

From Prioritization to Proof: The Capability–Deployment Gap in Digital Health and Medical AI

Andreas Keck, MD¹  and Jefferson G. Fernandes, MD, PhD, MBA² ¹Researcher, Syte Institute, Hamburg, Germany; ²Visiting Professor, Federal University of Health Sciences of Porto Alegre, Porto Alegre, Brazil, THMT editorDOI: <https://doi.org/10.30953/thmt.v11.744>Corresponding Author: Jefferson G. Fernandes, Email: jeffernandes@specismed.com

Keywords: AI governance, clinical deployment, digital health, medical artificial intelligence, organizational readiness

Abstract

Drawing on findings from a survey of multinational pharmaceutical companies operating in Brazil and recent evidence on agentic and conversational medical AI, this editorial examines the capability–deployment gap: the distance between prioritizing technological capabilities and integrating them into routine, accountable, and sustainable practice.

Submitted: July 20, 2026; Accepted: September 1, 2026; Published: September 30, 2026

In a recent survey of 37 multinational pharmaceutical affiliates operating in Brazil, we reported a striking pattern: 65% of respondents prioritized artificial intelligence (AI) and machine learning, while only 8% reported meaningful telemedicine or electronic health record (EHR) integration.¹ The survey treats AI as a transversal enabling capability rather than a discrete application—which is what AI should be. But capability is transversal only when it connects to workflows, data, and evidence. Prioritized at 65% and connected at only 8%, it is transversal in ambition, not in operation. The survey did not read the situation as a deficit. This editorial does.

The deficit is not a lack of interest; interest is abundant. Nor is it a lack of imagination; healthcare and life sciences have no shortage of pilots, dashboards, and digital transformation claims. What is scarce is the connective tissue that turns capability into operational value: integrated data pathways, workflow integration, outcomes measurement, governance, and reimbursement. This is the capability–deployment gap—the distance between what organizations prioritize and what they can actually deploy, govern, measure, and sustain in routine operation.

The Brazilian survey makes this gap visible, but it is far from a Brazilian anomaly. Across healthcare systems, digital health and medical AI typically emerge first as strategic intent and only later, if at all, as operational capabilities—capabilities procured before organizations decide how to absorb them into workflows, evidence generation, and accountability.

The Brazilian data show the problem at the commercial and organizational level. AI, patient engagement, and data analytics ranked among the most frequently cited priorities, while telemedicine, EHR integration, and dedicated governance remained less mature.¹ The asymmetry matters because high-value pharmaceutical use cases require precisely the missing layer—feedback loops, longitudinal data, reliable outcome measurement, and integration into care pathways. This is a gap in data and evidence infrastructure, not in the quality of care.

Although the underlying challenge is similar, pharmaceutical companies and healthcare providers face fundamentally different deployment responsibilities. Pharmaceutical companies rarely operate hospital workflows or carry direct responsibility for live clinical escalation; their deployment problem sits in evidence generation, patient support, pharmacovigilance, market access, and partner-mediated integration with care. Providers own the clinical deployment problem directly. The shape is the same; the responsibilities are not.

The same structure is now visible on the clinical plane through agentic medical AI. As systems move from answering questions to coordinating tasks, generating recommendations, and managing parts of clinical workflows, the relevant unit of evaluation changes. It is no longer sufficient to ask whether a model performs well; the more important question is whether an AI-supported system can operate safely, accountably, and sustainably in live care. In medicine, ungoverned capability is not progress; it is a liability masquerading as progress. Nor is the gap confined to medicine: in a global survey of over 3,000 enterprises, only 21% reported a mature governance model for autonomous AI agents, even as most planned to deploy them within 2 years.²

Recent studies show both the promise and the current limits of the evidence. Medical intelligence for reasoning and action (MIRA), an autonomous EHR-integrated agent, performed strongly in a sandboxed EHR environment, yet its authors stressed the need for prospective real-world studies to establish generalization, safety, and governance.³ Articulate Medical Intelligence Explorer (AMIE), a conversational diagnostic AI system, was non-inferior to primary-care physicians in management reasoning in a virtual Objective Structured Clinical Examination (OSCE) format—but its authors likewise noted that real-world translation requires further research.⁴ Topol, a leading cardiologist and digital-medicine scholar, described these as important advances while stressing that simulated,

text-only settings are not real medical practice.⁵ Importantly, this limitation is not an external criticism. The investigators themselves explicitly acknowledge this limitation, and Topol independently reinforces the same conclusion.

The distinction is substantive: model performance differs from system readiness, simulation differs from deployment, and robustness differs from governance. A system may perform well under test conditions and still lack escalation pathways, liability allocation, auditability, and reimbursement mechanisms. Better models reduce some technical risks; they do not determine accountability, what data can be used, or when a human must intervene. This is where the debate stays too shallow: deployment is treated as the step after capability. In healthcare, however, it is part of the evidence problem itself. A system that cannot be governed, reimbursed, audited, and integrated into clinical workflows has not merely failed implementation; it has failed the conditions under which performance becomes clinically meaningful.

The pace of model development makes this more acute, not less. Benchmark performance has risen sharply year over year,⁶ and clinical literature is expanding just as fast. A large language model (LLM)-assisted review identified 4,609 peer-reviewed studies on LLMs in clinical medicine from January 2022 to September 2025; of these, just 19 were prospective randomized trials, and where models outperformed clinicians, the advantage shrank as the task moved closer to real medicine.⁷ Four thousand papers and 19 randomized trials do not yet constitute a mature clinical evidence base; they describe a field whose publication volume has outpaced its deployment proof. The faster models improve, the greater the temptation to confuse technical progress with deployment readiness. Organizations can therefore become technologically current while remaining operationally obsolete.

A second constraint is structural and will outlast any particular model or vendor: deployment capacity depends not only on internal readiness but also on model availability—a function of procurement, contract architecture, jurisdiction, and policy, not of performance alone. Organizations reasonably restrict AI use to enterprise-approved systems; platforms such as Microsoft Purview, Google's Gemini Enterprise governance layer, and Amazon Web Services (AWS) Bedrock AgentCore provide the security, audit, and compliance controls that ad hoc access cannot match.⁸ However, governance convenience can harden into strategic narrowing. An organization that uses only the model family that fits its contract stack lets procurement dictate its AI strategy—creating the kind of vendor dependence agile, supplier-diverse procurement is meant to avoid. In doing so, it risks surrendering the operational independence on which durable deployment depends.⁹

A single approved tool is not a strategy; it is a procurement state. Operational independence in medical AI is not independence from advanced models but from any single model, vendor, contract, or regulatory assumption.

None of this is a failure of will. Enterprise structures were built for a rate of change that AI now outpaces—planning cycles, governance, and accountability assume a slower clock than the technology runs on.¹⁰ The more difficult diagnosis is organizational rather than technological. The limiting factor is usually absorptive capacity, not innovation—the finding,

across public and private organizations, that scaling stalls more on skills, data access, legacy systems, and the absence of impact-measurement frameworks than on technology.⁹ And the gap is not new. A systematic review had already found only six of 66 economic-impact studies to be usable, none of which included a full cost–benefit analysis, 6 years before that review identified 19 randomized trials among 4,609 studies.¹¹ Two measurements, 6 years apart, point to the same conclusion: despite an explosion in technological capability, the evidence base for deployment has remained structurally thin. That is why it will not resolve itself as models improve and why so many AI strategies remain pilot graveyards with a budget line.

The path forward requires reversing the diagnosis. If the constraint is a mismatch of cadence, the remedy is not faster adoption but greater organizational readiness: assess where AI is competitively decisive, build the readiness to act, then implement iteratively, aligning internal cadence to the market rather than chasing each release.¹⁰ For pharmaceutical companies, this is concrete: the differentiator will rarely be the most advanced algorithm, but the ability to embed digital capability into credible evidence and operating models that survive technological and regulatory change. The threshold is crossed not by better tools but by an operating rhythm built to move with them.

Closing the gap requires a different evidence discipline: evaluating not only whether a tool works, but also the conditions under which it can be operated. Several questions become fundamental.

- Who controls the system and its output?
- How is liability distributed?
- How does escalation to a human occur?
- Can it be reimbursed or economically justified?
- Is the evidence drawn from real-world operations or from simulation alone?
- Can the organization compare and replace models without losing governance?
- What if a model becomes unavailable under policy or procurement?

These questions are not brakes on innovation but the conditions under which it becomes usable—the transition from potential to proof.¹² A proof-oriented strategy does not stop at naming promising technologies; it specifies the evidence needed to cross the threshold and the responsibilities that follow once deployment begins.

The Brazilian survey therefore points beyond Brazil. It captures a broader transition: from interest, experimentation, and strategic signaling toward evidence infrastructure, deployment discipline, and operational independence. The transition grows more urgent as medical AI moves from advisory functions toward autonomous participation in care: the more capable the system becomes, the less acceptable it is to evaluate it only as a tool rather than as a component of care.

The discipline that matters is asking, before deployment, whether capability has been matched by the conditions for responsible operation. The decisive question is no longer what technology can do, but what healthcare and life sciences organizations can prove, operate, govern, source, and scale.

Closing the capability–deployment gap will determine whether digital health and medical AI remain on the near

side of operational proof or become reliable, accountable, value-generating practices within care.

Funding

This work did not receive any funding.

Financial and Non-Financial Relationship and Activities

Not applicable.

Data Availability Statement (DAS), Data Sharing, Reproducibility, and Data Repositories

Not applicable.

Application of AI-Generated Text or Related Technology

ChatGPT v.5.0 was used for English language consistency review.

Contributions

This article involves researcher contributions and authorship criteria in multi-region collaborations (Brazil and Germany). Each author made substantial contributions to the conception, design of the presentation; or the acquisition, analysis, or interpretation of data for the work; and drafting the article or revising it critically for important intellectual content. Both authors approved of the final version to be published. There is agreement for accountability for all aspects of this work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Fernandes JG, Keck A. Digital health: a market survey of multinational companies and lessons from global peers. *Telehealth Med Today*. 2026;11:689. <https://doi.org/10.30953/thmt.v11.689>
2. Deloitte AI Institute. State of AI in the enterprise 2026: the untapped edge [Internet]. Deloitte; 2026 [cited 2026 Jul 20]. Available from: <https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html>
3. Ferber D, Hilgers L, Höper C, Kinny-Köster B, Eckardt JN, Egger-Heidrich K, et al. Towards autonomous medical artificial

intelligence agents. *Nature*. 2026;655:1282–91. <https://doi.org/10.1038/s41586-026-10675-5>

4. Liévin V, Palepu A, Weng WH, Saab K, Stutz D, Cheng Y, et al. Towards conversational AI for disease management. *Nature*. 2026;655:1292–9. <https://doi.org/10.1038/s41586-026-10764-5>
5. Topol E. Agentic AI comes to medicine. *Ground Truths* [Internet]. 2026 [cited 2026 Jul 20]. Available from: <https://erictopol.substack.com/p/agentic-ai-comes-to-medicine>
6. Stanford Institute for Human-Centered Artificial Intelligence. The 2026 AI index report [Internet]. Stanford University; 2026 [cited 2026 Jul 20]. Available from: <https://hai.stanford.edu/ai-index/2026-ai-index-report>
7. Chen SF, Alyakin A, Seas A, Yang E, Choi JJ, Lee JV, et al. LLM-assisted systematic review of large language models in clinical medicine. *Nat Med*. 2026;32:1152–9. <https://doi.org/10.1038/s41591-026-04229-5>
8. Microsoft. Use Microsoft Purview to manage data security and compliance for ChatGPT Enterprise [Internet]. Microsoft Learn; 2026 [cited 2026 Jul 20]. Available from: <https://learn.microsoft.com/>
9. Organisation for Economic Co-operation and Development. Governing with artificial intelligence: the state of play and way forward in core government functions. Paris: OECD Publishing; 2025. <https://doi.org/10.1787/795de142-en>
10. Gretscher P, Keck A, Wolff J, König A, Graf-Vlachy L. Strategische Herausforderungen für etablierte Unternehmen im Zeitalter der generativen KI. In: Obermaier R, editor. *Handbuch Industrie 4.0 und digitale Transformation*. Wiesbaden: Springer Gabler; 2026. https://doi.org/10.1007/978-3-658-36874-6_26-1
11. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. *J Med Internet Res*. 2020;22(2):e16866. <https://doi.org/10.2196/16866>
12. Keck A, Wolff J, Gretscher P. From potential to proof: how artificial intelligence in healthcare now delivers measurable return on investment – and why many companies risk falling behind. In: Kilic A, editor. *Artificial intelligence and machine learning in healthcare*. London: Academic Press/Elsevier; 2026. p. 197–206.

Copyright Ownership: This is an open-access article distributed in accordance with the Creative Commons Attribution Non-Commercial (CC BY-NC 4.0) license, which permits others to distribute, adapt, enhance this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0>. The authors of this article own the copyright.