Medicine After the Model: From Language to Symbiosis

30 June 2026 · Nikon Sugar

Between 2021 and 2026 artificial intelligence in medicine crossed three thresholds in quick succession: language models that could answer, agents that could act, and the first systems built to reason alongside a clinician rather than beneath one. This essay traces that acceleration, sets out the mathematics that made it possible, and argues that the destination is neither automation nor replacement but symbiosis — with embodied, humanoid care as the next frontier. The question for the next generation of doctors is not whether the machine is good enough, but what a clinician is for once it is.

For most of its history, computational medicine was a study in narrowness. A model could read a mammogram (1) or grade a skin lesion (2) at the level of a specialist, and each such system was a monument to a single task — trained at great cost, deployed in a single clinic, blind to everything outside its frame. The promise was real and the ceiling was low. What changed, beginning around 2021, was not that the models got slightly better. It is that they stopped being narrow.

This essay is about that change and where it leads. The short version of the argument is captured in Figure 1: in five years medical AI climbed from systems that *answer*, through systems that *act*, toward systems that *co-reason* with a clinician — and the slope does not flatten where the diagram runs out of ink.

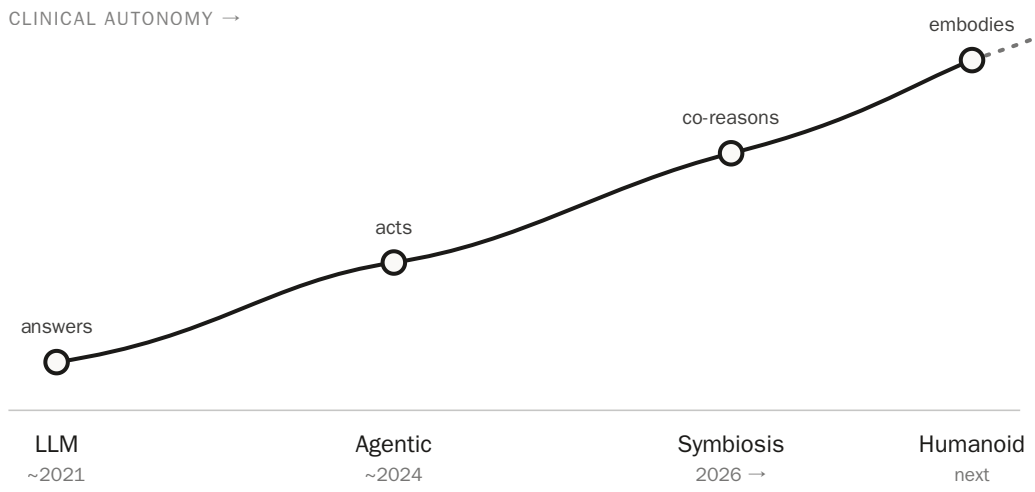
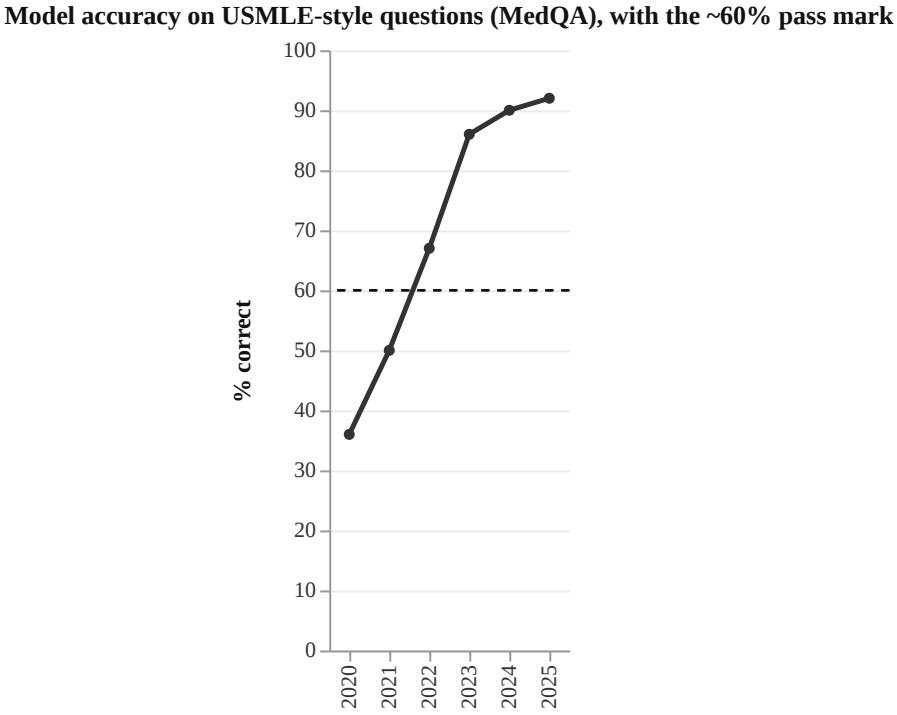


Figure 1. Four eras of medical AI on an ascending curve of clinical autonomy: language models that answer, agentic systems that act, human–AI symbiosis that co-reasons, and embodied humanoid care.

The five-year jump

It is easy to forget how fast this was. As late as 2020, a large model attempting the United States Medical Licensing Examination scored in the high thirties — below chance for a motivated student, nowhere near the ~60% pass mark. By late 2022 a medically tuned model cleared the bar for the first time (3); within a year, general-purpose systems were scoring in the high eighties on the same question banks (4), and the curve has been grinding toward its ceiling ever since.



Why so fast? The honest answer is that medicine got the spillover from a more general discovery. The architecture underneath all of this — the transformer — replaced recurrence with a single operation, attention, that lets every token weigh every other directly (5):

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{1}$$

What Equation 1 bought was not medical knowledge but *scalability* — a model whose quality improves predictably as you add parameters and data. That predictability was itself measured: across many orders of magnitude, loss falls as a power law in model size,

$$L(N) \;\approx\; \left(\frac{N_c}{N}\right)^{\alpha_N} \tag{2}$$

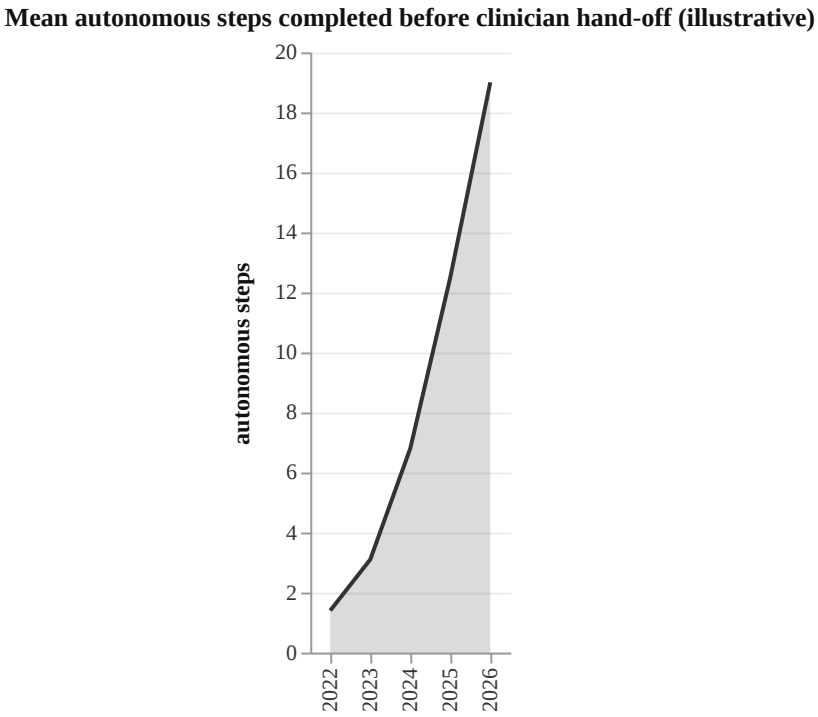
with a small exponent α_N (6). Equation 2 is the quiet engine behind the steep line above: medicine did not need a bespoke breakthrough each year, only a share of a curve that the whole field was climbing. A model trained on the open record of human language had, almost incidentally, read

enough medicine to reason about it — and a generalist substrate, fine-tuned lightly, began to outperform the narrow specialists it was supposed to merely assist (7).

From answers to actions

An examination rewards a system that *knows*. A clinic rewards a system that *does* — that orders the test, reconciles the medication list, drafts the referral, flags the interaction, and chases the result three days later. The second wave of medical AI, roughly 2023 onward, was the move from a model you query to an **agent** that pursues a goal across many steps, calling tools and reading their outputs as it goes.

The capability that defines the agentic era is not eloquence but *task length*: how much of a real workflow a system can carry before it must hand back to a human. That number has grown sharply.



This is also where the stakes change. A model that answers a question wrongly produces a bad sentence; an agent that acts wrongly produces a bad *event* — an order placed, a message sent.[1] The engineering discipline of the agentic clinic is therefore less about raw accuracy than about *bounded autonomy*: explicit tool permissions, reversible actions, and a human checkpoint placed exactly where the cost of error spikes.

What the machine changes about diagnosis

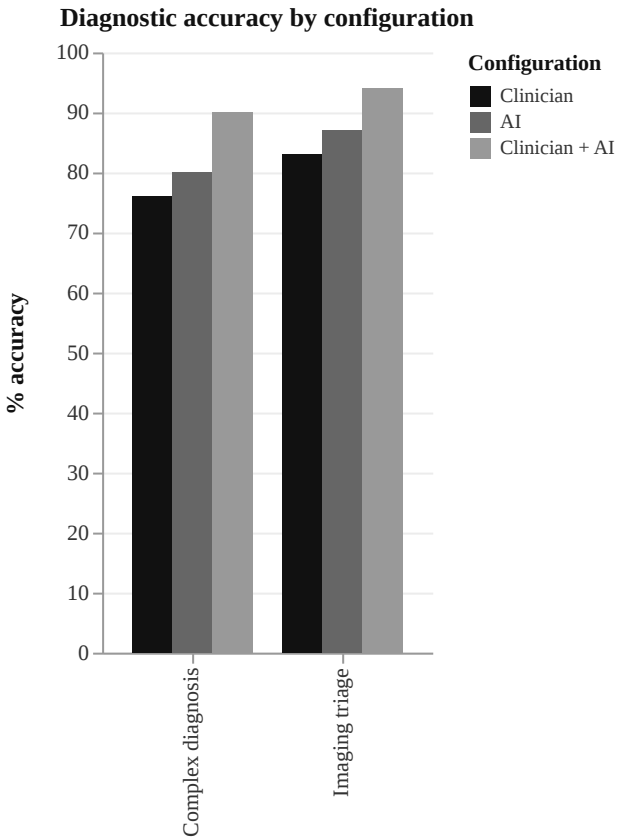
Diagnosis has always been probabilistic, even when clinicians did the arithmetic in their heads. Bayes' rule in odds form is the whole of it: a test updates a pre-test belief by its likelihood ratio,

$$O(D \mid +) = O(D) \times \mathrm{LR}_{+}, \qquad \mathrm{LR}_{+} = \frac{\text{sensitivity}}{1 - \text{specificity}} \tag{3}$$

The trouble has never been Equation 3 itself; it is that a human juggling a differential cannot hold twenty competing priors and update them all faithfully as each result lands. A language-grade model can — and, prompted as a conversational partner rather than an oracle, it can also *ask the next question*. The first systems built explicitly to take a clinical history this way have matched or exceeded primary-care physicians on diagnostic accuracy and on the softer measures of a good consultation in controlled studies (8). The lesson is not that the machine is a better doctor. It is that the machine is a tireless Bayesian, and that this is a narrow and complementary virtue.

Symbiosis, not substitution

Here is the finding that should shape the next generation of practice: across task after task, the clinician-plus-AI pairing beats either the clinician or the AI alone. The machine's recall and consistency cover the human's fatigue and anchoring; the human's context, accountability, and bedside judgement cover the machine's confident errors. Topol called this convergence years before it was deployable (9); we can now measure it.



The eras are not replacements for one another so much as accumulating layers, each adding a capability without discarding the last. Table 1 lays them side by side.

Table 1. The accumulating eras of medical AI — each layer adds a capability and redefines, but does not erase, the clinician's role.

Era	The system…	Clinical example	The clinician is…
LLM	answers	summarises a chart, drafts a note	the editor
Agentic	acts	orders, reconciles, follows up	the supervisor
Symbiosis	co-reasons	runs the differential *with* you	the accountable partner
Humanoid	embodies	examines, assists, performs at the bed	the director of care

The embodied turn

Everything so far lives behind a screen. The last column of Table 1 is where the next frontier is — and it is the one the curve in Figure 1 points toward but does not yet reach. Medicine is irreducibly physical: a pulse is palpated, a wound is dressed, a patient is turned. A reasoning system that cannot touch the world is a consultant, not a caregiver.

The humanoid turn closes that gap. Pair a model that can reason about a case with a body that can take vitals, fetch and position equipment, assist a turn or a transfer, and stand the long bedside watch that exhausts human staff, and the symbiosis of the previous section acquires hands. The near-term reality is unglamorous and valuable: the embodied system as the tireless nurse's aide and the surgeon's third arm, not the autonomous physician. The constraints are no longer mostly cognitive — they are mechanical, tactile, and above all a matter of safety in a setting that does not forgive a dropped patient.

What the next generation of practice looks like

Put the pieces together and the shape of the next clinic is legible. The model removes the documentation tax that drives clinicians to burnout. The agent carries the long, dull tail of coordination that consumes a third of a clinical day. The symbiotic system makes the clinician a better diagnostician than either party alone. And the embodied system, when it arrives, returns to the bedside the hours that paperwork and logistics had stolen from it.

None of this retires the doctor. It relocates the doctor — upward, from executor to director; from the one who remembers every guideline to the one who decides when to break one; from the scarce expert rationed across a panel of thousands to the accountable human at the centre of a system that finally scales. The machine became a tireless Bayesian, an indefatigable agent, and soon a pair of steady hands. What it cannot become is the person who sits with the frightened patient and owns the decision. That, increasingly, is the job — and it is a better one.

References

1. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89–94.
2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.
3. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172–80.
4. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on Medical Challenge Problems. arXiv preprint arXiv:230313375. 2023;
5. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. In: Advances in Neural Information Processing Systems. 2017.
6. Kaplan J, McCandlish S, Henighan T, Brown TB, Chess B, Child R, et al. Scaling Laws for Neural Language Models. arXiv preprint arXiv:200108361. 2020;
7. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259–65.
8. Tu T, Palepu A, Schaekermann M, Saab K, Freyberg J, Tanno R, et al. Towards conversational diagnostic artificial intelligence. Nature. 2025;642:442–50.
9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019;25(1):44–56.

Footnotes

1. This is why the regulatory centre of gravity is shifting from *model approval* to *workflow approval* — certifying the loop in which a model acts, not just the weights in isolation. ↩