https://doi.org/10.1038/s41568-024-00685-8

Guardrails for the use of generalist AI in cancer care

Stephen Gilbert & Jakob Nikolas Kather

Artificial narrow intelligence models, trained for specific intended purposes, have gained approval and recommendation for cancer treatment. Generalist medicial artificial intelligence (GMAI) models are now being developed for cancer treatment. Policy makers have a stark choice: radically adapt frameworks, block generalist approaches or force them onto narrower tracks.

Approved artificial intelligence (AI)-enabled approaches for cancer diagnosis and therapy have been specifically designed for relatively narrow tasks such as detecting abnormalities on X-rays or detecting polyps in colonoscopy videos. There is a shift towards 'multi-purpose', or even 'generalist' models, which are trained on and can flexibly respond to a broad range of diverse data, spanning various types of images and even non-image data, including text[1,2]. For example, this means for a patient with colorectal cancer, a single GMAI model could interpret their endoscopy videos, pathology slides and electronic health record (EHR) data.

Although some generalist AI approaches have existed for some time[3], there has been marked recent progress enabled by a relatively new type of neural network, called a transformer, which underlies most natural language processing applications and many computer vision tasks today[4]. Another technical innovation that enables this progress is self-supervised learning, by which AI models can be trained on large quantities of unlabelled data[1]. Large language models (LLMs) are notable examples of transformers trained with self-supervised learning[1,2,5]. LLMs can process text at human level but have also been extended to be able to process images such as the recently released model Generative Pre-trained Transformer-4-Vision (GPT4-V). GPT4-V can interpret medical text, medical images and tabular medical data[4].

Generalist models and regulation

Current international regulation requires that from conception, AI-enabled medical devices are designed for a defined and fixed purpose, set of clinical indications and target population[5] (Fig. 1). Regulations rigidly stipulate that these are prespecified in the phase of development, and then remain as-specified after approval. Adaptation or extension is only possible through following detailed quality management and regulatory processes and is administratively burdensome and time-consuming[5].

GMAI models are already used in on-market clinical charting and EHR products, which are not considered medical devices in the USA[5]. Early research shows some promise for the use of GMAI models in decision support for physicians, by providing information on diagnoses and therapeutic options to consider, however the clinical potential and regulatory implications of GMAI models remain uncertain[5] (Fig. 1). Nonetheless, these models have shown remarkable flexibility in that they have large potential in the personalization of predictions, and even in prediction without specific training examples in the domain of application. This characteristic, termed 'zero shot' reasoning, refers to the ability to carry out multi-step reasoning processes on tasks that have not been provided as hand-crafted examples in model training. It enables the models to generalize knowledge to a degree from their wide training data, and then apply this to new and unseen situations, although it must be acknowledged that current validation of these approaches is qualitative and preliminary. This could enable decision support tuned to the unique symptoms and findings of the individual patient[1,2,4,6]. At the same time, because of this flexibility on inputs and application, quantifying the reliability of advice from these models is immensely challenging[5].

Although there are international differences in the approach to AI regulation, GMAI approaches are excluded by all international frameworks and no GMAI-enabled device has been authorized by the US Food and Drug Administration (FDA), or, to our knowledge, by any other national regulator. This is likely because GMAI approaches are fundamentally incompatible with standard device or medical device regulations[1]. Whereas a device is regarded as something made or adapted for a particular purpose, especially a piece of mechanical or electronic equipment, both in the common usage of the term and in all international medical device frameworks, 'general intelligence' is neither adapted for a particular purpose nor a thing truly 'made' through a human design process. As such, GMAI models are not devices and are instead highly adaptive frameworks, closer to human thought processes than a narrow task device. Therefore, from first principles, 'true' GMAI will never be approvable under regulatory frameworks developed solely for discrete and narrow-by-design devices[5]. If the regulatory framework remains unchanged, it may be possible to develop hybrid approaches that could be approved as medical devices, which use some of the flexibility of GMAI, and then constrain these through limiting the range of clinical prompts that can be provided through the user interface. Early studies of GMAI approaches have benchmarked performance across a limited number of diverse yet specific tasks[2]. It is not inconceivable that the first regulatory authorization of GMAI models will be for a limited form of generality. These approaches would, to a greater or lesser degree, constrain GMAI technologies to be non-generalist, and they would therefore not be true GMAI at all. Thus, the potential transformative benefits of a wider concept of GMAI would not be realized in either cancer therapy or research.

The medical information accuracy of LLMs and vision models is likely to increase as model training sets grow and GMAI approaches are refined. This improvement will probably occur even in generic tools that are not specifically designed or marketed for medicine, however these will not be refined and tested for medical accuracy and will probably remain patchy with respect to the safety of the information they provide.

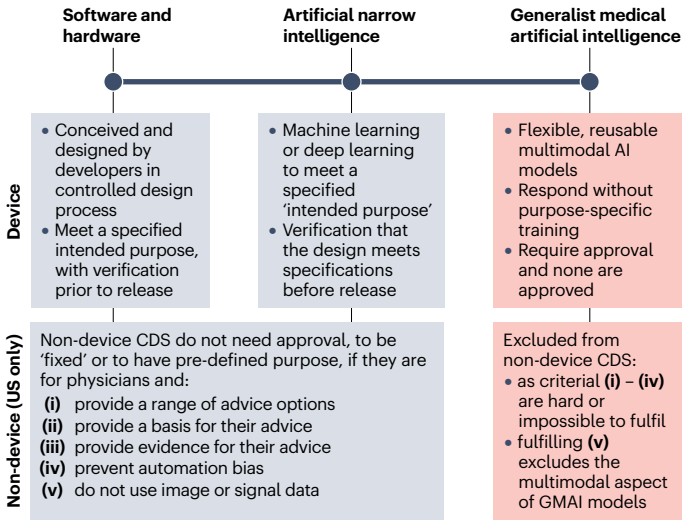

Fig. 1 | The spectrum of technologies for clinical decision support, classified as device or non-device for regulatory purposes. The spectrum displayed as a horizontal bar shows the range of approaches from pre-machine learning to artificial narrow intelligence, and finally to generalist medical artificial intelligence (GMAI). Approaches possible within current compliance frameworks are shown in grey and those that are likely excluded are shown in red. AI, artificial intelligence; CDS, clinical decision support. L. Crow/Springer Nature Limited.

Patients increasingly have access to their EHR and could in principle upload it to generalist AI tools[5]. It will be impossible to stop patients and physicians using such generic models or unapproved medical decision support systems. Indeed, they may be driven towards untested and probably unsafe GMAI model-based tools as they have highly accessible chat interfaces, and as current regulations exclude access to approved tools based on the same technologies. Unapproved tools also bring with them greater concerns over the safeguarding of private data, the degree of the control of bias, and medical liability. Excluding physician access to GMAI models could undermine them in their role of leading the diagnostic process. If the patient can come prepared with an unapproved but comprehensive GMAI-based analysis of their medical history, and the physician is barred from such tools, they are underinformed and underpowered compared with the patient they are treating. If these tools continue to grow in their performance and flexibility, then sticking rigidly to current regulatory frameworks, which forbid this type of flexibility, will likely lead to the undermining of public and physician support for regulation and limited compliance with regulations, as has already been seen with LLMs[5].

Novel regulation of generalist models

Could regulatory frameworks make room for highly versatile multimodal GMAI-based decision support tools in cancer? We can draw inspiration from other medical sectors in which similar regulatory evolution has occurred. For example, the FDA introduced a new flexible and adaptive regulatory approach for genetic and genomic tests, which made publicly accessible genomic database evidence applicable instead of insisting on new analytical validation in every scenario[7]. At least in the medium term, GMAI models will have variable accuracy and will need the insights of well-trained and well-adapted human physicians. The authorization of true GMAI models might be possible if there was better co-development of medical practice and regulation to acknowledge the crucial role of physicians as empowered information interpreters. Oncologists undergo extensive training at universities and clinics to be able to effectively diagnose and treat cancer. They routinely order a suite of diagnostic and monitoring tests that together can produce a vast amount of data that is time consuming and challenging to interpret[8]. It is here that GMAI models are particularly well-suited to assist physicians, keeping physicians in the central role as most patients desire. Regulation could ensure human-centric GMAI integration into clinical workflows in a manner that integrates the holistic decision-making power of both human clinicians and AI. GMAI could have a role analogous to an additional source of information on tumour boards[9] and to the junior healthcare provider in a referral cascade[10]. Decision support approaches without a physician in the loop are probably unsuitable to the currently conceived GMAI models, which often provide partial or incorrect interpretations. It is important that physicians are appropriately trained to avoid automation bias (that is, being lulled over time into accepting bad advice from automated systems, a particular concern with highly plausible GMAI models) and be provided with explanations of the medical context of the outputs of GMAI models to assist their own critical decision making. This is challenging as explanations can lead to greater automation bias or blind trust in AI decision support systems[11]. Further research is needed on how to best train physicians to make good judgments alongside GMAI advice, and how GMAI advice should be best presented to physicians. Any regulatory framework for GMAI use in hospitals must be linked to careful expert oversight that is enforced through legislation.

Conclusions

Current regulations can block healthcare system provision of GMAI models but cannot effectively block unofficial doctor or patient access. "You shall not pass!" is not the best approach for ensuring patient safety, supporting doctors or delivering the best care. Instead, non-device approaches applied to multimodal GMAI models, as an extension of the current FDA approach to clinical decision support systems, would better enable the new role of the physician as empowered information interpreter.

Stephen Gilbert & Jakob Nikolas Kather
Else Kröner Fresenius Center for Digital Health, TUD Dresden University of Technology, Dresden, Germany.
e-mail: stephen.gilbert@tu-dresden.de

Published online: 16 April 2024

References

1. Moor, M. et al. Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
2. Tu, T. et al. Towards generalist biomedical AI. *NEJM AI* **1**, AIoa2300138 (2024).
3. Liang, M., Li, Z., Chen, T. & Zeng, J. Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **12**, 928–937 (2015).
4. Bubeck, S. et al. Sparks of artificial general intelligence: early experiments with GPT-4. Preprint at https://doi.org/10.48550/arXiv.2303.12712 (2023).
5. Gilbert, S., Harvey, H., Melvin, T., Vollebregt, E. & Wicks, P. Large language model AI chatbots require approval as medical devices. *Nat. Med.* **29**, 2396–2398 (2023).
6. Singhal, K. et al. Large language models encode clinical knowledge. *Nature* **620**, 172–180 (2023).
7. FDA. Use of public human genetic variant databases to support clinical validity for genetic and genomic-based in vitro diagnostics. https://www.fda.gov/media/99200/download (2018).
8. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J. & Shah, S. P. Harnessing multimodal data integration to advance precision oncology. *Nat. Rev. Cancer* **22**, 114–126 (2022).
9. Cohen, I. G. et al. How AI can learn from the law: putting humans in the loop only on appeal. *npj Digit. Med.* **6**, 1–4 (2023).
10. Morrell, W., Shachar, C. & Weiss, A. P. The oversight of autonomous artificial intelligence: lessons from nurse practitioners as physician extenders. *J. Law Biosci.* **9**, lsac021 (2022).
11. Balagopalan, A. et al. The road to explainability is paved with bias: measuring the fairness of explanations. In *2022 ACM Conference on Fairness, Accountability, and Transparency* 1194–1206 (ACM, 2022).

Competing interests

S.G. has or has had consulting relationships with Una Health GmbH, Lindus Health Ltd, Flo Ltd, Thymia Ltd, FORUM Institut für Management GmbH, High-Tech Gründerfonds Management GmbH and Ada Health GmbH; and he holds share options in Ada Health GmbH. J.N.K. provides consulting services for Owkin, France; DoMore Diagnostics, Norway and Panakeia, UK, he holds shares in StratifAI GmbH and he has received honoraria for lectures by Bayer, Eisai, MSD, BMS, Roche, Pfizer and Fresenius.

Additional information

Peer review information *Nature Reviews Cancer* thanks Olivier Elemento and the other anonymous reviewers for their contribution to the peer review of this work.