Chapter

Ethical Hazards of Large Language Models in Primary Care: A Clinician-Focused Update

Michelle Lu, Justin J. Gillette and Thomas F. Heston

Abstract

Large Language Models (LLMs) are transforming clinical workflows in primary care through capabilities like diagnostic support, clinical documentation, and simulated empathetic engagement. Yet, these advancements bring underappreciated ethical hazards that directly impact front-line physicians. Unlike general discussions of AI ethics, this chapter focuses on dilemmas arising with the use of LLMs in real-world primary care: questions of liability when LLM suggestions influence clinical decisions; risks to confidentiality when LLMs interact with protected health information; cognitive offloading that may erode diagnostic skills; disruptions in patient trust when LLMs simulate empathy; and the growing reality of patients turning to generative models as informal therapists. As LLMs are embedded into electronic medical records and clinical apps, physicians must become active ethical agents in how these tools are used. These challenges arise in contexts where regulatory frameworks lag behind technological deployment, placing responsibility squarely on individual clinicians to navigate uncertain ethical terrain. Drawing on current literature and real-world clinical challenges, this chapter proposes a clinician-focused ethical framework to guide the responsible use of LLMs in primary care. This framework addresses both immediate practical concerns—such as informed consent for LLM-assisted care and appropriate documentation of AI involvement—and longer-term questions about professional identity and diagnostic autonomy in an AI-augmented practice environment. The goal is not to demonize these powerful tools but to equip physicians with the necessary conceptual tools, awareness, and decision-making strategies for safe and ethical integration. Ultimately, the foundation of primary care—human judgment, presence, and trust—must remain at the center of clinical decision-making, even in an era augmented by LLMs.

Keywords: Artificial Intelligence, large language models, primary care ethics, clinical decision-making, Generative AI, confidentiality, empathy

1. Introduction

Large language models (LLMs), such as ChatGPT, represent a recent paradigm shift in clinical informatics, with immediate and far-reaching implications for the practice of primary care. These generative models operate by leveraging transformer-based neural network architectures, trained on extensive medical and general-domain corpora, enabling outputs that resemble human reasoning and dialog [1]. In the primary care setting, characterized by longitudinal relationships, diagnostic ambiguity, and high cognitive load, LLMs are already being embedded in tools for documentation automation, differential diagnosis support, prior authorization, and patient communication interfaces.

Despite these utilities, the ethical hazards introduced by LLM deployment are underappreciated by clinicians and underrepresented in guidelines. Unlike algorithmic tools that are rule-based or supervised, LLMs operate in stochastic and opaque ways, with outputs influenced by training biases, prompt phrasing, and model-specific heuristics. In this sense, they blur traditional ethical boundaries: authorship becomes ambiguous, accountability is diffused, and the illusion of understanding may obscure diagnostic uncertainty [2, 3].

Furthermore, their integration into real-time clinical workflows introduces tensions specific to primary care. These include liability ambiguity when clinical actions are influenced by LLM-generated text, erosion of patient trust when empathy is simulated, confidentiality risks arising from model-data interaction, and the potential atrophy of clinical reasoning due to cognitive offloading. For example, recent studies indicate that patients are already using generative models as informal therapists, prompting complex medico-ethical considerations about the scope of practice, consent, and monitoring [4, 5].

This chapter examines these ethical fault lines from the standpoint of the practicing primary care physician. It does not offer a general LLM ethics theory, nor does it advocate for or against the use of LLMs per se. Instead, it identifies five concrete domains where dilemmas arise at the interface of LLMs and clinical reasoning: liability and clinical judgment, confidentiality and trust, erosion of clinical thinking, the physician-patient relationship, and patients' use of LLMs as informal therapists. For each domain, we provide a pragmatic ethical framework for physicians confronted with LLM-integrated tools. The aim is to reinforce the clinician's role as a moral agent, not merely a user, who must evaluate, interpret, and sometimes resist technological interventions. As LLMs become fixtures of the clinical landscape, the ethical center of gravity must remain rooted in human judgment, professional accountability, and the physician-patient relationship.

2. Liability and clinical judgment

"If I Follow the LLM and Get Sued, Who Pays?"

The use of LLMs in clinical reasoning presents a profound ethical and legal inflection point for primary care. At the heart of this transformation arises a novel ethical dilemma: the diffusion of accountability. Historically, medical liability has been grounded in a clinician's autonomous judgment, informed by best practices, clinical training, and professional discretion. However, when a physician acts based on LLM-generated suggestions – particularly from a generative model that offers probabilistic rather than deterministic guidance – the lines of responsibility become blurred [6]. LLMs do not possess agency, yet their outputs may materially influence decision-making in diagnosis, prescribing, and triage. This further demonstrates the profound influence of LLMs in primary care, but one whose accountability remains unassigned.

Unlike traditional clinical decision support systems, which follow transparent if-then logic, LLMs produce outputs based on statistical associations derived from opaque training data [7]. Thus, even when an LLM generates a plausible diagnostic or therapeutic recommendation, its underlying rationale cannot be reliably interrogated or reproduced in clinical practice [8]. Consequently, physicians remain fully liable for medical harm – even when decisions are materially shaped by algorithmic suggestions whose basis cannot be interrogated. This leads to liability asymmetry: the model may substantially inform a clinical decision, yet the legal burden of error and ethical accountability remains exclusively on the clinician. Furthermore, commercial LLM providers often explicitly disclaim medical responsibility in their terms of service, reinforcing that physicians bear full legal exposure – even when model outputs are directly integrated into electronic health record (EHR) interfaces or clinical workflows [9–11]. While regulatory bodies and professional organizations are beginning to develop LLM-specific guidance for healthcare settings, these frameworks remain nascent and have not yet resolved the fundamental question of liability distribution when LLMs materially influence clinical decisions.

Compounding this legal uncertainty is implicit delegation. LLMs generate highly fluent and confident-sounding text, masking significant errors, biases, or omissions rooted in training data skew [11]. In such settings, clinicians may unconsciously defer to LLM suggestions, even if they conflict with their clinical judgment and intuition [9, 12].

This dynamic is particularly hazardous in high-stakes scenarios such as chest pain triage, antibiotic stewardship, or prescribing in polypharmacy contexts. A physician who defers to an LLM-generated differential diagnosis may do so without recognizing the hidden assumptions embedded in the model's output. This risk is heightened under time constraints and cognitive overload [13]. If harm ensues, standard malpractice frameworks may not adequately capture the contributory role of the LLM, and existing tort law offers little clarity on LLM-influenced negligence [13].

Therefore, clinicians must retain active interpretive control over LLM-generated content. The impacts of generative AI in primary care are profound, but without institutional safeguards, legal reform, and ethical clarity, its integration can obscure medical responsibility [14, 15]. Ethical integration requires that outputs be treated as unvalidated suggestions – not surrogate reasoning. Embedding LLMs into clinical workflows without concurrent safeguards for accountability risks not only legal liability but also professional disempowerment [12, 14]. The clinician must remain the ultimate interpreter of clinical truth, regardless of whether that truth originates from human reasoning or LLM-generated algorithmic synthesis.

3. Confidentiality and trust

"Can I Trust the LLM with My Patient's Secrets?"

Primary care is grounded in trust, with confidentiality serving as a cornerstone of the physician-patient relationship. The integration of LLMs into clinical environments raises serious ethical concerns about the security of sensitive health data [16]. Unlike traditional software tools, LLMs operate in ways that make it difficult to guarantee data boundaries. Model outputs may reflect, store, or infer protected health information (PHI) in ways that are not transparent to the end user [16].

Current implementations of LLMs in healthcare often rely on cloud-based APIs managed by third-party vendors. When clinicians input identifiable or clinically sensitive information into these systems – whether for documentation, summarization, or patient communication – the data traverses insecure external networks to external servers for processing. Although many vendors claim to anonymize or secure this information, recent disclosures reveal that LLMs can retain and inadvertently regurgitate data fragments from previous inputs, including personal identifiers, posing significant risks to confidentiality [17, 18].

This technical vulnerability is compounded by regulatory ambiguity. In the United States, HIPAA compliance frameworks were established before the advent of the LLM era. Questions remain about whether LLMs operated by non-covered entities qualify as business associates, and whether their training, fine-tuning, or caching mechanisms are compatible with existing privacy mandates [16, 19]. Furthermore, when patients independently use public-facing LLMs for health queries, those interactions fall entirely outside HIPAA protections and, thus, beyond the reach of clinical oversight or informed consent.

Trust erosion is not only a legal issue but also a relational one. Patients may reasonably question whether LLM-augmented clinical notes or chatbot interactions compromise their privacy. Studies reveal that patient trust in LLM-mediated care declines when clinicians are unable or unwilling to explain how their data is used, stored, or potentially leveraged for model training [20, 21]. A clinician who cannot confidently explain where a patient's data goes or how it might be used undermines the fiduciary basis of care [20]. In primary care settings, where longitudinal relationships and trust hinge on disclosure, perceived breaches or data misuse may cause lasting harm and have durable effects [21].

Thus, any implementation of LLMs in clinical workflows must be accompanied by explicit data governance protocols. These should include limits on identifiable input, strict audit trails, model-use boundaries, and informed consent where applicable. Until these protections are formalized, the ethical imperative remains non-negotiable but straightforward: patient secrets must not become model fodder.

4. Erosion of clinical thinking

"Will LLMs Atrophy My Clinical Skills?"

LLMs provide powerful cognitive support by generating differential diagnoses, treatment plans, and documentation in real-time. However, their utility as clinical extenders comes with a potential cost: erosion of the clinician's diagnostic reasoning [22, 23]. In primary care, where uncertainty is the norm and presentations often lack pathognomonic clarity, cognitive vigilance is critical. The routine deferral of foundational clinical thinking to LLMs may gradually displace active clinical reasoning with passive acceptance of LLM-generated suggestions.

This hazard is not hypothetical. Analogous phenomena have been documented in aviation and radiology, where over-reliance on decision-support systems has led to skill degradation – a phenomenon termed "automation bias" [22, 23]. LLMs deliver authoritative, confident-sounding suggestions despite the underlying mechanism being purely statistical. In the context of medicine, clinicians exposed to LLMs may unconsciously accept well-framed but fundamentally flawed suggestions, even when these suggestions conflict with clinical findings and knowledge [22, 23]. This anchoring effect is hazardous in primary care, where diagnosis often requires pattern recognition over time and contextual inference unavailable to the model.

Moreover, LLMs introduce a cognitive asymmetry: their fluency and confidence can mask their underlying lack of understanding. LLMs are text predictors trained to mimic coherence, which can obscure the fact that the model lacks an accurate understanding or access to patient-specific context [24]. When these outputs are accepted uncritically, the clinician risks substituting statistical mimicry for mechanistic reasoning.

The implications for medical education are equally concerning. Trainees increasingly encounter LLMs as clinical resources, potentially shaping epistemic habits that prioritize information retrieval over diagnostic synthesis, thus preventing the development of foundational skills [22, 24]. If differential construction becomes outsourced, the iterative process of refining diagnostic hypotheses through active synthesis, comparison, and testing may atrophy [23, 24]. In this sense, LLMs could function not as knowledge extenders but as cognitive substitutes – eroding the very competencies they are meant to support.

To mitigate this risk, clinicians must preserve diagnostic authorship. This requires deliberate friction in the use of LLM outputs: pausing to question, verify, and reinterpret rather than accepting suggestions at face value [24]. Documentation generated by LLMs should be treated as a draft – not a diagnostic endpoint – and must be critically edited to reflect the physician's reasoning process [24]. The long-term integrity of clinical judgment depends not on resisting technology but on integrating it without surrendering cognitive agency.

5. The physician-patient relationship in the LLM era

"Does My Patient Want Me or the Machine?"

Empathy is not ancillary in primary care; it is essential to effective therapy. A physician's ability to engage, actively listen, and emotionally respond to their patient is not merely an interpersonal skill but rather a key tenet of effective healing. It promotes adherence, reduces diagnostic error, and directly improves clinical outcomes. Yet LLMs, now embedded in clinical communication tools, have introduced a new variable: simulated empathy [25]. These models generate seemingly compassionate responses through syntactically tailored phrases, contextual memory, and reflective language [25]. These capabilities raise foundational questions about the nature of the therapeutic alliance. If patients feel heard and understood by a model, what remains uniquely human in the clinician's role?

Recent studies have revealed that patients often rate LLM-generated responses as more empathetic and easier to understand than those provided by physicians, particularly in asynchronous messaging or decision-aid scenarios [26]. This observation not only reflects the persuasive fluency of generative text but also highlights the system-level limitations – time, emotional fatigue, documentation burdens, and volume – that pose significant barriers to sustained real-time human empathy in practice [27]. As a result, AI-mediated empathy can be experienced as more consistent, more immediate, and, in some cases, more emotionally validating.

However, the ethical risk lies in the mismatch between emotional resonance and actual relational presence that arises from the erosion of relational accountability when a physician's core interpersonal responsibilities are delegated to a computational model [26]. LLMs do not possess internal states, moral commitments, or memory beyond session context (unless artificially engineered) [27]. Their empathy is computationally constructed from learned linguistic patterns rather than being based on affective presence. Patients using LLMs may feel heard and emotionally validated, while the system's response is unaware, unfeeling, and uninterested; there is no memory continuity beyond the chat encounter [26]. When patients receive emotionally resonant messages from a non-sentient model, the resulting relationship is asymmetrical and potentially deceptive [25]. This misalignment may erode informed consent, especially if the model's non-human status is unclear.

Further, reliance on LLMs for relational labor may change physician behavior. In high-volume practices, LLMs are increasingly proposed as front-line communicators for delivering lab results, routine counseling, and even diagnostic explanations [26]. Using an LLM as a substitute rather than a supplement may cause physicians to offload core obligations – presence, accountability, and moral responsibility – resulting in depersonalized care [27]. Delegating these interactions risks narrowing the clinician's role to technical oversight, thereby commodifying what was once a core dimension of healing: presence [27].

Physicians must therefore reassert the ethical centrality of relational labor. While LLMs can augment communication, they cannot replace the embodied moral presence that undergirds trust [25]. Transparency is essential – patients must know when they are interacting with a machine. A recent review found that patient trust in LLMs was highest when a physician was present, mediating communication, and lowest when the LLM was used as a substitute for a human moderator [28]. Thus, further proof is provided that the use of LLMs in patient communication should be additive, not substitutive: enhancing access or efficiency without replacing the bidirectional, meaning-making dialog that defines human care [25, 28].

In this LLM era, the clinician's role may evolve – but it cannot be reduced. The therapeutic value of being known, remembered, and cared for cannot be emulated by even the most linguistically adept machine [25].

6. Counseling patients about LLM therapists

"My Patient Is Using LLM as a Therapist - What Now?"

In under-resourced settings, or among patients facing financial, geographic, or cultural barriers to mental health care, LLM-based tools are increasingly used as informal therapists [29]. Accessible through mobile devices, chat interfaces, or integrated health apps, LLMs offer patients a semblance of listening, reflection, and even problem-solving [28]. This emergent phenomenon introduces a complex ethical terrain for primary care clinicians, who often serve as the default mental health providers for their patients.

LLMs can produce therapeutic-sounding responses that emulate techniques from cognitive behavioral therapy, motivational interviewing, or psychodynamic reflection [29]. Yet, they are not trained mental health professionals, nor are they capable of clinical assessment, emotional attunement, or crisis intervention. LLMs create the illusion of professional support when none exists [29]. Patients who rely on LLMs for psychological support may unknowingly receive content that is unvalidated, non-specific, or even harmful in the context of suicidality, trauma, or psychosis [29].

Moreover, these interactions occur outside any formal clinical setting. There is no therapeutic contract, no confidentiality in the traditional sense, and no legal or ethical framework governing the exchange [16]. Data shared with LLM platforms may be stored, monetized, or repurposed without the patient's awareness [14, 16]. Some models lack guardrails to prevent the creation of suggestive or misleading statements, especially in emotionally charged contexts. In this way, LLMs can act as pseudo-clinicians without bearing any of the fiduciary, legal, or ethical obligations of care.

Clinicians must be prepared to acknowledge and address LLM-integrated care [30]. They must recognize that patients may seek LLM-based support due to accessibility barriers, stigma, or prior negative experiences with traditional healthcare. In contrast, others may engage due to a lack of awareness that the tool is non-human or unregulated [30]. Validation is essential. Patients should not be shamed or dismissed for seeking support where it is available. Physicians should assess patients' perceived benefits and risks associated with LLM usage and, if necessary, clarify the scope and limitations of LLM use [30]. Moreover, physicians should guide patients to safer alternatives, such as support group networks and teletherapy. Another option may be to utilize a mixed-use model, where clinicians use LLM as care supplementation for early reflections, while decisions involve face-to-face shared decision-making [30]. Most importantly, physicians must engage in shared follow-up not only to establish physical presence and build patient rapport but also to assess the emotional impacts of LLM use, identify adverse effects of LLM usage, and clarify any concerns or questions that may arise from patient use [29, 30].

Ethically, the clinician's role is twofold: to validate the patient's pursuit of support while clarifying the limitations and potential harms of LLM-based therapy [29]. This conversation should be approached with neutrality and framed within the broader commitment to safety, privacy, and continuity of care [30]. Where appropriate, referrals should be made to qualified human providers, but the clinician must recognize that LLM-based support may persist regardless [30]. The goal is not to eliminate its use, but to contextualize it – and ensure it occurs within a framework of informed awareness. Regardless of whether future LLMs demonstrate high diagnostic and therapeutic efficacy, profound ethical issues of trust, liability, and human relational presence remain.

7. Conclusion

The integration of large language models into primary care is not a distant prospect – it is an active transformation unfolding in clinical practice. These systems offer undeniable efficiencies in documentation, diagnostics, and communication [31]. Yet, these benefits also introduce profound ethical hazards that are distinctively shaped by the nature of frontline medicine, including time constraints, ambiguous symptoms, emotional fatigue, and relational continuity dilemmas, which abstract AI ethics frameworks do not encounter [15, 20, 32].

This chapter outlines several emergent hazards, including liability displacement, confidentiality risks, cognitive offloading, empathy simulation, and the rise of LLM-as-therapist dynamics. Across these domains, one constant remains: the clinician is the final ethical filter [25]. Regardless of how persuasive, efficient, or accurate LLMs appear, their outputs require physician judgment, skepticism, and contextualization [33].

Therefore, ethical practice in the LLM era demands more than technical competence [27]. Physicians must remain critical of well-written but potentially flawed LLM outputs, vigilant in protecting privacy, and unapologetic in defending the relational core of care [27]. Practical skepticism means questioning confident-sounding LLM outputs, especially when they conflict with clinical intuition or patient-specific context. Accountability must adapt to the growing role of LLMs, yet medicine is an inherently human duty that cannot be abdicated to a machine [34].

LLMs and generative AI models will continue to evolve, further embedding into electronic health records, patient portals, and clinical workflows, reshaping the clinical landscape. However, the ethical center of medicine – its accountability, empathy, and clinical reasoning – must not be outsourced. In the presence of transformative technology, the task is not to retreat but to remain: fully present, fully responsible, and fully human. **Table 1** summarizes these principles into actionable recommendations for clinical practice.

Ethical Domain	Core Principle	Practical Actions
Liability & Clinical Judgment	Maintain diagnostic authorship	• Treat LLM outputs as unvalidated suggestions. • Pause to question and verify before accepting. • Document your reasoning process, not just the LLM output. • Never delegate final decisions to the LLM.
Confidentiality & Trust	Protect patient data boundaries	• Use only HIPAA-compliant, on-premises LLM tools whenever possible. • Avoid inputting identifiable information into cloud-based LLMs. • Establish clear data governance protocols. • Inform patients about the use of LLM and data.
Clinical Reasoning	Preserve cognitive vigilance	• Use LLMs as a supplement, not a substitute for thinking. • Deliberately question LLM-generated differentials • Edit LLM documentation to reflect your reasoning. • Maintain diagnostic skills through active practice.
Physician-Patient Relationship	Prioritize authentic human connection	• Be transparent about LLM involvement in communication. • Use LLM to enhance, not replace personal interaction. • Preserve time for direct patient engagement. • Clarify when patients are interacting with an LLM versus a human.
Patient AI Use	Validate and guide safely	• Ask non-judgmentally about patient LLM use. • Clarify limitations and risks of LLM therapy. • Provide alternative resources when appropriate. • Follow up on the emotional impacts of LLM interactions

Ethical Domain	Core Principle	Practical Actions
General Principles (All Domains)	Core principles for all domains	• Transparency: Always disclose LLM involvement to patients. • Accountability: Physicians remain fully responsible for all clinical decisions. • Critical Evaluation: Question LLM outputs rather than accepting them uncritically. • Human-Centered Care: Preserve the relational core of medicine • Continuous Learning: Stay informed about LLM capabilities and limitations.

Table 1.
Key ethical recommendations for LLM use in primary care.

Acknowledgments

The authors acknowledge support from the Department of Family Medicine at the University of Washington and the Department of Medical Education and Clinical Sciences at Washington State University. The authors thank all clinical colleagues who offered case insights and discussions during the preparation of this chapter.

Conflict of interest

The authors declare no conflicts of interest

Author details

Michelle Lu[1], Justin J. Gillette[2] and Thomas F. Heston[2]*

1 School of Medicine, University of Washington, Seattle, Washington, USA

2 Elson S. Floyd College of Medicine, Washington State University, Spokane, Washington, United States of America

*Address all correspondence to: theston@uw.edu

References

[1] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Precise4Q consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making. 2020;**20**(1):310

[2] Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. The Lancet Digital Health. 2023;**5**(6):e333–5

[3] Currie GM, Hawk KE, Rohren EM. Generative artificial intelligence biases, limitations and risks in nuclear medicine: An argument for appropriate use framework and recommendations. Seminars in Nuclear Medicine. 2024;**55**:423–436

[4] Heston TF. Safety of large language models in addressing depression. Cureus. 2023;**15**(12):e50729

[5] Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Internal Medicine 2023;**183**:589–596

[6] Gabison GA, Xian RP. Inherent and emergent liability issues in LLM-based agentic systems: A principal-agent perspective. In: Proceedings of the 1st Workshop for Research on Agent Language Models (REALM 2025). Stroudsburg, PA, USA: Association for Computational Linguistics; 2025. p. 109–130

[7] Hao Y, Qiu Z, Holmes J, Löckenhoff CE, Liu W, Ghassemi M, et al. Large language model integrations in cancer decision-making: A systematic review and meta-analysis. NPJ Digital Medicine 2025;**8**:450

[8] Mesinovic M, Watkinson P, Zhu T. Explainability in the age of large language models for healthcare. Communications Engineering. 2025;**4**(1):128

[9] Gerke S, Minssen T, Cohen G. Artificial Intelligence in Healthcare. In: Bohr A, Memarzadeh K, eds. Chapter 12 - Ethical and legal challenges of artificial intelligence-driven healthcare. Academic Press; 2020. p. 295–336. DOI: 10.1016/B978-0-12-818438-7.00012-5

[10] Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? Frontiers in Surgery. 2022;**9**:862322. DOI: 10.3389/fsurg.2022.862322

[11] Lee B, Naftalovich R, Ali S, Chaudhry FA, Tewfik GL. Liability exposure of clinicians in Artificial Intelligence-Driven healthcare. Biomedical Instrumentation & Technology. 2024;**58**(2):39–42

[12] Warraich HJ, Tazbaz T, Califf RM. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. JAMA. 2025;**333**(3):241–247

[13] Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. Heliyon. 2024;**10**(4):e26297

[14] Price IIWN, Gerke S, Cohen IG. Liability for use of artificial intelligence in medicine. In: Solaiman B, Cohen IG, editors. Research Handbook on Health, AI and the Law. Cheltenham, UK: Edward Elgar Publishing Ltd; 2024. p. 150–166

[15] Habli I, Lawton T, Porter Z. Artificial intelligence in health care: Accountability and safety. Bulletin of the World Health Organization. 2020;**98**(4):251–256

[16] Murdoch B. Privacy and artificial intelligence: Challenges for protecting health information in a new era. BMC Medical Ethics. 2021;**22**(1):122

[17] Dennstädt F, Hastings J, Putora PM, Schmerder M, Cihoric N. Implementing large language models in healthcare while balancing control, collaboration, costs and security. Npj Digital Medicine. 2025;**8**(1):143

[18] Privacy International. Large language models and data protection [Internet]. Privacy International. 2024. Available from: https://privacyinternational.org/explainer/5353/large-language-models-and-data-protection [Accessed: 2025-August-19]

[19] Ziller A, Mueller TT, Stieger S, Feiner LF, Brandt J, Braren R, et al. Reconciling privacy and accuracy in AI for medical imaging. Nature Machine Intelligence 2024;**6**:764–774

[20] Zondag AGM, Rozestraten R, Grimmelikhuijsen SG, Jongsma KR, van Solinge WW, Bots ML, et al. The effect of Artificial Intelligence on patient-physician trust: Cross-sectional vignette study. Journal of Medical Internet Research. 2024;**26**: e50853

[21] Esmaeilzadeh P, Mirzaei T, Dharanikota S. Patients' perceptions toward Human-Artificial Intelligence interaction in health care: Experimental study. Journal of Medical Internet Research. 2021;**23**(11):e25856

[22] Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments. 2024;**11**(1):28

[23] Jones N. Does using ChatGPT change your brain activity? Study Sparks debate. Nature. 2025;**643** (8070):15–16

[24] Choudhury A, Chaudhry Z. Large language models and user trust: Consequence of Self-Referential learning loop and the deskilling of health care professionals. Journal of Medical Internet Research. 2024;**26**:e56764

[25] Montemayor C, Halpern J, Fairweather A. In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. AI & Society. 2022;**37**(4):1353–1359

[26] Karoulla E, Matragkas N, Islam SU, Epiphaniou G. Challenges and risks of using synthetic data for AI-Driven healthcare applications. Studies in Health Technology and Informatics. 2025;**328**:81–85

[27] Allen MR, Webb S, Mandvi A, Frieden M, Tai-Seale M, Kallenberg G. Navigating the doctor-patient-AI relationship - a mixed-methods study of physician attitudes toward artificial intelligence in primary care. BMC Primary Care. 2024;**25**(1):42

[28] Steerling E, Siira E, Nilsen P, Svedberg P, Nygren J. Implementing AI

in healthcare-the relevance of trust: A scoping review. Frontiers in Health Services. 2023;**3**:1211150

[29] Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S. To chat or bot to chat: Ethical issues with using chatbots in mental health. Digital Health. 2023;**9**:20552076231183542

[30] Schiff D, Borenstein J. How should clinicians communicate with patients about the roles of artificially intelligent team members? AMA Journal of Ethics. 2019;**21**(2):E138–145

[31] Sallam M. Chatgpt utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. Healthcare (Basel). 2023;**11**(6):887

[32] Cestonaro C, Delicati A, Marcante B, Caenazzo L, Tozzo P. Defining medical liability when artificial intelligence is applied on diagnostic algorithms: A systematic review. Frontiers in Medicine (Lausanne). 2023;**10**:1305756

[33] Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello C-P, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digital Medicine 2023;**6**:111

[34] Price WN, Cohen IG. Privacy in the age of medical big data. Nature Medicine. 2019;**25**(1):37–43