

AI Is Already Here

Large Language Models, Agents, and the
Future of Psychiatric Practice

Mina Boazak, MD, MMCI

NCPA Annual Meeting · September 26, 2026

animosanopsychiatry.com

About me and disclosure

Mina Boazak

MD, MMCI

Founder/CEO and Medical Director
Animo Sano Psychiatry

Medical Assistant Professor
Duke Department of Psychiatry
and Behavioral Sciences

Mina Boazak, M.D., has no relevant
financial relationships to disclose.



Four agents working during this talk

- 01 Prepare an administrative decision and draft reply
- 02 Draft and review a note from a fictional encounter
- 03 Research a clinical technology question
- 04 Build a small clinician application

All patient and practice data in this demonstration are synthetic.

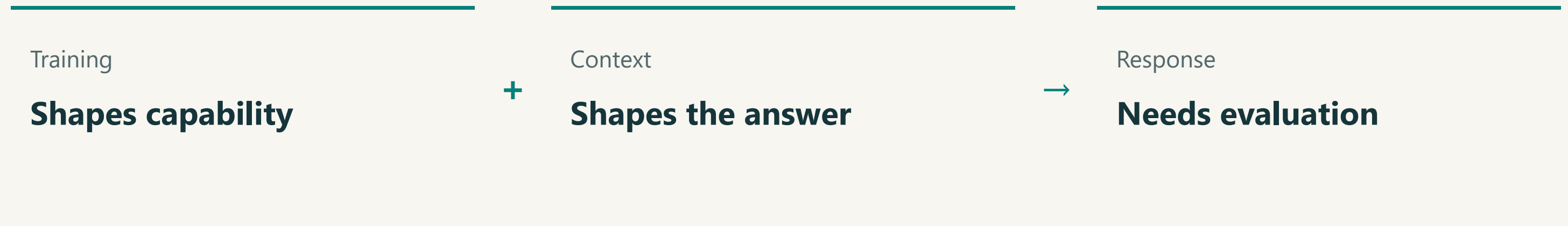
Why learning the tools matters

We cannot reliably predict every job outcome.

We can learn to use, evaluate, and supervise the tools shaping our work.

What an LLM does

A large language model learns patterns from data and generates responses using the context available to it.



Neither training nor context guarantees truth.

What makes an agent different

An agent pursues a goal through repeated actions.

Read
information



Choose and
use tools



Check
results



Continue or
ask for help

Access and permissions determine what it can change.

A larger range of work

Text and documents

Images and audio

Research and data analysis

Software creation

Connected workflows

One encounter, several useful tasks

BEFORE

Prepare the chart

Live: Ambience Patient Summary at St. Luke's.

DURING / AFTER

Draft the note

In clinical use: ambient scribes, including tools tested in the 238-physician trial.

BETWEEN VISITS

Follow up and escalate

Deployed: UHS / Hippocratic AI discharge calls.

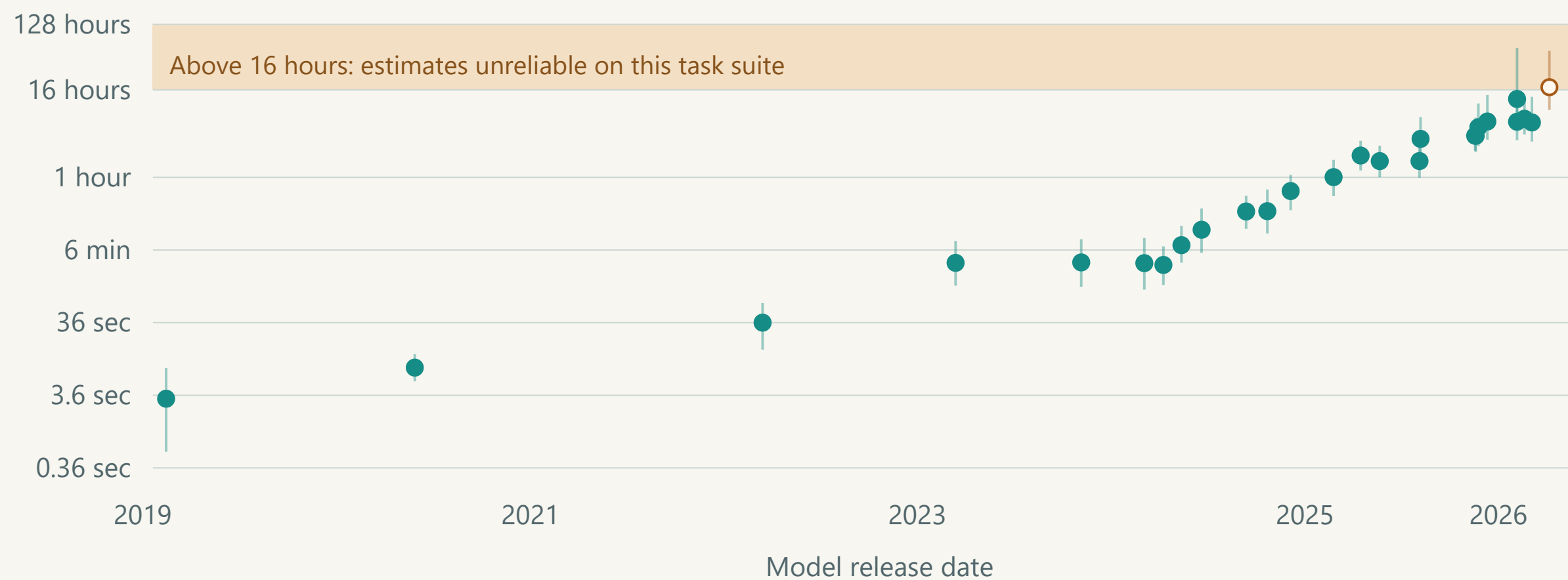
Clinicians confirm the facts, interpretation, and proposed clinical actions.



AI-generated illustration

METR: the length of tasks agents can complete

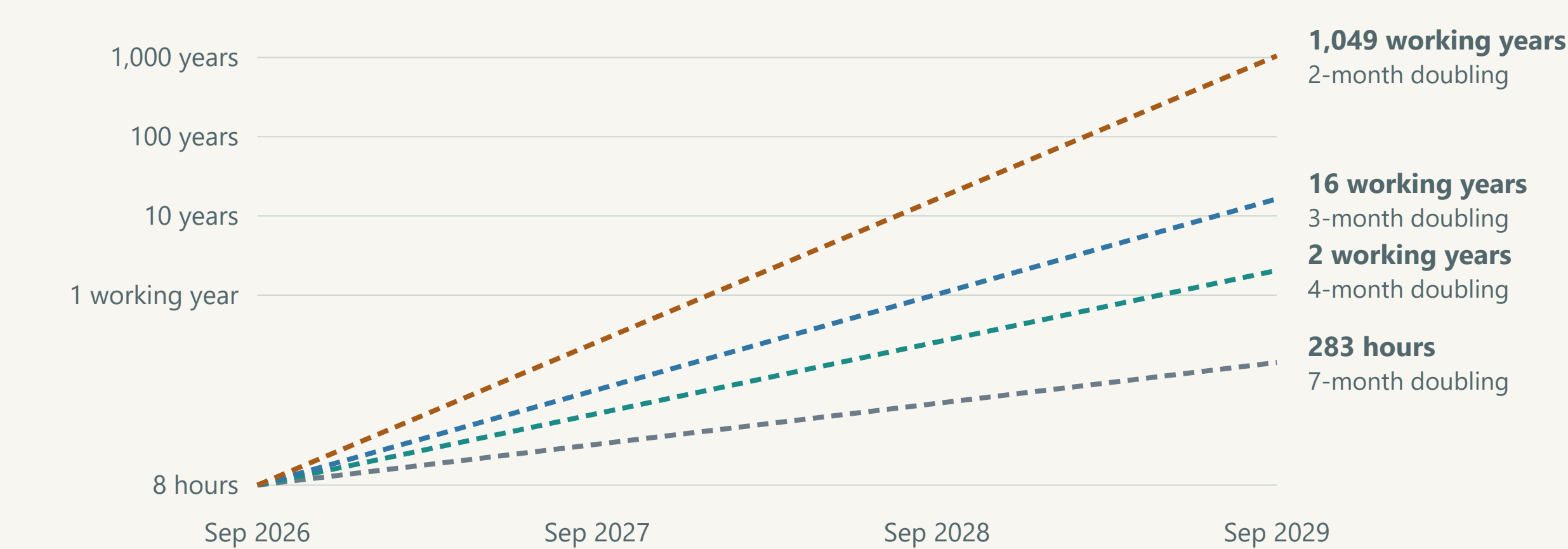
Human task duration at a **50% agent success rate**



Software, machine learning, and cybersecurity tasks · Logarithmic scale · Bars: reported 95% confidence intervals

What continued growth would imply

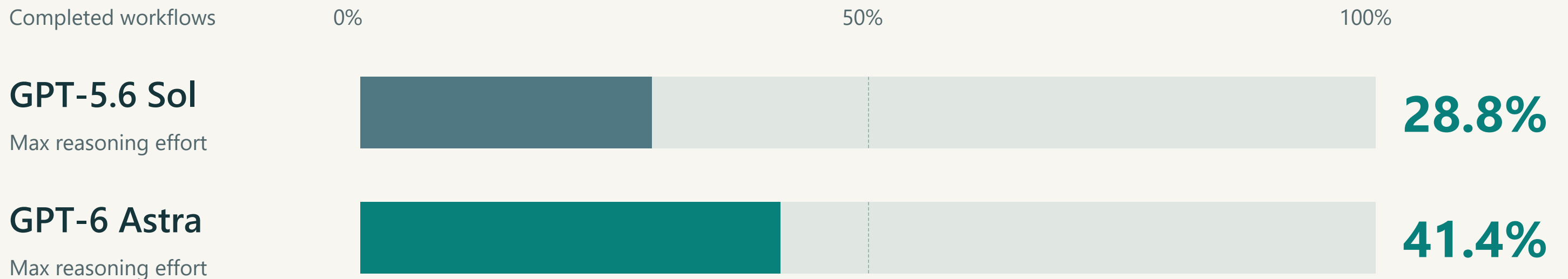
Illustration: **8 human hours** in September 2026 → three years of constant doubling



Scenario arithmetic, not a METR forecast. One working year = 2,000 hours. Extrapolates far beyond measured tasks.

Can an agent finish the job?

Find the right record → follow the policy → update the right systems.



AutomationBench grades the final state of a simulated business workflow. Success requires the correct actions across connected applications.

Human baseline: not reported. The 100% endpoint means every task passes; it is not a measured human score.

Work AI systems are already doing in 2026

INSURANCE COMPANIES

Support member service

Humana / Google Cloud

Call summaries and benefits retrieval for human advocates.

Deployed staff assistance

HEALTH SYSTEMS

Contact patients

UHS / Hippocratic AI

Discharge calls, symptom questions, and clinician escalation.

Deployed patient outreach

EARLY-ADOPTER PRACTICES

Handle routine calls

Main Street Medical / healow Genie

Scheduling, refill requests, and after-hours calls.

Live practice deployment

AMBIENT SCRIBES

Draft clinical notes

Nabla / DAX Copilot

Turn encounter conversations into notes for clinician review.

In clinical use; randomized trial

PRESCRIBING TOOLS

Renew medications

Doctronic / Legion Health

Selected existing prescriptions within Utah’s restricted protocols.

Limited, supervised state pilots

MENTAL-HEALTH CHATBOTS

Provide therapeutic support

Wysa: available wellbeing support.

Therabot: treatment studied in a randomized trial.

Deployed support / research intervention

2027–2029: from tasks to running care workflows



AI-generated illustration

Run the routine episode

Intake, record synthesis, draft assessment and plan, documentation, scheduling, and follow-up.

Manage a panel continuously

Repeated outreach, symptom tracking, care-gap closure, and escalation of exceptions.

Operate the practice around it

Carry authorizations and appeals to resolution; maintain analyses and internal software.

Likely direction: agents do most routine digital work; clinicians direct care and handle consequential decisions.

What companies are trying to build next

01 **Abridge + Availity**

Connect the clinical conversation to prior-authorization requirements, review, and submission.

02 **Hippocratic AI**

Coordinate teams of voice agents around ongoing care goals.

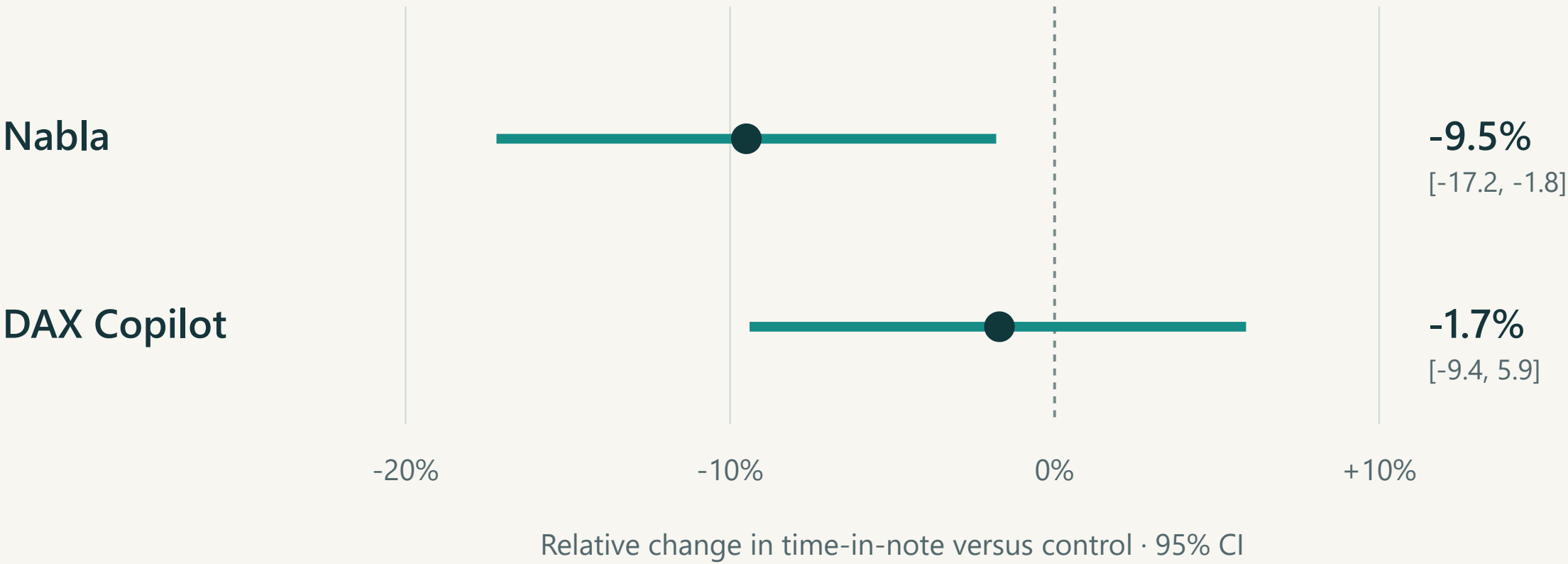
03 **Lotus**

Build an AI-led entry point to medical care backed by human physicians.

The ambition is a continuous service that carries work forward across encounters.

Ambient scribes: measured benefit, variable results

Randomized trial · 238 outpatient physicians



Clinicians reported occasional inaccuracies. Documentation time and patient benefit are different outcomes.

Model knowledge does not guarantee useful assistance

1,298

participants in a randomized study

2026 study · models tested at study time

Model given a complete scenario

Strong performance

Person using the model

**No improvement in decisions
about what care to seek**

The interaction needs evaluation too.

Therapeutic AI: early clinical evidence

210

adults in the Therabot randomized trial

A specialized therapeutic chatbot.

106

Therabot

104

Waitlist control

Symptom improvement versus waitlist.

Four-week treatment phase;
follow-up through week eight.

Results apply to the tested intervention and population.

Insurers are already offering AI-supported care

HEALTH INSURANCE

Highmark + Sword

Virtual physical care offered through health benefits, including expansion to fully insured and individual ACA members.

AI-guided exercises and feedback
combined with a clinical team.

LIFE-INSURER WELLNESS BENEFIT

MassMutual + Wysa Assure

Free AI-guided mental-health support for eligible policyowners.

On-demand conversational support
offered through a wellness program.

Payment and distribution channels for AI-assisted care already exist.

What this could change about care

VERIFIED PAYMENT SIGNAL

CMS ACCESS + payer alignment

Technology-supported chronic care, including depression and anxiety.

Ten-year test began July 5, 2026. Major payers pledged aligned options by January 1, 2028.

LIKELY IMPLICATIONS

Care becomes more continuous

Plans can offer AI-assisted services directly; clinicians may receive patients after substantial digital care.

Outcome-based payment could favor routine AI follow-up with human care for complex needs.

For practices: the referral pathway, service mix, and unit of payment could all change.

Utah: two AI medication-renewal pilots

DOCTRONIC

Approved-formulary renewals

Renew existing prescriptions within an approved formulary.

Phased physician oversight and escalation; later protocol phases permit direct renewal submission.

LEGION HEALTH

Psychiatric medication renewals

Stable patients on existing non-controlled psychiatric medications.

No diagnosis or new treatment. Risk findings trigger clinician review, with phased oversight and reporting.

Shared limits: no new medications, dose changes, or controlled substances.

What does “insured AI doctor” mean?

Health insurance

Who pays for a covered service?

Authorization

What activity may the service perform?

Malpractice insurance

Who covers specified liability?

Clinical evidence

How well does it work?

AI use belongs in the clinical conversation

"Have you used an AI chatbot to discuss your symptoms or treatment?"

"What did it suggest?"

"Did it change what you did or how you felt?"

Psychiatric risks extend beyond wrong facts

- 01 Agreement that reinforces a harmful belief
- 02 Unsafe responses to crisis or vulnerability
- 03 Stigma and inconsistent treatment of users
- 04 Emotional dependence or misplaced trust
- 05 A confident answer that misses the clinical context

A fair comparison includes existing care

For the same population and task:

01 What benefit changes?

02 What harm changes?

03 What human effort remains?

04 Who receives less access or worse results?

Scale magnifies both benefit and harm.

What meaningful review looks like

- 01 Facts linked to their sources

- 02 Missing information made visible

- 03 Proposed actions separated from completed actions

- 04 A named reviewer with time and authority to intervene

- 05 Errors recorded and used to improve the workflow

What the agents completed

Task	Time and cost
Administrative task	Not yet measured
Encounter note	Not yet measured
Evidence brief	Not yet measured
Clinician application	Not yet measured

Jobs, training, and the distribution of gains

Tasks and staffing may change at different speeds.

Better access

Less burden

More throughput

More capacity creates choices about how the gains are used.

Training still needs opportunities to build judgment.

A practical first month

- 01 **Week 1** · Try one low-risk task in an approved tool

- 02 **Week 2** · Improve the context and instructions

- 03 **Week 3** · Compare quality and review time

- 04 **Week 4** · Review your successes and failures and repeat

Living fully in a changing world

AI VOICE / AVATAR DEMONSTRATION

“let that bomb when it comes find us
doing sensible and human things—
praying, working, teaching, reading,
listening to music...”

C. S. Lewis
“On Living in an Atomic Age” (1948)

The demonstration voice and avatar are synthetic.



AI-generated illustration

Questions

What would you delegate first?

What evidence would change your view?

Where do you want the profession to set boundaries?

The capabilities deserve
serious attention.

Their limits require
serious evaluation.

Learning the tools gives clinicians
more influence over their use.

animosanopsychiatry.com

CMS, FDA, state authority, and payment

Arrangement	Question it addresses
CMS ACCESS	How can a participating organization receive payment for technology-supported care?
FDA TEMPO	Under what pilot conditions may certain digital devices receive enforcement discretion?
Utah mitigation agreement	Which activities receive specific, temporary state regulatory relief?
Health plan contract	Which services does this plan cover, under what conditions?
Liability policy	Which claims and losses does the policy cover?

Early Doctronic results and their limits

Utah public report · May 19, 2026 · Phase 1



91% First-reviewer agreement *within recommended renewals.*

Company physician reviews · Limited sample · Overall sample count not stated

Agreement is not a patient-outcome measure.

How to read the thousand-year scenario

$$H = 8 \times 2^{36 / d}$$

H = human task hours
d = doubling interval in months

At d = 2: 2,097,152 hours
÷ 2,000 = 1,048.576 working years

- 01 Starting horizon: 8 human hours
- 02 Elapsed time: 36 months
- 03 Constant doubling interval
- 04 Working year: 2,000 hours
- 05 Same task concept beyond the measured range

This does not estimate agent wall-clock time or clinical readiness.

Specialized clinical tools and general-purpose models

JUNE 2026

Three general-purpose frontier models outperformed two specialized clinical tools on the study's evaluations.

SEPTEMBER 2026

A published critique questions the reach of the conclusions.

Selection also requires workflow, privacy, safety, and support evidence.

A useful instruction identifies...

01 The goal

02 The available information

03 The output and acceptance criteria

04 The permitted actions

05 When to ask for help

A useful answer to the strongest safety objection

The failure deserves investigation.

The comparison needs a defined intervention and population.

The response should change the system, not merely add a disclaimer.