// ORTENT WHITEPAPER · AUGUST 2026

The three maturities of healthcare AI

Seventeen categories of AI in healthcare, clinical and administrative, scored on evidence, deployment and economics. A category only scales when all three line up. This paper explains the method, gives the scored read of every category, and names what moves next.

Companion to the interactive Healthcare AI Radar at ortent.co/tools/healthcare-ai-radar

Most maps of healthcare AI answer the wrong question

They list companies by specialty, which tells you where the regulatory clearances are but nothing about whether anyone is being paid. Or they group products by technology, generative, predictive, agentic, which tells you about the model and nothing about the workflow it has to survive in. Either way the map looks busy and tells a buyer, a board or an investor almost nothing about what is actually ready.

A better question is: what does it take for a category of healthcare AI to scale? The record of the last decade gives a consistent answer. Three things, all at once.

Evidence. Independent, prospective, ideally randomised results, on a population like the buyer's. Not vendor benchmarks, not internal validation, not a press release with an AUROC in it.

Deployment. Routine use at scale in real services. Not pilots, and especially not pilots that have run for over a year without converting.

Economics. A repeatable way to get paid: a reimbursement code, a national procurement route, or a return-on-investment case a chief financial officer has actually signed. In one word, a budget.

The pattern across every category in this paper is the same: **a category only scales when all three maturities line up, and the categories that stalled were missing exactly one.** Radiology AI scaled because clearance, payment codes and a radiologist shortage converged. Ambient scribes became the fastest adoption curve healthcare IT has seen because they need no device clearance at the basic tier and pay for themselves against the workforce's loudest complaint. Sepsis prediction, with obvious clinical value, stalled for a decade because the evidence leg was contested. AI drug discovery has raised billions against a category where no leg is yet proven end to end.

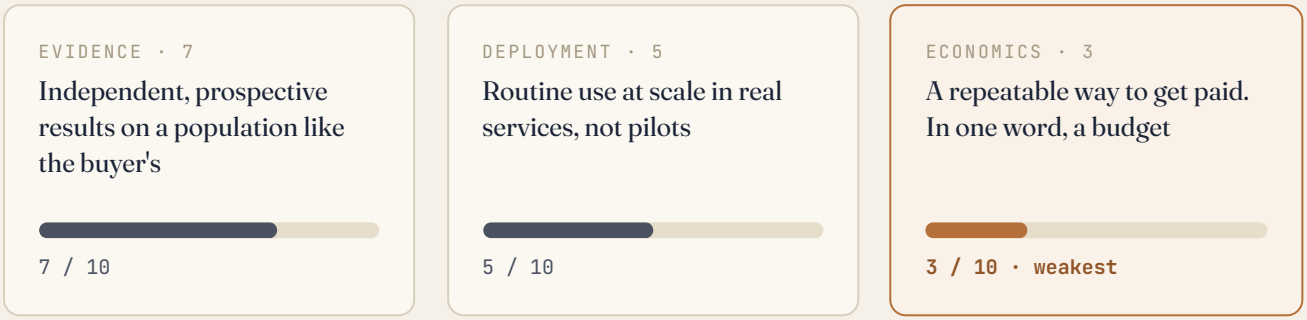
So the radar scores each category three times, separately, out of ten. The status of a category, Landed, Scaling, Emerging or Watch, is set by its weakest score, not its average, because the weakest leg is the one that decides. The gaps between a category's three scores are not noise. They are the diligence questions.

> Score evidence, deployment and economics separately, take the minimum, and read the gaps. That is the whole method. Everything else in this paper is the evidence for why it works.

The method in one figure

FIGURE 1 · ORTENT

Three scores, and the weakest one decides



→ **Status = the minimum, not the average.** Worked example: sepsis and deterioration, E7 · D5 · Ec3. Proven and deployed, still unpaid. The gap between 7 and 3 is the diligence question.

EMERGING · SET BY EC3

Notation used throughout this paper: E evidence, D deployment, Ec economics, each scored 0 to 10. Status bands on the weakest score: 7+ Landed, 5 to 6 Scaling, 3 to 4 Emerging, under 3 Watch.

Same problem, opposite methods: MASAI and the Epic Sepsis Model

The field ran its own controlled experiment on the value of evidence, in two acts.

Act one. The Epic Sepsis Model shipped bundled into the dominant American electronic health record, deployed at enormous scale on the strength of internally reported performance of 0.76 to 0.83 AUROC. Independent external validation found 0.63, with 33 percent sensitivity at operational alert thresholds. Later emergency-department validations found sensitivity as low as 14.7 percent and a positive predictive value of 7.6 percent. A proprietary black-box model, adopted because it was already in the system, missed most of the sepsis it was supposed to catch. Deployment had run years ahead of evidence, and clinicians paid for the gap in alert fatigue and missed cases.

Act two. The same clinical problem, done in the right order. TREWS, developed at Johns Hopkins and commercialised by Bayesian Health, went through prospective multi-site evaluation showing roughly a 20 percent reduction in mortality through earlier antibiotics, was deployed at Cleveland Clinic, MemorialCare and Rochester, and in May 2026 received FDA approval as one of the first cleared AI early-warning systems for sepsis. The category is now rebuilding on proper foundations, and the honest score reflects both acts: evidence repaired at 7, deployment mid at 5, economics still unformed at 3.

The counterpart on the other side of the ledger is MASAI: a randomised controlled trial of AI-supported breast screening inside the Swedish national programme, more than 105,000 women, using ScreenPoint Medical's Transpara, published in The Lancet Digital Health. The results: 29 percent more cancers detected with no increase in false positives, a 44 percent reduction in radiologist screen-reading workload, and in the final results a 12 percent reduction in interval cancers with 27 percent fewer aggressive interval cancers. A named commercial product, tested at national-programme scale, in the world's most credible format.

MASAI is the bar. The Epic story is the warning. Between them they define the evidence axis of this radar: every category is scored on how close its proof stands to the first, and how much of its deployment rests on the pattern of the second.

> The single most useful diligence question in healthcare AI: is there independent, prospective evidence on a population like the buyer's, or only internal benchmarks? Whole categories divide on the answer.

Autonomy climbs a ladder, regulation draws a line, and in 2026 the economics turned

The ladder. Every product in this paper sits on a four-step autonomy ladder: flag (draws human attention), triage (re-orders a human's queue), draft (produces work a human signs off), act (executes without per-case review). Almost everything deployed at scale today stops at draft. The step to act is the commercial and regulatory cliff-edge, and the largest unclaimed prize on the map.

The line. In July 2026 the MHRA, working with NHS England, drew the clearest regulatory line the sector has had: ambient voice products that transcribe, summarise, draft correspondence or suggest clinical codes for human review are not medical devices; anything supporting diagnosis, treatment or automated clinical action is. That single distinction creates a two-lane road, an unregulated-but-governed lane for draft-for-review tools and a device lane for everything clinical, and the EU AI Act, fully applicable in 2026, adds a heavier European third lane. Companies that misjudge their lane lose a year. Roadmaps that cross the line, and most ambitious ones do, need to know at which feature they cross.

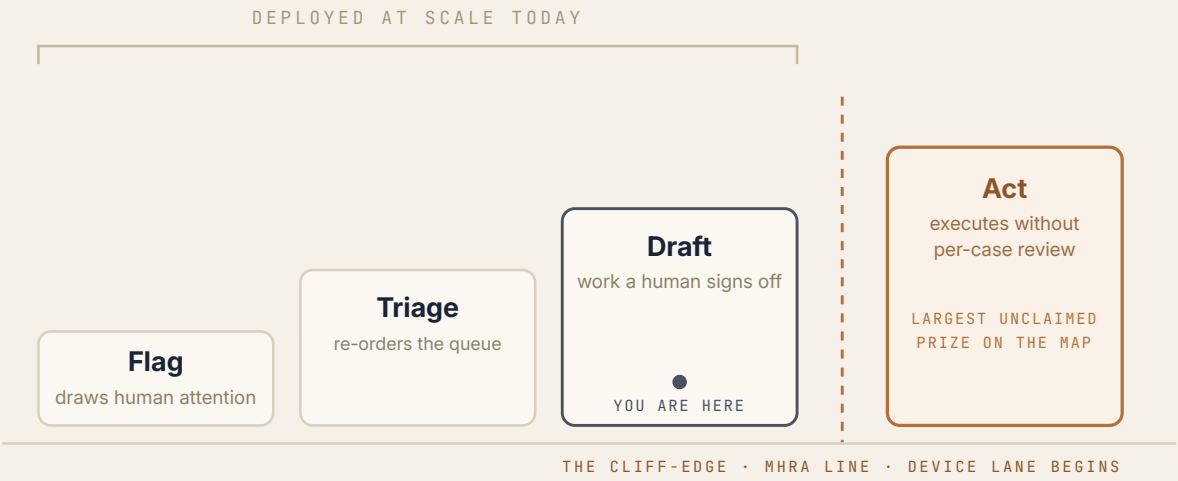
The money. For a decade the honest criticism of clinical AI was that nobody pays for it. In 2026 that measurably changed. The United States now has 26 CPT codes for clinical AI. The 2026 Hospital Outpatient rule established national reimbursement for AI-assisted cardiac analysis. From 1 October 2026, New Technology Add-on Payments open for FDA-cleared inpatient AI, with Aidoc's foundation-model triage suite and InVision's cardiac amyloid detection, worth up to $2,275 per stay, among the first through. In England the lever is central: £10 billion over three years for NHS technology, with AI triage in the NHS App and ambient scribes named as priorities and an expectation of adoption from April 2026. When a category's economics score moves, it usually moves because of one of these mechanisms, which is why the radar tracks them by name.

The solvent. One more force is dissolving the map itself: generalist medical foundation models. Radiology models now draft full reports from images, priors and clinical context; pathology equivalents are close behind. Each generalist model does tasks that currently support a single-product company each. The unit of analysis is shifting from algorithm-per-finding to platform-per-workflow, and thin single-finding point solutions are the most exposed layer in the sector. Funding is concentrating to match: $4 billion into digital health in the first quarter of 2026, with twelve companies taking 59 percent of it.

The ladder in one figure

FIGURE 2 · ORTENT

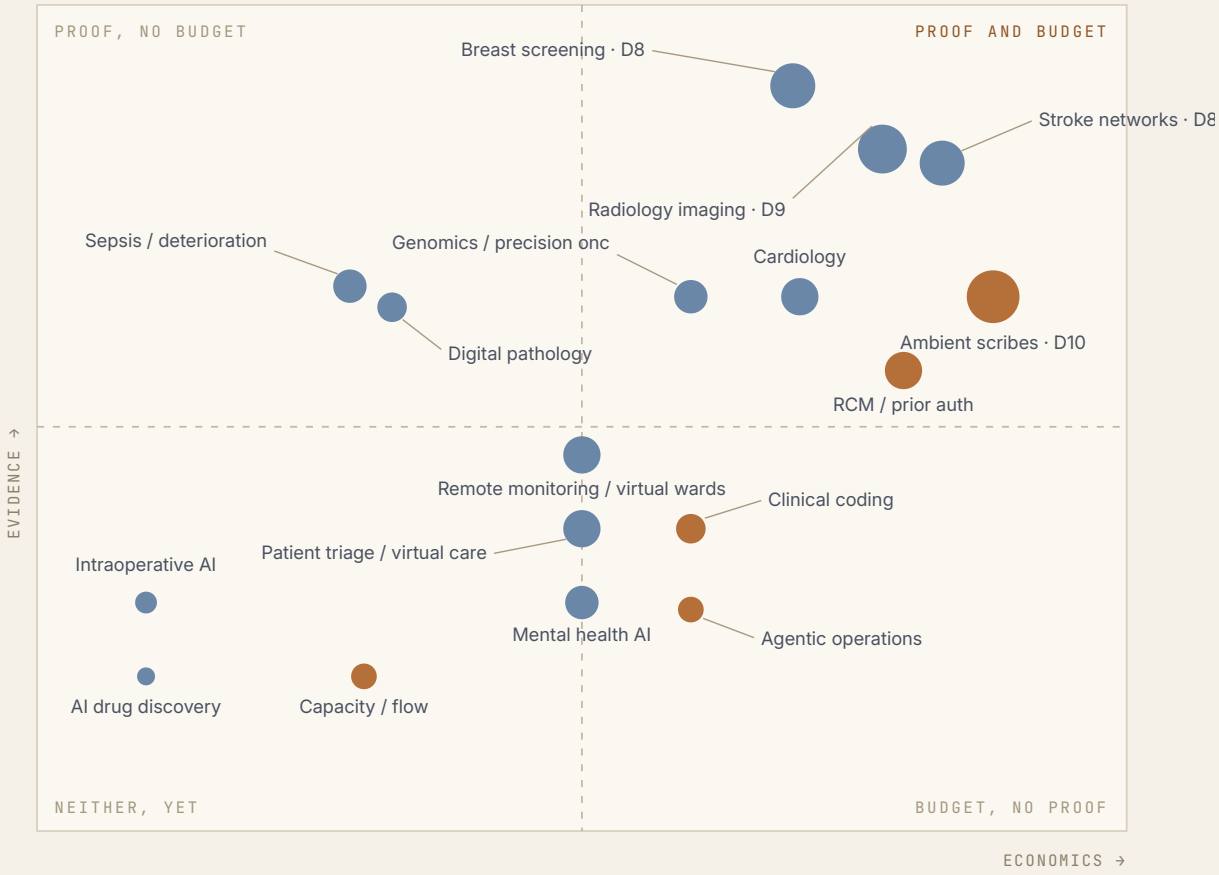
The autonomy ladder, and where everything stops



Roadmaps that cross the line need to know at which feature they cross. Everything scaled so far stops at draft; whoever cracks act-level autonomy, first in administration, later in care, takes the largest prize in the sector.

Proof against budget, seventeen categories

The chart plots evidence against economics; bubble size is deployment. The four corners name themselves: **proof and budget** is where categories scale; **proof, no budget** is proven work waiting for a payment route; **budget, no proof** is the caution corner, money flowing ahead of independent evidence, which is where buyers get burned; **neither, yet** is early, which is not an insult, the operating theatre lives there today.



CLINICAL · ADMINISTRATIVE · D2 · PILOTS · D10 · AT SCALE · BUBBLE SIZE = DEPLOYMENT

Evidence (vertical) against economics (horizontal); bubble size is deployment scale. Blue clinical, orange administrative. Positions as scored August 2026. Interactive version with the evidence behind every score: ortent.co/tools/healthcare-ai-radar

The scorecard

CATEGORY	EVIDENCE	DEPLOYMENT	ECONOMICS	STATUS
Radiology imaging	9	9	8	Landed
Stroke networks	9	8	8	Landed
Ambient scribes	7	10	9	Landed
Breast screening	10	8	7	Landed
Cardiology	7	6	7	Scaling
Revenue cycle and prior authorisation	6	6	8	Scaling
Genomics and precision oncology	7	5	6	Scaling
Remote monitoring and virtual wards	5	6	5	Scaling
Patient triage and virtual care	4	6	5	Emerging
Clinical coding from notes	4	4	6	Emerging
Sepsis and deterioration	7	5	3	Emerging
Digital pathology	7	4	3	Emerging
Mental health AI	3	5	5	Emerging
Agentic operations	3	3	6	Emerging
Capacity and patient flow	2	3	3	Watch
Intraoperative AI	3	2	1	Watch
AI drug discovery	2	1	1	Watch

Scores 0 to 10, August 2026. Status bands on the weakest score: 7+ Landed, 5 to 6 Scaling, 3 to 4 Emerging, under 3 Watch. Blue clinical, orange administrative. Where a dated, linked source exists it is in the interactive radar and the open dataset. The other scores are category readings and are marked as such there: 25 of 51 as at 3 October 2026.

Clinical, part one

Radiology imaging

E9 · D9 · Ec8 · LANDED

The proven beachhead. 1,163 FDA-cleared algorithms, 76 percent of all cleared AI devices, with roughly 30 more clearing each month. Overwhelmingly flag-and-triage autonomy: the algorithm highlights, the radiologist signs. The one clinical category where all three maturities have converged, which makes its question consolidation: foundation models are starting to absorb single-finding algorithms into platforms, and the first foundation-model triage suite has already won CMS reimbursement.

Breast screening

E10 · D8 · Ec7 · LANDED

The strongest evidence in clinical AI: the MASAI randomised trial, 105,000 women in the Swedish national programme, 29 percent more cancers found with no rise in false positives, 44 percent less reading workload, 12 percent fewer interval cancers. Economics run through programme-level procurement rather than per-case codes, so the score waits on more national programmes committing. Every other category is judged against this one.

Stroke networks

E9 · D8 · Ec8 · LANDED

Proof that the value is often in the pathway, not the pixel. Brainomix 360, deployed across 26 English hospitals, is associated in a five-year prospective study with a doubling of thrombectomy rates; Viz.ai cuts door-to-puncture times by 11 to 25 minutes in US networks. The algorithms detect large vessel occlusions, but what they sell is minutes across a hub-and-spoke network. A template waiting to be applied to other time-critical pathways.

Cardiology

E7 · D6 · Ec7 · SCALING

The second wave, and the category whose economics turned in 2026 with national outpatient reimbursement for AI cardiac analysis. Three businesses inside it: AI-ECG reading cheap ubiquitous signals for expensive structural disease, automated echocardiography now in NHS community diagnostic hubs, and coronary CT analysis moving into procedure planning. Watch the cheap-signal pattern: wearable ECGs as population-scale screens for amyloidosis and heart failure.

Digital pathology

E7 · D4 · Ec3 · EMERGING

Where radiology was around 2016. The algorithms are cleared and clinically evaluated, Paige prospectively assessed in routine NHS workflows across 1,000-plus prostate biopsies, Ibex cutting stain use 38 percent in routine practice, but you cannot run software on glass. Scanner adoption is the bottleneck and the leading indicator, and the structural driver underneath, pathologist shortage plus precision oncology demanding more from every slide, is not going away.

Sepsis and deterioration

E7 · D5 · Ec3 · EMERGING

The cautionary tale that finally matured, told fully in section two. The Epic model showed what deployment without evidence costs; TREWS showed the repair, prospective multi-site proof of a roughly 20 percent mortality reduction and FDA approval in May 2026. Payment remains unformed, which is why a category with repaired evidence still sits at Emerging. Whether approval unlocks a payment route is the thing to watch.

Clinical, part two

Intraoperative AI

E3 · D2 · Ec1 · WATCH

The newest clinical frontier and the highest-stakes white space on the map. Stimulated Raman histology plus models like FastGlioma give surgeons molecular-level tumour classification in minutes during the operation instead of days from the lab, directly informing how much tissue to take. Around it: augmented-reality margins, workflow-phase recognition, automated operative notes. Evidence emerging, deployment early, economics unformed, ceiling highest of any clinical category.

Genomics and precision oncology

E7 · D5 · Ec6 · SCALING

Shifted from variant pipelines to multimodal foundation models over proprietary data at a scale competitors cannot assemble: Tempus holds 500-plus petabytes across 45 million de-identified patient journeys. Published work shows tumour-normal matching, RNA sequencing and liquid biopsy reflex finding actionable results standard testing misses. The moat is data exhaust from care delivery, not model architecture, and payment rides established test reimbursement.

AI drug discovery

E2 · D1 · Ec1 · WATCH

Heavily funded, unproven end to end. As of mid-2026: 117 AI-enabled assets in human trials across 63 companies, only 8 past Phase 2, zero approvals of an AI-designed drug. Insilico's rentosertib entered Phase III for pulmonary fibrosis in July 2026 and is the bellwether. The technology demonstrably compresses discovery, but discovery was never the binding constraint. Trials are, and the next 24 months of readouts decide the category's claim.

Patient triage and virtual care

E4 · D6 · Ec5 · EMERGING

The Babylon collapse defines the category's scar tissue: peak-hype symptom checking, safety questions, gone by 2023. What changed in 2026 is the buyer. The NHS is embedding AI triage in the NHS App, 200,000 patients in the first year, all users by April 2028. When the system operator makes triage infrastructure, the standalone consumer symptom checker dies; the opportunity moves to powering the infrastructure and the regulated conversational diagnostics coming behind it.

Remote monitoring and virtual wards

E5 · D6 · Ec5 · SCALING

The substrate of hospital-at-home, now converging with language models: monitoring streams plus conversational check-ins plus automated documentation. Deployment ran ahead of outcome evidence during the post-pandemic push and the evidence is catching up. Converges with scribes and agents into a coordination layer around the clinician, which is where its next value lies.

Mental health AI

E3 · D5 · Ec5 · EMERGING

The top-funded clinical indication seven years running, which is exactly why it needs the evidence test most. The access gap is real and the funding durable, but clinical-grade autonomous therapy remains evidence-thin relative to the money raised. That gap between capital and proof is itself the diligence signal, and regulated digital therapeutics with randomised evidence are what would move the score.

Administrative

Ambient scribes

E7 · D10 · Ec9 · LANDED

The fastest adoption curve healthcare IT has seen: over 70 percent of large US systems deployed or piloting, Abridge in 300-plus systems including Kaiser Permanente, and the NHS mandating ambient voice adoption from April 2026 with a national supplier registry. Randomised studies now back the time-saving claims. Strategically a wedge, not an endpoint: the ladder runs scribe, coding, order drafting, care navigation, and every rung climbs further into regulated territory across the MHRA line.

Clinical coding from notes

E4 · D4 · Ec6 · EMERGING

The next rung on the scribe ladder: codes suggested from the ambient note for human confirmation. Economically attractive because coding backlogs are measurable money; regulatorily sensitive because code suggestion sits close to the MHRA line and payment integrity. Expect the scribe leaders to absorb this category rather than standalone winners to emerge.

Revenue cycle and prior authorisation

E6 · D6 · Ec8 · SCALING

Direct, CFO-signed returns in a US market where denial pain compounds yearly: a $20.6 billion market heading for $70 billion by 2030, mature deployments cutting denial rates 30 to 40 percent. The structural caveat: payers automate denials as fast as providers automate appeals, and an AI-versus-AI arms race compresses margin on both sides. The UK translation runs through clinical coding and the productivity agenda rather than prior authorisation.

Agentic operations

E3 · D3 · Ec6 · EMERGING

The frontier of administrative AI: systems that execute multi-step workflows across records, billing, communication and compliance rather than producing one output, landing first where work is repetitive and rules-heavy. Four in five healthcare executives expect significant value. The gate is governance: hospital security frameworks were not built for autonomous systems holding human-level privileges, which makes zero-trust agent architecture an unsolved, and investable, problem.

Capacity and patient flow

E2 · D3 · Ec3 · WATCH

Perennially promised, stubbornly hard. Bed, theatre and discharge prediction demos well, but the binding constraints are physical and political rather than informational, and predictions that cannot move a constraint do not move a metric. Held at Watch until deployments show sustained flow outcomes rather than dashboard adoption.

Where the opportunity sits

For investors and operating partners. The pattern to back: categories where evidence and deployment exist and the economics just turned, cardiology after the outpatient rule, inpatient imaging after the October add-on payments; and proven wedges climbing the autonomy ladder, scribes moving into orders and coding. The pattern to price cautiously: high deployment on thin evidence, the Epic shape; single-finding point solutions exposed to foundation-model consolidation; and any drug-discovery valuation not yet underwritten by Phase 2 data. Five diligence questions fall straight out of the framework: which regulatory lane, and where does the roadmap cross the line; independent prospective evidence on the buyer's population, or internal benchmarks; the specific payment mechanism, not "reimbursement tailwinds"; exposure to a generalist model absorbing the feature; and for anything agentic, the governance story.

For companies selling into the NHS. The NHS has, unusually, pre-announced its demand: AI triage in the App, ambient scribes from April 2026, £10 billion over three years, a supplier registry with Class 1 device accreditation and DTAC as the entry ticket. The near-term market is those named priorities plus the proven pathway plays: stroke-style network coordination, screening, community diagnostics. The Ortent NHS Board Radar measures the other side of the same coin, what 122 NHS organisations' own board papers show them actually delivering, and the gap between national announcement and board-level delivery is where sales cycles go to die.

For the clinical decade. In rough order of arrival: foundation-model radiology reporting going mainstream at draft autonomy; pathology repeating radiology's curve as scanners land; conversational diagnostics moving from feasibility studies to regulated products; the operating theatre acquiring real-time tissue and workflow intelligence; and agents quietly automating the coordination layer around the clinician. The biggest unclaimed territory on the whole map is trusted autonomy. Everything that has scaled so far stops at draft. Whoever cracks the evidence, liability and governance stack for act-level autonomy, first in administration, later in care, takes the largest prize in the sector.

> The radar is refreshed quarterly and the events that would move each category are listed on its scorecard. Company-level assessment against this framework, which lane, whose evidence, what payment route, is what Ortent does in diligence and advisory work. ortent.co/contact

How the scores are made

Each category is scored 0 to 10 on three separate maturities, from published evidence gathered and verified in August 2026. Evidence: independent, prospective, ideally randomised results, not vendor benchmarks; vendor-run studies score lower than independent ones. Deployment: routine use at scale, not pilots. Economics: a repeatable way to get paid, a code, a procurement route, a national mandate, or a repeatedly signed ROI case. Status is the minimum of the three: 7 and above Landed, 5 to 6 Scaling, 3 to 4 Emerging, below 3 Watch. Scores are editorial judgement applied to cited evidence, and the evidence is published alongside every score in the interactive radar and the open dataset so the judgement can be challenged. Example players named in the radar are illustrative, not a ranking, not exhaustive, and imply no endorsement. Categories are held constant between quarterly refreshes so movement is real.

Principal sources

MASAI randomised controlled trial, The Lancet Digital Health; final results via ScreenPoint Medical, 2024-2026.

Epic Sepsis Model external validations, JAMIA Open and related literature, 2024-2025.

TREWS prospective evaluation and FDA approval, Johns Hopkins University and NEJM AI, May 2026.

FDA AI-enabled device clearance counts, IntuitionLabs analysis of the FDA list, March 2026.

Brainomix 360 five-year prospective observational study, The Lancet Digital Health, 2025.

Viz.ai VALIDATE study and stroke network experience, 2025-2026.

Cardiology clearance counts, Cardiovascular Business, March 2026; 2026 Hospital OPPS Final Rule.

Paige Prostate NHS evaluation, NCBI Bookshelf; Ibex implementation study, ScienceDirect, 2025-2026; KLAS digital pathology reports, 2026.

Tempus multimodal foundation-model results, ASCO 2026; JCO Precision Oncology, March 2026.

AI drug discovery pipeline analysis presented at ASCO 2026; Insilico rentosertib Phase III announcements, July 2026.

NHS England, AI rollout announcement and £10bn technology programme, July 2026; ambient scribing guidance, 2026.

MHRA, regulatory status of ambient voice technologies, 29 July 2026.

KLAS ambient speech reports, 2025-2026; Abridge, Microsoft DAX and Heidi deployment announcements.

CMS New Technology Add-on Payment approvals including Aidoc CARE and InVision cardiac amyloid, effective October 2026; AMA CPT clinical AI codes, January 2026.

Rock Health digital health funding reports, Q1 and H1 2026.

Revenue cycle AI market and denial statistics, Healthcare Finance News and industry analyses, 2025-2026.

Full dated links for every score: the open dataset at ortent.co/tools/healthcare-ai-radar

© 2026 Ortent Advisory Ltd. Ortent is a registered trade mark of Ortent Advisory Ltd (UK00004376535). This paper is a market read, not investment, legal, clinical or regulatory advice. August 2026 edition; next review November 2026.