EDITORIAL

DOI: http://dx.doi.org/10.15403/jgld-6827

Artificial Intelligence in Medicine. Systems Anatomy, Decision Physiology, Hygiene of Use

Bogdan Surdea-Blaga

Independent AI Strategy Consultant; Computer Engineer; PhD in International Business, Cluj-Napoca, Romania

Address for correspondence:
Bogdan Surdea-Blaga
bogdan@vucamix.com

Received: 04.01.2026
Accepted: 18.01.2026

Artificial intelligence (AI) is no longer a distant or experimental concept in medicine. From radiology and pathology to research writing and clinical documentation, AI-based systems are already embedded into everyday medical practice [1]. Yet, despite increasing exposure, there remains substantial confusion about what AI actually is, what it can reasonably do, and where its limitations begin. This gap between availability and understanding carries clinical, ethical, and professional risks [2].

A useful starting point is to clarify AI systems anatomy. Artificial intelligence is not a single tool but an umbrella term covering multiple system types. Traditional rule-based expert systems operate on fixed rules and do not learn. Machine learning systems learn from examples, while deep learning models identify complex patterns using layered neural networks [3, 4]. Generative AI, and particularly large language models (LLMs), represent a recent subset capable of producing new text, images, video, or code based on probabilistic pattern recognition. Increasingly, medical AI applications combine generative models with direct access to curated medical information or databases, allowing outputs to be informed by retrieved, task-specific content rather than by model knowledge alone. Fig. 1 provides a schematic overview of these AI system categories and their relationships.

These systems do not "understand" medicine in a human sense; rather, they predict plausible outputs based on prior data exposure. While these categories often overlap in practice rather than existing as strictly separated system types, confusing these main categories leads to misplaced expectations and inappropriate use [5].

In clinical contexts, generative AI is best understood not as an autonomous decision-maker, but as a decision support environment. Its strength lies in synthesis, summarization, pattern exploration, and scenario generation, not necessarily in diagnosis or judgment. This distinction becomes clearer when considering the physiology of AI-based reasoning. Large language models do not reason through causal chains as clinicians do; instead, they assemble responses by estimating

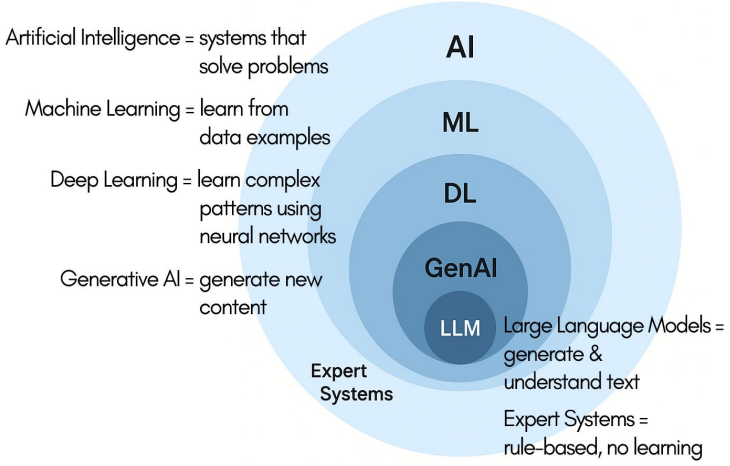

Fig. 1. Schematic representation of major artificial intelligence system categories.

likelihood using patterns learned from large collections of text. Without contextual understanding, which is often limited or incompletely provided, AI output can be fluent, confident, and occasionally wrong in subtle ways. The risk is not random error, but plausible error, sometimes called hallucination [6].

It is also important to acknowledge that, in selected and narrowly defined tasks, recent AI systems have demonstrated performance comparable to, or in some cases exceeding, that of clinicians. Large language models have shown strong results in structured medical question-answering and knowledge retrieval tasks, particularly when evaluated against standardized clinical benchmarks [7]. However, these gains are largely confined to controlled settings and well-specified problems, and do not translate into autonomous clinical reasoning or decision-making in real-world, context-rich environments. Recognizing both the strengths and the boundaries of such performance is essential to avoid inappropriate extrapolation of AI capabilities into domains where clinical judgment remains indispensable.

For this reason, the metaphor of AI as a cothinker or co-pilot [8] is more appropriate than that of a replacement. As a co-pilot, AI assists with information processing: summarizing literature, structuring manuscripts, drafting protocols, or generating alternative phrasings. As a co-thinker, it can help explore hypotheses, outline different considerations, or examine whether the clinical reasoning remains sound and consistent when different assumptions or scenarios are considered. In both roles, responsibility remains firmly with the clinician. The system assists cognition; it does not own it.

The growing integration of AI into academic and clinical workflows raises the issue of hygiene of use. Hygiene begins at the personal level. Clinicians should define the role of AI before each task, select platforms that allow data control, and avoid sharing identifiable patient information. Effective use of AI also requires a basic form of user training, sometimes referred to as prompt engineering, which involves learning how to frame questions and instructions clearly and consistently. Repeated tasks benefit from standardized prompts, while critical outputs should be verified either by the AI system itself, by a second AI system, or by independent sources. In critical cases, all these layers of verification should be employed. At minimum, the human-in-the-loop verification is not optional; it is critical to success [2, 9]. Although AI may reduce certain aspects of cognitive load, this benefit is partly offset by the need for careful review and verification of outputs, particularly in clinical or academic contexts. In addition, clinicians should be aware that AI systems evolve over time. While such changes are generally intended to improve performance, they may alter behavior and output quality, underscoring the importance of ongoing attention and periodic reassessment of how these tools are used.

Beyond personal practice, contextual hygiene must also be addressed. Existing guides and regulations governing confidentiality, authorship, accountability, and clinical responsibility continue to apply, regardless of AI involvement [10, 11]. These regulatory frameworks are evolving rapidly and can vary substantially across jurisdictions, adding further complexity to the use of AI in clinical and academic settings. Transparency is essential in academic work, where disclosures such as "AI-generated draft, clinician-verified" help maintain trust and accountability [12]. Another concept, explainability, matters when an AI system's reasoning cannot be meaningfully interrogated. If the reasoning of an AI system cannot be examined, its output should not be used for high-stakes decisions [13]. Opaque systems, sometimes called black boxes, in opaque contexts are unsafe by definition, as their use to inform clinical decisions, even if clinicians retain responsibility for the final judgment, carries an additional risk of liability when harm occurs. Additionally, AI systems may reflect or amplify biases present in their training data, with potential implications for diagnostic accuracy and treatment recommendations across different patient populations or groups.

An additional and often underestimated dimension is the patient perspective. Patients are already using AI systems to interpret symptoms, laboratory values, and imaging reports, often without sufficient medical knowledge. Clinicians must assume this reality and be prepared to contextualize, correct, and integrate AI-informed patient questions into consultations. Importantly, patients differ widely in health literacy and digital literacy, which can amplify misunderstandings when AI-generated information is interpreted without guidance. In this setting, AI may unintentionally lead to misinformation, overconfidence in preliminary interpretations, or anxiety driven by plausible but incorrect outputs. These dynamics have direct implications for shared decision-making, as AI-derived information may influence patients' expectations, risk perception, and treatment preferences. Ignoring patient-facing AI use does not prevent it; it only shifts these discussions outside the clinical relationship, where misinterpretation and loss of trust are more likely to occur [14, 15].

Despite these challenges, the potential value of AI in medicine is substantial. When used thoughtfully, AI can reduce cognitive load, improve access to information, and support reflective practice [1, 3]. However, its value is constrained less by technological capability than by human judgment. Over-automation risks deskilling, as continuous reliance on AI for routine tasks may reduce opportunities for clinicians, particularly trainees, to practice and maintain core clinical skills [16]. In addition, automation bias, defined as the tendency of clinicians to over-rely on automated suggestions, even when they are incorrect, has been well documented and may lead to reduced independent verification and missed errors [17]. The challenge is not rapid adoption of AI in medicine, but the sensible integration of these systems into day-to-day practice, given that sufficient guardrails are put in place.

In conclusion, AI in medicine should be approached as a powerful but carefully controlled tool. Understanding its systems anatomy clarifies what it can and cannot do. Appreciating its decision physiology explains why fluent outputs require careful scrutiny. Practicing hygiene of use, both personal and contextual, helps ensure safety, accountability, and trust. As with any transformative technology in medicine, progress will depend not on novelty and immediate integration, but on balanced and responsible adoption. As Arthur C. Clarke observed, "any sufficiently advanced technology is indistinguishable from magic" [18], a reminder that the responsibility of medicine is not to be impressed by technological sophistication, but to remain accountable for its use in the best interests of patients.

Conflicts of interest: None to declare.

REFERENCES

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56. doi:10.1038/s41591-018-0300-7
2. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380:1347–1358. doi:10.1056/NEJMra1814259
3. Obermeyer Z, Emanuel EJ. Predicting the future-big data, machine learning, and clinical medicine. N Engl J Med. 2016;375:1216–1219. doi:10.1056/NEJMp1606181
4. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–444. doi:10.1038/nature14539
5. Bommasani R, Drew A. Hudson, Ehsan Adeli, et al. On the opportunities and risks of foundation models. arXiv. 2021;2108.07258. doi:10.48550/arXiv.2108.07258
6. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots. Proc ACM FAccT. 2021:610–623. doi:10.1145/3442188.3445922
7. Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. Nat Med. 2024;30:309–318. doi:10.1038/s41591-023-02761-0
8. Farri E, Rosani G. Generative AI for Managers. Boston: Harvard Business Review Press; 2024.
9. Faraj S, Pachidi S, Sayegh K. Working and organizing in the age of the learning algorithm. Inf Organ. 2018;28:62–70. doi:10.1016/j.infoandorg.2018.02.005
10. European Commission. Ethics guidelines for trustworthy AI. Brussels; 2019.
11. European Parliament, Council of the European Union. Regulation (EU) 2024/1689. Artificial Intelligence Act. Off J Eur Union. 2024;L1689:1–144.
12. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364-1374. doi: 10.1038/s41591-020-1034-x
13. Jabbour S, Fouhey D, Shepard S, et al. Measuring the impact of artificial intelligence in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. JAMA. 2023;330(23):2275–2284. doi:10.1001/jama.2023.22295
14. Nundy S, Montgomery T, Wachter RM. Promoting trust between patients and physicians in the era of artificial intelligence. JAMA. 2019;322(6):497–498. doi:10.1001/jama.2018.20563
15. Gundlack J, Thiel C, Negash S, Buch C, Apfelbacher T, Denny K, et al. Patients' perceptions of artificial intelligence acceptance, challenges, and use in medical care: qualitative study. J Med Internet Res. 2025;27:e70487. doi:10.2196/70487
16. Saroha S. Artificial intelligence in medical education: promise, pitfalls, and practical pathways. Adv Med Educ Pract. 2025;16:1039–1046. doi: 10.2147/AMEP.S523255
17. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigation strategies. J Am Med Inform Assoc. 2012;19:121–127.
18. Clarke AC. Hazards of prophecy: the failure of imagination. In: Profiles of the Future. Rev ed. London: Victor Gollancz; 1973. p.14.