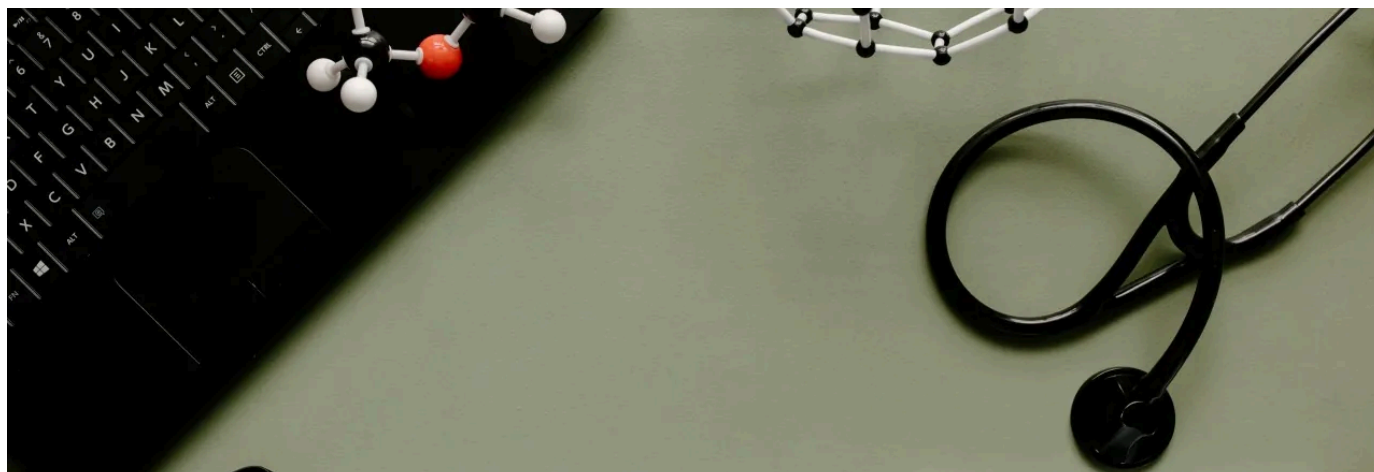


CodeCaffe

// Technology and society, brewed from the evidence.

Agentic AI in Healthcare: Balancing Autonomy, Policy, Governance & Regulation

Agentic AI can set goals and act on its own in clinical settings, raising new questions of trust, control and responsibility.



From the archive: Season 1, first published on LinkedIn on July 7, 2025. Reformatted for the website; the text is as first published unless marked.

Barista's Note

Agentic AI is emerging as one of the most significant developments in the future of healthcare. These are not just tools that follow instructions, but intelligent systems that can set goals, make decisions, and adapt in real-time. In clinical settings, this opens the door to faster workflows, better coordination, and new possibilities for patient care. It also raises new challenges around trust, control, and responsibility that existing systems were never designed to handle.

In this issue, we examine the implications of introducing autonomy into healthcare through agentic AI. We highlight practical use cases, assess the role of policy and governance, and ask what happens when these systems fail. We also examine the impact on the healthcare workforce, which is already under strain. To conclude, we propose ideas for integrating regulation, governance, and innovation in a manner that fosters accountability, transparency, and equitable outcomes in the medical sector.

ON THE MENU

Barista's Note A warm, personal message from the editor that frames the theme.

The Deep Drip Our long-form feature story.

Espresso Shots Quick jolts of insight in facts, stats or visuals.

Crème de la Crème Standout startups, tools, ideas or innovations.

Grind Articles Research papers, whitepapers and case studies that dig deeper.

The Last Sip Further reading, external links and hidden gems.

Beyond the Brew Closing thoughts and an invitation to connect.

Total

13 min read

The Deep Drip

Anything that can go wrong will go wrong

Murphy's law

AI agents are autonomous systems that respond to specific triggers or tasks and combine different algorithmic tools to achieve broader goals. These systems have emerged from the evolution of transformer-based large language models, which provide the reasoning and language capabilities that power modern agentic AI. The idea of autonomous agents performing and streamlining routine tasks has since taken the tech world by storm. The global agentic AI industry is currently valued at \$13.8 billion and projected to reach \$140.8 billion by 2032, according to [Markets and Markets](#). However, real-world implementations of these systems come with practical limitations. A recent [study](#) found that the best-performing AI agent failed to complete real-world office tasks nearly 70% of the time. In a separate evaluation, [Apple researchers](#) tested large language models on multi-step reasoning tasks such as the Tower of Hanoi and River Crossing puzzles, and found that the models struggled with complexity, often failing to maintain coherent reasoning as task difficulty increased.

Several promising applications have been identified for agentic AI in healthcare delivery and administration. Large language models (LLMs) have demonstrated strong potential in mining vast amounts of unstructured medical data (Figure 1), which is often scattered across various formats, including clinical notes, imaging reports, and laboratory records. Other notable use cases include automated clinical note-taking, diagnostic decision support, and medical education. A recent study published in [Nature Cancer](#) developed and validated an agentic AI system for clinical decision-making in oncology. The proposed system incorporates vision transformers for biomarker analysis in histopathology slides, MedSAM for radiological image segmentation, and web-based search tools for cross-referencing. The system was shown to improve decision-making accuracy by over 50% compared to using GPT-4 alone. However, real-world adoption of these systems remains limited. This is due mainly to the highly sensitive and heavily regulated nature of the healthcare industry, where the risk of costly errors, data breaches, and privacy violations requires a cautious and measured approach to integration.

Model	What it does
BioGPT 	Trained on biomedical literature to support tasks like question answering and text generation in the medical domain.
PubMedGPT 	Fine-tuned on PubMed abstracts to handle medical Q&A and summarization tasks.
GatorTron 	Trained on over 90 billion words from clinical records, enabling EHR comprehension and clinical decision support.
Med-PaLM 	Developed by Google, fine-tuned for medical question answering with performance nearing expert-level on some benchmarks.
ClinicalBERT 	Adapted from BERT and trained on MIMIC-III clinical notes for tasks like clinical entity recognition and patient risk prediction.
MedAlpaca 	An open-source medical chatbot fine-tuned from Alpaca using medical dialogue data.
PMC-LLaMA 	Trained on biomedical research from PubMed Central using the LLaMA architecture, supports biomedical reasoning tasks.

Figure 1: Overview of leading medical LLMs and their core functions. Each model is trained or fine-tuned on domain-specific datasets such as PubMed abstracts, clinical notes, or biomedical literature to support tasks including question answering, clinical decision support, summarization, and biomedical reasoning.

Autonomous Clinical Workflows

Agentic AI use-cases in healthcare can be meaningfully segmented based on the severity of their potential outcomes, which helps guide how they should be governed, monitored, and deployed. At the high-severity end are life-critical applications such as oncology treatment planning, emergency triage, and diagnostic support in radiology or pathology. These systems operate in contexts where errors can lead to serious harm or death, requiring rigorous oversight, real-time auditing, and human-in-the-loop validation. Medium-severity use cases involve systems that support clinical decision-making or workflow efficiency, such as drug dosing recommendations, disease progression predictions, or care coordination tools. While not directly life-threatening, mistakes in this area can still have a negative impact on patient outcomes and require structured monitoring and policy alignment. Low-severity use cases focus on administrative or operational tasks, such as scheduling, billing, and patient intake. These carry minimal clinical risk and prioritize reliability, usability, and basic data protection. Categorizing agentic AI in this way ensures that safeguards are proportionate to the level of risk, supporting responsible innovation across all levels of healthcare delivery.

The implementation framework for agentic systems in clinical workflows can be broadly classified as either open or closed. In open systems, agentic AI components operate with transparent protocols, modular interfaces, and auditable decision pathways. These systems enable interoperability across clinical tools, provide real-time oversight, and facilitate integration with institutional policies. They enable hospitals and regulators to customize rules, trace decisions, and intervene when necessary, thereby supporting higher trust and adaptability. Closed systems offer quick deployment and strong performance on specific tasks, but they are often proprietary and opaque. The underlying architecture, decision logic, and data handling methods are not easily accessible or modifiable by end users. This limits visibility into how outputs are generated or updated, making it difficult to align the system with institutional policies or evolving regulations. In environments with weak governance, closed systems can increase the risk of untraceable decisions and accountability gaps.

The deployment of agentic workflows in tech ecosystems has triggered a quiet reshaping of labor markets. Tech giants have laid off thousands, citing efficiency gains from AI agents that now replace roles once held by humans. This trend is now making its way into healthcare, a sector already under strain. The global health workforce is projected to face a shortfall of over 10 million workers by 2030, according to the [World Health Organization](#). Agentic systems are being positioned as the solution, particularly for low- to medium-severity tasks such as documentation, scheduling, and diagnostics. But if healthcare workers are still expected to verify every AI-generated output, this is not progress. It is a transfer of responsibility to unaccountable systems, masking decades of neglect and underinvestment in human care. The real question is whether these tools will ease the burden on workers or boost profits for those already benefiting from a system stretched to its limit.

Policy, Governance and Regulation

The global agentic AI healthcare landscape remains largely unregulated, with policy and governance frameworks lagging significantly behind the rapid pace of innovation in this field. Regulatory institutions often lack the manpower and financial resources available to the tech giants driving these advancements, creating a dangerous imbalance. This disconnect raises serious concerns, especially in the Global South, where regulatory structures are either underdeveloped or non-existent. In many leading African nations, national AI strategies are still in their infancy and tend to exist more on paper than in practice, with little implementation, enforcement, or monitoring.

While the growing tech ecosystems in these regions are increasingly adopting closed-source AI tools to power healthcare innovations, this progress is a double-edged sword. On one hand, it promises improved healthcare access and innovation; on the other, it poses serious risks due to the opacity of proprietary models and the absence of robust regulatory safeguards. Moreover, in many low-resource countries, AI regulation in healthcare is often lumped with other sectors, resulting in vague, non-actionable language that lacks sector-specific clarity. To prevent potential harm and ensure the ethical deployment of AI, there is an

urgent need for digital governance tools that can enforce policies, monitor AI behaviour in real-time, and bridge the gap between innovation and oversight, especially in underserved regions where populations are most at risk.

The burden of accountability

Accountability is not a feature to add later. It must be part of the foundation.

Much of the excitement around agentic AI in healthcare focuses on its ability to make autonomous decisions, adapt to patient data, and streamline clinical workflows. However, there is far less attention paid to a critical issue: accountability. As these systems become more involved in sensitive and personalised care, we must ask who is responsible when they make mistakes. If an AI agent misdiagnoses a condition, overlooks a critical symptom, or recommends an unsafe treatment, who should be held accountable? Is it the hospital that deployed the system, the company that developed it, or the vendor that adapted it for clinical use? These systems often continue to learn and evolve after deployment, which makes their actions difficult to trace and predict.

In high-stakes environments, such as healthcare, this uncertainty poses a significant risk. Without clear guidelines, hospitals may bear the consequences of failures they did not create, while developers and vendors avoid responsibility by pointing to the complexity of the models. Holding AI agents accountable requires more than technical audits or vague oversight. We need clear legal frameworks that define liability, systems that log decision-making steps, and governance structures that support shared responsibility. Developers must be required to provide transparency and ensure their models are safe and interpretable before deployment. Hospitals and care providers must retain the authority to override AI decisions when necessary. Accountability is not a feature to add later. It must be part of the foundation. Trust in agentic AI will depend not only on how well it performs, but also on how clearly we can understand its actions when things go wrong.

Companies that embed policy compliance, governance observability, and system interoperability from day one will be best positioned for market access and long

term defensibility

As healthcare AI systems become more autonomous and agentic, their increased complexity introduces new challenges related to safety, oversight, and accountability. Drawing inspiration from the airline industry, where system failures are treated with urgency, transparency, and procedural rigour. Healthcare stakeholders must adopt a similarly mature and structured approach to AI reliability in the following ways:

- **Real-Time Failure Reporting:** In aviation, every critical failure is logged, investigated, and reported in real time. This culture of incident reporting is embedded into aircraft systems and reinforced by strict regulatory oversight. Agentic healthcare systems need the same philosophy. AI failures, whether due to algorithmic drift, data mismatch, or coordination errors between agents, must trigger automated logs, notify relevant stakeholders, and be fed into an incident registry that regulatory bodies can monitor in real-time.
- **Local Operating Systems for Decentralized Control:** To support scalable reliability, healthcare institutions should run decentralized AI agent operating systems. These act as local execution environments, managing all agent activity at the provider level. Each system enforces local interoperability, orchestrates multiple agent modules, and applies custom policies aligned with clinical workflows. This decentralized model reduces the risk of centralized bottlenecks and supports local accountability while maintaining a national or regional standard.
- **Built-In Safeguards for Data Privacy and Leakage Prevention:** Every local agent operating system must prioritize data security by design. This includes isolating patient-sensitive data, enforcing encrypted communication between modules, and applying appropriate safety protocols where necessary. Such safeguards are essential for maintaining public trust, especially when autonomous agents handle real-time patient information.
- **Orchestrated Agent Interoperability within Healthcare Workflows:** Similar to how cockpit systems coordinate avionics, propulsion, and navigation software, agentic healthcare systems must enable seamless coordination between diagnostic, predictive, and administrative agents. The operating system functions as the orchestrator, ensuring each module communicates effectively and respects workflow hierarchies. This coordination prevents conflicts and reduces the chance of contradictory or unsafe decisions.
- **Periodic Updates and Dynamic Policy Integration:** Just as aircraft systems are updated with new protocols and software upgrades, agentic healthcare platforms must support scheduled, verifiable updates to algorithms, operating systems, and security protocols. These updates ensure the systems stay current with evolving medical standards, cybersecurity threats, and regulatory expectations.
- **Embedded Governance and Remote Oversight:** Governance policies should be programmable and downloadable into each institution's operating system. This allows health authorities and regulators to have direct oversight of system behaviour across distributed deployments. Regulatory bodies can release periodic policy updates that are pushed to local systems, ensuring alignment without requiring complete reinstallation or manual review.

Espresso Shots

The infographic below provides summary statistics on global adoption, policy, governance and regulation in healthcare.

Figure 2: Overview of AI in healthcare, highlighting agentic AI market growth, failure rates in routine tasks, and the presence of governance structures like approval processes and AI councils. It also compares global AI legislation and underscores regional gaps between the Global North and South. The figure concludes with trends in physician adoption, industry use, and time saved through AI-powered documentation.

Crème de la Crème

This segment showcases standout companies and startups that utilise agentic AI tools to develop innovative solutions for the healthcare industry.

Nabla

Led by Alex LeBrun | Delphine Groll | Martin Raison

Providing agentic AI solutions that streamline clinical documentation across electronic health records so providers can focus on patient care. Raised \$70M in Series C funding round.

Toothy(YC W25)

Led by Johnny Chen | Matthew Kerrigan

Toothy is an AI Assistant for Dental Revenue Cycle Management (RCM), providing automations for insurance verifications, claims filing, explanation of benefits (EOB) processing and much more.

BitBoard(YC X25)

Led by Connor Jones | Ambar Choudhury

Providing agentic AI workers for healthcare operations.

Hippocratic AI

Led by Munjal Shah | Vishal Parikh | Meenesh Bhimani MD, MHA | Subhabrata (Subho) Mukherjee | Saad Godil | Alex Miller | Kim Parikh

Launched with a mission to close the gap in the worldwide healthcare staffing shortage. Recently launched an App Store for AI agents in January 2025. Raise \$53M in Series A funding in March 2024, preceded by a seed round of \$50M in the previous year.

Health Force

Led by Fadi Haddad | Juan Sebastián Suárez Valencia

Provides agentic AI automation for daily administrative tasks in hospitals.

Notable Health

By Pranay Kapadia | Adam Ting | Justin White

Providing an agentic AI workforce for administrative tasks in clinical workflows.

Amelia AI

Led by Michael Anderson | Uday Chinta

Providing industry-agnostic agentic platforms for conversational AI, knowledge management, and IT workflow automations. Their solutions cover healthcare, finance, insurance, telecommunications, and travel use cases. Business applications in healthcare cover digital health services for patients, providers and big pharma.

Inbenta

Led by Melissa Solis | Merlin Bise

Provides industry-agnostic AI solutions for knowledge management, search and enterprise workflow automation. In healthcare, these solutions power patient-centric services, documentation and predictive analytics.

Grind Articles

A closer look at the research, strategies, and challenging questions shaping policy, governance and applications of autonomous AI agents in healthcare.

- [The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity](#)
- [TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks](#)
- [Towards a HIPAA Compliant Agentic AI System in Healthcare](#)
- [Next-generation agentic AI for transforming healthcare](#)
- [LLM-based agentic systems in medicine and healthcare](#)
- [International partnership for governing generative artificial intelligence models in medicine](#)
- [Human–robot collaboration in surgery at the nexus of knowledge, agency, and ownership](#)
- [Stakeholder-centric participation in large language models enhanced health systems](#)
- [Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology](#)
- [Implementing large language models in healthcare while balancing control, collaboration, costs and security](#)
- [Generalist medical AI reimbursement challenges and opportunities](#)
- [Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis](#)
- [Large language model agents can use tools to perform clinical calculations](#)

The Last Sip

These articles and blogs offer additional context and perspectives on the ideas we explored.

- [2025 global health care outlook](#)
- [What Is Agentic AI In Healthcare And Why It's The Next Frontier For Innovation?](#)
- [The Percentage of Tasks AI Agents Are Currently Failing At May Spell Trouble for the Industry](#)
- [Inside Agentic AI: Reshaping Decisions and Orchestration in Life Sciences](#)
- [Artificial Intelligence for Drug Development](#)
- [FDA plans to roll out AI agency-wide for reviews in June](#)

Beyond the Brew

The future of health will not be built by machines alone. It will be built by people who choose to lead with intention.

The future of healthcare is being co-authored by algorithms and clinicians, shaped by lines of code and decisions made in policy rooms. As generative AI tools promise efficiency, accuracy, and scale, we must ask: Are we designing systems that truly serve human well-being, or simply deploying technology faster than we can govern it? Responsible AI in healthcare is not just a technical pursuit. It is a societal contract that demands clarity in legislation, equity in access, and intentional safeguards. Regional disparities in governance and staggering model failure rates remind us that innovation without structure can leave the most vulnerable behind.

If your work intersects with digital health, responsible AI, or the governance of emerging technologies, let's [connect](#). I am open to collaboration, editorial partnerships, and consulting engagements that challenge the status quo and shape technology with care.

If you've made it this far and you're hungry for more...

Gain early access to *AI in Healthcare — Innovations, Opportunities & Challenges*, the first release in the Research Intelligence series. This premium, research-backed document examines the tools, regulations, adoption frameworks, and global disparities that are shaping the future of AI in clinical environments.

What's inside:

- 60+ pages of expertly curated, visually engaging content
- Detailed benchmarking of AI tools and models used in clinical settings
- Global case studies from North America, Africa, Asia, and Europe
- Practical frameworks for evaluating and adopting AI in hospitals
- Policy and governance analysis of Health Canada, FDA, and the EU AI Act
- AI adoption checklist for healthcare leaders

 [Join the waitlist now to be notified before launch](#)

Next in the archive: No. 3, Digesting the Planet

Get the next CodeCaffe

One hard question about technology and society, every second Tuesday. Free, no ads.

Working on this question? [Be our Barista](#) in a future issue. Download the [PDF edition](#) or browse [every issue](#).

How this issue was made: First published on LinkedIn in 2025 and reformatted for the website in 2026. The text is as first published.