

Why some hospitals are making their own ChatGPTs for patient records

Stanford Health Care and Penn Medicine have developed chart chat tools for doctors

By [Brittany Trang](#) – STAT News

Jan. 28, 2026

Health Tech Reporter

To make the most of his 30-minute appointments with patients, Penn Medicine Chief Health Information Officer Srinath Adusumalli goes through his patients' charts the day before to figure out why they are seeing him, a cardiologist. To do that, he has to navigate multiple tabs in the electronic health record: prior appointments, prior labs and imaging tests, as well as scanned documents from other hospitals.

Earlier this month, Penn Medicine introduced a tool called Chart Hero that aims to cut this work by letting clinicians query and summarize the patient's medical record with an artificial intelligence chatbot-like interface. Currently, around 70 Penn clinicians in different specialties are testing it out.

Stanford Health Care last year launched ChatEHR, a similar, internally built tool that integrates with Stanford's Epic electronic health record. As of mid-January, it was being used by 1,450 clinicians out of roughly 7,000 eligible users. The two health systems are not alone: Cliniques universitaires Saint-Luc Hospital in Brussels, Belgium, implemented a similar homegrown system, and teams at Duke Health and the Children's Hospital of Philadelphia are developing similar applications.

Health systems are building these tools from scratch because they are already managing too many AI tools developed by different companies, executives told STAT. Having access to a product's behind-the-scenes

tech also allows health systems to interrogate the error rate for the AI-generated summaries. But while building these tools in-house imparts the flexibility to iterate quickly and tailor for the health system's unique needs, whether such an approach is scalable is an open question.

Nigam Shah, chief data scientist at Stanford Health Care, thinks building in-house is the only way out of an increasingly complicated hamster wheel of tech solutions.

Stanford's health system already runs about 1,500 IT systems and doesn't need more to manage. "Every new application comes with its own integration challenges, identity and access management, bug fixes, updates. They all have their roadmap. And they all want us to co-innovate with them. It is utter madness," Shah said. "If you want agency and [to] not be buffeted around by the commercialization aims and payback plans of all of these entities that have raised VC money, we've got to take the steering wheel in our own hands." Shah is also the co-founder of Atropos Health, a startup that uses AI to compile real-world medical evidence.

Penn's Adusumalli thinks that over time, chatbots that ease the friction of engaging with the EHR will be combined with other AI products as a unified clinical, ambient intelligence platform for clinicians. In the ambulatory setting, for example, he imagines that a chart chat solution would merge with existing ambient scribes and other products. Such a product would "understand the patient that you're about to see, help document their care, explore downstream information, including that conversation that you just had, code and bill that visit appropriately, and then follow up with the patient," he said.

But such a product didn't exist over a year ago, when Penn Medicine started building Chart Hero. At the time, all of the chart-summarization tools available were very rigid, generating only fixed summaries, said Adusumalli. To his knowledge, Ambience is currently the only vendor with a chat-like feature for exploring the clinical record, and that was only announced in August.

Plus, the team's knowledge of its health system's data has allowed it to build something useful to the team immediately, and to iterate quickly

on feedback. “We wanted to get started and start learning, versus waiting for a product to become available,” he said.

But what about less-resourced health systems? What can they do, if they don’t have the resources to build a chatbot for their EHR?

“We invented the heart transplant, and we taught a lot of people,” said Shah. He sees this as the same thing: If another hospital has the personnel, Stanford will help them figure out how to do it at their health system, which he thinks will save money compared to buying a third-party solution. “The Stanfords, Mayos, and Dukes of the world, we have the resources to prove that this is possible. We can share our recipes,” he said.

Ashley Beecy, chief AI officer at California-based Sutter Health, hesitated to say that health systems should land on the “build” side of the “build vs. buy” paradigm. “Building conversational search internally is resource intensive. It requires significant engineering and continuous maintenance,” she said in an email to STAT.

At Stanford, it took a team of three full-time employees about four months to prototype ChatEHR, and then another three to four months of piloting the product. Those three employees are maintaining the product as well. But Shah noted that those three people are embedded within a 900-person IT department and also have an ecosystem of clinicians, postdocs, and students who contributed in some manner.

Beecy also expects similar features to be built into the existing EHR software soon. “As EHR vendors and ecosystem partners mature these capabilities, I expect conversational search to become a standard EHR primitive, much like in-basket, chart summaries, or decision support, because it fundamentally changes how clinicians retrieve and act on patient information,” she said. Shah isn’t so sure that tools from EHR companies or other third parties would materialize, or be able to meet Stanford’s specific needs so easily.

“Every family has their own dinner ritual and the way they make their favorite dish — pasta, dal, whatever,” said Shah. “Yeah, you can buy it premade. But it’s a lot easier if you make your own.”

For example, Stanford has its own criteria for when a surgery patient is eligible for an internal medicine doctor to co-manage their surgical hospital stay — a program not every hospital has. With Stanford’s ChatEHR, the department can run a batch of patients through the tool to see which ones meet the criteria, instead of manually identifying which patients might benefit.

Another popular use for ChatEHR is assisting with reviewing “hospital account letters of agreement” — a negotiation between a hospital and an insurer when an out-of-network patient gets referred for care. The approval for that care requires someone to read the patient’s chart and summarize what care is needed and why — a responsibility that falls on the provider.

If tools built into Stanford’s Epic EHR are good enough and don’t cost any extra money, “we’re not going to build our own,” said Shah. “But they’re not going to build us that letter of agreement thing or the surgical co-management thing.”

In a soon-to-be-released preprint, Stanford researchers explain several reasons why solutions from Epic seem inadequate. But an Epic spokesperson said that within the next few months, the company will roll out conversational chat for its “Art” and “Emmie” clinician and patient AI models, respectively.

“Conversational queries, such as identifying patients who meet specific eligibility criteria, are available today. In addition, Art can create Inpatient and Outpatient Patient Summaries, which surface key information for clinicians’ pre-visit chart review and are now used over 16 million times each month,” said the spokesperson. They noted that other AI features similar to those in the preprint are on Epic’s roadmap as well, like AI that identifies patients eligible for transfers, and an Agent Factory that will allow Epic customers to build their own custom automations.

The Stanford study also details the results of a number of analyses Shah and his team ran on the ChatEHR tool’s first year of use. The team estimated that different ChatEHR automations would increase revenue by approximately \$6 million a year, largely by more quickly identifying which patients are eligible for transfers to lower-acuity facilities thus

freeing up beds faster. The team also estimated some modest time savings, though they couldn't fully quantify the productivity boost or the benefit to patients.

More interestingly, the team adapted a recent Stanford tool called VeriFact to judge the rate of errors in ChatEHR's AI-generated summaries. The tool checks the summary against the source record to see if the statement is supported by the underlying evidence.

The team found that on average during the pilot phase, there were 2.49 unsupported claims per summary. Of those mistakes, 0.74 were true fabrications that were unsupported by the record, while 1.75 were inaccuracies — the model getting the temporal sequence of events wrong, misattributing who did an action, or getting terms or numeric values mixed up. During a three-month deployment phase in September to early December 2025, a sampling of the conversations showed a decrease in unsupported statements — to 2.33, 0.73, and 1.60 unsupported claims, fabrications, and inaccuracies, respectively, per summary.

Until this point, there haven't been many quantified error checks for real-world applications of generative AI, whether compared to ground truth or against human error. Most health care generative AI tools rely on users catching errors and leaving feedback. "I don't know whether those [error rates] are considered high, low, good, bad, ugly," said Shah. "Nobody knows."

Yevgeniy Gitelman, head of custom software at Penn Medicine's tech innovation center and the health system's associate chief health information officer, said that this kind of quantification is very difficult to define and measure. The Penn team provides citations for the facts that appear in Chart Hero's summaries — links that go back to the progress note or imaging study or lab result the fact originated from. The team is also looking at implementing a tool that will check the answer and verify that the fact is supported by the citation. But adding that verification agent would increase the response time for the tool, which might create a barrier to adopting the tool, said Gitelman.

Chart Hero and ChatEHR have undisputable benefits: helping prep a clinician for a patient that signed up for an appointment half an hour before it starts; letting inpatient clinicians quickly ask who placed a

patient's tube or catheter; surfacing information from further back in a patient's history than the doctor would have looked by themselves; giving imaging specialists looking at scans back-to-back the chance to quickly get context on why a patient needed a scan.

But as health systems weigh those benefits alongside guardrails like how to best train the workforce to use AI and optimizing the human-AI interaction, Gitelman still doesn't know when he'd feel great saying that people don't have to double-check the output. "And it's not just with Hero," he said. "I think this is a general generative AI challenge."