

AI HEALTH WEEKLY

What the week's evidence supports, and what it does not.

THE PRESENTING WEEK

The week the evidence asked harder questions

A diagnostic model beat doctors on paper, a meta-analysis showed humans rarely correct a wrong machine, and the FDA's first generative-AI docket closed under a pile of letters.

ASSESSMENT

The strongest items this week cut in opposite directions. A generalist diagnostic AI posted its best benchmark yet. The week's best oversight evidence found the odds of a correct clinical decision about ten times higher when automated advice was correct than when it was wrong.

FINDINGS OF THE WEEK, GRADED BY EVIDENCE

BENCHMARK

A diagnostic AI out-scores doctors on the hardest written cases

Microsoft's MAI-DxO, published in Science, solved 83% of NEJM clinicopathological cases against 36% for experienced physicians - written puzzles, not patients.

Why it matters: the strongest public benchmark yet - and still a benchmark.

Source: Science, doi:10.1126/science.aeh7930

POLICY

The FDA's inbox closes with more than 500 opinions inside

The comment window on the generative-AI device discussion paper (docket FDA-2026-N-7874) shut on 19 October; the AHA, ATA and AMA all filed.

Why it matters: the first US rule sketch for general-purpose medical AI now has a clock on it.

Source: AHA comment letter to FDA

SYSTEMATIC REVIEW

Human oversight catches fewer errors than assumed

An npj Digital Medicine meta-analysis found the odds of a correct clinician decision about ten times higher with correct than incorrect automated advice (OR 10.19, 95% CI 7.43-13.98); in four heterogeneous real-world ECG studies, 26.6% of erroneous computer interpretations were never corrected (95% CI 6.4-65.6%).

Why it matters: oversight must be measured as error interception, not assumed, the authors argue.

Source: npj Digital Medicine

PROSPECTIVE

Blind-sweep ultrasound generalises to new populations

Google's AI-guided gestational-age assessment from novice-acquired 'blind sweep' ultrasounds held up across new cohorts, reports JAMA Network Open.

Why it matters: prenatal imaging where no sonographer exists moves closer.

Source: JAMA Network Open

INSIDE THIS ISSUE

- 2 Editorial: the test set and the bedside
- 3 The deep read: 500 comments later
- 3 Mechanism of action
- 4 The radar: what we follow, and where it stands
- 5 Rapid rounds, by domain
- 5 Vital signs and follow-up

EVIDENCE KEY

- PROSPECTIVE real patients, planned
- FEASIBILITY small, no comparison arm
- BENCHMARK tested on cases, not care
- RETROSPECTIVE past data, reviews
- POLICY / PRESS rules and reports
- COMPANY self-reported, unchecked

10×

higher odds of a correct decision when automated advice was right than when it was wrong: the week's most uncomfortable number

QUOTE OF THE WEEK

“Gen-AI policy is a White House priority.”

FDA CDRH DIGITAL HEALTH LEAD, OCTOBER BRIEFING

NEXT ISSUE • SUNDAY, 1 NOVEMBER 2026

The test set is not the bedside

BY DR. HEMANT L. PATEL

Eighty-three per cent against thirty-six. On the hardest written diagnostic cases in medicine, a machine now out-performs experienced physicians by more than two to one. Published in *Science*, the result is real, and it is the strongest public benchmark a generalist diagnostic model has posted.

Now place beside it the week's other number. Across the studies pooled in *npj Digital Medicine*, the odds of a correct clinician decision were about ten times higher when automated advice was correct than when it was wrong. In four real-world ECG studies, more than a quarter of erroneous computer interpretations were never corrected. The first number says the models are getting good on paper. The second says the safeguard we cite most is the one we measure least.

The distance between the two numbers is the distance between the test set and the bedside. A

written case waits politely, contains every clue, and carries no consequence for a wrong answer. A patient does none of these things. Benchmarks like this week's are valuable precisely because they are clean. The mistake is to read them as evidence about care.

None of this argues for slowing the models down. It argues for speeding the measurement up. If a tool performs at superhuman level on curated cases, the urgent question is how it performs in a noisy clinic with a tired human reading its output — because that pairing, not the model alone, is what treats patients.

The FDA closed its first generative-AI comment docket this week and signalled draft guidance within months. The letters it received will say many things, but the week supplied the agency's hardest problem for free: how to clear tools that ace the exam, knowing the exam is not the patient.

WHAT TO WATCH

- Does the FDA's promised draft guidance adopt the 'evaluate like a physician' framing - and who carries liability when it does?
- Independent replication of MAI-DxO on real clinical cohorts rather than written cases.
- Whether the Mosaic petition on FDA jurisdiction over vision-language models gets an early answer.
- India's proposed public register of AI health systems: does it move from proposal to rule?
- Johns Hopkins Bayview's remediation - and whether other systems disclose similar AI data incidents.

FROM THE EDITORIAL

"The safeguard we cite most is the one we measure least."

HOW THIS ISSUE WAS MADE

- Every item is drawn from the six daily editions of AI Health Times published 19-24 October 2026, with fresh primary-source checks on 25 October.
- Each item carries an evidence grade: PROSPECTIVE, FEASIBILITY, BENCHMARK, RETROSPECTIVE, POLICY, COMPANY or PRESS-REPORTED. Company-reported figures are labelled as such and are not audited outcomes.
- Facts appear once. Where a claim rests on a secondary report, the item says so.

THE DEEP READ • STORY OF THE WEEK

Five hundred comments later: what the FDA heard

The first US docket on generative-AI medical devices closed with hospitals, doctors and telehealth on record, and the agency's digital-health lead promising draft guidance in months.

WHO SAID WHAT

The American Hospital Association asked for risk-tiered oversight and clarity on where hospital-tuned tools sit. The American Telemedicine Association argued for room to deploy under supervision. The AMA set its four-question governance test for clinical agents. Underneath them all sat the unanswered Mosaic petition: whether vision-language models that merely inform doctors are devices at all.

WHAT IS CONFIRMED

The window on docket FDA-2026-N-7874 closed on 19 October with more than 500 comments filed. The CDRH's digital-health lead said at a briefing that generative-AI policy is a White House priority, that draft guidance is months rather than years away, and that such systems should be assessed the way a physician would be - on performance, not on resemblance to old devices.

WHAT IS UNSETTLED

Jurisdiction remains the live wire: nothing the agency said this week answers the petition's practice-of-medicine argument. Accountability is unsettled too - several

Source: AHA comment letter on docket FDA-2026-N-7874

THE ARGUMENT IN ONE LINE

"The gap between correct and incorrect automated advice makes error interception a clinical outcome to measure, not a safeguard to assume."

THE PAPER TRAIL, IF YOU READ ONE THING

- The docket itself, FDA-2026-N-7874, for the raw comment record.
- The AHA's filing, the fullest provider-side brief.
- The AMA's four-question framework, the profession's yardstick.
- The Mosaic citizen petition, the jurisdiction challenge underneath everything.

MECHANISM OF ACTION

From docket to document

A closed window starts the slowest machine in regulation: staff sort every submission, draft guidance may follow - itself open to comment - then final guidance, which still binds no one. A binding rule needs the full notice-and-comment cycle. Months, at best.

Locked versus adaptive models

A locked model is frozen at clearance; any change needs a new submission. An adaptive model can keep learning, which current device law barely accommodates. Predetermined change-control plans are the proposed bridge - pre-agreed rules for how a model may change without fresh review.

comment letters put the burden on providers who deploy tools, which hospitals called unworkable without vendor disclosure. And the Medical Device Development Tools programme's new analysis of agentic systems is a framework, not a rule.

WHY IT MATTERS NOW

Every stakeholder now agrees on the destination - some form of lifecycle oversight for models that change - and on nothing else. The comment record is the raw material from which the draft guidance will be built, and this week's filings are the map of the coming fight: hospitals want safe harbour, vendors want the exemption, doctors want disclosure.

THE VIEW FROM INDIA

India is solving the same problem from the other end. Rather than one agency's docket, the week brought a proposed public register of AI health systems, fresh MeitY engagement on AI governance, and ICMR funding seven institutions to build indigenous devices and diagnostics - building the state capacity first, and the rules alongside it.

Special controls and the De Novo path

A genuinely novel low-to-moderate-risk device with no predicate goes De Novo: the agency writes 'special controls' - performance tests, labelling, monitoring - that future devices can then claim as a predicate. Generative devices have no such class yet; the first one through defines the mould.

Predicate creep

Each 510(k) clearance cites an earlier device as predicate, stretching the comparison a little further each time. Over a decade, a chatbot can descend from a filing cabinet. Critics say generative AI is where the chain stops.

What we are following, and where it stands

Updated every Sunday. Each item carries a stage and a change flag: **NEW** **ADVANCED** **UNCHANGED** **CLOSED** . Flags show movement since the 18 October issue.

ITEM	WHAT IT IS	WHERE IT STANDS	WHAT DECIDES IT
STORY FDA generative-AI device paper ADVANCED	FDA discussion paper on evaluating foundation-model and agentic devices.	COMMENTING CLOSED Window shut with 500-plus comments; AHA, ATA and AMA on record; draft guidance promised in months. Source: AHA letter	Substance of the draft; any early answer to the Mosaic jurisdiction petition.
INNOVATION Diagnostic-LLM benchmarks NEW	Generalist language models tested against diagnostic cases.	BENCHMARK MAI-DxO's Science result - 83% versus 36% on NEJM cases - resets the public benchmark; tested on written cases, not care. Would change our view: replication on real cohorts. Source: Science	Prospective evaluation on live patients, and cost claims tested in situ.
STORY Human oversight of clinical AI NEW	Whether human review catches wrong AI output, as deployment rules assume.	MEASURED npj meta-analysis: odds of a correct clinician decision about ten times higher with correct than incorrect automated advice (OR 10.19, CI 7.43-13.98); 26.6% of erroneous ECG interpretations uncorrected (CI 6.4-65.6%). Source: npj Digital Medicine	Whether regulators write measured oversight failure into guidance.
STORY Epic Deterioration Index v2 UNCHANGED	FDA-cleared inpatient deterioration model, deployed across Epic hospitals.	CLEARED No new facts; still Exhibit A in the enforcement-gap debate. Would change our view: an FDA decision document or Epic statement on the locked-model claim. Source: FDA 510(k) K260008	Subgroup validation by hospitals, published monitoring.
STORY Epic Connectors NEW	A channel for third-party apps, including AI tools, to exchange data with Epic systems.	LAUNCHING Epic opened Connectors for apps to communicate with its EHR; AI vendors are the early audience. Would change our view: independent read of the data-flow and liability terms. Source: Fierce Healthcare	Which AI tools integrate first, and on what data terms.
STORY AI data governance in hospitals NEW	Staff use of unapproved AI tools on patient data.	INCIDENT Johns Hopkins Bayview alerted 3,182 patients after a staffer entered their data into an AI tool outside governance. Source: WMAR Baltimore	Other systems' disclosures; whether OCR treats AI tools as a distinct breach class.
INNOVATION Supervised patient-facing AI ADVANCED	Chatbots and agents that meet patients before or instead of a clinician.	DEPLOYING OpenEvidence embedded at Northwestern and Johns Hopkins; Abu Dhabi's M42 launched its Ossi agent. Source: Fierce Healthcare	Post-market safety data and consent standards for record access.
STORY India: AI rules and capacity ADVANCED	Register proposal, MeitY engagement, ICMR funding for indigenous devices.	BUILDING A public AI-register proposal circulated; ICMR picked seven institutions with Rs 105 crore for priority devices and diagnostics. Source: Times of India	Whether the register proposal becomes a rule, and what the ICMR hubs ship.
STORY UK AI device rulemaking UNCHANGED	MHRA Airlock sandbox, change-control guidance, classification consultation.	IN PROGRESS Airlock open, cohort unchosen; change guidance due around December; classification consultation next year. Source: GOV.UK	Whether the UK publishes methods other regulators adopt.

RAPID ROUNDS, BY DOMAIN

REGULATION & POLICY

REPORT

MDDI analysis: the FDA's Medical Device Development Tools programme published a framework for evaluating agentic AI components - a methods document, not a rule.

REPORT

Provider accountability: several docket letters put liability on deploying providers; hospital groups called that unworkable without vendor disclosure duties.

REPORT

AHA position: risk-tiered oversight and clarity on hospital-tuned tools; the fullest provider-side filing in the docket.

REPORT

Statute watch: no movement toward a US AI-in-care law; states keep passing disclosure patches.

MARKET & FUNDING

The pattern: deployment news outnumbered funding news, and most figures stayed company-reported.

COMPANY

M42's Ossi: Abu Dhabi's M42 launched an agentic health AI for its network; regional scale, company-reported.

COMPANY

SimBioSys catch-up: TumorSight Risk's FDA 510(k) (K260023) clears the first MRI-derived breast-cancer recurrence-risk AI; all figures company-reported.

COMPANY

Krsnaa-AZmed: the Indian diagnostics chain will deploy AZmed's fracture-detection AI across its imaging network.

COMPANY

OpenEvidence-Doximity: the espionage suit settled; terms undisclosed.

EVIDENCE & TRIALS

STUDY

Sepsis RL discordance: a npj Digital Medicine study found reinforcement-learning sepsis policies often disagree with clinicians on fluids and vasopressors - and disagreement is not always error.

STUDY

Diabetic-eye equity: the ACCESS2 analysis found AI screening cut disparities when paired with same-day referral paths; the tool alone changed less.

STUDY

Liver-scanning robot: a 588-gram autonomous ultrasound robot, trained on sonographer trajectories, matched clinical references for standard liver views (Nature Communications).

STUDY

Synthetic images: Stanford's generalist generator, trained on 1.26M public images, aims to sidestep privacy walls; validation on real data still owed.

STUDY

Three-lab score: the PROGRESS score stratifies metastatic prostate-cancer risk from PSA, ALP and AST alone; validated across cohorts (JCO CCI, via briefing).

INDIA

REPORT

ICMR's seven: seven institutions picked to build priority devices and diagnostics, Rs 105 crore behind them.

REPORT

AI register: a proposal for a public register of AI systems in Indian health circulated among policymakers.

REPORT

Aarthi Scans: the Tamil Nadu chain added AI chest-X-ray triage across centres.

REPORT

CanAssist Breast: OncoStem's prognostic test added an AI-based recurrence module for HR+/HER2- disease.

REPORT

BHASHINI at JIPMER: MeitY's language mission ran a health-AI workshop with JIPMER Puducherry - scribing, voice assistants, multilingual guidance.

REPORT

Robotic surgery week: Mysuru reported AI-guided IVUS cardiac care; a Chennai team completed a robotic knee series; SurgiOne launched a surgery-planning tool.

VITAL SIGNS

83% vs 36% MAI-DxO versus physicians on NEJM diagnostic cases	500+ comments in the FDA's generative-AI docket at close	OR 10.19 odds of a correct decision, correct vs incorrect automated advice	26.6% erroneous computer-ECG readings uncorrected (CI 6.4-65.6%)	3,182 patients notified in the Bayview AI data incident	Rs 105 cr ICMR backing for seven indigenous-device hubs
---	--	--	--	---	---

FOLLOW-UP: THE WEEK AHEAD

COMING WEEKS	1 NOV	DECEMBER	2027
FDA synthesis of the docket; draft gen-AI guidance is promised in months, and the Mosaic petition still needs an answer.	Next weekly issue; daily editions resume Monday.	MHRA's expected draft on changing AI devices; UK classification consultation to follow.	A formal FDA rule on generative devices, if it comes, lands next year at the earliest.

AI Health Weekly is synthesised from the week's daily editions of 19-24 October 2026, with fresh checks on 25 October. Offers, targets and company-reported figures are not audited outcomes. Regulatory status as of this edition.