



# Situating governance and regulatory concerns for generative artificial intelligence and large language models in medical education



Michael Tran<sup>1</sup>✉, Chinthaka Balasooriya<sup>1</sup>, Jitendra Jonnagaddala<sup>1</sup>, Gilberto Ka-Kit Leung<sup>2</sup>,  
Neeraj Mahboobani<sup>3</sup>, Subha Ramani<sup>4</sup>, Joel Rhee<sup>1</sup>, Lambert Schuwirth<sup>5</sup>,  
Neysan Sedaghat Najafzadeh-Tabrizi<sup>1</sup>, Carolyn Semmler<sup>6</sup> & Zoie SY Wong<sup>1,2,7</sup>

Generative artificial intelligence (GenAI) and large language models represent gains in educational efficiency and personalisation of learning. These are balanced against the considerations of the learning process, authentic assessment, and academic integrity. A pedagogical approach helps situate these concerns, and informs various types of governance and regulatory approaches. In this review we identify current and emerging issues regarding GenAI in medical education including pedagogical considerations, emerging roles, and trustworthiness. Potential measures to address specific regulatory concerns are explored.

Generative artificial intelligence (GenAI), refers to deep-learning models that can generate high-quality content based on the data on which it was trained<sup>1</sup>. Large language models are a category of foundation models trained on immense amounts of data. Although not inherently able to understand text and data, these models are able to generate natural, human-like language<sup>2</sup> which is perceived as “conversational” by users. GenAI differs from extractive AI technologies that excel in accessing, collating, prioritizing, adapting, and using information under narrow circumstances<sup>3</sup>.

GenAI including large language models (LLMs), provides learners and educators in medicine and health sciences with previously-unimaginable opportunities for teaching and learning. The scope of potential applications, and the efficiency with which this could be completed, is rapidly increasing. The utility of its capacity to utilize multi-modal approaches has been demonstrated in cardiac electrophysiology education<sup>4</sup> and digital pathology<sup>5,6</sup>.

AI in general is a complex social, cultural, and material artifact whose meaning and place continue to be constructed by different stakeholders. There remains a paucity of information regarding the development, deployment and commercialization<sup>7</sup> of these models and their applications and services based upon them<sup>8</sup>. Unsurprisingly, there has been growing consternation among educators, professional bodies, and governments regarding the potential need for regulation, in its various forms, and control of the influence of this technology. It often feels like attempts to provide regulatory frameworks and

legislation are reactionary and ineffectual, in the face of rapid global progress unbounded by any specific institutional or sovereign authority. Several guidelines, ethical considerations<sup>9</sup>, statement papers, and recommendations have been published in recent years<sup>10</sup>, including primers for AI<sup>3</sup>, recommendations for workforce implications<sup>11</sup>, and considerations, especially ethical<sup>12</sup>, regarding the integration of AI in medical curricula<sup>13</sup>.

Enthusiasm for, and trepidation of the future role of GenAI in medical education must to be considered in the context of our evolving understanding of pedagogical principles and best practices. The impact of GenAI cannot be ignored, as it risks multi-level harms ranging from a lack of structures to ensuring scholarly integrity, stagnation, and irrelevance of learning approaches<sup>14</sup>. There is likely a need for sustainable and adaptable responses to GenAI in learning, teaching and assessment<sup>15</sup>. This review seeks to situate our current understanding of the impact of GenAI in undergraduate medical education, within a pedagogical framework, to inform regulatory concerns. A narrow focus recognizes the differences between undergraduate, postgraduate, and continuing professional education requirements. We outline the concerns of GenAI and LLMs in medical education, pedagogical considerations, and emerging roles, and present a discussion regarding the regulation and preservation of academic integrity. A summary of the key considerations and concerns regarding GenAI and LLMs is provided in Fig. 1.

<sup>1</sup>University of New South Wales, Kensington, NSW, Australia. <sup>2</sup>The University of Hong Kong, Hong Kong, PR China. <sup>3</sup>Department of Imaging and Interventional Radiology, Faculty of Medicine, The Chinese University of Hong Kong (CUHK), Hong Kong, PR China. <sup>4</sup>Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. <sup>5</sup>Flinders University, Adelaide, SA, Australia. <sup>6</sup>The University of Adelaide, Adelaide, SA, Australia. <sup>7</sup>St Luke's International University, Chuo, Japan. ✉e-mail: [Michael.m.tran@unsw.edu.au](mailto:Michael.m.tran@unsw.edu.au)

## Pedagogical considerations for generative AI in medical education

### GenAI and learning

The learning process is perhaps the biggest consideration when situating GenAI within a pedagogical approach. Educability, that is, the ability of learners to utilize any and all previously-learned information in meaningful ways, distinguishes their learning capacity from machines<sup>16</sup>. GenAI is likely to be a useful adjunctive tool in medical education but is unlikely to replace all the experiences and social interactions that are important for the development of empathetic and contextually aware learners in constructivist and experiential frameworks. It is important to include information on GenAI in education for all students and educators. Without the opportunity to learn about the ethical use of GenAI, learners are more susceptible to engaging in inappropriate use of GenAI<sup>17</sup>. A recent scoping review identified the need for further research in three key areas to improve our understanding of the role of GenAI in medical education. These include (i) developing learners' skills to evaluate GenAI critically, (ii) rethinking assessment methodology, and (iii) studying human–AI interactions<sup>18</sup>.

What learning processes, in the medical student “journey,” are likely to be impacted (adversely or positively) by LLMs and GenAI? Students learn through a combination of different means, and the theoretical approaches to classify these have revolved around cognitive psychology, humanistic psychology, and social anthropology. Social constructivism, as an epistemological framework, outlines learning through the construction of knowledge and interactions with others by linking new information to that previously learned and incorporating new experiences into a knowledge base. It is not simply the transmission of knowledge from the external world to the

learner<sup>19</sup>. A complementary learning theory, experiential learning, defines learning as a process whereby knowledge is created through the transformation of experience. Different learners pass through phases of reflective observation, abstract conceptualization, and active experimentation in their own preferred order<sup>20</sup>. Therefore, the context in which learning is experienced and knowledge is acquired is critical. Whilst offering “efficiency,” how GenAI is situated with respect to the “context” of learning and the transition from “novice” to “expert” is not yet fully understood. Cognitive psychology offers many theories and explanations of how expertise develops, and is fostered via learning<sup>21</sup>. As our understanding and experience of GenAI grows, it may well be that we find it fits within existing frameworks or demands a novel approach to understanding its role in this transition. The extant literature identifies instances where GenAI can accelerate learning in novices<sup>22</sup> but may also accelerate skill decay and hinder skill acquisition<sup>23</sup>. A theoretical counterargument posits that GenAI itself could play the role of “the more knowledgeable other” in the social constructivist framework<sup>24</sup>. Operating in a metaphorical contextual vacuum, GenAI is unlikely to bypass the process of exposure and experience in learning. Its inability to teach the integration of contextual and external information, comprehend sensory and nonverbal cues, cultivate rapport and interpersonal interaction, and align with overarching medical education and patient care goals<sup>25</sup> remains a key limitation for the current generation of LLMs.

### Clinical reasoning and GenAI

The key difference between extractive and GenAI is that the latter leverages machine learning, such as neural networks, to generate new content. This method is based on the relationships and patterns found in existing datasets,



**Fig. 1** | Summary of the key concerns and considerations for GenAI and LLMs from medical education, pedagogy, and regulatory perspectives. The overlap in concerns regarding GenAI and LLMs between each perspective is demonstrated in the Venn

diagram. Unknown biases and trustworthiness are concerns shared across all three perspectives.

which can be broad and varied. GenAI can autonomously and rapidly produce large volumes of content and has the capacity to be “imaginative” and “disruptive” in its innovation. Via a “black box” mechanism involving multiple layers of neural networks (which remains opaque even to developers and difficult for computer scientists to explain), GenAI can create substantially larger output than the input provided<sup>10,26</sup>. The concerns are slightly different for extractive AI compared to GenAI. The former, with more utility for diagnostic processes, has a stronger expectation of being correct, while the latter must be plausible and useful. The strengths and weaknesses of GenAI and extractive and algorithmic forms of AI are distinct.

This “black box” component of GenAI may be problematic if trying to teach clinical reasoning and decision-making skills in medical education, which means necessarily involving opaque, partial and ambiguous situations<sup>27</sup>. When prompted, GenAI might offer a plausible explanation of its decision-making process, but this explanation is not necessarily an accurate, or comprehensible representation of how GenAI *actually* made its decisions. It could also be argued that the ways in which humans recognize and solve problems and engage in clinical reasoning are unclear. This opacity is problematic when considering that the formative process of learning clinical decision-making, including understanding when and why errors occur, has implications for legal liability and accountability in medical practice<sup>14</sup>. LLMs can impact critical evaluation and analytical thinking and, when used inappropriately, could negatively impact students’ ability to discriminate valuable information from inaccurate and irrelevant inputs<sup>28</sup>. Just as the Internet has led to the externalization of factual knowledge, there are concerns that LLMs could externalize medical reasoning. The response to the readily-available nature of a vast amount of information was to place greater emphasis on debate and discussion and knowledge “management,” rather than memorization<sup>28</sup>. With an improvement in the capacity to assist with clinical reasoning<sup>28</sup>, LLMs are likely to promote further changes to educational methods and the need to reconceptualize assessment.

In being unable to account for patient context in its formulation, there is a risk that GenAI may reduce complex patient experiences to linear problem-solving interventions, promising “solutionism,” and risking objectifying patients, based on potentially-biased learning of patient populations that are “most studied” or “most prevalent in the literature. Recalculating patient illness experiences into solution-based computational terms risks ignoring the benefits of dialog and the complex and often unpredictable patient experiences<sup>29</sup>. Algorithmic and extractive AI technologies may excel at diagnostic components of consultations that are akin to data reduction and categorization tasks. By contrast, developing a treatment plan is context-dependent, imbued with uncertainty and more nuanced, a scenario more suited to GenAI capabilities. Uncertainty in its many guises cannot be avoided in medical education and clinical practice. In synthesizing large bodies of knowledge, GenAI may obfuscate or overstate uncertainty, and learners will need to develop skills to understand not only their own but also technological reactions<sup>30</sup>. Although GenAI and LLMs, as adaptive educational systems, may improve the efficiency and interactivity of the learning experience, their unknown impact on learners’ attention and other cognitive and metacognitive abilities need to be considered<sup>31</sup>.

Here we highlight the need to acknowledge both the strengths and limitations of GenAI in medical education. Educators and learners are well-advised to consider the implications of the output of GenAI will be unscaffolded and not peremptorily verified. This presents a unique facet to the constructivist and experiential approaches that underpin much medical education pedagogy.

### Assessment and GenAI

Despite the benefits of LLMs being able to synthesize and personalize information for learners<sup>3</sup>, there is much consternation with respect to the use of LLMs to subvert current assessment processes<sup>32</sup>. Using GenAI in this manner (when explicitly disallowed in the task description) is an academic dishonesty. A more vexing concern is whether learners will become overly or completely reliant on these technologies and what can be gained or

sacrificed using GenAI as an educator. There is a potential risk of denying learners the formative experiences and important skills such as critical thinking necessary in the journey from novice to “expert.”

GenAI has progressed to the point where LLMs can pass licensure examinations in many undergraduate<sup>33</sup> and postgraduate specialty training programs<sup>34,35</sup>. However, ongoing challenges to medical education from LLMs include ensuring the accuracy and contemporaneity of information, reducing bias<sup>36</sup>, ensuring accountability<sup>28</sup>, minimizing learner over-reliance, preventing patient privacy exposure, safeguarding data security, enhancing the cultivation of empathy, and maintaining academic integrity<sup>37</sup>. With existing and potential changes for learners and educators, it remains important to consider what place GenAI has within broader educational aims and pedagogy in the context of its potential, limitations, and boundaries. We need to consider what makes educators and learners unique and whether GenAI can support or supplant this in working towards the goal of creating competent and empathetic doctors. Empirical studies evaluating the use of GenAI and LLMs in medical education and their efficacy in developing competencies in health professional training are scarce. These studies have focused on the use of GenAI for learning support<sup>38</sup> or automated assessments of clinical skills, but there has been limited use of theory or conceptual frameworks<sup>39</sup>.

Curriculum and assessment redesign encompassing future-focused competencies recognizes that new skills will be required for novel models of care. Learners should be proficient in understanding the origins and development of technologies that they will be using in their clinical work, in research, and in continuing learning and professional development. New areas of technical competence will be essential for learners to work in AI-integrated healthcare environments to deliver patient care, communicate with other health professionals, and effectively manage large amounts of population-wide data that will become increasingly available<sup>40</sup>. It is impossible to address the potential impacts of GenAI use in medical education without acknowledging the intersection with clinician training and clinical care.

## Emerging roles of GenAI in medical education

### For learners

Artificial intelligence is likely to impact medical education methods by producing intelligent and personalized systems to identify and respond to gaps in students’ knowledge, adaptable virtual facilitators in constructivist learning approaches, mining data, and providing intelligent feedback to learners<sup>3,41</sup>. GenAI not only delivers content but also enables adaptive learning, provides information and feedback, creates individualized learning pathways, supports competency-based assessment, and potentially provides and manages programmatic assessment data<sup>42</sup>. For learners, this individualized learning can be customizable in depth, tone, and style of output, making it an ideal personalized teaching assistant<sup>28</sup>.

Students can benefit from improved practical skills<sup>43</sup>, robust selection processes and research assistance<sup>44</sup>. Recent research on the medical student perspective suggests that GenAI is good at facilitating differential diagnosis brainstorming, providing interactive practice cases, and aiding in multiple-choice question review<sup>25</sup>. LLMs can be used to create interactive and engaging simulations. For example, students may use LLMs to have conversations with simulated patients, allowing them to practice taking patient histories or assessing diagnoses and discussing treatment plans<sup>28</sup>.

### For educators

From an educator’s perspective, LLMs may help shape medical curriculum development and engender changes in teaching methodologies<sup>41</sup>. It has been demonstrated that GenAI with expert human guidance can also produce assessment items for medical examinations<sup>45</sup>. Human-developed questions still retain a higher discriminatory power<sup>46</sup>. This is potentially due to human assessors being more adept at generating items with higher construct validity and also being more closely aligned with a priori knowledge such as lecture material. GenAI may help reduce the administrative burden on educators<sup>42</sup>, with help in assessment and attendance tracking<sup>3</sup>. Analogous to

**Table 1 | Key opportunities, challenges, and pedagogical considerations of GenAI and large language models**

Learners	
Positive Impacts	Challenges
<ul style="list-style-type: none"> <li>• Personalized learning systems that provide intelligence feedback and virtual facilitation in constructivist learning approaches</li> <li>• Creation of interactive and engaging learning simulations and interactive practice cases</li> <li>• Brainstorming of differential diagnoses</li> <li>• Application of AI in MCQ review and exam preparation</li> </ul>	<ul style="list-style-type: none"> <li>• Need to ensure accuracy and contemporaneousness of information, reduce bias, ensure accountability, minimize learner over-reliance, retain empathy</li> <li>• The “black box” output mechanics of neural networks is difficult to explain and so the basis of any decision-making process is opaque and will impact critical reasoning, thinking and decision-making skills</li> </ul>
Educators	
Positive Impacts	Challenges
<ul style="list-style-type: none"> <li>• Reduced administrative burden with help in assessment and attendance tracking</li> <li>• Ability to manage large volumes of programmatic assessment data</li> <li>• Use of GenAI to produce assessment items for examinations</li> </ul>	<ul style="list-style-type: none"> <li>• GenAI can pass licensure examinations and could potentially subvert assessment processes and contribute to academic dishonesty</li> <li>• In using GenAI, there is a need to safeguard data security and patient privacy, and help maintain academic integrity</li> <li>• Teaching and assessment will need to change to keep assessment original and authentic</li> <li>• Future assessment may need to be developed to evaluate whether students can safely and knowledgeably use GenAI</li> </ul>
Pedagogical Considerations	
<ul style="list-style-type: none"> <li>• It is unclear if GenAI can provide the social context and experiences often thought to be required to acquire knowledge and develop understanding</li> <li>• Efficiency could contribute to expertise by providing augmented capacity to access information.</li> <li>• GenAI at this stage cannot teach the integration of contextual and external information, comprehend sensory and nonverbal cues, cultivate rapport and interpersonal interaction, and align with overarching medical education and patient care goals</li> <li>• Algorithmic approaches in AI risks objectifying patients and experiences to simple problem-solving approaches, when this is rarely straightforward in healthcare</li> <li>• It is unclear if GenAI can accurately describe its approach to uncertainty</li> <li>• Curriculum redesign is required to capture future-focussed competencies including working with GenAI</li> <li>• Formal instruction regarding GenAI in medical education is likely to be necessary to consider the operation, features, limitations, and ethical issues with emerging technology</li> <li>• The role of educators in learning is likely to change and may involve reinforcing uniquely human skills</li> </ul>	

“precision medicine,” educators can foster “precision education” by leveraging data to individualize training and assessment. Data can inform the strategic deployment of educational resources and strengthen the link between practice and education, and educators can advocate the development of appropriate tools<sup>42</sup>.

Traditional assessment methodologies are increasingly at risk of obsolescence<sup>12</sup>, necessitating a paradigm shift towards assessment modalities that are more resistant to unapproved GenAI assistance, such as continuous in-person assessment of practical or clinical skills, and oral examinations<sup>47</sup>. A contrasting and perhaps more realistic view accepts that students are more likely to use GenAI and will be working in healthcare environments that have been transformed by the integration of GenAI. Assessment may need to be better at evaluating whether students can use GenAI with a complete understanding of its strengths and limitations, demonstrating its effective and safe use in their own learning and in patient care. The reliance on traditional written tasks, which are susceptible to completion by GenAI without genuine student engagement or learning, underscores the urgency for educators to redesign assessments<sup>41</sup>. Educators may need to rethink and re-define “authenticity” and “originality” in assessment that incorporates the use of GenAI<sup>48,49</sup>. Competency frameworks need to be reconsidered and updated to consider 21st century realities. The abilities that students and future clinicians require to adequately meet patients’ healthcare needs will be impacted by AI-enabled systems<sup>50</sup>. There remains a need to improve the digital literacy of future physicians while incorporating patients’ views with the increasing use of GenAI technologies<sup>50</sup>. This highlights the need for new assessment strategies which permit authenticity of the learner’s voice, discourage over-reliance on GenAI for completion and which prepare the future workforce for workplaces where they will need to navigate GenAI and technology competently.

### Impacts on the educator–learner relationship

Learners have expressed a lack of confidence in being able to inform others about the features and risks of GenAI applications due to a lack of formal instruction about the use of such programs<sup>51</sup>. Unsurprisingly, there is a demand for structured GenAI training, particularly in terms of reducing medical errors and ethical issues<sup>51</sup>. One apparent deficiency of GenAI and

LLMs identified by medical students was the reduction in the humanistic aspect of medicine. The nature of learning is likely to evolve with the introduction of GenAI in education, as well as the roles of educators, what is demanded of them, and the relationships they have with students. It may well be that there is greater emphasis on reinforcing human skills, communication, empathy, professionalism, and contextualizing and individualizing treatment strategies for patients. The opportunities, challenges, and considerations for GenAI are summarized in Table 1.

### Trustworthiness and the intersection of GenAI use in medical education and clinical practice

Underlying some concerns about GenAI is perhaps the belief that seeking autonomous input from GenAI will necessarily result in nefarious outcomes<sup>52</sup>. In healthcare settings, one factor delaying the translation of GenAI and its potential benefits to patient care and education is whether learners, educators, clinicians, and patients would trust it. The patient’s voice regarding their needs and expectations is not always fully considered in the application of GenAI in healthcare<sup>53</sup>. Studies of patient perceptions about the use of GenAI in healthcare have concluded that most are comfortable with its involvement but would prefer the final plans and management to be approved and delivered by humans. Trusting the decision-making capacity of the clinician is based on the pre-existing trust that patients have with their physicians<sup>54</sup>. The mistrust of GenAI is perhaps secondary to its inability to explain its rationale<sup>55</sup> and decisions, that are not fully transparent<sup>42</sup>. This is complicated by the understanding that GenAI technology has inherent biases that join human biases in shaping the diagnostic process, in a potentially non-neural manner<sup>56</sup>. Where GenAI and LLMs perform with some degree of autonomy, from an ethical perspective, this gives them a moral agency that needs to be accounted for<sup>57</sup>. Blind acceptance of AI decisions is another potential source of mistrust, where GenAI output replaces, rather than augments, human decision making. A greater understanding of the “permissible” ways in which GenAI could augment human processes may help learners, users, and patients to ensure that the use of GenAI remains responsible.

When decisions are subjective or the variables change, human judgment is trusted more because of people’s capacity for empathy. Even when



GenAI systems outperform human doctors, trust in GenAI does not increase<sup>58</sup>. This mistrust is even greater where factors affecting diagnoses may be behavioral, and case-specific, such as mental health<sup>59</sup>. Another driver of consumer resistance to medical AI is the phenomenon of “*uniqueness neglect*.” This indicates that GenAI systems are less able than human providers to account for consumers’ unique characteristics and circumstances, and drive consumer resistance to medical AI<sup>60</sup>. The human ability to combine contextual awareness with knowledge leads to the perception of superiority in planning, managing, and achieving favorable results<sup>59</sup>. To this end, GenAI and LLMs should retain an assistive role in clinical encounters, and medical education needs to adapt to ensure that future doctors are prepared for an GenAI-assisted work environment to preserve doctor-patient relationships<sup>61</sup>. GenAI and LLMs use quantifiable datasets, and there is a risk that patients themselves are reduced to data points, neglecting their experiences and individual context. Patients and healthcare professionals should consider this and encourage patient empowerment by expressing their individual circumstances<sup>62</sup>. Medical training should prepare doctors to operate dynamically in being able to adapt to a range of technologically-enabled, or not, environments. There will be patients who may lack digital literacy with whom the nature of interaction would differ to those who operate in more AI-enabled environments and who may have consulted independently with GenAI technologies to understand their medical conditions. The digitally-literate physician will be able to navigate and address the diverse needs of all patient groups.

A further consequence of using GenAI as an adjunct to a healthcare provider’s work, is that administrative tasks may be made less onerous for clinicians, reducing burnout, and allowing for greater time and connection with the humanistic side of medicine. The counterargument is that there exists a potentially increased burden with higher throughput of patient consultations and increased cognitive load with monitoring and correcting the output of GenAI processes. Democratization of patient care could also be achieved by providing patients with access to their information in a timely and comprehensible fashion<sup>63</sup>. Considering the doctor-patient-AI “triad” relationship represents a paradigm shift and calls for further research to better understand how GenAI influences the doctor-patient relationship and respective autonomies, to ensure that ethical practice remains present<sup>64</sup>. Technological mediation may inhibit the development of trust in a doctor-patient relationship, and this too, would benefit from further research and understanding. As a mediator placed between the doctor and patient, GenAI systems can inhibit tacit understanding of the patient’s health and well-being and encourage both clinicians and patients to discuss health solely in measurable quantities or machine interpretable terms<sup>65</sup>.

Often, models are developed without input from the people who will ultimately use them, namely students, practitioners, and patients. GenAI models also have no intrinsic ability to use context or meaning to inform output and decisions, which is problematic because context critically determines the quality of outcomes for patients<sup>52</sup>. This contextual awareness is likely to improve with newer generations of GenAI, but it will be critical for any underlying bias within the material that the GenAI has “learned” from to be either eliminated or mitigated, as this will inform the technology’s capacity to make inferences about the patient context. A codesigned approach that considers which tasks are more efficient with GenAI elements, which should be learner and student-led, is likely to be more productive<sup>66</sup>.

The U.S. Department of Health and Human Services has identified six principles of trustworthy AI<sup>67</sup>, including LLMs being robust and reliable, fair and impartial, transparent and explainable, responsible and accountable, safe and secure, and ensuring privacy and consent. However, it is unclear what would make GenAI trustworthy in clinical practice and without a clear understanding, the development of effective implementation strategies will be impaired in the healthcare setting<sup>68</sup>. GenAI and LLMs are likely to continue to develop in ways that benefit particular groups (especially commercial), but without a high level of trustworthiness, they are unlikely to be acceptable to all aspects of health professions. Evaluating and ensuring the presence of these underlying “foundations” of the trustworthiness of GenAI

technologies by health professionals, possibly as part of the responsibility for self-regulation, may be required to shape GenAI development in equitable and acceptable ways. These are considerations that those involved in medical education, particularly learners and educators, need to heed and develop personal approaches to. Further research including all stakeholders, especially patients, learners, and educators, into the foundations of trustworthiness and how these features in future AI-enabled workspaces will be critical.

## Is regulation necessary?

Given the speed and unpredictability of innovation, quantum of investment, and lack of technical information, it is almost impossible to forecast the opportunities and risks of GenAI accurately. LLMs raise questions about the opportunities and risks of widespread adoption; scope and adequacy of national strategic planning and policies; fitness of legal and regulatory approaches, and implications of increasing geopolitical competition and geo-specific regulations<sup>8</sup>. Regulation needs to be defined within the context of this review.

- (i) With regard to the medical device functionality of GenAI in clinical work, a legal definition of regulation is appropriate where it represents rules, or directives, designed to control and govern conduct. Oversight would be the domain of government departments responsible for the implementation and use of therapeutic goods and devices.
- (ii) Regarding medical education, regulation refers to accreditation and validation, that is, formal processes to ensure that standards for quality and competency are met. Oversight would be local and context-dependent.

There are several challenges associated with attempts to regulate technology. The perceived risks of harm are tempered by social norms, market pressure, and the coding architecture (design, structure, and organization of the codebase). Adapting formal regulation may be one element for ensuring safe and ethical GenAI use. A stepped approach to GenAI regulation recognizes that a new technology does not necessarily imply the need for new rules. Where there are risks from the use of GenAI that warrant some form of regulation, identifying which component or process requires regulation will be important and the codesign of any framework with all stakeholders will be critical<sup>69</sup>. Existing legal frameworks may address and mitigate some risks of patient-facing GenAI use<sup>69</sup> however specific contexts for GenAI and LLM use will require may require specific regulatory attention.

## Regulation and GenAI use in medical education

### Preserving academic integrity

Detecting the misuse of LLMs for plagiarism where there is no augmentation of learner abilities<sup>44</sup> remains challenging given the lack of transparency, from both GenAI programs and the algorithms used by detection tools. The ability of GenAI to pass high-stakes examinations<sup>70</sup> highlights an issue with reliance on “single-shot” examinations and their inherent difficulties with generalizability<sup>71</sup>, being limited assessments of knowledge. Programmatic assessment<sup>72</sup>, which focuses on a wide variety of assessment tasks, including workplace-based assessments, is potentially more resistant to the unauthorized use of GenAI. GenAI can help organize the wealth of performance evidence that accompanies programmatic assessment, visualizing and interpreting it in a manner that informs future learning and identifying signals in performance evidence that would steer additional diagnostic assessments or learning experiences<sup>42</sup>. There will be an onus placed on educators to rethink how the utility of GenAI can be maximized<sup>73</sup> while mitigating concerns about its potential misuse.

GenAI, students, and clinicians are likely to have an interdependent relationship. Bearman and Ajjawi provided a framework to work with, and not fear, “black boxes”<sup>27</sup>. Orienting students to quality standards and providing meaningful interactions with GenAI systems would (i) permit an understanding of the social regulating boundaries around GenAI (ii) promote learner interactions with GenAI while building evaluative judgment in

weighing GenAI's contribution and (iii) encouraging understanding of the evaluative, ethical, and practical necessities of working with "black boxes"<sup>27</sup>. Just as learners will use GenAI to an increasing degree, it will continue to rely on high-quality input from users<sup>74</sup>, including students and clinicians. Initial training with GenAI is unlikely to be sufficient as a standalone endeavor and additional training is likely to be required as the field evolves. By establishing frameworks for its adoption and education early, this process becomes more feasible in the future.

As with any source material, understanding the nature of veracity and applicability of information to their own learning and eventually patient care needs to be emphasized. Learners should continuously critique and question GenAI-generated outputs in biomedical knowledge and the pathophysiology of the disease<sup>13,75</sup>. This would prevent GenAI-generated information from acting as an automated crutch for clinical decision-making<sup>76</sup>, which could hamper the development of clinical reasoning abilities<sup>77</sup>.

Specific concerns regarding GenAI use in medical education include algorithmic bias, overreliance, plagiarism, misinformation, inequity, privacy, and copyright concerns<sup>9,41</sup>. Many practical guidelines regarding the regulation of GenAI agree on factors that require regulatory oversight including: transparency, bias, content validity, data protection, excessive (and non-consensual) data collection, data ownership, informed consent, ensuring that users remain empowered and establishing accountability<sup>9,78</sup>. Where information is drawn from by GenAI programs is also of importance with respect to intellectual property and copyright protection.

### Regulation and GenAI use in clinical practice

If GenAI and LLMs are used in clinical settings, there is ambiguity regarding the responsibility for medical diagnoses, whether it is a GenAI or a healthcare professional. Calls have been made for users of GenAI to be guided by ethical principles, which practically and legally may involve reforming the categories of medical malpractice, vicarious liability, and product liability, as well as the ancillary duties healthcare providers<sup>62</sup>. Many of these recommendations fall under the rubric of "soft law," presenting self-regulating obligations and codes of conduct that are not legally enforceable but are considered "good practice." With the introduction of GenAI systems, there is potentially an argument for some aspects, such as the duty to warn of limitations and obtain informed consent, to be reallocated to "hard law," becoming legal obligations related to disclosure of information.

Although regulations regarding therapeutic goods and devices focus mostly on patient safety, they do not necessarily guarantee it<sup>79</sup>. Not all AI tools with regulatory authorization are not necessarily clinically validated<sup>80</sup> and if GenAI is implemented poorly, it may add to doctors' burden of responsibility and potentially expose doctors to the risks of poor decision-making. Alternatively, GenAI implemented with a responsible design, informed by cognitive science, would allow doctors to offload many of their cognitive tasks to GenAI when appropriate and focus their attention on patients<sup>52</sup>. Responsible GenAI requires the development of legal frameworks to protect patients and consumers from potential harm arising from poorly developed GenAI and inappropriate deployment in socio-technical systems. Most importantly, patients and consumers have the right to be informed about the limitations of GenAI to allow them to decide which aspects of their lives could benefit from it<sup>52</sup> and the choice to opt-out of systems employing GenAI.

As previously discussed, the use of GenAI during medical training may result in inadequate development of critical thinking and clinical reasoning skills, which may threaten patient safety several years later, as the learner starts to take on greater responsibility for patient care. On the other hand, training of the future generation without adequate recognition of the role of GenAI in their future practice and the new competencies that are required is likely to result in graduates who are underprepared for their clinical roles which may ultimately adopt such technologies. Clinicians are likely going to need to understand and keep pace with patient use of GenAI as well.

GenAI currently operates in a regulatory framework that is patchwork, at best. One call for legislation is based on human-rights, with concerns for emerging harms from GenAI centered on privacy, algorithmic

discrimination, automation bias, misinformation, and disinformation<sup>81</sup>. Legislation does exist to regulate GenAI usage in specific settings or circumstances; however, many gaps still exist.

Regulation, applied with the intent of supporting safe innovation, may also to some degree, incur human and economic opportunity costs in also potentially restricting progress and innovation. This reinforces the overall message that regulation, in whichever form it is present, needs to be with purpose and should assure educators, learners, medical professionals, and patients that LLMs can be used without causing harm or compromising data or privacy<sup>82</sup>.

### Levels of regulation

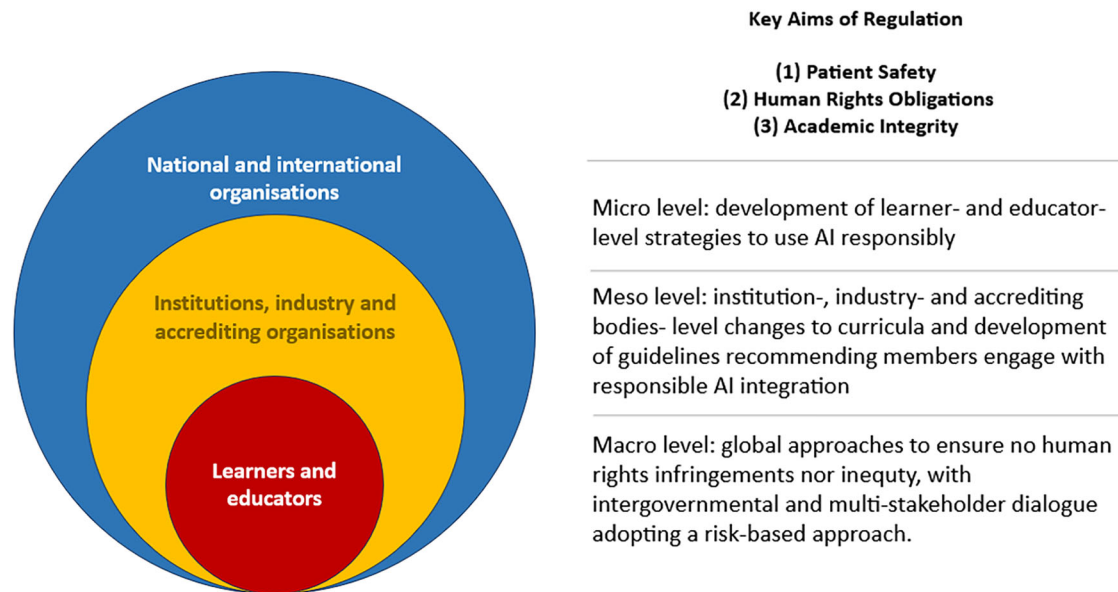
Consideration of the different types of, and levels at which regulation may apply will inform how individuals, institutions, accrediting bodies, national governments, and global organizations manage the establishment of the acceptable use of GenAI in medical education to ensure safe and ethical practice. The ecological framework allows the consideration of regulatory principles at the micro, meso, and macro levels and has been used to synthesize and unify existing learning theories to model the roles of artificial intelligence in promoting learning processes<sup>83</sup>. The ecological framework not only identifies increasingly broader levels of influence but also considers the relationships across different levels. Any regulatory effort is unlikely to succeed without all levels interacting to some degree. At this nascent stage however, the most readily-applicable action is likely to be at the micro level, that of the individual learner and the educator. This framework is summarized in Fig. 2.

#### Micro: individual learners and educators

Regulation at the micro level, including individual learners and educators, would predominantly involve degrees of self-regulation. Regulatory responsibility for the use of GenAI in medical education will likely need to focus on developing robust strategies to counter or address opacity and inexplicability, data privacy and security, fairness and bias, reliability<sup>84</sup>, protection of intellectual property, assurance of quality control and standardization, informed consent, data ownership, over-reliance on GenAI models, and continuous monitoring and validation<sup>82</sup>. Educators and learners should be encouraged to develop personal and morally-informed strategies, akin to a personal code of conduct, to manage these issues and be ready to state how these have been addressed, or not, when using GenAI. Ideally, learners will be empowered to increase their knowledge and skills to use a range of emerging digital health systems, analyze the data emanating from them, and evaluate information for trustworthiness and relevance<sup>47</sup>. This would not only ensure that students are adept at leveraging GenAI in their future careers but also emphasize the importance of critical thinking and maintaining integrity and professional standards in their work<sup>47</sup>, despite the convenience of readily-generated information<sup>14</sup>.

#### Meso: institutions and accrediting bodies

At this level, there is the intersection of regulatory processes governing therapeutic goods and devices as well as educational accreditation. Professional health education curricula will need to evolve to include comprehensive teaching on the ethical and appropriate use of GenAI<sup>9</sup> and critical appraisal of information created with it. Institution-level approaches to GenAI may be retrofitted to existing national guidelines<sup>85</sup>. Similarly, institutional policies may be developed based on guidelines that have been developed. It would be the responsibility of tertiary educational institutions and professional colleges overseeing pre-vocational and postgraduate vocational training to develop frameworks appropriate to their accreditation and validation processes. Individual professional organizations such as the Royal Australian College of General Practitioners, have also developed evolving position statements to guide clinicians<sup>86</sup>. The latter position statement outlines various concerns and issues and makes legally non-binding recommendations but calls on general practitioners to be cognizant of technological advances and their ethical and clinical implications, calling for individual responsibility with GenAI use. These reminders should be



**Fig. 2** | A depiction of regulatory levels for GenAI and LLMs within the ecological framework. The micro-level consists of learners and educators, the meso-level of institutions, industry bodies and accrediting organizations and the macro-level of

national and international organizations. There is intercalation of the regulatory concerns, strategies and frameworks across these different levels.

reinforced at the medical student and learner levels to encourage forward thinking about the ethical challenges that GenAI systems will pose. Enabling this self-reflection would emphasize on faculty development and educator's upskilling to safely and productively engage with GenAI<sup>87</sup>.

The Australian Health Practitioner Regulation Agency, responsible for clinician accreditation in Australia, reminded practitioners to consider their professional obligations when using GenAI in practice, particularly with respect to accountability, understanding, transparency, and informed consent<sup>88</sup>. Some recommendations have called for self-regulation at an industry level, with a codesign process between stakeholders, developers and users encouraging transparency and potentially increasing public trust<sup>89,90</sup>.

### Macro: national and international organizations

Most LLMs have been released globally. Ideally, a global approach from regulators is required; however, proactive regulation is impossible with the proverbial cat being already out of the bag. Broader regulation at the national and international macro level is challenging and likely lags significantly behind GenAI research and development. The first international convention was recently signed by the Council of Europe<sup>91</sup>. The Bletchley Declaration, signed by 28 countries and the European Union, at an AI Safety Summit, establishes a shared understanding of the opportunities and risks posed by frontier artificial intelligence. The aim of this declaration was to promote increased transparency by private actors developing frontier AI capabilities, appropriate evaluation metrics, tools for safety testing, and developing relevant public sector capability and scientific research while acknowledging that approaches would "differ" with respect to applicable legal frameworks<sup>92</sup>. A similar declaration from the United Nations<sup>93</sup> has cited concerns about human rights infringements and inequity with GenAI technology, but apart from development of an independent international scientific panel, intergovernmental and multi-stakeholder policy dialog tacitly acknowledges the difficulties in enforcement. Some regulatory approaches include risk-based approaches, where compliance obligations are proportionate to the level of risk, medical or otherwise, posed by the use of GenAI technology. These include sector-agnostic and sector-specific rules and regulations, depending on a particular sector's use of GenAI; and policy alignment, incorporating GenAI-related rule making within existing frameworks for cybersecurity, data privacy, and intellectual property protection.

National governments have recognized that there is low public trust in GenAI systems which can in turn slow adoption and public acceptance. The risk-based approach seeks, through greater testing, transparency, and oversight, to pre-emptively mitigate potential negative impacts from GenAI and LLMs that could be difficult or impossible to reverse<sup>94</sup>. GenAI systems are being developed and deployed at a speed and scale that will outpace the capacity of the legislative frameworks. A map of potential GenAI risks may need to be developed to be answered by future GenAI regulations to ensure that it can account for and handle new risks, potential, and actual alike<sup>95</sup>. Government-level organizations are calling on those developing and deploying GenAI in high-risk contexts to take their own proactive steps to ensure user and consumer safety<sup>94</sup>.

Other national approaches have included mooted the legal protection of human rights in the USA<sup>96</sup>, national AI strategies in the UK<sup>97</sup>, Hong Kong<sup>98</sup>, white papers in Japan<sup>99</sup> and voluntary standards in Australia<sup>100</sup>.

Legal regulatory approaches for therapeutic devices are required to account for the unique differences in the development and distribution of LLMs compared to other existing medical technologies. To safeguard patient care, Mesko and Topol suggested that a regulatory body only has to design regulations for LLMs if either the developers of LLMs claim that their LLM can be used for medical purposes, or if LLMs are developed for, adapted, modified, or directed toward specific medical purposes<sup>82</sup>. Such adaptation or use of LLMs for medical purposes may not always be explicitly stated, or even intended, by developers. Even if the currently widespread LLMs do not fall into either category, further iterations of the medical alternatives of LLMs specifically trained on medical data and databases will probably occur. A participatory approach to AI governance, informed by the micro and end-user levels will be more effective than overarching top-down regulations.

There is little by way of international oversight governing the use of GenAI in medical education. An initial advancement of the meso-level approaches would see institutions and accrediting bodies collaborating and adopting shared strategies at a national level. It remains to be seen if global governance would be necessary, or even feasible, in medical education.

Underpinning regulatory concerns is an understandable focus on patient safety, privacy, transparency, and ongoing trust in the healthcare profession. However, this safety cannot be guaranteed if learners, the future workforce, are deficient in clinical reasoning and critical thinking skills because of, or when operating within, GenAI-integrated environments.



Accountability lies with the end-user of any GenAI technology, as it would with any therapeutic good or device, and navigating the challenges that GenAI represent is an important learning skill. A broader view of “regulation” with participation from all stakeholders, will help ensure that accrediting bodies, education providers and students will understand and consider how GenAI and LLMs affect learning, development of knowledge and skills and attainment of competency in practice.

## Conclusion

The intersection of the role of GenAI in medical education and clinical use hinges on issues of governance and regulation. Currently, GenAI is unlikely to become fully autonomous and unreservedly accepted by the wider medical community and by patients due to issues of trustworthiness and the as-yet unknown impacts on the doctor-patient relationship, despite promised gains in efficiency and personalization of outcomes. Its use and inputs need to be constantly moderated and updated by humans to ensure the veracity and utility of its output, retain its generative capacity and prevent model collapse. The implications of this in medical education should be considered in the context of the learning process, authentic assessment, and the preservation of academic integrity. Regulation, in its different guises, applied thoughtfully at different levels, will guide users towards safe, appropriate, and equitable use of these technologies. In place of blind acceptance, a balanced and considered collaboration between humans, GenAI, and governance will permit advancements in learning possibilities and efficiencies without over-regulation stifling innovation and progress.

## Data availability

No datasets were generated or analyzed during the current study.

Received: 27 September 2024; Accepted: 14 May 2025;

Published online: 27 May 2025

## References

1. IBM. What is Generative AI? [cited 2024 10 Sep 2024]; Available from: <https://research.ibm.com/blog/what-is-generative-ai> (2024).
2. IBM. What are Large Language Models (LLMs)? [cited 2024 10 Sep 2024]; Available from: <https://www.ibm.com/topics/large-language-models> (2024).
3. Masters, K. Artificial intelligence in medical education. *Med. Teach.* **41**, 976–980 (2019).
4. Hanycz, S. A. & Antipervitch, P. A practical review of generative AI in cardiac electrophysiology medical education. *J. Electrocardiol.* **90**, 153903 (2025).
5. Lu, M. Y. et al. A Multimodal generative AI Copilot for human pathology. *Nature* (2024).
6. Wang, X. et al. Foundation model for predicting prognosis and adjuvant therapy benefit from digital pathology in GI cancers. *J. Clin. Oncol.* p. JCO2401501 (2025).
7. Eynon, R. & Young, E. Methodology, legend, and rhetoric: the constructions of AI by academia, industry, and policy groups for lifelong learning. *Sci., Technol., Hum. Values* **46**, 166–191 (2020).
8. Bell, G., Burgess, J., Thomas, J. & Sadiq, S. Rapid Response Information Report: Generative AI—Language Models (LLMs) and Multimodal Foundation Models (MFM), (Australian Council of Learned Academies, 2023).
9. Masters, K. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. *Med Teach.* **45**, 574–584 (2023).
10. Tolsgaard, M. G. et al. The fundamentals of Artificial Intelligence in medical education research: AMEE Guide No. 156. *Med. Teach.* **45**, 565–573 (2023).
11. Reznick, R. Harris, K., Horsley, T. & Mohsen, H. Task force report on artificial intelligence and emerging digital technologies, Royal College of Physicians and Surgeons of Canada (2020).
12. Gordon, M. et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Med Teach.* **46**, 446–470 (2024).
13. Lee, J., Wu, A. S., Li, D. & Kulasegaram, K. M. Artificial intelligence in undergraduate medical education: a scoping review. *Acad. Med.* **96**, S62–S70 (2021).
14. Alam, F., Lim, M. A. & Zulkipli, I. N. Integrating AI in medical education: embracing ethical usage and critical understanding. *Front. Med.* **10**, 1279707 (2023).
15. AAIN Generative AI Working Group, *AIN Generative Artificial Intelligence Guidelines* (Austrian Academic Integrity Network, 2023).
16. Valiant, L. *The Importance of Being Educable: A New Theory of Human Uniqueness* (Princeton University Press, 2024).
17. Foltyniek, T. et al. ENAI recommendations on the ethical use of artificial intelligence in education. *Int. J. Educ. Integr.* **19** (2023).
18. Preiksaitis, C. & Rose, C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. *JMIR Med Educ.* **9**, e48785 (2023).
19. Dong, H., Lio, J., Sherer, R. & Jiang, I. Some learning theories for medical educators. *Med. Sci. Educ.* **31**, 1157–1172 (2021).
20. Kolb, D. A. *Experiential Learning: Experience as the Source of Learning and Development* (Prentice-Hall, 1984).
21. Ericsson, K. A. & Staszewski, J. J. skilled memory and expertise: mechanisms of exceptional performance. In: Klahr, D., Kotovsky, K. Editors *Complex Information Processing*, 235–267 (Lawrence Erlbaum Associates, 1989).
22. Wang, S. et al. Artificial intelligence in education: a systematic literature review. *Expert Syst. Appl.* **252** (2024).
23. Macnamara, B. N. et al. Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers’ awareness? *Cogn. Res. Princ. Implic.* **9**, 46 (2024).
24. Tran, M., Balasooriya, C. & Semmler, C., Rhee, J. Generative artificial intelligence: the ‘more knowledgeable other’ in a social constructivist framework of medical education. *npj Digit. Med.* Under Review (2025).
25. Safranek, C. W., Sidamon-Eristoff, A. E., Gilson, A. & Chartash, D. The role of large language models in medical education: applications and implications. *JMIR Med Educ.* **9**, e50945 (2023).
26. Linardatos, P., Papasteranopoulos, V. & Kotsiantis, S. Explainable AI: a review of machine learning interpretability methods. *Entropy* **23** (2020).
27. Bearman, M. & Ajajawi, R. Learning to work with the black box: Pedagogy for a world with artificial intelligence. *Br. J. Educ. Technol.* **54**, 1160–1173 (2023).
28. Clusmann, J. et al. The future landscape of large language models in medicine. *Commun. Med.* **3**, 141 (2023).
29. van der Niet, A. G. & Bleakley, A. Where medical education meets artificial intelligence: ‘Does technology care?’ *Med Educ.* **55**, 30–36 (2021).
30. Reddy, S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implement Sci.* **19**, 27 (2024).
31. Mykhailov, D. Philosophical dimension of today’s educational technologies: framing ethical landscape of the smart education domain. *NaUKMA Res. Pap. Philo. Relig. Stud.* 68–75 (2023).
32. Moritz, S., Romeike, B., Stosch, C. & Tolks, D. Generative AI (gAI) in medical education: Chat-GPT and co. *GMS J. Med. Educ.* **40**, Doc54 (2023).
33. Brin, D. et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Sci. Rep.* **13**, 16492 (2023).
34. Alfertshofer, M. et al. Sailing the seven seas: a multinational comparison of ChatGPT’s performance on medical licensing examinations. *Ann. Biomed. Eng.* **52**, 1542–1545 (2024).
35. Lucas, H. C., Upperman, J. S. & Robinson, J. R. A systematic review of large language models and their implications in medical education. *Med. Educ.* (2024).
36. Zack, T. et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. *Lancet Digit. Health* **6**, e12–e22 (2024).



37. Li, Z. et al. Large language models and medical education: a paradigm shift in educator roles. *Smart Learning Environ.* **11**(2024).
38. Chan, K. S. & Zary, N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. *JMIR Med. Educ.* **5**, e13930 (2019).
39. Tolsgaard, M. G., Boscardin, C. K., Park, Y. S., Cuddy, M. M. & Sebok-Syer, S. S. The role of data science and machine learning in Health Professions Education: practical applications, theoretical contributions, and epistemic beliefs. *Adv. Health Sci. Educ. Theory Pract.* **25**, 1057–1086 (2020).
40. Balasooriya, C. et al. Learning, teaching and assessment in health professional education and scholarship in the next 50 years. *FoHPE* **25** (2024).
41. Abd-Alrazaq, A. et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med. Educ.* **9**, e48291 (2023).
42. Lomis, K. et al. Artificial Intelligence for Health Professions Educators. *NAM Perspect.* **2021** (2021).
43. Nagi, F. et al. Applications of artificial intelligence (AI) in medical education: a scoping review. *Stud. Health Technol. Inf.* **305**, 648–651 (2023).
44. Boscardin, C. K., Gin, B., Golde, P. B. & Hauer, K. E. ChatGPT and Generative artificial intelligence for medical education: potential impact and opportunity. *Acad. Med.* **99**, 22–27 (2024).
45. Artsi, Y. et al. Large language models for generating medical examinations: systematic review. *BMC Med. Educ.* **24**, 354 (2024).
46. Laupichler, M. C., Rother, J. F., Grunwald Kadow, I. C., Ahmadi, S. & Raupach, T. Large language models in medical education: comparing ChatGPT- to human-generated exam questions. *Acad. Med.* **99**, 508–512 (2024).
47. Scott, K. & Hart, J. Digital technologies in health: implications for health professional education. *FoHPE.* **25** (2024).
48. Pearce, J. & Chiavaroli, N. Rethinking assessment in response to generative artificial intelligence. *Med. Educ.* **57**, 889–891 (2023).
49. Fawns, T. & Schuwirth, L. Rethinking the value proposition of assessment at a time of rapid development in generative artificial intelligence. *Med. Educ.* **58**, 14–16 (2024).
50. Rampton, V., Mittelman, M. & Goldhahn, J. Implications of artificial intelligence for medical education. *Lancet Digit. Health* **2**, e111–e112 (2020).
51. Jackson, P. et al. Artificial intelligence in medical education - perception among medical students. *BMC Med. Educ.* **24**, 804 (2024).
52. Australian Academy of Technological Sciences and Engineering (ATSE) and A.I.f.M.L. (AIML), Responsible AI: Your questions answered. 2023: Canberra, Adelaide.
53. Moy, S. et al. Patient perspectives on the use of artificial intelligence in health care: a scoping review. *J. Patient Cent. Res. Rev.* **11**, 51–62 (2024).
54. Mikkelsen, J. G., Sorensen, N. L., Merrild, C. H., Jensen, M. B. & Thomsen, J. L. Patient perspectives on data sharing regarding implementing and using artificial intelligence in general practice - a qualitative study. *BMC Health Serv. Res.* **23**, 335 (2023).
55. Khullar, D. et al. Perspectives of patients about artificial intelligence in health care. *JAMA Netw. Open* **5**, e2210309 (2022).
56. Kudina, O. & de Boer, B. Co-designing diagnosis: Towards a responsible integration of Machine Learning decision-support systems in medical diagnostics. *J. Eval. Clin. Pract.* **27**, 529–536 (2021).
57. Mykhailov, D. A moral analysis of intelligent decision-support systems in diagnostics through the lens of Luciano Floridi's information ethics. *Hum. Aff.* **31**, 149–164 (2021).
58. Juravle, G., Boudouraki, A., Terziyska, M. & Rezlescu, C. Trust in artificial intelligence for medical diagnoses. *Prog. Brain Res.* **253**, 263–282 (2020).
59. Candelon, F., di Carlo, R. C., De Bondt, M. & Evgenious, T. AI regulation is coming. *Harvard Bus. Rev.* (2021).
60. Longoni, C., Bonezzi, A. & Morewedge, C. K. Resistance to medical artificial intelligence. *J. Consum. Res.* **46**, 629–650 (2019).
61. Sauerbrei, A., Kerasidou, A., Lucivero, F. & Hallowell, N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. *BMC Med. Inf. Decis. Mak.* **23**, 73 (2023).
62. de Boer, B. & Kudina, O. What is morally at stake when using algorithms to make medical diagnoses? Expanding the discussion beyond risks and harms. *Theor. Med. Bioeth.* **45**, 245–266 (2021).
63. Kingsford, P. A. & Ambrose, J. A. Artificial intelligence and the doctor-patient relationship. *Am. J. Med* **137**, 381–382 (2024).
64. Lorenzini, G., Arbelaes Ossa, L., Shaw, D. M. & Elger, B. S. Artificial intelligence and the doctor-patient relationship expanding the paradigm of shared decision making. *Bioethics* **37**, 424–429 (2023).
65. Mittelstadt, B. *The Impact of Artificial Intelligence on the Doctor-Patient Relationship* (Council of Europe 2021).
66. Mittermaier, M., Raza, M. & Kvedar, J. C. Collaborative strategies for deploying AI-based physician decision support systems: challenges and deployment approaches. *NPJ Digit. Med.* **6**, 137 (2023).
67. U.S. Department of Health and Human Services. Artificial Intelligence (AI) at HHS. [cited 10 Sep 2024]; Available from: <https://www.hhs.gov/programs/topic-sites/ai/index.html> (2024).
68. Jonnagaddala, J. & Wong, Z. S. Privacy preserving strategies for electronic health records in the era of large language models. *NPJ Digit. Med.* **8**, 34 (2025).
69. Productivity Commission, Australian Government. Making the most of the AI opportunity: The challenges of regulating AI: Canberra (2024).
70. Nikolic, S. et al. ChatGPT, Copilot, Gemini, SciSpace and Wolfram versus higher education assessments: an updated multi-institutional study of the academic integrity impacts of Generative Artificial Intelligence (GenAI) on assessment, teaching and learning in engineering. *Austr. J. Eng. Educ.* **29**, 1–28 (2024).
71. Schuwirth, L. The need for national licensing examinations. *Med. Educ.* **41**, 1022–1023 (2007).
72. Schuwirth, L. W. & Van der Vleuten, C. P. Programmatic assessment: from assessment of learning to assessment for learning. *Med. Teach.* **33**, 478–485 (2011).
73. Bhanji, F. et al. Competence by design: the role of high-stakes examinations in a competence based medical education system. *Perspect. Med. Educ.* **13**, 68–74 (2024).
74. Shumailov, I. et al. AI models collapse when trained on recursively generated data. *Nature* **631**, 755–759 (2024).
75. De Angelis, L. et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front. Public Health* **11**, 1166120 (2023).
76. Xu, X., Chen, Y. & Miao, J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. *J. Educ. Eval. Health Prof.* **21**, 6 (2024).
77. Ngo, B., Nguyen, D. & vanSonnenberg, E. The cases for and against artificial intelligence in the medical school curriculum. *Radio. Artif. Intell.* **4**, e220074 (2022).
78. Franco D'Souza, R., Mathew, M., Mishra, V. & Surapaneni, K. M. Twelve tips for addressing ethical concerns in the implementation of artificial intelligence in medical education. *Med. Educ. Online* **29**, 2330250 (2024).
79. Fleisher, L. A. & Economou-Zavlanos, N. J. Artificial Intelligence can be regulated using current patient safety procedures and infrastructure in hospitals. *JAMA Health Forum* **5**, e241369 (2024).
80. Chouffani El Fassi, S. et al. Not all AI health tools with regulatory authorization are clinically validated. *Nat. Med.* **30**, 2718–2720 (2024).

81. Australian Human Rights Commission. Australia Needs AI Regulation [cited 3 Dec 2024]; Available from: <https://humanrights.gov.au/about/news/australia-needs-ai-regulation> (2023).
82. Mesko, B. & Topol, E. J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit. Med.* **6**, 120 (2023).
83. Glbson, D., Kovanovic, V., Ifenthaler, D., Dexter, S. & Feng, S. Learning theories for artificial intelligence promoting learning processes. *Br. J. Educ. Technol.* **54**, 1125–1146 (2023).
84. Yu, H. & Guo, Y. Generative artificial intelligence empowers educational reform: current status, issues, and prospects. *Front. Educ.* **8** (2023).
85. Gniel, H. *AI: A Regulatory Perspective* (Australian Government Tertiary Education Quality and Standards Agency, 2023).
86. The Royal Australian College of General Practitioners. Artificial intelligence in primary care [cited 2024 20/09/2024]; Available from: <https://www.racgp.org.au/advocacy/position-statements/view-all-position-statements/clinical-and-practice-management/artificial-intelligence-in-primary-care> (2024).
87. Knopp, M. I. et al. AI-enabled medical education: threads of change, promising futures, and risky realities across four potential future worlds. *JMIR Med Educ.* **9**, e50373 (2023).
88. Australian Health Practitioner Regulation Agency. Meeting your professional obligations when using Artificial Intelligence in healthcare [cited 3 Dec 2024]; Available from: <https://www.ahpra.gov.au/Resources/Artificial-Intelligence-in-healthcare.aspx> (2024).
89. Australian Government Digital Transformation Agency, Policy for the Responsible Use of AI in Government, Commonwealth of Australia (Digital Transformation Agency) (2024).
90. The White House. Fact Sheet: Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence [cited 13 Dec 2024]; Available from: <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/> (2023).
91. Council of Europe, Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Council of Europe Treaty Series No. 225 (2024).
92. Department for Science, I.T., The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023, Department for Science, Innovation & Technology (2023).
93. United Nations AI Advisory Body, Governing AI for Humanity (2024).
94. Australian Government Department of Industry, S.a.R., Safe and responsible AI in Australia consultation: Australian Government's interim response. Commonwealth of Australia (2024).
95. Wellner, G. A postphenomenological guide to AI regulation. *J. Hum.-Technol. Relat.* **2**, 1–18 (2024).
96. The White House. Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People [cited 20 Sep 2024]; Available from: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> (2024).
97. Government of the United Kingdom, National AI Strategy, HM Government (2021).
98. Digital Policy Office: The Government of the Hong Kong Special Administrative Region of the People's Republic of China, Ethical Artificial Intelligence Framework (2024).
99. Ministry of Education, C., Sports, Science and Technology - Japan, White Paper on Science, Technology, and Innovation: How AI will transform Science, Technology and Innovation. (Ministry of Education, Culture, Sports, Science and Technology, 2024).
100. Australian Government Department of Industry, S.a.R., Voluntary AI Safety Standard. (Commonwealth of Australia, 2024).

## Author contributions

Each author, M.T., C.B., J.J., G.L., N.M., S.R., J.R., L.S., N.S., C.S. and Z.W. contributed to the planning, writing and review of the manuscript. All authors have read and approved the final manuscript.

## Competing interests

The following authors are affiliated with this journal (npj digital medicine) as Editors. They have excused themselves from the editorial process and handling of this article to remain impartial. (1) J.J. (2) Z.W.

## Additional information

**Correspondence** and requests for materials should be addressed to Michael Tran.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025