Society of Medical Education & Research

2024, Volume 4, Issue 1, Page No: 171-180
ISSN: 3108-5059

Asian Journal of Ethics in Health and Medicine

Integrating Ethics into Medical AI Development: A Proposed Framework

Saira Malik[1], Imran Qureshi[2]*, Hina Shah[1]

[1]Department of Bioethics, Faculty of Health Sciences, University of Sargodha, Sargodha, Pakistan.
[2]Department of Medical Ethics, Faculty of Medicine, King Edward Medical University, Lahore, Pakistan.

***E-mail** ✉ imran.qureshi@gmail.com

Abstract

The rise of artificial intelligence (AI) has brought a surge of ethical concerns, prompting numerous public and private organizations to publish high-level ethical guidelines. However, there remains a pressing need to explore practical ways to help AI developers anticipate, recognize, and address ethical challenges in AI systems. This is particularly critical in healthcare, where AI applications frequently interact directly with patients who may be in vulnerable states. This paper proposes an 'embedded ethics' approach, in which ethicists collaborate closely with developers throughout an iterative and continuous development process, as a potential method for integrating robust ethical considerations into the practical design and deployment of medical AI.

Keywords: Ethics, Medical, AI, Ethical guidelines

Introduction

Advances in computing power, the availability of extensive digital datasets, and the creation of sophisticated algorithms have generated substantial opportunities for AI and its subfields, including machine learning, natural language processing, and robotics. AI technologies are already prevalent in daily life, and their influence on society and the economy is expected to grow significantly [1, 2].

Despite the potential benefits, AI raises numerous ethical concerns, including issues related to privacy, data protection, transparency, explainability, data bias, accountability, and the societal impacts of automation, among others [3-5]. These challenges have fueled the rapid emergence of "AI Ethics" initiatives aimed at ensuring that AI is developed and implemented responsibly. A recent scoping review found 84 publications issuing ethical principles or guidelines for AI, produced by a wide array of organizations [6]. Analyses of these documents reveal a convergence around principles similar to traditional medical ethics, such as beneficence, non-maleficence, transparency, justice, and responsibility [6-8]. While these foundational principles are helpful in guiding biomedical research and clinical practice, it remains uncertain how effectively high-level ethical guidelines can inform the technical development of AI. Some scholars argue that relying solely on these principles may be ineffective [9], or may even serve as a strategic attempt by the tech industry to avoid enforceable regulation [10, 11].

In the absence of binding legal frameworks, AI developers are largely responsible for interpreting high-level ethical principles as they see fit [9]. Even where regulations exist, complex ethical dilemmas often remain, similar to clinical ethics, where legal guidance can direct conduct but cannot resolve all moral questions. Although many technology companies are genuinely interested in addressing AI ethics [12], most developers

Access this article online https://smerpub.com/

Received: 01 February 2024; **Accepted:** 29 April 2024

How to cite this article: Malik S, Qureshi I, Shah H. Impact of Undergraduate Medical Ethics Education on Students: A Quasi-Experimental Study in Rural India. Asian J Ethics Health Med. 2024;4(1):171-80. https://doi.org/10.51847/sEFn97X9zz

lack the training necessary to operationalize abstract ethical principles. Developers come from diverse professional backgrounds, typically without formal ethics education, while few ethicists are employed in tech settings, and there is limited culture of practical collaboration between these fields.

To address this gap, initiatives have sought to enhance the ethical awareness of technical teams. For instance, Floridi and Strait note that many technology companies now employ "ethical foresight analysis" to anticipate potential ethical issues and assess the impacts of specific technologies [13]. Additionally, several leading universities and research institutions have incorporated ethics into technical curricula to cultivate developers' capacities for ethical reasoning and awareness [14, 15]. Nonetheless, fully mastering the application of ethical principles to complex, real-world issues requires extensive training and a diverse methodological toolkit. Without substantial transformation of educational programs and corporate training, it is unrealistic to expect all developers to independently identify and address the ethical challenges posed by the AI systems they create.

The need for improved integration of ethics in medical AI development

Governments worldwide continue to struggle with regulating AI technologies [16], highlighting the importance of providing practical support to AI developers in identifying ethical challenges and determining how to address them. While the necessity of ethical guidance applies across a wide spectrum of AI applications—from smartphone apps to self-driving vehicles and autonomous weapons—the high-stakes nature of healthcare makes it particularly crucial to embed ethics within medical AI development. Nevertheless, the broader concerns around ethics in technology, and the framework we propose here, are applicable across multiple domains.

AI is anticipated to revolutionize healthcare [17, 18]. By leveraging extensive clinical datasets, medical AI systems can enhance a variety of functions, including diagnosis, clinical decision-making, personalized treatment, research, drug development, and administrative operations [19–24]. Certain medical AI tools also utilize "embodied AI," which interact with patients and their surroundings through physical forms, such as intelligent robotic assistants or smart prosthetics [25–28]. These applications hold the potential to improve care quality and efficiency, reduce costs, expand access to underserved populations, and alleviate pressures on overburdened healthcare systems [29–31].

However, medical AI can pose significant risks when interacting with patients in vulnerable states, including those with diminished health or decision-making capacity, potentially threatening patient preferences, safety, and privacy [32]. Studies indicate that some medical AI systems are developed without explicit ethical considerations, and a persistent gap exists between AI development and successful integration into clinical practice [29, 33, 34]. Rapid deployment of these tools can leave patients as "unwitting guinea pigs," as medical AI often lacks the rigorous testing applied to other medical devices [35]. Such shortcomings have already resulted in serious errors, including algorithms affecting millions of patients that demonstrated significant racial bias [36]. Ethical considerations made during the design and development stages fundamentally shape the final product's form and functionality. Therefore, robust ethical evaluation must occur well before clinical testing and deployment. As AI becomes increasingly integrated into healthcare, a multi-layered approach—incorporating mechanisms for translating ethical principles into AI systems and supporting "bottom-up" AI ethics—is likely to be most effective [1, 9].

Ethics and technology development often reflect divergent cultural approaches. Ethics tends to explore issues through analogies and thought experiments, whereas technical fields emphasize concrete cases and actionable results. Ethical analysis can be time-intensive, while technology development is generally guided by efficiency and practical constraints. Given these differences, and the reality that AI developers are unlikely to become ethics experts, we argue that the recently proposed 'embedded ethics' approach offers a promising strategy for integrating comprehensive ethical considerations into the development of medical AI [37].

Embedded ethics

"Embedded ethics" has been used in past work to describe involving philosophers in computer science or data science education, as well as programming ethical principles directly into AI systems to guide algorithmic behavior [14, 38, 39]. Here, we expand the concept to encompass a continuous, iterative integration of ethics throughout the entire AI development lifecycle, transforming ethical reflection into a fully collaborative, interdisciplinary process. While this approach aligns in some ways with socio-technical frameworks like Responsible Research and Innovation (RRI), it also differs in key aspects. In the sections below, we outline the defining elements of embedded ethics, including its objectives, methods of incorporation, typical practices, and the knowledge and training required **(Table 1).**

Table 1. Summary of guidance

Domain	Guidance
Aims	1. The purpose of embedded ethics is to foresee, recognize, and address ethical challenges that emerge throughout the development of medical AI. 2. Embedded ethics should engage collaboratively with the development team, addressing ethical concerns through an ongoing, iterative process.
Integration	3. Embedded ethics should be integrated through regular interactions—formal or informal—between ethicists and technical team members.
Practice	4. The ethical frameworks applied should be clearly articulated and transparent. 5. The rationale for selected frameworks and ethical positions should be justified in relation to specific project objectives. 6. The team's decision-making process should be defined and communicated at the outset of the project. 7. Embedded ethics should explore approaches to maintain transparency of ethical analyses during AI development, while respecting confidentiality and intellectual property constraints.
Expertise and Training	8. Effective embedded ethics requires proficiency in ethical analysis within practical contexts, along with a foundational understanding of the relevant AI technology and its clinical application. 9. Projects should provide opportunities before and during development for participants to gain the necessary knowledge and skills to practice embedded ethics effectively.

Aims

The primary goal of an embedded ethics approach is to support the development of medical AI technologies that are ethically responsible and socially beneficial, ensuring that these technologies serve people without causing harm. To achieve this, ethical considerations are incorporated into every stage of development—from planning and ethics approval to design, programming, piloting, testing, and eventual implementation. Positioning ethicists throughout the development lifecycle also has the potential to advance scholarship that anticipates, rather than merely reacts to, the ethical and social tensions arising from medical AI applications.

An “ethical issue” encompasses areas of uncertainty or disagreement—whether within the development team, the literature, or society at large—about the appropriate course of action or the interpretation of ethical principles in relation to a medical AI technology [40]. Common examples include clarifying the goals of a device and its intended beneficiaries, identifying potential risks and harms and determining how these can be mitigated through software or hardware adjustments, addressing algorithm-specific concerns such as data biases and the need for explainability, evaluating the implications of physical embodiments like robotic interfaces, and considering longer-term societal consequences, such as labor displacement versus improvements in safety and efficiency.

Embedded ethics involves ongoing, iterative collaboration between ethicists and developers to address these concerns, drawing on approaches such as clinical ethics advisory in hospitals [41] or ethical, legal, and social aspects (ELSA) research in biomedical settings [42]. As development proceeds, previously unforeseen issues are likely to emerge, requiring ad hoc analyses and recommendations formulated through direct observation and consultation with developers and stakeholders. In this way, embedded ethics promotes transparency regarding uncertainties or disagreements and contributes diverse, ethically defensible options for addressing pressing challenges.

In its objectives, embedded ethics bears resemblance to socio-technical frameworks like Responsible Research and Innovation (RRI) [43–47]. Both emphasize early consideration of potential impacts, inclusion of multiple stakeholder perspectives, and broad reflection on emerging technologies. However, important distinctions exist. First, RRI largely emphasizes governance and policy, aiming to improve anticipation, reflexivity, inclusion, and responsiveness within institutional frameworks [48]. In contrast, embedded ethics works directly with developers to identify ethical considerations from the bottom up, without attempting to control or govern technological development. This approach fosters open, collaborative interdisciplinary dialogue.

Second, while RRI is sometimes linked to a precautionary principle and often seeks to ensure that innovations are ethically acceptable and marketable [45, 49], embedded ethics deliberately separates ethical integration from concerns about product acceptance or marketability. Combining these aims can risk “ethics washing” [50, 51]. While developers naturally desire their technologies to be accepted and usable, embedded ethics prioritizes identifying and addressing ethical issues independently of assumptions that the technology itself is the optimal solution to societal challenges.

Integration

The implementation of an embedded ethics approach can take several forms depending on the project’s context and available resources. Ideally, the most effective integration involves having one or more ethicists embedded as permanent members of the development team. In this setup, collaboration is maintained through frequent interactions—both formal and informal—between ethicists and technical team members. A notable example of this model can be seen in genomics research, where Jeantine Lunshof has acted as an embedded ethicist at Harvard’s Wyss Institute for Biologically Inspired Engineering, engaging in regular lab meetings, co-authoring publications, and contributing to protocol development for ongoing research [52].

When embedding an ethicist full-time is not feasible due to resource limitations, an alternative is to have ethicists join development meetings on a regular basis, either in person or via digital communication platforms. Crucially, these interactions must be scheduled consistently from the start of the project rather than being invoked only when ethical dilemmas or social conflicts arise. Sporadic or reactive involvement risks overlooking ethical concerns and undermines the development of genuine ethical awareness and critical reasoning within the team. Development teams should avoid consulting ethicists solely in response to identified harms or external pressures, as this compromises the authentic integration of ethical reflection into the development process.

In many projects, a single embedded ethicist may suffice, but larger initiatives—especially those involving multiple organizations—may require a team of ethicists. Regardless of scale, a clear protocol should be established at the outset, outlining the logistics of engagement, including frequency, timing, participants, and methods of interaction. Periodic evaluation and reflection on this process are also essential to ensure its effectiveness and to identify opportunities for improvement.

Practice

The integration of ethics into AI development should involve a clear and rigorous normative evaluation of the issues that emerge throughout the development process.

This entails elucidating and clarifying complex ethical challenges to facilitate a deeper understanding of them, as well as applying ethical reasoning to defend or critique a specific position or course of action [40]. Currently, there is no universally accepted framework or standardized methodology for AI ethics. Employing a variety of ethical perspectives is common across most areas of applied ethics and can help highlight diverse viewpoints. While embedded ethics should not dictate a prescriptive method for analyzing medical AI ethically, it is essential to ensure transparency and justification. Specifically, embedded ethicists should (1) clearly state the theoretical ethical frameworks guiding their normative analysis and (2) justify why these frameworks are appropriate for achieving the project's objectives [40]. Regardless of the chosen analytical approach, embedded ethics should involve explicitly articulating the nature of the identified ethical issues and situating them within existing literature addressing similar or comparable concerns [40]. Although literature may be limited on newly emerging issues, drawing on previous discussions in related areas can often inform the team's normative reasoning.

Ideally, embedding ethics entails the involvement of ethicists throughout the entire development process, from early planning, design, and programming stages to supporting the regulatory pathway when developments advance—for instance, by facilitating meaningful compliance with ethical review boards and established guidelines. At each stage, potential ethical challenges should be collaboratively analyzed and addressed, seeking joint solutions. For this process to be effective, the team's decision-making structure and allocation of responsibility must be clearly defined from the outset. This may follow a clinical ethics model, where ultimate responsibility for development decisions remains with the lead developer, or it could adopt a more horizontally distributed decision-making approach [37]. Nevertheless, due to the substantial financial investments in AI healthcare technologies, embedded ethicists often operate in environments with significant power imbalances. Such hierarchies may risk critical decisions being made without fully considering ethical dimensions or allowing decision-makers to evade accountability. Conversely, there is industry concern that ethics considerations could hinder development. To mitigate decision-making conflicts and other tensions among stakeholders, a structured process for resolving disagreements between developers and ethicists should be established.

Similar to other members of the development team, embedded ethicists must adhere to confidentiality and intellectual property rules. This requirement can create tension, given that embedded ethics promotes transparency in decision-making and ethical evaluation. Nonetheless, an embedded ethics framework should explore how transparency in reporting the analytic processes behind AI healthcare technology development can be achieved within these constraints. Ideally, such reporting would document the primary ethical issues identified and addressed, the theoretical frameworks applied, the reasoning behind the chosen course of action, any relevant disagreements, and unresolved matters. This documentation would not only advance the practice of embedded ethics but also contribute to the broader technological field by enabling learning, debate, and the creation of additional embedded ethics literature and case studies.

Expertise and training

Embedded ethicists involved in medical AI development should have solid expertise in ethical analysis. This may include not only graduates from dedicated ethics programs but also researchers trained in social sciences and humanities disciplines that examine ethical issues in science, technology, and medicine, such as medical ethics, science and technology studies, sociology, anthropology, or philosophy of science. In addition to ethical expertise, embedded ethicists need domain-specific knowledge of the particular technology they are engaging with. Similar to clinical ethicists, who must understand the nuances of clinical contexts, embedded ethicists require relevant technical knowledge and skills to effectively contribute to an AI project. While the specific requirements will vary by project, in the context of AI in healthcare, this typically involves a foundational understanding of machine learning principles, robotic design, and the clinical setting in which the technology will be applied. Early adopters of embedded ethics likely already possess such foundational knowledge from prior involvement in the field. Nevertheless, opportunities should be made available both prior to and during projects to allow participants to gain relevant skills and knowledge—for example, through dual-training programs or internships—particularly for junior researchers. In the longer term, developing training modules that build interdisciplinary foundational knowledge will be important; these could be incorporated into emerging interdisciplinary courses and curricula at universities and research institutions.

The merits of embedding ethics in medical AI

Although embedded ethics as an interdisciplinary, collaborative approach to AI is still emerging in both academic and industry settings, it holds strong potential to ensure that robust ethical considerations are incorporated into medical AI development. This section highlights three principal advantages of embedding ethics in medical AI: addressing the unique nature of medical AI, mitigating limitations of traditional medical ethics principles, and addressing current regulatory gaps.

Addressing the uniqueness of medical AI

AI applications in medicine present distinctive challenges and potential risks. Unlike traditional medical tools, AI systems—particularly those based on machine learning and artificial neural networks—operate using hidden or complex data correlations that may not be fully understood, while increasingly performing roles previously handled by human experts [53, 54]. Consequently, healthcare professionals utilizing AI, as well as the developers, may struggle to anticipate system behaviors or explain specific outputs, such as diagnostic results or treatment recommendations. This can also raise questions about shared clinical responsibility with a non-human "colleague." The introduction of such technologies is fundamentally reshaping medical practice, including critical relationships between patients and practitioners and between practitioners and the broader scientific and technical community [17, 36]. Embedded ethics provides a framework for continually adapting to these evolving technological landscapes, helping ensure that AI is developed and implemented in ways that respect the range of social and ethical values involved.

While embedded ethics is not a comprehensive solution to all these complex challenges, it fosters ongoing dialogue between ethicists and AI developers, promoting awareness of the ethical, social, and political dimensions of medical AI. Ethicists with experience in medicine and technology can collaborate closely with developers to facilitate communication with practitioners, patients, and other stakeholders, ensuring that values such as transparency, patient autonomy, and equitable resource allocation are considered. By working together, embedded ethicists and AI developers can better understand the ethical culture present in clinical environments and tailor AI solutions accordingly, for instance, through value-sensitive design [55]. This approach represents a meaningful advance toward creating medical AI that is responsive to ethical considerations.

Mitigating limitations of traditional ethical principles

Currently, one common approach in AI ethics relies on the adaptation of classic medical ethics principles—such as autonomy, non-maleficence, beneficence, and justice [56]. While these principles are undeniably valuable, their direct application to AI development may not be the most effective means of guiding ethical practice. As Mittelstadt notes, AI development generally lacks "(1) common aims and fiduciary duties, (2) professional history and norms, (3) proven methods to translate principles into practice, and (4) robust legal and professional accountability mechanisms" [9]. Consequently, the straightforward use of traditional medical ethics may be insufficient, and companies advocating such principles may do so more to preempt regulation than to genuinely embed ethics. Moreover, even well-intentioned AI developers often lack the necessary resources, experience, and training to address ethical challenges comprehensively.

Embedded ethics offers a promising way to navigate these limitations. While Mittelstadt highlights the structural differences between medical practice and AI development, these issues can be partially mitigated when attention is narrowed to AI applications in medicine and when the values of medical communities are integrated early in the development process. In this context, AI teams collaborating with embedded ethicists can establish clearer shared goals, recognize relevant histories, and align with commonly held clinical values. This approach allows traditional ethical principles to serve as flexible guides or starting points, adaptable throughout the stages of a project. Nonetheless, consistent with critiques of principlism, it should not be assumed that these principles can be applied rigidly in AI contexts. As AI presents novel ethical dilemmas, it requires innovative resolutions and precise norm-setting [57]. Embedded ethicists play a key role in equipping developers with the skills and frameworks needed to identify, interpret, and address these ethical considerations, enabling principles and ethical reasoning to be more effectively integrated into the development of medical AI.

Addressing regulatory gaps

Another critical advantage of embedded ethics is its ability to compensate for the current lack of specific regulation for medical AI. Unlike other medical products, AI technologies are often not subject to rigorous testing during development, and existing ethics guidelines frequently intervene too late to influence design meaningfully [58]. Typically, assessment occurs only during clinical trials or through review by ethics committees, by which point significant development decisions have already been made. While progress has been made in establishing transparency, explainability standards [54, 59], and stronger data protection mechanisms such as the EU's General Data Protection Regulation [60, 61], comprehensive regulatory frameworks for medical AI are still emerging.

Embedded ethics can help bridge these regulatory gaps and complement emerging policies, including the EU's Artificial Intelligence Act [62]. Universities are increasingly developing interdisciplinary programs and research initiatives to prepare ethicists for these challenges, while forward-looking corporations are demonstrating a growing commitment to ethical foresight [12, 13]. This combination creates opportunities to train embedded ethicists effectively and position them at the core of AI development teams, where ethical guidance is most needed. By embedding ethicists directly within development processes, this approach provides a proactive and practical response to the pressing demand for ethical oversight in the design and deployment of medical AI technologies.

Conclusion

Medical AI holds substantial promise for enhancing healthcare delivery, reducing costs, and expanding access to underserved populations. However, the field is broad and rapidly evolving; some innovations never reach clinical practice due to regulatory hurdles, while others transition from development to patient care quickly, sometimes carrying unresolved or unforeseen ethical and social implications. Despite numerous recommendations from ethicists, researchers, and technology developers, there remains no unified framework to systematically address these issues and fully harness the potential of medical AI.

To realize the benefits of medical AI, ethical considerations must be systematically anticipated and addressed early—well before products enter clinical trials or routine practice. The embedded ethics approach, while most readily applied in academic settings or public–private partnerships, is versatile and can be integrated across various environments, including commercial development. Embedding ethicists within development teams facilitates the creation of equitable and safe AI applications, ensuring that ethical, social, and political dimensions are considered from the outset.

Embedded ethics represents one critical component in a broader strategy for addressing emerging challenges in medical AI. It can complement methods such as ethical foresight analysis, educational initiatives to train AI developers and engineers, and formal regulations or voluntary industry commitments, including AI ethics pledges [63]. The strength of embedded ethics lies in its capacity to create a systematic, iterative, and integrated ethical framework within the development process. It is flexible, applicable across diverse settings, and can be implemented immediately, adapting to the specific needs of development teams, technologies, or processes.

Despite its promise, several challenges remain. First, AI development—even in publicly funded research—is highly competitive and prioritizes speed and efficiency; in commercial contexts, profit motives may further overshadow ethical considerations. Conflicts between ethical guidance and business interests, such as tensions between transparency and intellectual property, are likely. As Metcalf *et al*. note, embedding ethics within corporate environments risks being co-opted by agendas emphasizing technological solutionism, market efficiency, or meritocracy [12]. Power imbalances, as illustrated by high-profile cases like Timnit Gebru's departure from Google, underscore the need for enforcement mechanisms—whether regulatory, certification-based, or voluntary—to prevent ethics from being reduced to mere "ethics washing" [50]. Lessons from other industries, such as ethics certifications or compliance programs, could inform these safeguards. Ethics must be treated not as self-regulation, but as an integral part of healthcare technology development [64]. Second, funding embedded ethics roles remains an open question. Pilot programs could initially rely on public funding within academia, while commercial adoption may encounter initial resistance. Nonetheless, the fact that major tech companies are already hiring ethicists suggests that organizational buy-in may be achievable [12]. Moreover, given the financial and reputational risks associated with poorly developed AI technologies, companies have a strong incentive to invest in embedding ethics to prevent harm and ensure responsible innovation. Third, robust training for both ethicists and AI developers is essential. Interdisciplinary education, including programs for professionals already in the field, will support the collaborative exchange required for embedded ethics to succeed. Training in multicultural and interdisciplinary contexts can also increase awareness of potential biases, fostering open discussion of diverse perspectives and reducing the impact of individual preconceptions on technology development.

Finally, establishing clear standards and methodological practices is crucial for embedded ethics to mature as a recognized community of practice. Referenceable standards, case studies, and theoretical frameworks will benefit all stakeholders, including educators designing training programs, professionals currently working in medical AI, ethicists transitioning into the field, and researchers or public stakeholders concerned with the ethical, social, and political dimensions of AI in healthcare.

Acknowledgments: None

Conflict of Interest: Alena Buyx is Chair of the German Ethics Council. Sami Haddadin is member of the European High-Level Group on Artificial Intelligence, and is a founder of robotics company Franka Emika. The other authors have no competing interests to declare.

Financial Support: Open Access funding enabled and organized by Projekt DEAL. The support by the Bavarian Institute for Digital Transformation is gratefully acknowledged. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Ethics Statement: None

References

1. Craglia M, Annoni A, Benczur P, Bertoldi P, Delipetrev B, De Prato G, Feijóo C, Fernandez Macias E, Gomez Gutierrez E, Iglesias Portela M, Junklewitz H, Lopez Cobo M, Martens B, Sanchez Martin JI, Tolan S, Tuomi I, Vesnic-Alujevic L. Artificial Intelligence—a European perspective. Luxembourg: European Commission Publications Office; 2018.
2. Crawford K, Dobbe R, Dryer T, Fried G, Green B, Kaziunas E, Kak A, Mathur V, McElroy E, Nill Sánchez A, Raji ID, Rankin J, Richardson R, Schultz J, West SM. AI Now 2019 report. New York: AI Now Institute; 2019.
3. Coeckelbergh M. Ethics of artificial intelligence: some ethical issues and regulatory challenges. Technol Regul. 2019;2019:31–4.
4. Bostrom N, Yudkowsky E. The ethics of artificial intelligence. In: Frankish K, Ramsey WM, editors. The Cambridge handbook of artificial intelligence. Cambridge: Cambridge University Press; 2014. p. 316–34.
5. Whittlestone J, Nyrup R, Alexandrova A, Dihal K, Cave S. Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. Oxford: Nuffield Foundation; 2019.
6. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1:389–99.
7. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018;28:689–707.
8. Hagendorff T. The ethics of AI ethics: an evaluation of guidelines. Minds Mach. 2020;30:99–120.
9. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1:501–7.
10. Ochigame R. The invention of “Ethical AI”: how big tech manipulates academia to avoid regulation. The Intercept. 2019. Available from: https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/
11. Dignam A. Artificial intelligence, tech corporate governance and the public interest regulatory response. Camb J Reg Econ Soc. 2020;13(1):37–54.
12. Metcalf J, Moss E, Boyd D. Owning ethics: corporate logics, Silicon Valley, and the institutionalization of ethics. Soc Res Int Q. 2019;82:449–76.
13. Floridi L, Strait A. Ethical foresight analysis: what it is and why it is needed? Minds Mach. 2020;30:77–97.

14. Grosz B, Grant D, Vredenburgh K, Behrends J, Hu L, Simmons A, Waldo J. Embedded EthiCS: integrating ethics broadly across computer science education. Commun ACM. 2018;62(8):54–61.
15. Fiesler C. Tech ethics curricula: a collection of syllabi. Medium. 2018. Available from: https://medium.com/@cfiesler/tech-ethics-curricula-a-collection-of-syllabi-3eedfb76be18
16. OECD. Recommendation of the Council on Artificial Intelligence. 2019. Available from: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
17. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. New York: Basic Books; 2019.
18. Celi LA, Fine B, Stone DJ. An awakening in medicine: the partnership of humanity and intelligent machines. Lancet Digit Health. 2019;1(6):e255–7.
19. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271–97.
20. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320(21):2199–200.
21. Schork NJ. Artificial intelligence and personalized medicine. Cancer Treat Res. 2019;178:265–83.
22. Woo M. An AI boost for clinical trials. Nature. 2019;573(7774):S100–2.
23. Fleming N. How artificial intelligence is changing drug discovery. Nature. 2018;557(7707):S55–7.
24. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94–8.
25. Shibata T, Wada K. Robot therapy: a new approach for mental healthcare of the elderly—a mini-review. Gerontology. 2011;57(4):378–86.
26. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res. 2019;21(5):e13216.
27. Johannes MS, Bigelow JD, Burck JM, Harshbarger SD, Kozlowski MV, Van Doren T. An overview of the developmental process for the modular prosthetic limb. Johns Hopkins APL Tech Dig. 2011;30(3):207–16.
28. Jahn E, Reindl A, Müller M, Haddadin S. Roboterassistenten als Helfer der Senioren im Alltag der Zukunft? Altenheim. 2019 (accepted).
29. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res. 2018;20(6):e10410.
30. Davenport TH, Glover WJ. Artificial intelligence and the augmentation of health care decision-making. NEJM Catal. 2018;4(2):16.
31. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216–9.
32. Rigby MJ. Ethical dimensions of using artificial intelligence in health care. AMA J Ethics. 2019;21(2):E121–4.
33. Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical design of intelligent assistive technologies for dementia: a descriptive review. Sci Eng Ethics. 2018;24(3):1035–55.
34. Diehl JJ, Schmitt LM, Villano M, Crowell CR. The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord. 2012;6(1):249–62.
35. Szabo L. Artificial intelligence is rushing into patient care—and could raise risks. Scientific American. 2019. Available from: https://www.scientificamerican.com/article/artificial-intelligence-is-rushing-into-patient-care-and-could-raise-risks/
36. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.
37. McLennan S, Fiske A, Celi LA, Müller R, Harder J, Ritt K, Buyx A. An embedded ethics approach for AI development. Nat Mach Intell. 2020;2(9):488–90.
38. Bezuidenhout L, Ratti E. What does it mean to embed ethics in data science? An integrative approach based on microethics and virtues. AI Soc. 2021;36(4):939–53.

39. Bonnemains V, Saurel C, Tessier C. Embedded ethics: some technical and ethical challenges. Ethics Inf Technol. 2018;20(1):41–58.
40. Ives J, Dunn M, Molewijk B, Schildmann J, Breuer E. Standards of practice in empirical bioethics research: towards a consensus. BMC Med Ethics. 2018;19:68.
41. Jonsen AR, Siegler M, Winslade WJ. Clinical ethics: a practical approach to ethical decisions in clinical medicine. 8th ed. New York: McGraw Hill; 2015.
42. Hilgartner S, Prainsack B, Hurlbut JB. Ethics as governance in genomics and beyond. In: Felt U, Fouché R, Miller CA, Smith-Doerr L, editors. The handbook of science and technology studies. Cambridge: MIT Press; 2017. p. 823–52.
43. Hellström T. Systemic innovation and risk: technology assessment and the challenge of responsible innovation. Technol Soc. 2003;25(3):369–84.
44. Guston DH. Responsible knowledge based innovation. Society. 2006;43(4):19–21.
45. Owen R, Bessant JR, Heintz M, editors. Responsible innovation: managing the responsible emergence of science and innovation in society. Chichester: Wiley; 2013.
46. Van den Hoven J, Doorn N, Swierstra T, Koops BJ, Romijn H, editors. Responsible innovation. Dordrecht: Springer; 2014.
47. Dignum V. Responsible artificial intelligence: how to develop and use AI in a responsible way. Cham: Springer Nature; 2019.
48. Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. Res Policy. 2013;42(9):1568–80.
49. Von Schomberg R. A vision of responsible innovation. In: Owen R, Bessant JR, Heintz M, editors. Responsible innovation: managing the responsible emergence of science and innovation in society. Chichester: Wiley; 2013. p. 51–74.
50. Bietti E. From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. New York: ACM; 2020. p. 210–9.
51. Floridi L. Translating principles into practices of digital ethics: five risks of being unethical. Philos Technol. 2019;32(2):185–93.
52. Begley S. In a lab pushing the boundaries of biology, an embedded ethicist keeps scientists in check. Stat News. 2017. Available from: https://www.statnews.com/2017/02/23/bioethics-harvard-george-church/
53. Matthias A. The responsibility gap: ascribing responsibility for the actions of learning automata. Ethics Inf Technol. 2004;6(3):175–83.
54. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-López S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82–115.
55. McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. J Med Ethics. 2019;45(3):156–60.
56. Beauchamp TL, Childress JF. Principles of biomedical ethics. 8th ed. New York: Oxford University Press; 2019.
57. Richardson HS. Specifying, balancing, and interpreting bioethical principles. J Med Philos. 2000;25(3):285–307.
58. Ferretti A, Ienca M, Hurst S, Vayena E. Big data, biomedical research, and ethics review: new challenges for IRBs. Ethics Hum Res. 2020;42(5):17–28.
59. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60.
60. Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv JL Tech. 2018;31:841–87.
61. Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Bus Law Rev. 2019;2019(2):494–620.
62. European Commission. Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. 2021. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELLAR:e0649735-a372-11eb-9585-01aa75ed71a1
63. Sample I. Maths and tech specialists need Hippocratic oath, says academic. The Guardian. 2019. Available from:

https://www.theguardian.com/science/2019/aug/16/mathematicians-need-doctor-style-hippocratic-oath-says-academic-hannah-fry
64. Wagner B. Ethics as an escape from regulation: from ethics-washing to ethics-shopping? In: Bayamlıoğlu E, Baraliuc I, Janssens LAW, Hildebrandt M, editors. 10 years of profiling the European Citizen. Brussels: Vrije Universiteit Brussel; 2018. p. 108–15.