Patient Information Forum

Balancing the risks and benefits of AI in the production of health information

Published April 2024
Review date April 2025

Position statement

Position statement

Balancing the risks and benefits of AI in the production of health information

Use of artificial intelligence (AI) in healthcare is on the rise. Bodies including UK Governments, the National Institute for Health and Care Research and the NHS AI Lab are all investing in developing and deploying the technology.

As the landscape evolves, health information producers are investigating the risks and benefits of using AI in their everyday work.

Developed in collaboration with PIF's AI working group (see page 9), this position statement aims to help members understand the AI landscape and how to manage it.

A full framework for policy creation is in development and will be published in the autumn of 2024.

This position statement includes:

1. What is AI? P3
2. The benefits of using AI P4
3. The risks presented by AI P5
4. Risk management: The case for AI use policies P7
5. Transparency P9
6. The AI Working Group P9
7. Glossary of terms P10
8. References and further reading P11

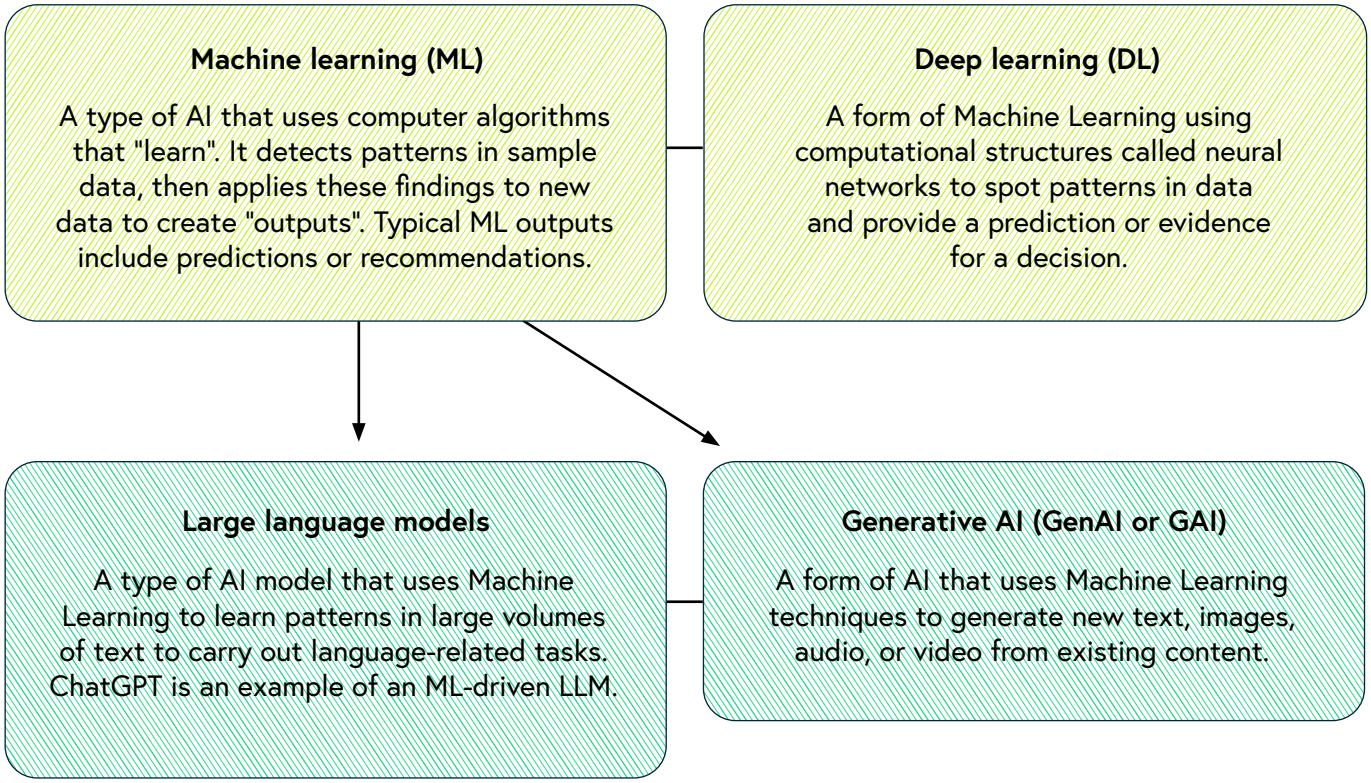
Patient Information Forum

1. What is AI?

AI is a catch-all term covering diverse technology, from self-driving cars to analysis of large scientific data sets.

The UK Government defines AI as "the use of digital technology to create systems capable of performing tasks commonly thought to require human intelligence"[1].

There are different subsets of AI[2]. The most relevant to health information is **machine learning** (ML). **Generative AI** (GAI) is based on ML techniques.



ML works by using statistical techniques to spot patterns in large data sets. It then uses these patterns to create "outputs". Typical ML outputs include predictions or recommendations.

For example, website "live chat" and **chat bot** functions learn to spot patterns in the questions people commonly ask a health charity. They can then direct people with similar queries to the relevant information.

Deep learning (DL) is a more sophisticated form of ML. It "layers" **algorithms** to make more accurate predictions. ML and DL models tend to be either supervised or unsupervised.

In supervised learning, programmers tell the model what to look for in the training data. Unsupervised models are left to spot patterns for themselves.

Generative AI uses ML and DL to create new text, images, video, audio, or other content. When related to text, you might also see these referred to as **Large Language Models** (LLMs). Examples include user interfaces such as ChatGPT and BLOOM.

For example, if ChatGPT is asked to summarise bowel cancer symptoms it will search its training datasets for content. The model will then use its probability distributions and internal control mechanisms to generate a text summary.

---

2. Benefits of AI

Used as part of a robust information production process, AI may streamline the development and delivery of health information.

Applications in use and under consideration include:

- Automatic translation of materials, including leaflets and videos
- Chatbots to respond to online information requests
- Accelerated data analysis, including the guided analysis of unstructured, qualitative data
- GAI-generated health information and plain language summaries
- Accelerated, AI-guided research analysis
- ML and DL-driven "search engines" to recommend the most appropriate best practice guideline, research paper, or patient information content.

This could help healthcare information providers to "do more with less". Delegating tasks such as data analysis to computers, for example, can free staff to work on other activities.

AI-enabled solutions could also help extend an organisation's reach. For example, AI translation can help teams serve seldom-heard communities and automated **chat bots** can provide support outside of office hours. Generative AI (GAI) is based on ML techniques.

3. Risks presented by AI

There are risks associated with the use of AI which health information providers should be aware of.

These challenges relate to AI in all its uses. But, arguably, they are of critical importance in the health information space where we strive to produce accurate, unbiased, inclusive materials.

Data privacy and security

AI runs on data. Whether developing models in-house or outsourcing to developers, organisations need to ensure personal information is secure and used in line with relevant regulations[3].

Bias in datasets

ML and DL algorithms learn from the data they are given. If the training data contains bias, this is likely to be reflected or compounded in the AI model's outputs. AI has been shown to mirror existing misinformation and prejudices[3].

Hallucinations

Hallucinations are when AI generates convincing but completely made up content producing incorrect or misleading results[4]. This can include fake references.

Poor quality sources

Some AI includes relevant content, regardless of source or accuracy, in its analysis.

For example, GAI models will trawl their training data or the internet looking for answers to the questions they have been given.[5] If the information found is biased or incorrect, the AI model may spot patterns that do not exist.

Models also tend not to have access to paywall or otherwise protected content[4]. This means primary sources are often excluded from analysis.

Out of date sources

GAI models are trained on pre-existing data, meaning they do not have access to the latest information or research.

For example, the first free-to-use version of ChatGPT was trained on data published before 2021[6]. This meant any searches relating to COVID-19 returned drastically out-of-date results.

Lack of individualised or specific information

GAI tends to over simplify health topics because it lacks the ability to apply context or nuance to its results, or to understand the meaning behind the data[7].

While it can be useful for researching general information on a specific health condition, it is unable to interpret and relay how different health conditions or interventions may impact on each other.

The "Black Box" transparency problem

AI cannot show its workings. We may know or be able to control what data goes in, and the results that come out, but not how the model comes to its conclusions.

This makes it challenging to know where the AI-generated information came from, and whether it is accurate or biased[8].

Public trust and understanding

A recent survey of more than 17,000 people from 17 countries found 61% were wary about trusting AI systems[9]. It is vital the use of AI does not undermine public trust in vital health information resources.

Liability

Currently AI has a high risk of inaccuracy. This means using AI to produce health information comes with complex and unanswered questions around the liability of AI-generated output used by an organisation[10].

Loss of website traffic

AI tools make use of content from health charities, the NHS and other trusted sources. This could lead to a reduction in direct traffic to websites which place information in its full context. This could have wider impacts on an organisation's sustainability.

Copyright breach

The use of AI tools creates a risk of copyright breach[5]. In January 2024, UK Parliament government confirmed AI training data will infringe copyright unless permitted under licence or an exemption[11].

GAI tools which search the internet for answers can pull large sections of copyrighted text from sources including charity and commercial websites. This can have both ethical and legal implications.

4. Risk management: The case for AI use policies

The challenges and opportunities of AI are issues attracting attention at the highest levels.

- A United Nations Educational, Scientific and Cultural Organization (UNESCO) report has emphasised the need to build public awareness of the technologies. Greater AI literacy, it argues, will help people to embrace the advantages of AI while avoiding the pitfalls[12].
- The European Union (EU) AI Act is currently making its way through the European Parliament, and aims to become the global standard on responsible AI use[13]. It will include codes of practice on individual organisations' obligations, and techniques for labelling content as artificially generated[14].
- In the UK, the Information Commissioner's Office has acknowledged that GAI is changing the way people access information, which carries inherent risks around transparency and data privacy. The body has committed to work with other regulators to "bring more definition and regulatory clarity"[15].
- The Charity Commission has published advice for charities on the use of AI. This includes options to consider and how to manage risks[16].

Taking no action is not an option. AI is here to stay, and forbidding its use is unfeasible.
We need to manage the risks, not ignore them.

What should my policy cover?

PIF recommends health information providers create AI usage policies. Clear policies will provide teams with the guardrails they need to use AI responsibly and with confidence.

A policy should cover:

- Use of AI by type
- Risk control strategies
- Staff training programmes that give staff a grounding in available tools and learning on AI
- User involvement and engagement procedures
- Content and accessibility standards
- Transparency policies to ensure users understand how AI is being used and what actions have been taken to mitigate and manage the risks
- A feedback loop process, including how users can submit feedback and how it is collated and acted upon
- Evaluation of the impact of the tools, in terms of efficiency, scale of use, and user uptake
- An organisation-level acknowledgement or statement of liability risks, and how to mitigate them.

Many of these checks align with the PIF TICK criteria.

Initial considerations

PIF's AI working group is working on a framework for the development of AI usage policies. In the meantime, the following considerations are designed to help members get started:

- Currently, GAI is not suitable for the creation of health information and content in isolation. The risk of inaccurate, biased outputs that lack the necessary nuance and context to provide individualised or specific health information are too high. GAI may be used to assist in generating concepts for health content, but only with human oversight.

  This approach, sometimes referred to as a "human in the loop", is necessary to review, interpret and validate the accuracy, relevance and inclusivity of outputs. Explainable AI, or AI that is able to "show its workings", could overcome the Black Box paradox and move the dial. However, this is still a long way off.

- Applications such as data analysis, including the analysis of unstructured, or "free text", have been used to streamline processes. Organisations wishing to use this approach should be aware of the risks. They include replicating or compounding bias in the dataset and protecting data privacy. Mitigation strategies include ensuring the quality of input data, securing informed consent for its use in an AI model, and complying with relevant data privacy regulation.

- Automated translation software can expand the reach of multilingual content, but also carries significant risks. AI-based language translation models might not accurately convey important medical terms or instructions, leading to misinterpretation of information, and ultimately risking patient safety. They might also struggle to express cultural and linguistic nuances, leading to inappropriate or misleading content. However, accuracy issues can also apply to human translation services.

  PIF recommends all translation processes include sense-checking for accuracy, appropriate use of language, and cultural sensitivity. When used on websites, auto translation should contain an accuracy warning. See our How to Guide to Translating Health Information.

- Regular reviews should be carried out. AI, and the way we use it, is evolving fast. Policies need to keep pace with changes in technology and our learning curve.

5. Transparency

Transparency should be central to all applications of AI technology. The public needs to be able to trust our members if they are to trust the information they provide.

Being open and upfront about how and why the organisation is using AI is key to building and maintaining that trust.

Trust will also be central in helping us to learn from each other as we continue on our mission to balance the risks and benefits of AI. This technology is rapidly evolving and here to stay.

Our role at PIF is to work with members to harness the opportunity, manage the risk and provide education and support to the members and the public.

---

6. The AI Working Group

This position statement has been developed in collaboration with the PIF AI Working Group. The group is made up of members and non-members representing the charity sector, public health bodies, NHS Trusts, regulatory bodies and commercial companies.

The position statement is the result of a webinar and a member round table. We would like to thank members who took part in both events. A recording of the webinar is available **here**.

Contributors
Trishna Bharadia, The Spark Global
Sheena Campbell, PIF
Caroline De Brun, UK Health Security Agency
Sue Farrington, PIF
Lynsey Hawker, The Kings Fund
Steph Jury, Guy's and St Thomas' NHS Foundation Trust
Raisa McNab, Association of Translation Companies
Dr Knut Schroeder, Expert Self Care
Julie Smith, EIDO Systems
Sophie Randall, PIF

Author: Amanda Barrell, medical and life sciences writer.

7. Glossary of terms

Algorithm: The sequence of rules a computer uses to convert an input, or the dataset into an output, or a pattern in the data.

Artificial intelligence (AI): Machines that can perform tasks previously carried out by human intelligence.

Chat bot: Software designed to mimic human conversation that "talks" to users via speech or text.

Deep learning (DL): A form of Machine Learning (ML) using computational structures called neural networks to spot patterns in data and provide a prediction or evidence for a decision.

Generative AI (GenAI or GAI): A form of AI that uses Machine Learning (ML) techniques to generate new text, images, audio, or video from existing content.

Human in the loop: AI systems that combine the power of human and artificial intelligence. The human can intervene by training, fine-tuning or testing the system's algorithm, for example, to help it produce more useful results.

Large Language Model (LLM): A type of AI model that uses Machine Learning (ML) to learn patterns in large volumes of text to carry out language-related tasks.

ChatGPT is an example of an ML-driven LLM. LLMs are notable for their ability to achieve general-purpose language generation and understanding. Modern LLMs can produce entire sentences, paragraphs, or even entire documents.

Machine learning (ML): A type of AI that uses computer algorithms that "learn". It detects patterns in sample data, then applies these findings to new data to create "outputs". Typical ML outputs include predictions or recommendations.

8. References and further reading

1. *Gov.uk. 2019. A guide to using artificial intelligence in the public sector. Accessed March 2024.*

2. *The Alan Turing Institute. Defining data science and AI. Accessed March 2024.*

3. *World Health Organization. 2021. Ethics And Governance Of Artificial Intelligence For Health. Accessed March 2024.*

4. *Athaluri SA et al. 2023. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. Cureus. 15(4). Accessed cureus.com March 2024.*

5. *Cornell University. 2023. Generative AI in Academic Research: Perspectives and Cultural Norms. Accessed March 2024.*

6. *Aberystwyth University. 2024. Utilising AI in the Library: A Student's Guide: The Limitations of AI. Accessed March 2024.*

7. *Hoffman, B. 2023. Leaders looking to leverage AI need to think about context. Accessed March 2024.*

8. *Wadden, J.J. 2022. Defining the undefinable: the black box problem in healthcare artificial intelligence. Journal of Medical Ethics. 48(10), pp 764-768. Accessed jme.bmj.com March 2024.*

9. *KMPG. 2023. Shifting public perceptions of AI. Accessed March 2024.*

10. *Center for International Governance Innovation. 2023. Who Is Liable for AI-Driven Accidents? The Law Is Still Emerging. Accessed March 2023.*

11. *UK Parliament. 2024. Connected tech: AI and creative technology: Government Response to the Committee's Eleventh Report of Session 2022–23. Accessed April 2024.*

12. *UNESCO. 2024. User empowerment through media and information literacy responses to the evolution of generative artificial intelligence (GAI). Accessed March 2024.*

13. *EU Artificial Intelligence Act. N.d. Up-to-date developments and analyses of the EU AI Act. Accessed March 2024.*

14. *EU Artificial Intelligence Act. 2024. The AI Office: What is it, and how does it work? Accessed March 2024.*

15. *Information Commissioner's Office. N.d. Next Generation Search. Accessed March 2024.*

16. *Charity Commission. Charities and Artificial Intelligence. Accessed April 2024.*

Join us

Patient Information Forum

PIF is the independent UK membership body for people working in health information and support. We also run the only independently assessed quality mark for health information – the PIF TICK.

PIF represents more than 300 organisations across the NHS, voluntary, academic, freelance and commercial sectors. Our expert guidance on the production of high-quality health information supports an improved healthcare experience for patients and the public.

Our vision

Everyone has access to personalised health information and support to enable them to make informed decisions about their health, wellbeing and care.

Our Partners

abpi
ATC Association of Translation Companies DEFINING STANDARDS OF EXCELLENCE
AstraZeneca
Cuttsy &Cuttsy
FCBHEALTH EUROPE
International CLINICAL TRIALS
HIFA .org
janssen
MACMILLAN CANCER SUPPORT
MSD
National Voices
NICE National Institute for Health and Care Excellence
Pfizer
PRSB Professional Record Standards Body
Self Care Forum
texthelp

Published by the Patient Information Forum
483 Green Lanes, London, N13 4BS
pifonline.org.uk
info@pifonline.org.uk
@PIFonline
@PIFonline