further consideration and research, while also offering a potential way forward through the realignment of mental models. The concept of the gungfly is being shared as an intuitive way for everyday users to potentially reappraise their understanding of LLM functionality. The aim is to better align everyday user perception of these systems as probabilistic, by clearly situating AI systems in relation to conventional computing. In addition, realigned mental models for LLMs might encourage different forms of interaction that lean into the creative potential within probabilistic outputs. This could help to position LLMs within different areas of HCI, aligned with their probabilistic functionality, including moving away from concerns about hallucinated content as a possibly unresolvable problem.

6. REFERENCES

[1] Bender, E.M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Conference on Fairness, Accountability,

and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922

[2] Blomberg, J., Burrell, M., and Guest, G. (2002) An Ethnographic Approach to Design. In Jacko, J.A. and Sears, A. (eds) The Human-Computer Interaction Handbook. Lawrence Erlbaum Associates, New Jersey.

[3] Cena, F. and Grasso. F. (2025) Exploring Users' Mental Models of Conversational Agents: A Systematic Review. Behaviour & Information Technology, 45(10), 2189–2212. https://doi.org/10.1080/0144929X.2025.2573436

[4] Chatterji, A., Cunningham, T., Ong, C., Shan, C., Deming, D., Hitzig, Z. and Wadman, K. (2025) How People Use ChatGPT. National Bureau of Economic Research: Working Papers. https://doi.org/10.3386/w34255

[5] Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., Jurafsky, D. (2026) Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence. Science 391, eaec8352 (2026). https://www.science.org/doi/epdf/10.1126/science.aec8352

[6] Craik, K.J.W. (1943) The Nature of Explanation. Cambridge University Press, Cambridge.

[7] Dawkins, R. (2026) When Claudia Met Claudius: So Are They Really Conscious? UnHerd. 5 May 2026. https://unherd.com/2026/05/when-claudia-met-claudius/

[8] de Seta, G. (2024) Synthetic Probes: A Qualitative Experiment in Latent Space Exploration. Sociologica, 18, 2, 9–23. https://doi.org/10.6092/issn.1971-8853/19512

[9] de Seta, G., Pohjonen, M., and Knuutila, A. (2024) Synthetic Ethnography: Field Devices for the Qualitative Study of Generative Models. Big Data & Society, 11, 4. https://doi.org/10.1177/20539517241303126

[10] Ehsan, U., Watkins, E.A., Wintersberger, P., Manger, C., Kim, S.S.Y., van Berkel, N., Riener, A. and Riedl, M.O. (2024) Human-Centered Explainable AI (HCXAI): Reloading Explainability in the Era of Large Language Models (LLMs). In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3613905.3636311

[11] Farrell, H., Gopnik, A., Shalizi, C., Evans. J. (2025) Large AI Models are Cultural and Social Technologies. Science. 13 Mar 2025, Vol 387, Issue 6739: pp. 1153-1156. https://doi.org/10.1126/science.adt9819

[12] Gautam, S., Venkit, P.N., and Ghosh, S. (2024) From Melting Pots to Misrepresentations: Exploring Harms in Generative AI. In CHI '24: Generative AI and HCI workshop in CHI '24, May 11–16, 2024, Honululu, Hawaii. ACM, New York, NY, USA, 7 pages.

[13] Holmquist, L.E. and Nemeth. S. (2025) Don't Believe Anything I Tell You, It's All lies!": A Synthetic Ethnography on Untruth in Large Language Models. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3706599.3719939

[14] Hughes, N., Jia, Y., Sujan, M., Lawton, T., Habli, I. and McDermid. J. (2026) Design Principles for Human-Centred Explainable AI: A Scoping Review. ACM Trans. Interact. Intell. Syst. 16, 1, Article 4 (January 2026), 26 pages. https://doi.org/10.1145/3771720

[15] Johnson-Laird, P.N. (1983) Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Harvard University Press, Cambridge, Mass.

[16] Jones, N.A, Ross, H., Lynam, T., Perez, P. and Leitch, A. (2011) Mental Models: An Interdisciplinary Synthesis of Theory and Methods. Ecology and Society, 16, 1 (46). http://www.ecologyandsociety.org/vol16/iss1/art46/

[17] Kalai, A.T., Nachum, O., Vempala, S.S., and Zhang, E. (2025) Why Language Models Hallucinate. OpenAI. https://doi.org/10.48550/arXiv.2509.04664

[18] Mickle, T., Metz, C., Freedman, D., Terol, T.M., and Collins, K. (2026) How Accurate Are Google A.I.s Overviews. New York Times, April 7, 2026: https://www.nytimes.com/2026/04/07/technology/google-ai-overviews-accuracy.html

[19] Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S. and Farajtabar, M. (2024) GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models. ArXiv: https://doi.org/10.48550/arXiv.2410.05229

[20] Norman, D.A. (1988) The Design of Everyday Things. Basic Book, New York.

[21] Placani, A. (2024) Anthropomorphism in AI: hype and fallacy. AI Ethics 4, 691–698 (2024). https://doi.org/10.1007/s43681-024-00419-4

[22] Reeves, B. and Nass, C.I. (1996) The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Cambridge University Press, Cambridge.

[23] Rezwana J, and Maher, M.L. (2025) An Exploration of Mental Models of AI in Human–AI Co-Creativity: A Framework and Insights. ACM Trans. Interact. Intell. Syst. 15, 4, Article 24 (December 2025), 26 pages. https://doi.org/10.1145/3769072

[24] Roig, J.V. (2026) How Much Do LLMs Hallucinate in Document Q&A Scenarios? A 172-Billion-Token Study Across Temperatures, Context Lengths, and Hardware Platforms. ArXiv: https://doi.org/10.48550/arXiv.2603.08274

[25] Schneider, J. (2025) Mental Model Shifts in Human-LLM Interactions. J Intell Inf Syst 63, 1737–1752 (2025). https://doi.org/10.1007/s10844-025-00960-6

[26] Smith, D.S. (2013) Phenomenology. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/phenomenology/

[27] United Nations (2026) AI Deception: Brief of the Scientific Advisory Board. UN Secretary-General's Scientific Advisory Board. 19 March 2026. Brief #11. 9 Pages. https://www.un.org/scientific-advisory-board/sites/default/files/2026-03/260317_AI%20Deception%20Brief%20%284%29.pdf

[28] Venkit, P.N., Laban, P., Zhou, Y., Mao, Y., Wu, C-S. (2025) Search Engines in the AI Era: A Qualitative Understanding to the False Promise of Factual and Verifiable Source-Cited Responses in LLM-based Search. In The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3715275.3732089

[29] Wang, Q., Walsh, S.E., Si, M., Kephart, J.O., Weisz, J.D. and Goel, A.K. (2024) Theory of Mind in Human-AI Interaction. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3613905.3636308

[30] Xiong, H., Bian, J., Li, Y., Li, X., Du, M., Wang, S., Yin, D. And Helal, S. (2024) When Search Engine Services Meet Large Language Models: Visions and Challenges. In IEEE Transactions on Services Computing, vol. 17, no. 6, pp. 4558-4577, Nov.-Dec. 2024. https://doi.org/10.1109/TSC.2024.3451185