THE ARISING

Clinical AI

Better Information, Better Care?



NOVEMBER 2026

Clinical AI: Better Information, Better Care?

Patient conversations, overnight heart signals, and the evidence needed to connect them with better decisions.

The Arising · November 2026

A good medical appointment begins with a person trying to explain what is wrong. Sometimes the story comes easily. Sometimes an important detail arrives on the way home: a symptom that seemed too small to mention, a medicine left off the list, a question that disappeared when the conversation moved on.

Artificial intelligence might help organize some of that information. It might also find patterns in measurements that already exist. Both possibilities are worth exploring. The useful question is what happens next: whether a clinician receives a clearer history, whether a risk estimate leads to an appropriate investigation, and whether the person receiving care is better off.

Two recent studies approach that question from different directions. One tested supervised AI conversations before urgent primary-care visits. The other examined cardiovascular prediction using overnight sleep-study recordings. Neither establishes that an AI system can independently take responsibility for care. Together, they help explain what a promising beginning looks like, and what still needs to be demonstrated. [1,2]

Before the appointment

Imagine having time to describe a concern before arriving at the clinic, with prompts that help organize the sequence of events. A record of that conversation could give the clinician a starting point. It could also contain errors, missing details or an emphasis that does not match what matters to the patient.

The AMIE study investigated an early version of this idea. AMIE stands for Articulate Medical Intelligence Explorer. English-speaking adults interacted with the system up to five days before an urgent primary-care appointment for a single complaint. Physician safety supervisors monitored every interaction and could intervene using predefined criteria. Transcripts and summaries were shared with the primary-care clinicians. [1]

This supervision belongs in any account of the results. The researchers tested a system that included people watching for problems. Removing them would change the intervention.

Of 114 enrolled patients, 98 completed both the AI interaction and the clinical appointment. None of the conversations reached the prespecified threshold for a safety stop. Supervisors nevertheless recorded one hallucination and supplied additional clinical information in five interactions. A conversation can avoid a stopping rule while still needing correction. [1]

That distinction is useful beyond this particular system. A safety measure answers the question it was designed to answer. It cannot cover every possible way a conversation could confuse someone, omit information or influence a later decision.

What the clinicians thought

Clinicians completed post-visit surveys in 60 of the 98 cases. Of the 60 surveyed cases, 44 involved a clinician who reviewed the transcript before the appointment. Clinicians reported that it helped preparation in 33 of those 44 cases and might have changed their behavior in 25. [1]

Those reports are encouraging, but they belong to a particular subset. It would be misleading to describe them as the views of all participating clinicians across all 98 completed encounters. It would also go too far to turn “helpful for preparation” into proof of better diagnosis or recovery.

The study was prospective, meaning the workflow was followed as it happened. It was conducted at one center, with one study group and no randomized comparison against usual care. Its main purpose was feasibility: could this supervised approach function in a real clinical setting, and how did participants experience it? [1]

A next study could ask whether a comparable group receiving ordinary preparation had different outcomes. It could measure clinician time, missed information, unnecessary follow-up, patient understanding and diagnostic accuracy. Some benefits might come with new burdens. Someone has to review the output, resolve discrepancies and decide which details need attention.

The opportunity is a better encounter, not simply a longer transcript.

Information that is already being collected

The second study concerns a very different kind of information: electrical heart recordings obtained during overnight sleep testing.

Researchers combined single-lead ECG recordings with sleep-stage information. The model was trained using 15,809 patients at Massachusetts General Hospital and externally tested in cohorts of 9,810 at Emory and 12,576 at Beth Israel Deaconess. These are cohort totals, not the number contributing to every outcome analysis. [2]

The study examined 10-year risks of atrial fibrillation, stroke, heart attack, heart failure and death, with outcomes identified from diagnosis codes in electronic health records. These outcomes should not be treated as interchangeable. The primary record reports associations of differing strength, including a heart-attack estimate in one validation cohort whose confidence interval included no association. [2]

An overnight hospital sleep study also differs from a watch recording during an ordinary day. This research does not validate an interchangeable consumer-device score. Its inputs included sleep stages, not ECG alone. [2]

The possibility is attractive: information collected for one clinical purpose might help with another. Whether that additional information should change care is a separate question.

A forecast needs a useful response

Consider an illustrative forecast that labels a person “higher risk.” Before it can guide a decision, several questions need answers. Higher than whom? Over what period? For which condition? How uncertain is the estimate? What happens if the forecast is wrong?

A model might rank people correctly while giving poorly matched absolute probabilities. For example, a group described as having a one-in-five chance of an event might experience something very different. Checking how predicted probabilities compare with observed frequencies is called calibration. This explanation is a general statistical principle, not a new performance claim about the sleep-study model.

There is another step beyond prediction. A clinician could respond to a high score with additional testing, closer follow-up or a conversation about established risk factors. Each option has costs and potential consequences. Research would need to show that the proposed response produces enough benefit to justify unnecessary investigations, anxiety and missed cases.

A statistical association cannot supply that answer by itself. Nor does testing a model at a second hospital automatically establish that acting on it improves patients' lives. External validation is valuable because a model can behave differently outside its development setting. Clinical usefulness still requires a suitable decision and evidence about what follows.

For the reader, a sensible question is: “What would we do differently because of this result, and what evidence supports that change?”

Three ordinary situations

The following examples are illustrative. They are not participants from either study or descriptions of an existing Arising clinical service.

A symptom has been summarized incorrectly

A patient reviews a previsit summary and notices that an intermittent problem is described as constant. The useful feature is an easy way to correct that statement and make sure the clinician sees the correction. A polished summary is not enough if the original account cannot be checked.

Questions worth asking include whether the patient can read the summary, how corrections reach the clinical team, and who reviews information that could change the urgency of care. A patient should not assume that typing something into a system means a clinician has already seen it.

A sleep report contains a new risk score

A reader finds an unfamiliar cardiovascular estimate beside the usual sleep-study results. The first step is understanding the report: what it predicts, whether the estimate is validated for a similar population, and whether the clinician would recommend anything different after considering the whole history.

The score should not be treated as a diagnosis or as a reason to alter medication independently. It can be a starting point for a clinical discussion when its purpose and limitations are clear.

A clinic is considering a new tool

A clinician asks whether an AI interview could make visits easier. An evaluation could begin with a carefully defined task, such as organizing a history for review. The clinic would need to know how much checking is required, where mistakes occur and whether patients can use the system comfortably.

Success should include the experience of people who decline, cannot use the interface or need language support. A study in English-speaking adults does not settle those questions for everyone else. The people who have the most difficulty with a tool should not disappear from the assessment of its value.

Responsibility should remain visible

For Arising Health, the original proposal was a supervised information-gathering workflow evaluated for completeness, factual accuracy, clinician burden and patient understanding. That remains a proposal for research. The AMIE results do not validate an Arising platform, and this magazine does not announce that such a clinical service has been implemented.

Any proposed system should make its responsibilities legible. Who monitors it? Who acts on an urgent concern? How does a patient reach a person? Can the clinician inspect the information behind a summary or estimate? What is retained, and who can access it?

These are design questions as much as technical ones. They should be answered before people rely on a tool, rather than left for patients to discover when something goes wrong. An educational chatbot should not be assumed to provide the supervision described in a clinical study.

If symptoms suggest an emergency, seek urgent medical care rather than waiting for an AI response or a routine appointment.

What would make the evidence stronger?

A useful next step for conversational systems is a comparison with ordinary care that measures the full encounter. Better preparation is valuable, but it should be assessed alongside errors, workload and what patients understand afterward.

For prediction systems, the next questions concern how estimates travel between populations and how clinicians should respond. A forecast may be informative without yet supporting a new care pathway. A future study could test that pathway directly rather than assuming that improved prediction must produce improved health.

Both kinds of research need room for revision. A correction should be visible, and an uncertain result should remain uncertain when it is translated into a headline or a patient conversation.

AI's place in medicine will be earned through work like this: a specific task, a defined setting, an honest account of mistakes and a better question for the next study. The aim is to help people tell their stories and help clinicians use information well. That is a concrete ambition, and a worthwhile one.

Sources and further reading

1. Brodeur PG, Koshy JM, Palepu A, et al. Conversational diagnostic artificial intelligence in ambulatory primary care: a prospective feasibility study. *The Lancet*. 2026;408(10565):1732–1744. DOI: 10.1016/S0140-6736(26)01535-7. Primary publication record. Alphabet funded the study.

2. Koscova Z, Nasiri S, Reyna MA, et al. Prediction of cardiovascular outcomes using electrocardiography from overnight polysomnography. *Sleep*. Published online October 9, 2026:zsag229. DOI: 10.1093/sleep/zsag229. Primary publication record · Publisher article.

The sleep study's funding statement identifies the US National Heart, Lung, and Blood Institute; its disclosures identify an author's advisory, consulting and equity relationship with Beacon Biosignals. This is a focused educational account, not a systematic review or individualized medical advice.

Continue exploring

Wearables and health signals · Virtual cells and clinical AI · Sleep and circadian health