How should scientists navigate themselves in the chaotic AI era?

Thi Mai Anh Tran
Michigan Technological University
https://orcid.org/0000-0001-8113-9087

June 17, 2024

"Pressing the buttons has gradually become somewhat of a new technological ritual.

– In "Innovation"; *The Kingfisher Story Collection* (2022)

[**WORLDVIEW**]

When I was a first-year Ph.D. student, one of the courses that took a lot of my sweat was the Social Thought course. Every three days, we were expected to do critical reading of at least five papers or a book, which often amounted to 100-250 pages, summarize them in our own words, raise and discuss questions, and write a reflective essay. The course was very interesting but quite demanding, not only for me but for all my classmates. Every time someone came to the class with more than a 10-page summary note, they also came with dark circles under their eyes.

Nowadays, everything has changed; tasks these days have become so easy because of AI technologies, which instantly summarize any book or paper in a few seconds [1]. Not only that, this "Genie" can synthesize text to do literature reviews, identify research gaps to formulate research questions, analyze qualitative data, generate text, create presentation slides, and even generate media dissemination posts [2,3]. The popularity of AI can be seen as a tide of technology that rose rapidly and transformed the entire academic landscape.

According to Stubbs-Richardson and colleagues, since last year, 250 AI applications have been created to assist scientists in psychology, sociology, and communication to conduct their research [4]. For example, large language models have been used for hypothesis generation [5], synthesizing the information for literature review [6], and doing predictive analysis [7]. In a recent survey published by Turnitin, almost 800 out of 1600 postsecondary students and 200 out of 1000 faculty in higher education were found to be regular users of AI. As AI technologies continue to advance, more and more applications of AI will become increasingly prominent in research and our daily lives.

Illustration. Large language models like ChatGPT have been an exciting tool for both scientists and the public nowadays. Photo by Thi Mai Anh Tran

AI has the potential to profoundly impact the research landscape. However, they also have some potential risks and drawbacks that scientists should be aware of, including hallucinations, bias, and consent [9]. These risks can potentially cause harm to science and society due to the misuse or over-reliance on AI [10,11].

First is hallucination. Hallucination is a situation when AI, specifically large language models (LLM), makes up data, fabricates responses and sources of information, or creates nonsense information. For example, last year, ChatGPT created fake news about a real legal professor accused of sexual harassment, then made up its citation on the Washington Post [12]. Why does it happen? First, we need to understand how LLM works.

LLM is an instance of a foundation model, which is pre-trained on large amounts of unlabeled and self-supervised data [13]. It means that the model learns from patterns in the data in a way that produces generalizable and adaptable output. In this case, LLMs are trained on large human-like text datasets such as books, articles, and conversations [13]. ChatGPT, or generative pre-trained transformer, is one of the large language models that can generate human-like text by enabling the model to learn to predict the next word in a sentence while considering its relation to every other word [14]. As a result, ChatGPT predicts the next best syntactically correct word, but not necessarily accurate answers based on an understanding of what humans are asking for because ChatGPT does not truly understand the meaning of its answer [12].

The hallucination problem is more critical nowadays as people tend to trust machine outputs over human sources [15]. According to Kidd and Birhane [16], although humans sometimes communicate false information, they regularly use uncertainty phrases such as "I think" or make corrections. In contrast, LLM, like ChatGPT, generates confident and fluent responses, which causes greater distortion than humans. Further, the number of a person's exposures to fabricated information influences their belief in a false statement, even when the statement contradicts a person's prior knowledge [16]. In the long run, this problem can potentially lead to the spread of misleading information or even false information, which can harm science and society.

The second risk is bias. In a study conducted by Chen et al. [17], the authors used LLMs to understand financial reports and found that the accuracy of GPT-3 to give correct reasoning was below 50%. As GPT-3 ability lies on its training dataset, the authors explain, it might not have seen a similar paradigm as their task setting during pre-training [17]. In social science, this bias can produce unfair and non-objective results. For example, if I want to ask ChatGPT questions related to the worldview of Ojibwe people in hunting, it might only provide information based on Reddit's users' perspectives, which is not representative of my study group, while not providing me the source of information. According to Messeri and Crockett [9], when scientists are not fully aware of this bias, they can be vulnerable to an "illusion of objectivity," where they "falsely believe that AI tools do not have a standpoint or are able to represent all possible standpoints, whereas AI tools actually embed the standpoints of their training data and developers" (p.50).

The last one is consent. Since many AI tools were developed by private companies with closed training data and models, users are unable to know and verify where the data came from [10]. Was it gathered with consent or with copyright? This issue is one of the core concerns for users as the response they get from LLMs might be extracted from private messages or minors without consent. For scientists who use LLMs as a tool for research, consent is a critical issue that they must consider and address.

Besides the risks and concerns of AI mentioned above, too much reliance on AI tools can harm students' critical thinking ability and contribute to the reproducibility crisis. For example, after using ChatGPT for a long time, I might feel lazy reading or writing by myself or not confident doing research without ChatGPT. Recently, a study by Ahmad et al. [18] among university students in Pakistan and China found that AI significantly impacted the loss of human decision-making (27.7% of students) and induced human laziness (68.9% of students). Further, as AI is developing rapidly and unpredictably nowadays, using AI as an analysis tool can put researchers in an unstable position where they might not be able to recreate their own research [10]. These problems can precipitate the existing challenges for academia in gaining public trust [19,20].

Since it is important that we know the strengths and weaknesses of AI, we must ask ourselves what relationship we want to have with this powerful tool. And how should scientists navigate themselves in this chaotic AI era?

AI is evolving every day to perfect itself. For example, the updated ChatGPT 4 gives much more reasonable and logical responses than ChatGPT 3 when solving math problems [21]. Rather than prevention, researchers, editors, and reviewers should adapt and apply AI technologies in their works [21]. Thinking this way, the strengths and drawbacks of AI are like two sides of the same coin: AI's power of information seeking, synthesizing, and creation can be our strengths when we know how to effectively utilize them, whereas they become drawbacks if we cannot control or over-rely on them.

For effective adaptation and application of AI, researchers, editors, and reviewers need deep theory for conceptual thinking and being able to process, combine, and validate information provided by AI for their knowledge management and generation processes. One implication of this is the integration of deep theory and analytical framework to build logical foundation and reasoning in every research.

Illustration. The BMF analytics' book. Photo by Thi Mai Anh Tran

In social sciences and humanities, a deep theory like Mindsponge Theory proves particularly valuable for innovative thinking as it focuses on the underlying level of human psychology and behavior through the information-processing lens rather than focusing solely on the high-level observations at individual and societal levels. More specifically, this theory elucidates how human psychology and behavior are influenced by internal information processes within the mind and interactions with the external environment. Consequently, it does not contradict existing psychological and social theories and frameworks but rather elaborates, solves inconsistencies, and connects concepts through the dynamic view of information processing [22].

When operationalizing Mindsponge Theory with computational thinking, such as Bayesian inference, researchers can utilize both theoretical reasoning strengths and statistical advantages to make their research more efficient and accurate [23]. For example, researchers can use Mindsponge Theory to design survey methods more straightforwardly and effectively by identifying which factors are highly likely to be involved in the psychological process of a study. From that, they can easily apply suitable measurements and data types [23]. In addition, Bayesian Inference allows researchers to include reliable information before analyzing data, which solves the multicollinearity problem in statistical analysis. Bayesian inference is also more theoretically advantageous than the frequentist approach, which helps scientists easily define the credible region where the true parameter value has a high probability of being within [22]. Together, the Bayesian Mindsponge Framework (BMF) has been applied in multiple disciplines, including psychology, education, environment, healthcare, and social sciences.

As AI does not “think” like humans and still has limitations in the way it works, researchers, editors, and reviewers need deep theory for critical thinking and creativity. Rooted in deep theory, Mindsponge Theory and BMF have the potential to offer a promising path forward for social scientists to conduct research in a creative, responsible, and productive manner. By utilizing AI advantages while maintaining human oversight and theoretical grounding, people can enter a human-led AI era instead of an AI chaotic environment.

References

[1] Brown TB, *et al.* (2020). Language Models Are Few-Shot Learners. Arxiv. https://arxiv.org/abs/2005.14165

[2] Andreotta M, et al. (2019). Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. *Behavior Research Methods*, **51**. https://doi.org/10.3758/s13428-019-01202-8

[3] Davidson T. (2024). Start generating: Harnessing generative artificial intelligence for sociological research. *SocArXiv Papers*, https://osf.io/preprints/socarxiv/u9nft

[4] Stubbs-Richardson M, Brown L, Paul M, Brenner D. (2023). Artificial intelligence applications for social science research. *Artificial Intelligence Applications for Social Science Research*. https://scholarsjunction.msstate.edu/ssrc-publications/6/

[5] Park YJ, *et al.* (2023). Can ChatGPT be used to generate scientific hypotheses? *ArXiv*. https://doi.org/10.48550/arxiv.2304.12208

[6] Ayd?n Ö, Karaarslan E. (2022). OpenAI ChatGPT generated literature review: Digital twin in healthcare. In Ö Ayd?n (Ed.), *Emerging Computer Technologies 2* (pp. 22-31). ?zmir Akademi Dernegi. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4308687

[7] Lopez-Lira A, Tang Y. (2023). Can ChatGPT forecast stock price movements? Return predictability and Large Language Models. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4412788

[8] Shaw C, *et al*. (2023). GenAI in Higher Education: Fall 2023 Update of Time for Class Study. Tyton Partners. https://www.turnitin.com/whitepapers/tyton-partners-time-for-class-study

[9] Messeri L, Crockett MJ. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, **627**(8002), 49–58. https://www.nature.com/articles/s41586-024-07146-0

[10] Spirling A. (2023). Why open-source generative AI models are an ethical way forward for science. *Nature*, **616**(7957), 413–413. https://www.nature.com/articles/d41586-023-01295-4

[11] Yin W. (2024). Will our educational system keep pace with AI? A student's perspective on AI and learning. *EDUCAUSE Review*. https://er.educause.edu/articles/2024/1/will-our-educational-system-keep-pace-with-ai-a-students-perspective-on-ai-and-learning

[12] Verma P, Oremus W. (2023). ChatGPT invented a sexual harassment scandal and named a real law prof as the accused. *Washington Post*. https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/

[13] IBM. (2023). What Are Large Language Models? https://www.ibm.com/topics/large-language-models

[14] Guinness H. (2023). How does ChatGPT work? https://zapier.com/blog/how-does-chatgpt-work/

[15] Swayne M. (2019). People more likely to trust machines than humans with their private information. https://www.psu.edu/news/research/story/people-more-likely-trust-machines-humans-their-private-information/

[16] Kidd C, Abeba Birhane. (2023). How AI can distort human beliefs. Science, 380(6651), 1222–1223. https://www.science.org/doi/10.1126/science.adi0248

[17] Chen Z, *et al.* (2022). CONVFINQA: Exploring the Chain of Numerical Reasoning in Conversational Finance Question Answering. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 6279–6292. https://aclanthology.org/2022.emnlp-main.421.pdf

[18] Ahmad SF, *et al.* (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, **10**, 311. https://www.nature.com/articles/s41599-023-01787-8

[19] França TFA, Monserrat JM. (2019). Reproducibility crisis, the scientific method, and the quality of published studies: Untangling the knot. *Learned Publishing*, **32**(4), 406–408. https://doi.org/10.1002/leap.1250

[20] Ho MT, Vuong QH. (2019). The values and challenges of "openness" in addressing the reproducibility crisis and regaining public trust in social sciences and humanities. *European Science Editing*, **45**(1), 14-17. https://doi.org/10.20316/ESE.2019.45.17021

[21] Vuong QH, *et al.* (2023). Are we at the start of the artificial intelligence era in academic publishing? *Science Editing*, **10**(2), 158–164. https://doi.org/10.6087/kcse.310

[22] Vuong QH. (2023). *Mindsponge Theory*. Walter de Gruyter GmbH. https://www.amazon.com/dp/B0C3WHZ2B3

[23] Vuong QH, Nguyen MH, La VP. (2022). *The mindsponge and BMF analytics for innovative thinking in social sciences and humanities*. Walter de Gruyter GmbH. https://www.amazon.com/dp/8367405102/

[24] Vuong QH. (2022). *The Kingfisher Story Collection*. https://www.amazon.com/dp/B0BG2NNHY6