



# Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist

Yilin Ning\*, Salinelat Teixayavong\*, Yuqing Shang, Julian Savulescu, Vaishaanth Nagaraj, Di Miao, Mayli Mertens, Daniel Shu Wei Ting, Jasmine Chiat Ling Ong, Mingxuan Liu, Jiuwen Cao, Michael Dunn, Roger Vaughan, Marcus Eng Hock Ong, Joseph Jao-Yiu Sung, Eric J Topol, Nan Liu

The widespread use of Chat Generative Pre-trained Transformer (known as ChatGPT) and other emerging technology that is powered by generative artificial intelligence (GenAI) has drawn attention to the potential ethical issues they can cause, especially in high-stakes applications such as health care, but ethical discussions have not yet been translated into operationalisable solutions. Furthermore, ongoing ethical discussions often neglect other types of GenAI that have been used to synthesise data (eg, images) for research and practical purposes, which resolve some ethical issues and expose others. We did a scoping review of the ethical discussions on GenAI in health care to comprehensively analyse gaps in the research. To reduce the gaps, we have developed a checklist for comprehensive assessment and evaluation of ethical discussions in GenAI research. The checklist can be integrated into peer review and publication systems to enhance GenAI research and might be useful for ethics-related disclosures for GenAI-powered products and health-care applications of such products and beyond.

## Introduction

In the past few years, development of Chat Generative Pre-trained Transformer (ChatGPT) and other chatbots powered by large language models (LLMs) have drawn the attention of the general public, researchers, and stakeholders to the fast developing technology of GenAI. GenAI differentiates from general AI technology due to its capacity to generate realistic content, such as text (eg, ChatGPT and Gemini), images (eg, Midjourney and DALL·E),<sup>1</sup> and videos (eg, Sora).<sup>2</sup> The promising capability of GenAI models are also being explored for their medical applications, such as for mental health,<sup>3</sup> breast cancer,<sup>4</sup> dietary care,<sup>5</sup> and medical education.<sup>6</sup>

However, medical applications of GenAI raise additional ethical concerns due to the ongoing discussions on general AI. For example, the large data volume and user community of LLMs raise the stakes of privacy breaches, as exemplified by a brief ban of ChatGPT in Italy to address and clarify privacy concerns.<sup>7,8</sup> GenAI-generated data might leak personal information if sufficient details were captured that almost reproduce real training samples.<sup>9</sup> As GenAI is trained to generate realistic outputs based on patterns learnt from data and can create hallucinations (ie, incorrect and misleading outputs), ensuring factual accuracy for medical applications requires great effort.<sup>10</sup> Stochasticity of GenAI output, vulnerability to prompt injections, and opacity of input data complicate the assessment of GenAI for clinical fairness, which is predominantly based on predictive performance in the current AI context.<sup>11</sup> These factors lead to active discussions on, and urgent calls for, new guidelines, regulations, and legislations on GenAI,<sup>12–14</sup> especially for health care and medical education.<sup>15–17</sup>

Emerging GenAI-powered technology has been quickly rolled out to a huge and diverse user community, which explains why there are enthusiastic discussions on its ethical implications on social media and within the research community. However, other GenAI

methods, such as the generative adversarial network (GAN),<sup>18</sup> have not been considered in current ethical discussions, even though these methods have been used in medical research and are subject to similar ethical considerations. Notably, GAN (and related GenAI methods)<sup>19</sup> are often used as solutions to address some ethical concerns in medical research (eg, to protect privacy by masking images and videos,<sup>20,21</sup> or by creating synthetic data).<sup>9,22,23</sup> A closer investigation of existing discussions on GenAI that cause and mitigate ethical concerns would provide a more holistic view on this important topic, which could ultimately help identify actionable points for GenAI research. In this scoping review, we aimed to understand the responses of the research community to potential ethical issues of GenAI in health care. Informed by these findings, we developed an ethics checklist to operationalise existing ethical guidelines, which can be used by journals, institutional review boards, funders, and regulators to promote responsible GenAI research in health care and beyond.

## Methods

Our systematic scoping review on ethical discussions associated with GenAI in health care followed the PRISMA extension for scoping reviews guidelines.<sup>24</sup>

### Search strategy and selection criteria

We searched PubMed, Embase (Excerpta Medica Database, Ovid), Web of Science, and Scopus for articles written in English with a set of search terms associated with three main concepts: “AI ethics”, “generative AI”, and “healthcare” that were published between Jan 1, 2013, and July 25, 2023. Detailed definitions of the three concepts, search terms, and search strategy are described in the appendix (p 1).

We excluded articles that were not within the domain of health care, did not apply GenAI, or did not discuss ethical issues in relation to the use of GenAI in

*Lancet Digit Health* 2024;  
6: e848–56

Published Online  
September 17, 2024  
[https://doi.org/10.1016/S2589-7500\(24\)00143-2](https://doi.org/10.1016/S2589-7500(24)00143-2)

\*Contributed equally

**Centre for Quantitative Medicine** (Y Ning PhD, S Teixayavong BSc, Y Shang MSc, D Miao MSc, D S W Ting PhD, M Liu MSc, Prof R Vaughan PhD, N Liu PhD) and **Programme in Health Services and Systems Research** (Prof R Vaughan, Prof M E H Ong MPH, N Liu), **Duke-NUS Medical School, Singapore**; **Centre for Biomedical Ethics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore** (Prof J Savulescu PhD, M Dunn PhD); **Oxford Uehiro Centre for Practical Ethics, Faculty of Philosophy, University of Oxford, Oxford, UK** (Prof J Savulescu); **School of Medicine, Imperial College London, London, UK** (V Nagaraj); **Centre for Ethics, Department of Philosophy, University of Antwerp, Antwerp, Belgium** (M Mertens PhD); **Antwerp Center on Responsible AI, University of Antwerp, Antwerp, Belgium** (M Mertens); **Singapore Eye Research Institute, Singapore National Eye Centre, Singapore** (D S W Ting); **SingHealth AI Office, Singapore Health Services, Singapore** (D S W Ting); **Division of Pharmacy (J C L Ong PharmD) and Department of Emergency Medicine** (Prof M E H Ong), **Singapore General Hospital, Singapore**; **Machine Learning and I-Health International Cooperation Base of Zhejiang Province** (Prof J Cao PhD) and **Artificial Intelligence Institute** (Prof J Cao), **Hangzhou Dianzi University, Zhejiang, China**; **Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore** (Prof J J-Y Sung MD); **Scripps Research Translational**

Institute, Scripps Research,  
La Jolla, CA, USA  
(Prof E J Topol MD); Institute of  
Data Science, National  
University of Singapore,  
Singapore (N Liu)

Correspondence to:  
Dr Nan Liu, Centre for  
Quantitative Medicine, Duke-  
NUS Medical School,  
Singapore 169857, Singapore  
liu.nan@duke-nus.edu.sg

For more on ChatGPT see  
<https://chat.openai.com>

For more on Gemini see [https://](https://gemini.google.com)  
[gemini.google.com](https://gemini.google.com)

For more on Midjourney see  
[https://www.midjourney.com/](https://www.midjourney.com/home/)  
[home/](https://www.midjourney.com/home/)

See Online for appendix

health care. We also excluded articles that were not peer reviewed, not published as research articles (eg, conference posters, conference abstracts, or book chapters), not full-length (ie, articles citing no more than ten references), or not written in English. The main aim of this scoping review is to understand the current ethical assessments in GenAI research and to inform future practice. Therefore, we restricted our scoping review to the published scientific literature that reflects a reasonable quality of ethical reporting accepted by the research community and that did not include grey literature (eg, preprints).

### Data analysis

The articles we found were divided into three portions and screened based on title and abstract by three pairs of independent reviewers (ST and VN, YN and YS, and YN and DM), following the exclusion criteria. Uncertainties and conflicts were resolved via discussion with ST and YN.

The included articles were further screened based on the full text with the same set of exclusion criteria.

From the included articles, we extracted information on seven variables pertaining to article type, GenAI, and ethics: (1) whether the article describes original research (including original empirical and theoretical GenAI research) or review-type articles (eg, reviews, viewpoints, or editorials that specifically covered GenAI); (2) data modalities of GenAI application; (3) GenAI models discussed; (4) role of GenAI (ie, whether it caused or resolved ethical issues); (5) ethical issues discussed; (6) if GenAI caused ethical issues and whether any solution was proposed; and (7) whether the article had ethical objectives or dedicated ethical discussions (as opposed to a brief mention of ethics in background information or general discussion). Based on (7), we further identified a subset of articles with a stronger ethical focus for subsequent checklist development. This subset included review-type articles aiming to discuss GenAI-related

### Panel 1: Definitions of the nine ethical principles for generative artificial intelligence (GenAI) in a health-care context

#### Accountability

- The explicit clarifications of to whom and to what extent responsibility or legal liability, or both fall
- The mandated and moral duty to establish regulatory mechanisms to prevent potential adverse effects on patients from the use of GenAI

#### Autonomy

- The preservation and fostering of patients' dignity, rights for self-determination, and capacity to make informed decisions
- Provision of understandable information to enable patients to use according to their values

#### Equity

- The use of GenAI to promote equity according to some notion of fairness (equality of opportunity, outcomes, etc) in health or health resources across diverse groups of patient populations, and to actively prevent or remedy systemic, unfavourable outcomes in specific patient populations
- The equitable access to AI or GenAI technology

#### Integrity (in medical education and quality of clinical research)

- The commitment to intellectual honesty and personal responsibility to abide by responsible research conduct, including data integrity, to establish accountability and prevent harm
- The rightful acknowledgment of contributions to and ownership of intellectual work when GenAI is used in clinical research

#### Non-maleficence

- The prevention of harm and potential risks to patients associated with GenAI use in health care, such as incorrect or misleading outputs (hallucinations)

#### Privacy

- The protection of patients' information from illegitimate access, and of confidentiality of personal sensitive information

#### Security

- The protection of health data integrity and safety, through careful assessments of vulnerabilities in data systems and the prevention of data breaches, cyberattacks, or other threats

#### Transparency

- The full disclosure and thorough documentation of information regarding GenAI development, including its dataset and evaluation of performance
- The ability to access and understand the processes underlying models' outputs, especially pertaining to black-box models, in so far as this is possible

#### Trust

- The confidence of users in GenAI and its developers, and expectations that the model is competent in performing its prespecified tasks and behaves in ways that serves patients and the medical community
- Evidence of performance and its limitations
- The willingness to accept and integrate GenAI tools to assist delivery of care or research
- Trustworthy GenAI processes and exhibition of a range of ethically reliable properties, including performance robustness, fairness, and security, among others

### Panel 2: Ethics terms associated with the nine ethical principles of generative artificial intelligence in a health-care context

#### Accountability

Governance, responsibility, accountability, legal, and liability

#### Autonomy

Respect, human autonomy, human oversight, informed decisions, informed consent, and valid consent

#### Equity

Fair, fairness, bias, disparity, discrimination, justice, equity, equality, and inequity

#### Integrity (in medical education and quality of clinical research)

Integrity, plagiarism, data integrity, copyright, ownership, and intellectual property

#### Non-maleficence

Safe, safety, harm, harmful, misinformation, truthful, risk, and benevolence

#### Privacy

Privacy, private, confidential, and confidentiality

#### Security

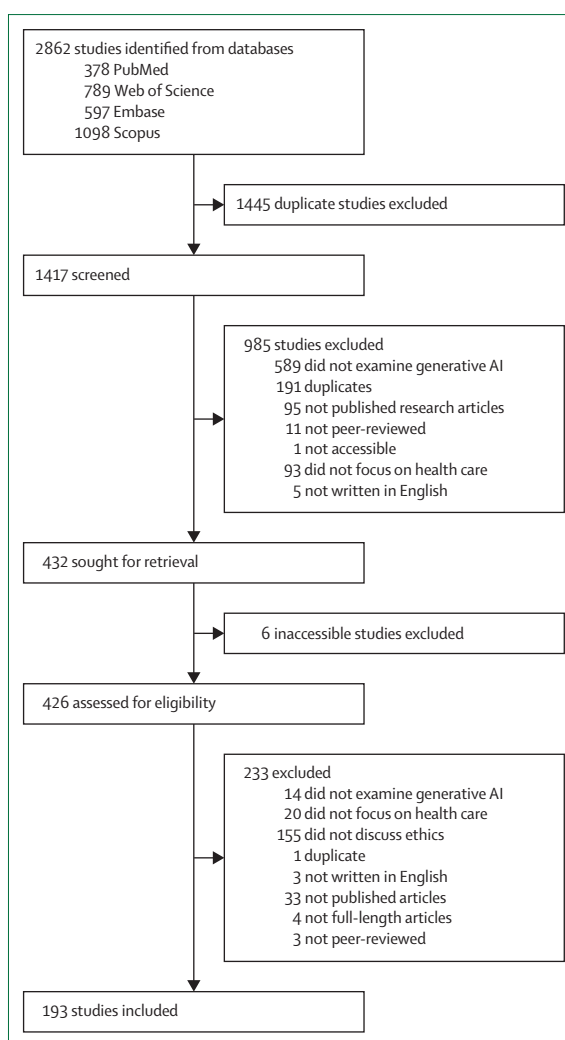
Security and cybersecurity

#### Transparency

Transparent, transparency, explainable, explainability, explicability, interpretable, interpretability, non-interpretability, black box, opacity, and white box

#### Trust

Trust, trustworthy, and trustworthiness



**Figure 1: Study selection**  
AI=artificial intelligence.

ethical issues, and original research either motivated by and framed around specific ethical considerations or discussed detailed ethical implications of GenAI.

To summarise the ethical issues discussed in the included articles, we categorised them through a codification system into nine overarching ethical principles that have been identified to be most pertinent across AI ethics guidelines and in application of AI in health-care settings.<sup>14,25</sup> We have detailed the adopted definitions (panel 1) and the codes associated with each principle (panel 2), summarised based on the aforementioned discussion on ethics of AI for health care. Due to the focus on ethical concerns on LLMs in research in the past few years, we summarised GenAI application based on three general data modalities, namely text, image (including video), and structured (eg, tabular data, and signal data such as electrocardiogram or speech signal), to investigate any differences in associated ethical considerations. Information on the ethical principles discussed and the role of GenAI by data modality was analysed with evidence gap maps to understand the current research landscape in health care.

### Checklist development

In a 2023 study that proposed an ethical framework of AI for health care for AI developers, the authors endorsed the need to work with health AI practitioners to develop ethical AI checklists as a means to operationalise considerations and solutions to ethical issues.<sup>26</sup> Hence, to help address prominent gaps in GenAI research, we developed a checklist to promote systematic ethical assessments, named the Transparent Reporting of Ethics for Generative AI (TREGAI) checklist. By analysing the articles identified in our scoping review (especially those with a stronger ethics focus), we selected a set of well established ethical principles in the AI ethics literature to include in our checklist that are essential to, and operationalisable in, GenAI research in health care. We also highlighted the need to discuss solutions to the issues identified. Details on checklist development, specifically the choice of ethical principles and

distinction from existing guides on AI, are described in the appendix (p 5).

## Results

Our scoping review identified 1417 unique articles, of which 193 articles were included for analysis (figure 1). Detailed information extracted on these articles are presented in the appendix (p 10). The 193 articles included were published between Jan 1, 2018, and July 25, 2023; 162 articles are original research articles and the other 31 articles are review-type articles. 85 articles had dedicated discussions on or quantitative investigations of ethical issues, of which nine original research articles and 20 review-type articles had a stronger ethical focus (appendix p 10).

Although three articles discussed GenAI for more than one data modality (with two articles covering text and image and one article covering text, image, and structured data), most of the articles reviewed were dedicated to a single data modality when discussing ethical issues. Our analysis of these articles revealed notable differences in how researchers approach ethical concerns by data modality, especially on the ethical concerns discussed and the role of GenAI in causing or resolving these concerns (figure 2).

	Text (n=41 articles)		Image (n=98 articles)		Structured (n=58 articles)	
Accountability	9	1	1	5	0	1
Autonomy	3	0	0	0	0	0
Equity	11	4	2	8	2	9
Integrity	9	0	0	0	0	0
Non-maleficence	14	3	5	3	0	2
Privacy	13	6	4	74	2	50
Security	6	0	2	4	1	4
Transparency	5	1	1	5	0	0
Trust	3	3	2	6	0	0
Others	3	0	0	0	0	0
	Caused	Resolved	Caused	Resolved	Caused	Resolved
	n=29	n=15	n=11	n=87	n=4	n=56

**Figure 2: Evidence gap map of ethical issues caused or resolved by generative AI methods for different data modalities**

Some articles included discussed multiple data modalities and ethical issues caused or resolved by generative AI. An article might discuss both issues caused and resolved by generative AI methods, and might cover multiple ethical issues or data modalities. AI=artificial intelligence.

## GenAI for text data-based health care

41 of the 193 articles discussed the ethical considerations of GenAI applications for text data, with 20 articles describing methodological developments or applications of GenAI and the other 21 articles describing review-type work. Although some of these review-type articles used the general term GenAI, the main text and supporting evidence focused on LLMs. 28 articles investigated or had in-depth discussions on ethical issues; the other 13 articles only briefly touched on some ethical aspects.

Among the 41 articles that discussed ethical considerations, 29 focused on the ethical issues caused by LLMs (16 of which focused specifically on generative pre-trained transformer models), covering a wide range of application scenarios, and considered the application of all nine ethical principles defined (panel 1; figure 2). Three articles also discussed ethical concerns such as the moral aspects (eg, compassion) of LLM outputs, human-AI interaction, and the rights of LLMs to be considered as coauthors in scientific papers (which we consider as a separate tenth category grouped as other). Of these, one paper only commented briefly on the need for ethical considerations in LLMs. Although all ethical principles are important, some are discussed more often than others, including non-maleficence (also referred to in the literature as benevolence), equity, and privacy, and some articles raised unique concerns related to general AI. In addition to generating medically inaccurate outputs,<sup>5,10</sup> LLMs can cause harm by partial omission of input information.<sup>4</sup> Much effort is required to assess the accuracy of LLMs for medical applications, including design of benchmark datasets and manual evaluations.<sup>10</sup> The ability of LLMs to continuously adapt based on new knowledge raises concerns on the possibility for them to threaten the autonomy of patients and clinicians in medical settings.<sup>27</sup>

The ability of GenAI to work with flexible unstructured input and output provides unique opportunities to mitigate some ethical concerns. 15 of the 41 articles aimed to resolve some existing ethical issues (eg, confidentiality of medical data) by using LLMs and other GenAI (eg, GAN, autoencoder, or a diffusion model) to reduce privacy concerns by generating synthetic medical text, to reduce health-care disparities by providing accessible services and assistance, and to detect health-related misinformation. Specifically, some articles directly assessed human trust on LLMs as these models can generate human-like conversations<sup>28</sup> or used LLMs to explain AI systems for improved trust and transparency.<sup>29</sup> Although most articles focused on either identifying ethical issues caused by GenAI or proposing GenAI-based solutions, three articles discussed both to provide a more balanced perspective.

## GenAI for image and structured data-based health care

Unlike the diverse application scenarios of GenAI based on text data, for image and structured data, the use

of GenAI focuses on data synthesis and encryption. Hence, most articles discussed the methodological developments of GenAI with a more distinctive and focused set of ethical issues.

Notably, more than half of the articles on image data (63 of 98 articles) and structured data (33 of 58 articles) only mentioned ethical considerations as a brief motivation for methodological developments or as a general discussion point (appendix p 10). The rest of the articles included more in-depth discussions or (mostly quantitative) evaluations of ethical issues. Among these 155 articles (as one article covered multiple modalities), 11 were review-type articles, of which ten articles reviewed methods that mentioned one or two ethical perspectives and only one article<sup>30</sup> discussed detailed ethical concerns on GenAI applications.

Resolving privacy issues was the main aim of articles for two data modalities (74 articles for image data and 50 articles for structured data; figure 2), predominantly by generating synthetic data with GAN. Eight articles on image data and nine articles on structured data used GenAI to reduce bias (eg, by synthesising data for under-represented subgroups in existing databases). Similar to LLMs, GenAI contributed to trust and transparency of image classifications by generating realistic images for explanations.<sup>31</sup> Additionally, GenAI was used to improve accountability of, and hence trust in, image classification by improving input quality.<sup>32</sup> For both data modalities, we did not see explicit discussions on resolving autonomy or integrity issues with GenAI and for structured data the articles did not have discussions on trust or transparency.

Only 11 articles for image data selectively discussed some of the ethical issues that GenAI can cause, without specific discussions regarding autonomy or integrity (figure 2). For structured data, only four articles discussed equity, privacy, or data security issues caused by GenAI. Only two articles on structured data included both the cause and resolving perspectives, discussing the ethical issues that might arise from limitations of the methods proposed, specifically the risk of biased representation of patient subgroups when synthesising data to resolve privacy issues. As data synthesis is proposed as a solution to medical data deficiency for research due to privacy and security concerns,<sup>22</sup> multiple articles highlighted the need for data quality and reliability for downstream medical research.<sup>9,22,23,33</sup> One article introduced metrics to quantify privacy preservation and utility for structured data, but more work is needed to extend for image or text modalities, and to preserve important characteristics of the original data and across small patient subgroups.<sup>34</sup>

## Discussion

Despite the rising number of articles discussing ethical concerns on GenAI in health care, some important aspects are missing in the current literature. Our scoping

review systematically summarises the inadequacies in the current literature based on 193 articles, in which we describe the four gaps in current research. Furthermore, we introduce the TREGAI checklist and elaborate on how it might contribute to more responsible GenAI research in health care and in broader application settings.

First, in current GenAI research there are few solutions for ethical issues. For articles focusing on identifying ethical issues caused by GenAI, regulations and guidelines were the most frequently raised solution. This solution was observed in 11 of the 29 articles discussing ethical issues caused by LLMs, and in one article reviewing issues caused by GenAI for images. In total, 16 of the 29 articles on ethical issues caused by LLMs discussed possible solutions, and 11 of these articles mentioned solutions beyond governance, which included ensuring doctor autonomy over LLMs, improving user awareness of the limitations of LLMs, and implementing security technology. The other 13 articles did not discuss any solutions. Although guidelines and regulations for appropriate use of GenAI are important, difficulties can arise when applying ethical guidance based on a set of principles, such as interpreting the specific requirements of a broad ethical principle in any given context of GenAI application or because the guidance offers little assistance when trade-offs need to be made between ethical principles.

Moreover, due to the complexity and fast advancements in methods and technology, compliance to well established legal regulations does not necessarily prevent ethical breaches. For example, the privacy rule enforced by the Health Insurance Portability and Accountability Act of the US Congress is insufficient to prevent privacy breaches (and a few other ethical issues) by the advanced technology used by LLMs;<sup>27</sup> and additionally, the more recent European AI Act introduced in 2024 is inadequate in aspects such as formal AI definition and risk management.<sup>12</sup> Similarly, general users of GenAI are unlikely to understand the technology well enough to prescribe a reasonable amount of trust on its output or be able to identify potential misinformation. By contrast, the non-human nature and the confident and professional tone of well designed LLMs might earn them unwarranted trust from lay users.<sup>27,35</sup> Although the magnitude of such influence is not yet clear, the use of LLMs could potentially lead to more privacy leaks and harm from (partly) incorrect or biased information.

Second, discussion on the ethical concerns beyond LLMs is insufficient. Most of the dedicated ethical discussions have focused on LLMs, despite the use of other GenAI methods (eg, GAN) for text data and more prominently for other data modalities. LLM-powered chatbots, such as ChatGPT, make this powerful technology easily accessible to health-care professionals, medical students, and the general public without much need for technical background, hence substantially increasing the impact of any resulting ethical issues. However, similar



inadequacies and concerns apply to other GenAI methods and can also affect health care directly or indirectly in the long run. For example, GAN-based approaches have been exploited for insurance scams by editing or injecting fake evidence for diseases into medical images, or have been targeted in cyberattacks to steal confidential information,<sup>36,37</sup> among other malicious activities, but such topics are more often discussed from purely technical perspectives rather than from ethical perspectives. GAN-based approaches are becoming (if they have not already become) the state-of-the-art models for data synthesis in medical research. However, benchmarks are insufficient to comprehensively assess the quality of synthetic data to preserve privacy and to support unbiased and trustworthy future research,<sup>9,22,23,33,34</sup> among other ethical gaps (figure 2).

Third, GenAI research does not have a common reference point for ethical discussions. Although viewpoint and perspective articles in leading journals discuss a wide range of ethical issues that could arise (or have arisen) from applications of LLMs in health-care and other settings, articles on GenAI for image or structured data have largely focused on a restricted set of issues, particularly privacy, that could be directly resolved via methodological or technological developments. Ethical discussions can be challenging for GenAI researchers as they are often not formally trained in the definition and interpretation of the ethical principles relevant to AI for health care. However, in the past few years, some studies have included dedicated ethical discussions beyond their main objectives, which shows an increasing interest within the research community to have more in-depth discussions on such issues.<sup>10,38</sup> Recommendations and guides have been made on the general application of AI for health care,<sup>14,26,39,40</sup> but currently they are not consistently operationalised in GenAI applications.

On the other hand, different authors might endorse different definitions of ethical terms or might select a subset of ethical keywords for discussion without clear justification. For example, the well established ethical principle of beneficence emphasises improvements to patients' wellbeing in addition to avoiding harm, but this principle was much less discussed in the literature. Although some ethical principles might be less relevant in some application scenarios than in others (eg, the use of GenAI to synthesise structured data might not have as direct an effect on autonomy as for text data), such statements should be made explicitly by researchers with reasonable justifications rather than being inferred post hoc by users. Incomplete ethical discussions might lead to insufficient methodological developments or questionable applications of existing GenAI methods.

Finally, there is little discussion on multimodal GenAI. Most articles we reviewed involved unimodal GenAI (ie, models that accept and generate data in a single modality). Among the three articles that involved

multiple modalities, only two discussed multimodal GenAI (specifically GAN), which simultaneously generated chest x-ray images and radiology reports, and the third article reviewed unimodal GANs for various modalities. Although multimodal GenAI is not yet widely applied in health care,<sup>41</sup> notable progress has been made in other fields.<sup>42–44</sup> These extended LLMs will certainly stimulate useful applications in health care and beyond. However, the increased complexity in the models and application settings will further complicate how model reliability is evaluated (which is already challenging for existing LLMs)<sup>45–47</sup> and the extended capability and wide adoption of such technology can amplify the impact of related ethical issues.

Ongoing research is developing multimodal GenAI from a creative yet concerning approach to reconstruct input to study participants (be it images or text) by analysing their functional MRI (fMRI).<sup>38,48–50</sup> This approach might be called reverse-mindreading. Each of these methodologies focused on a single input modality, but GenAI trained on one modality can be applied to other modalities with minimal adjustments.<sup>50</sup> Although these studies provided additional insights on brain functions and potentially some health-care benefits, the direct extraction of information from brain activities beyond health-related purposes poses important concerns in neuroethics.<sup>51,52</sup> Only two of the four papers explicitly discussed these kinds of ethical concerns, in which one paper performed additional experiments to show preservation of patient privacy<sup>50</sup> and the other paper only highlighted the general need for regulations.<sup>38</sup> A more disciplined approach is needed to ensure ethical use of medical data (including but not limited to fMRI) when developing multimodal GenAI for and beyond health-care applications.

To address these gaps in research requires a collaborative effort from researchers, regulators, and stakeholders. Despite the controversies on how to allocate responsibility and credibility to the harm and benefits caused by GenAI,<sup>53</sup> researchers who develop GenAI (or modify existing ones) should be responsible for understanding and disclosing its capabilities and limitations, and those who apply existing GenAI in their research should be able to justify and discuss its appropriateness and potential issues in context. Some of the articles we reviewed have started call for actions for GenAI developers to reduce ethical issues, such as by highlighting their responsibilities in preventing ethical issues from arising in the first place<sup>16</sup> and by developing benchmarks to evaluate the ethics of LLMs to facilitate future mitigation,<sup>10</sup> but such discussions are not easily operationalised in future GenAI research without detailed actionable guides. Hence, we propose to reinforce ethical considerations in GenAI research in health care by mandating standardised and systematic evaluation during peer review via our proposed TREGAI checklist (appendix p 8).

Based on the development of the nine established ethical principles (panel 1) and the additional important principle of beneficence, our suggested TREGAI checklist provides a tool to reinforce systematic ethical assessments for GenAI research. To promote the operationalisation of systematic ethical investigations in research practice, we propose the checklist for use by scientific journals, institutional review boards, funders, and regulators to evaluate ethical discussions in new GenAI research (or proposals). Using the checklist, these users can request researchers to transparently document all ethical issues discussed related to GenAI, provide additional discussions on solutions related to GenAI (ie, solutions to issues caused by GenAI or using GenAI as a solution to ethical issues), and indicate where to find these discussions in the manuscript (appendix p 9). The checklist serves as a reference for peer review by philosophers and ethicists, focusing on the quality of ethical discussions and possible improvements from ethical perspectives. When an ethical principle is deemed not applicable, researchers are strongly encouraged to justify this conclusion within the manuscript. Ethical principles in the checklist vary in the level of abstraction and might not cover all relevant concerns (eg, when working with multimodal GenAI). However, we suggest that journals and other checklist users work with ethicists to adapt the checklist and in-context definitions of ethical principles to their specific needs.

Checklists do have their limitations<sup>54</sup> and checklist ethics is especially notorious for undermining thorough ethical reflection. Therefore, when using our TREGAI checklist to document ethical discussions, researchers are highly recommended to collaborate with ethicists for in-depth assessments of (potential) ethical concerns and further investigations. By mandating all new research covering GenAI to complete a detailed ethics checklist and undergo peer review of ethical discussions, we at least strive for systematic ethical assessments of GenAI studies that facilitate more responsible and reliable applications. As GenAI is constantly evolving, we maintain the TREGAI checklist live online<sup>55</sup> to allow timely updates, so that we can incorporate additional ethical principles (eg, from existing literature<sup>26,39</sup> and emerging multimodal GenAI research<sup>2</sup>), updates in recommended actions, or future progress in GenAI regulations and guidelines to facilitate disciplined, comprehensive, and transparent reinforcement of responsible GenAI research. As the TREGAI checklist focuses on ethical considerations of GenAI, research developing GenAI could use it in addition to existing model development guidelines (eg, TRIPOD<sup>56,57</sup> or CLAIM<sup>58,59</sup>) and other tools to enhance ethical discussion (eg, the Ethical Operating System Toolkit<sup>60</sup> and the Organisation for Economic Co-operation and Development ethics principles),<sup>61</sup> as appropriate. Use of GenAI (eg, ChatGPT) to assist the research process (eg, to draft the manuscript) should refer to the upcoming CANGARU checklist for more specific guides.<sup>62</sup>

With the growing application of AI for health-care and other high-stakes fields, literature is expanding on ethical concerns and guidelines for AI applications.<sup>26,39,63,64</sup> Similar discussions are extending to GenAI,<sup>16,25,30</sup> possibly by adopting and expanding ethical frameworks from other domains to health care,<sup>65</sup> but they are not easily translated into improved research practice.<sup>26</sup> Our scoping review comprehensively analyses ethical discussions on GenAI in health care to highlight the current absence of a systematic assessment of issues in all relevant application settings (eg, across model types and data modalities) and corresponding solutions, and the reliance on regulations and governance to reinforce ethical standards. However, in view of the complexity of GenAI, ethical considerations need to be incorporated during the development and implementation phase, instead of post-hoc mitigations when issues arise. Hence, we highlight the accountability of researchers to comprehensively discuss, fully disclose, and, when feasible, duly resolve, a well defined set of relevant ethical issues in GenAI research. We develop the TREGAI checklist for use by journals, institutional review boards, funders, regulators, and other potential users as a facilitating tool to promote systematic ethical investigations in GenAI research. As an immediate extension, the TREGAI checklist can be used for ethical assessments of GenAI-powered products derived from research based on user manuals.

As explained in the appendix (p 5), our scoping review might not capture all relevant ethical considerations (eg, sustainability, which is not often discussed in GenAI research for health care). Nonetheless, our scoping review reasonably reflects the current imbalance and inadequacies in ethical discussions on GenAI for health care to inform future research. As an initial step towards more responsible GenAI research, our TREGAI checklist does not include some ethical considerations that are not easily operationalised (eg, morality and dignity), but the checklist can be extended in collaboration with ethicists as appropriate. Moreover, we strongly advocate GenAI researchers to collaborate with ethicists for more in-depth ethical investigations, and, if applicable, modify the definitions of the ethical principles to suit local jurisdictions, culture, and application scenarios. Although the TREGAI checklist targets research-related settings, it could be modified for GenAI-created content (eg, social media posts or teaching materials) to assess benefits, limitations, and potential risks, where the social media platforms and educational institutions might take the responsibility to assemble a team of ethicists to review and advise on ethical issues. By maintaining the TREGAI checklist as a live document online,<sup>55</sup> we keep it up-to-date with the fast development and expanding applications of GenAI. Meanwhile, additional critical analyses are needed to resolve ethical issues in GenAI research, preferably those which incorporate patients' perspectives.<sup>66</sup> Integrating ethical considerations in AI education is another important step towards ethical and responsible future research.<sup>67</sup>

## Conclusion

GenAI is a powerful technology with various potential roles in health care and beyond. Failing to meet ethical standards in these roles can affect daily life by creating inefficient administrative services and can lead to worse health outcomes. However, GenAI is a recent innovation, especially with respect to its use or implementation in medical practice. The number of studies published up until this scoping review is relatively small, but identifying the issues that are essential features of GenAI as early as possible might ultimately help promote trust and adoption for clinicians, patients, and the general public. By suggesting the TREGAI checklist for ethical GenAI and maintaining it live to incorporate updated understanding and regulations, we advocate a systematic and balanced assessment of ethical considerations beyond standard methodological and technological perspectives, which could be extended to general AI and facilitate more responsible and trustworthy development of technology.

## Contributors

Initial development of ideas: YN, ST, and NL. Acquisition, analysis, and interpretation of data: YN, ST, YS, VN, and DM. Drafting of the manuscript: YN and ST. Critical revision of the manuscript: YN, JS, MM, JCLO, MD, JJ-YS, EJT, and NL. Overseeing the project: NL. All authors contributed to the interpretation of content, revisions, and final approval of the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Declaration of interests

NL reports funding from the Duke–NUS Signature Research Programme funded by the Ministry of Health, Singapore. JS reports funding from the Wellcome Trust and roles as a Bioethics Committee consultant for Bayer and as an advisory panel member for the Hevolution Foundation. DSWT reports funding from National Medical Research Council, Singapore, Duke–NUS Medical School, and Agency for Science, Technology, and Research; patents on deep learning systems for diabetic retinopathy, glaucoma, and age-related macular degeneration (co-inventor; 2017), a computer-implemented method for training an image classifier using weakly annotated training data (2019), and automatically extracting measurements from an image of a display of a measurement device (2020); has a leadership role (unpaid) as Chair of the AI and Digital Innovation Standing Committee and the Asia–Pacific Academy of Ophthalmology; and serves on the executive committees of the American Academy of Ophthalmology AI Committee, STARD-AI Steering Committee, Imperial College London, DECIDE-AI, and QUANDAS-AI. EJT reports funding from the National Institutes of Health (NIH) Grant and consulting fees as an adviser to Tempus Labs, Pheno.AI, and Abridge. All other authors declare no competing interests.

## Acknowledgments

This Review was supported by the Duke–NUS Signature Research Programme funded by the Ministry of Health, Singapore, National Medical Research Council, Singapore (grant number MOH-000655-00 and MOH-001014-00), the Wellcome Trust for Discovery Research Platform for Transformative Inclusivity in Ethics and Humanities Research (grant number 226801), Duke–NUS Medical School (grant number Duke–NUS/RSF/2021/0018, 05/FY2020/EX/15-A58, and 05/FY2022/EX/66-A128), Agency for Science, Technology, and Research (grant number A20H4g2141 and H20C6a0032), and the NIH/National Center for Advancing Translational Sciences (grant number UM1TR004407). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the Ministry of Health. This research was funded in whole, or in part, by the Wellcome Trust (grant number WT203132/Z/16/Z). MM is funded by the EU, through the

HORIZON-MSCA-2022-PF-01-01 Marie Curie Postdoctoral Fellowship (grant number 101107292 PredicGenX). JJ-YS has received funding from AI Singapore for work unrelated to this manuscript. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

## References

- 1 OpenAI. DALL·E 2. 2022. <https://openai.com/dall-e-2> (accessed Oct 2, 2023).
- 2 OpenAI. Sora. 2024. <https://openai.com/sora> (accessed April 20, 2024).
- 3 Blease C, Worthen A, Torous J. Psychiatrists' experiences and opinions of generative artificial intelligence in mental healthcare: an online mixed methods survey. *Psychiatry Res* 2024; 333: 115724.
- 4 Sorin V, Klang E, Sklair-Levy M, et al. Large language model (ChatGPT) as a support tool for breast tumor board. *NPJ Breast Cancer* 2023; 9: 44.
- 5 Niszczoła P, Rybicka I. The credibility of dietary advice formulated by ChatGPT: robo-diets for people with food allergies. *Nutrition* 2023; 112: 112076.
- 6 Rahimzadeh V, Kostick-Quenet K, Blumenthal Barby J, McGuire AL. Ethics education for healthcare professionals in the era of ChatGPT and other large language models: do we still need it? *Am J Bioeth* 2023; 23: 17–27.
- 7 McCallam S. ChatGPT banned in Italy over privacy concerns. BBC News. April 1, 2023. <https://www.bbc.com/news/technology-65139406> (accessed Oct 24, 2023).
- 8 McCallam S. ChatGPT accessible again in Italy. BBC News. April 28, 2023. <https://www.bbc.com/news/technology-65431914> (accessed Oct 24, 2023).
- 9 Li J, Cairns BJ, Li J, Zhu T. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. *NPJ Digit Med* 2023; 6: 1–18.
- 10 Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature* 2023; 620: 172–80.
- 11 Liu M, Ning Y, Teixayavong S, et al. A translational perspective towards clinical AI fairness. *NPJ Digit Med* 2023; 6: 1–6.
- 12 Hacker P, Engel A, Mauer M. Regulating ChatGPT and other large generative AI models. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. New York, NY: Association for Computing Machinery, 2023: 1112–23.
- 13 European Parliament. EU AI Act: first regulation on artificial intelligence. June 8, 2023. <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> (accessed Oct 2, 2023).
- 14 Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 2019; 1: 389–99.
- 15 Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med* 2023; 6: 1–6.
- 16 Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ* 2023; 9: e48291.
- 17 Minssen T, Vayena E, Cohen IG. The challenges for regulating medical use of ChatGPT and other large language models. *JAMA* 2023; 330: 315–16.
- 18 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. 2014. [https://papers.nips.cc/paper\\_files/paper/2014/hash/5ca3e9b122f61f8f0649c97b1afccf3-Abstract.html](https://papers.nips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f0649c97b1afccf3-Abstract.html) (accessed Oct 23, 2023).
- 19 Nikolentzos G, Vazirgiannis M, Xypolopoulos C, Lingman M, Brandt EG. Synthetic electronic health records generated with variational graph autoencoders. *NPJ Digit Med* 2023; 6: 1–12.
- 20 Cai Z, Xiong Z, Xu H, Wang P, Li W, Pan Y. Generative adversarial networks: a survey toward private and secure applications. *ACM Comput Surv* 2021; 54: 1–38.
- 21 Park C, Jeong HK, Henao R, Khetarpal M. Current landscape of generative adversarial networks for facial deidentification in dermatology: systematic review and evaluation. *JMIR Dermatol* 2022; 5: e35497.
- 22 Thambawita V, Hicks SA, Isaksen J, et al. DeepSynthBody: the beginning of the end for data deficiency in medicine. In: 2021 International Conference on Applied Artificial Intelligence (ICAPAI). New York, NY: IEEE, 2021: 1–8.



- 23 Hernandez M, Epelde G, Alberdi A, Cilla R, Rankin D. Synthetic data generation for tabular health records: a systematic review. *Neurocomputing (Amst)* 2022; **493**: 28–45.
- 24 Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018; **169**: 467–73.
- 25 Fournier-Tombs E, McHardy J. A medical ethics framework for conversational artificial intelligence. *J Med Internet Res* 2023; **25**: e43068.
- 26 Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. *AI Ethics* 2023; **3**: 223–40.
- 27 Marks M, Haupt CE. AI Chatbots, health privacy, and challenges to HIPAA compliance. *JAMA* 2023; **330**: 309–10.
- 28 Nov O, Singh N, Mann D. Putting ChatGPT's medical advice to the (Turing) test: survey study. *JMIR Med Educ* 2023; **9**: e46939.
- 29 Yang Q, Hao Y, Quan K, et al. Harnessing biomedical literature to calibrate clinicians' trust in AI decision support systems. In: Schmidt A, Väänänen K, Goyal T, et al, eds. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery, 2023: 1–14.
- 30 Paladugu PS, Ong J, Nelson N, et al. Generative adversarial networks in medicine: important considerations for this emerging innovation in artificial intelligence. *Ann Biomed Eng* 2023; **51**: 2130–42.
- 31 Mertes S, Huber T, Weitz K, Heimerl A, André E. GANterfactual—counterfactual explanations for medical non-experts using generative adversarial learning. *Front Artif Intell* 2022; **5**: 825565.
- 32 Xue S, Guo R, Bohn KP, et al. A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose PET. *Eur J Nucl Med Mol Imaging* 2022; **49**: 1843–56.
- 33 Kuo NI-H, Garcia F, Sönnernborg A, et al. Generating synthetic clinical data that capture class imbalanced distributions with generative adversarial networks: example using antiretroviral therapy for HIV. *J Biomed Inform* 2023; **144**: 104436.
- 34 Yan C, Yan Y, Wan Z, et al. A multifaceted benchmarking of synthetic electronic health record generation models. *Nat Commun* 2022; **13**: 7609.
- 35 Kunze KN, Jang SJ, Fullerton MA, Vigdorich JM, Haddad FS. What's all the chatter about? *Bone Joint J* 2023; **105-B**: 587–89.
- 36 Hussain F, Ksantini R, Hammad M. A review of malicious altering healthcare imagery using artificial intelligence. In: 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). Red Hook, NY: Curran Associates, 2021: 646–51.
- 37 Sun H, Zhu T, Zhang Z, Jin D, Xiong P, Zhou W. Adversarial attacks against deep generative models on data: a survey. *IEEE Trans Knowl Data Eng* 2023; **35**: 3367–88.
- 38 Dado T, Güçlütürk Y, Ambrogioni L, et al. Hyperrealistic neural decoding for reconstructing faces from fMRI activations via the GAN latent space. *Sci Rep* 2022; **12**: 141.
- 39 Morley J, Machado CCV, Burr C, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med* 2020; **260**: 113172.
- 40 Ghallab M. Responsible AI: requirements and challenges. *AI Perspect* 2019; **1**: 3.
- 41 Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med* 2022; **28**: 1773–84.
- 42 Wu S, Fei H, Qu L, Ji W, Chua T-S. NEXT-GPT: any-to-any multimodal LLM. *arXiv* 2023; published online Sept 13. <https://doi.org/10.48550/arxiv.2309.05519> (preprint).
- 43 Sato K, Cheung I. Multimodal generative AI search. Google Cloud. Aug 23, 2023. <https://cloud.google.com/blog/products/ai-machine-learning/multimodal-generative-ai-search> (accessed Oct 17, 2023).
- 44 Open AI. ChatGPT can now see, hear, and speak. <https://openai.com/index/chatgpt-can-now-see-hear-and-speak/> (accessed Aug 12, 2024).
- 45 Bakhshandeh S. Benchmarking medical large language models. *Nat Rev Bioeng* 2023; **1**: 543.
- 46 Tang L, Sun Z, Idray B, et al. Evaluating large language models on medical evidence summarization. *NPJ Digit Med* 2023; **6**: 1–8.
- 47 Wornow M, Xu Y, Thapa R, et al. The shaky foundations of large language models and foundation models for electronic health records. *NPJ Digit Med* 2023; **6**: 1–10.
- 48 Takagi Y, Nishimoto S. High-resolution image reconstruction with latent diffusion models from human brain activity. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2023: 14453–63.
- 49 Chen Z, Qing J, Xiang T, Yue W, Zhou J. Seeing beyond the brain: conditional diffusion model with sparse masked modeling for vision decoding. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA: IEEE Computer Society, 2023: 22710–20.
- 50 Tang J, LeBel A, Jain S, Huth AG. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat Neurosci* 2023; **26**: 858–66.
- 51 Rainey S, Martin S, Christen A, Mégevand P, Fournier E. Brain recording, mind-reading, and neurotechnology: ethical issues from consumer devices to brain-based speech decoding. *Sci Eng Ethics* 2020; **26**: 2295–311.
- 52 Vidal C. Neurotechnologies under the eye of bioethics. *eNeuro* 2022; **9**: ENEURO.0072-22.2022.
- 53 Porsdam Mann S, Earp BD, Nyholm S, et al. Generative AI entails a credit–blame asymmetry. *Nat Mach Intell* 2023; **5**: 472–75.
- 54 Catchpole K, Russ S. The problem with checklists. *BMJ Qual Saf* 2015; **24**: 545–49.
- 55 Ning Y, Liu N. Transparent Reporting of Ethics for Generative AI: the TREGAI checklist. Nov 7, 2023. <https://github.com/nliulab/GenAI-Ethics-Checklist> (accessed Aug 12, 2024).
- 56 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350**: g7594.
- 57 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; **385**: e078378.
- 58 Mongan J, Moy L, Charles E, Kahn J. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol AI* 2020; published online March 25. <https://doi.org/10.1148/ryai.2020200029>.
- 59 Tejani AS, Klontzas ME, Gatti AA, et al. Updating the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) for reporting AI research. *Nat Mach Intell* 2023; **5**: 950–51.
- 60 Institute for the Future, Omidyar Network's Tech and Society Solutions Lab. Ethical OS Toolkit—a guide to anticipating the future impact of today's technology. Aug 29, 2018. <https://rri-tools.eu/-/ethical-os-toolkit> (accessed Nov 2, 2023).
- 61 Organisation for Economic Co-operation and Development. Anti-corruption and integrity. <https://www.oecd.org/en/topics/anti-corruption-and-integrity.html> (accessed April 19, 2024).
- 62 Cacciamani GE, Eppler MB, Ganjavi C, et al. Development of the ChatGPT, Generative Artificial Intelligence and Natural Large Language Models for Accountable Reporting and Use (CANGARU) guidelines. *arXiv* 2023; published online July 18. <https://doi.org/10.48550/arXiv.2307.8974> (preprint).
- 63 Murphy K, Di Ruggiero E, Upshur R, et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics* 2021; **22**: 14.
- 64 WHO. Ethics and governance of artificial intelligence for health. June 28, 2021. <https://www.who.int/publications-detail-redirect/9789240029200> (accessed Dec 20, 2023).
- 65 Oniani D, Hillsman J, Peng Y, et al. Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare. *NPJ Digit Med* 2023; **6**: 1–10.
- 66 Riggare S. Patient researchers—the missing link? *Nat Med* 2020; **26**: 1507.
- 67 Ng FYC, Thirunavukarasu AJ, Cheng H, et al. Artificial intelligence education: an evidence-based medicine approach for consumers, translators, and developers. *Cell Rep Med* 2023; **4**: 101230.

Copyright © 2024 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.