Artificial intelligence and health information literacy

Andrew Cox
Information school, University of Sheffield, Sheffield, UK

Abstract

The proliferation of generative AI is changing health information behaviour. But the problems of accuracy and lack of transparency it has require users to develop some degree of AI literacy as an aspect of their health information literacy. There are many models of AI literacy suggesting key potential components such as knowledge of AI technologies; how to use them and evaluate outputs; how to protect one's own safety; and ethical awareness, including of wider societal impacts. Conceiving these components as making up AI competency implies that it consists of the persistent attitudes and values of a critical information user, not the satisfied consumer that generative AI models try to create.

Key words: *AI (artificial intelligence); health literacy; patient education; libraries.*

Introduction

While Artificial Intelligence (AI) undoubtedly has huge potential to assist in improving healthcare, as information professionals we are much more ambivalent about the rise of mass generative AI and its impact on how health information is accessed. Generative AI's issues of low accuracy and lack of transparency suggest that we need to incorporate some elements of AI literacy into models of Health Information Literacy.

At the time of writing we are experiencing the rapid proliferation of generative AI across search and information use experiences in general. It now appears in AI chatbots, like ChatGPT, CoPilot, Gemini and the like, in their many versions, including their deep research agents; in search engines, such as Google overviews; in bespoke research tools like Consensus and ResearchRabbit; in library licensed databases; and in Retrieval Augmented Generation (RAG) applications.

Generative AI is proliferating partly because it genuinely makes it easier to find and use information. As well as doing lots of other useful things, such as generate images, code and checking grammar, generative AI offers a complete answer to a search query not just a list of resources. It summarises an individual source and allows us to pose questions to a collection of sources. It turns search into a conversation in natural language. These are attractive features making access to information easier.

Specifically for patients searching for health information, generative AI has considerable potential. For example, it can help explain diagnoses in non-technical language and translate information for audiences whose first language is not English. Indeed, generative AI is almost certainly reshaping information behaviour, including health information behaviour, although there seems to be scant research so far on how exactly behaviour is changing.

However, the use of mass AI systems like ChatGPT to discover health information is fraught with problems. Generative AI has a fundamental problem of inaccuracy. It often makes mistakes, is out of date and fails to cite its sources. Pushed to give sources it often invents them. Yet it presents its answers in such a confident tone that it promotes undue trust. AI lacks transparency. That is partly because it is based on hard-to-understand computation and statistics. As a result, it is difficult to form a clear mental model of how it

Address for correspondence: Andrew Cox, Information School, University of Sheffield, The Wave, 2 Whitham Road, The University of Sheffield, Sheffield S10 2AH. E-mail: a.m.cox@sheffield.ac.uk.

works and when it might be more or less reliable. Often its errors such as ignoring parts of prompts are hard to understand. It is also partly lacking transparency because of commercial secrecy. Big Tech deliberately withholds information on how their models are trained. More immediately at the point of use, although the chatbot interface to AI like ChatGPT is attractive, it answers as if it were a human, anthropomorphizing itself, promoting the wrong sort of trust.

At root AI chatbots are designed to create satisfied consumers not critical health information users. So, for example, they are prone to agree with you when you are wrong. The recent case of an AI model being withdrawn by ChatGPT because it was too sycophantic is just an extreme example of how AI Chatbots' "designed personality" and their invisible guardrails distort access to information (1). Furthermore, as generative AI functions proliferate into all search experiences it becomes less transparent to us that we are using AI, reinforcing the need for all users to have greater awareness of the structures within which they are searching. Such inaccuracy and lack of transparency is a source of hazard, particularly in the health context. Given that generative AI is not transparent, and governments and regulators have not yet forced providers to make it more transparent, so AI literacy becomes important for its safe and successful use. AI literacy needs to be incorporated in some way into Health Information Literacy programmes, including for those with low or no digital skills (2). This is a natural role for information professionals, not least because generative AI use appears in search behaviours alongside searching Google and more authoritative health resources.

Models of AI literacy

There are now no shortage of models of AI literacy, though most are intended for the educational domain: I have reviewed some of them in Cox (3). There have been more produced since that review. Two starting points for thinking about the makeup of AI literacy are Hibbert *et al.* from Educause (4) and Hervieux and Wheatley's recent synthesis (5). We also produced a generative AI specific model in Zhao, Cox and Cai (6). There are some common elements. Understanding of basic concepts of AI as a technology offers a foundation for AI literacy and this is referred to as "understand AI" in Hibbert *et al.* (4) and "know the basic principles" and "understand the fundamental different types of AI" in Hervieux and Wheatley (5). The assumption here is that we all need to know the basics of how AI works.

The second key element is to pick the right tool, know how to use it and evaluate its outputs for accuracy and bias. In the Hibbert *et al.* model this is covered by "Use and apply" and "analyse and evaluate" (4). In Hervieux and Wheatley it is "Experiment with AI tools" and "Review the outputs or outcomes of AI tools" (5). In our model we refer to it as pragmatic understanding (6). In the context of generative AI this is partly about understanding better prompt techniques to ask questions in ways that boost its richness and reliability. Some of the wisdom of prompt engineering could be encapsulated in the following points:

- define your question as precisely as possible;
- define the context for the question;
- upload training resources for the AI (if they are not copyright or confidential) to give it data or a model for its answer;
- define the character of the answer required, for example, in terms of word length or style;
- iterate the question and synthesise answers;
- ask for sources and check them (including that they actually exist);
- ask AI to define its confidence with its answers.

Good prompting is important, but AI literacy is as much about recognising that something is an output of AI and appreciating the consistent weaknesses in their outputs. Given what we know about bias in LLMs we should be actively anticipating them making biased assumptions (7).

Hervieux and Wheatley's emphasis on "experiment" usefully acknowledges the continuing evolution of the platforms and the need to constantly learn, as does their principle of "engage in the AI discourse" (5).

In the health context there is also the safety dimension. It is important to avoid sharing private information with generative AI. We refer to that as safety understanding (6).

The ethics of AI is sometimes embedded into these building blocks of AI literacy, sometimes separated out. In either case, it is important that generative AI is not seen merely as "a tool" and that consideration is given to wider ethical issues and social implications: such as work displacement or environmental impacts. In Hibwork displacement or environmental impacts. In Hibbert *et al*. it is part of “analyse and evaluate” (4). In Hervieux and Wheatley “Evaluate the impact of AI on a societal scale” (5). In Zhao *et al*. we talk about socio-ethical understanding (6).

Another AI literacy model that will undoubtedly be influential is the UNESCO AI competency framework (8). It is much broader in its scope than the information literacy aspects, but one element that could influence how we define AI literacy in Health Information Literacy is that it places emphasis on a human centred/ethical approach, prior to any consideration of more technical or pragmatic aspects. This seems useful.

The UNESCO framework also uses the term competencies rather than literacy. The term literacy aligns to our professional commitments in the tradition of developing AI literacy models. The term “skills” might be more accessible to many audiences. But the term “competencies” usefully implies persistent attitudes, values, identities. This is helpful in defining how AI is understood with commitment to being a critical information user and learner in tension with the generative AI driver to create a satisfied consumer.

Conclusion

There is little doubt that health information behaviour is being changed by generative AI. We do not know very clearly the extent or nature of these changes. However, it is apparent that in the health context generative AI’s low accuracy and lack of transparency are significant issues. Incorporating some forms of AI literacy into Health Information Literacy will be increasingly important.

Submitted on invitation.
Accepted on 30 May 2025.

REFERENCES

1. Gerken T. Update that made ChatGPT 'dangerously' sycophantic pulled. BBC [Internet]. 2025.
   Available from: https://www.bbc.co.uk/news/articles/cn4jnwdvg9qo
2. Good Things Foundation. Developing AI literacy with people who have low or no digital skills [Internet]. 2024.
   Available from: https://www.goodthingsfoundation.org/policy-and-research/research-and-evidence/research-2024/ai-literacy
3. Cox A. Algorithmic literacy, AI literacy and responsible generative AI literacy. Journal of Web Librarianship. 2024 Oct 9; 18(3): 93-110.
   Available from: https://doi.org/10.1080/19322909.2024.2395341
4. Hibbert M, Altman E, Shippen T, Wright MA. Framework for AI Literacy. EDUCAUSE Review. 2024.
   Available from: https://er.educause.edu/articles/2024/6/a-framework-for-ai-literacy
5. Hervieux S, Wheatley A. Building an AI literacy framework: perspectives from instruction librarians and current information literacy tools. CHOICE. 2024.
   Available from: https://www.choice360.org/research/white-paper-building-an-ai-literacy-framework-perspectives-from-instruction-librarians-and-current-information-literacy-tools/
6. Zhao X, Cox A, Cai L. ChatGPT and the digitisation of writing. Humanities Social Sciences Communications. 2024 April; 11: 482.
   Available from: https://doi.org/10.1057/s41599-024-02904-x
7. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digital Health. 2024 Nov; 3(11).
   Available from:
   https://doi.org/10.1371/journal.pdig.0000651
8. UNESCO. AI competency framework for students. 2024.
   Available from: https://www.unesco.org/en/articles/ai-competency-framework-students