LANDSCAPE REPORT

FROM BENCHMARK TO BEDSIDE: THE CURRENT STATE OF AI IN CLINICAL PRACTICE

May 2026

AMA AMERICAN MEDICAL ASSOCIATION

DiME DIGITAL MEDICINE SOCIETY

PURPOSE & USE OF THIS REPORT

This report has been developed to support the upcoming convening hosted by the American Medical Association Center for Digital Health and AI with support from the Digital Medicine Society (DiMe), *Defining the Role of the Physician in the Digital and AI Era of Medicine*, providing a focused pre-read that establishes a shared baseline on the current state of artificial intelligence in clinical practice and enables a more productive, action-oriented discussion across the sessions outlined in the agenda.

This report is intended to support that discussion by providing a shared baseline across participants.

The analysis draws on peer-reviewed literature, regulatory activity, and emerging payment models across the United States and comparable markets. It synthesizes where the evidence is strongest, where real-world deployments are showing measurable impact, and where implementation challenges persist. It also highlights patterns that cut across use cases, including the conditions under which AI performs well in practice and the factors that contribute to failure at scale.

This approach is shaped by the constraints of the current evidence base. The pace of peer-reviewed publication does not match the speed of technological development or real-world deployment, and many systems described in the literature no longer reflect what is in use today.

This report focuses on what holds across technologies: lessons from deployment, examples of measurable impact, persistent evidence gaps, and the questions the literature cannot yet resolve. When the evidence base cannot keep pace, informed judgment and cross-sector perspective become essential. This convening is designed to provide both.

The goal of the meeting is to co-create a clear, authoritative articulation of the physician's role in an AI-enabled healthcare system, grounded in both evidence and real-world experience. This report is intended to support that work by providing a common foundation, surfacing the most relevant signals from the current evidence base, and framing the questions that require expert discussion.

View report online:

EXECUTIVE SUMMARY

Artificial intelligence is already being used in clinical practice. Across imaging, documentation, predictive analytics, and patient-facing tools, the evidence shows that AI can deliver specialist-level performance on defined tasks and measurable improvements in efficiency and access.

These results do not consistently hold in real-world clinical settings, and the systems now in use represent early versions of what will follow.

Roughly 60% of AI pilots fail to reach sustained deployment. Performance degrades when systems are introduced into environments they were not designed for, when workflows are disrupted, or when monitoring and governance are insufficient to detect failure early. Reliable integration into real-world care delivery remains the primary constraint.

At the same time, AI capability is advancing rapidly. Systems are taking on more complex tasks across diagnosis, documentation, and patient interaction, and their influence on clinical decision-making is increasing. The evidence base used to evaluate these systems is not keeping pace, and regulatory, payment, and operational frameworks are evolving in parallel.

These dynamics are already shaping practice. Clinicians are incorporating AI tools into their work, and patients are arriving to care having already used them to interpret symptoms and guide decisions. In many cases, this use occurs outside formal governance structures. AI is influencing clinical decisions while the systems required to evaluate, integrate, and oversee its use are still being defined.

This creates a new set of requirements for clinical practice. Physicians remain responsible for decisions influenced by AI, even as they work in environments where system performance, limitations, and failure modes are not always visible. Clinical judgment is increasingly exercised in partnership with AI systems. At the same time, trust, communication, and accountability remain central to the physician’s role.

AI will become more capable, more embedded in care delivery, and more influential in clinical decision-making. At the same time, healthcare systems will continue to evolve in response, through changes in regulation, payment, infrastructure, and professional practice.

Health systems face increasing demand, workforce constraints, and rising costs. AI offers a path to improving access, quality, and sustainability, provided it is implemented in a way that supports patients and the clinicians who care for them.

1. WHAT PUBLISHED EVIDENCE SHOWS

Where AI is working today

Artificial intelligence can deliver strong performance in clinical settings when applied to clearly defined problems. In radiology, diagnostic systems reach 92 to 98% concordance with specialist reads. Autonomous screening tools demonstrate greater than 90% sensitivity for conditions such as diabetic retinopathy and glaucoma. In administrative workflows, ambient documentation tools reduce time spent on documentation by 31 to 47% in pilot implementations.

A small number of health systems have translated these capabilities into measurable clinical impact. Across 26 NHS hospitals, AI-supported stroke imaging doubled the rate of thrombectomy and improved functional outcomes at discharge. In urgent care settings, AI-assisted clinical workflows have reduced diagnostic errors by 16% and treatment errors by 13%. At Stanford Health Care, conversational tools for querying patient records have scaled to thousands of clinicians and patients, demonstrating that certain use cases can reach institutional adoption.

These successes share common characteristics. The task is clearly bounded. The system is deployed within an existing clinical workflow rather than requiring clinicians to adapt to a new one. Performance is validated against real-world conditions, and the organization has the infrastructure to monitor and manage the system over time.

Even in these settings, performance is bounded. AI systems perform best when retrieving, classifying, or summarizing information. They are less reliable when the task requires clinical judgment under uncertainty, integration of incomplete information, or adaptation to changing conditions.

AI performs at a specialist level on narrow, well-defined tasks

Real-world impact is measurable but concentrated in a small number of settings

Success depends on workflow fit, validation in practice, and supporting infrastructure

Where AI is currently falling short

The same evidence base shows that these successes are difficult to reproduce at scale. Roughly 60% of clinical AI pilots do not reach sustained deployment, and only 15 to 25% of clinical departments move beyond evaluation into routine use.

Failures are rarely driven by model performance alone and emerge at the point of implementation. Systems that require changes to established workflows, depend on manual data entry, or operate on fragmented data infrastructure struggle to gain traction. Clinician trust remains a consistent barrier, particularly when system outputs are difficult to interpret or verify.

Performance also degrades outside controlled settings. Models that score 84 to 90% on medical knowledge benchmarks fall to 45 to 69% on clinical decision-making tasks and to 40 to 50% on patient safety evaluations. Systems that perform well in simulated environments have not been consistently shown to maintain that performance under real-world conditions, where time pressure, interruptions, and incomplete information shape clinical decision-making.

Narrowly defined performance gains are often accompanied by tradeoffs. Documentation tools that reduce clinician time can introduce factual errors into the medical record. Systems designed to improve efficiency can shift risk rather than reduce it, moving the burden from clinician effort to patient safety if outputs are not carefully reviewed.

Finally, the environments in which AI succeeds are not evenly distributed. Health systems demonstrating consistent results have dedicated implementation teams, interoperable data infrastructure, and leadership willing to absorb the cost of failed pilots. Most health systems do not operate under those conditions, limiting the generalizability of current successes.

Most AI pilots do not reach sustained clinical use

Performance degrades in real-world conditions

Failures are driven by workflow, data, and trust—not just model capability

2. AI IS CHANGING CARE BEFORE ROLES & RULES ARE DEFINED

Why this gap matters

Artificial intelligence can deliver strong performance in controlled settings and measurable impact in select clinical environments. But most efforts to deploy these systems do not achieve sustained use, and performance often degrades in real-world conditions. This reflects a mismatch between what the technology can do and how healthcare systems are structured to use it. Clinical workflows are not designed for continuous decision support, data infrastructure remains fragmented, and oversight mechanisms are not equipped to monitor systems that evolve over time.

These dynamics are already shaping practice. AI systems are taking on more complex tasks across diagnosis, documentation, and patient interaction. Clinicians are incorporating these tools into their work, and patients are arriving to care having already used them, often outside formal governance structures.

The result is a healthcare system in which AI is already influencing clinical decisions without a consistent framework for how it should be evaluated, integrated, or overseen. As these systems become more capable and more embedded in care delivery, that gap becomes more consequential. Ensuring their use remains safe, effective, and accountable will require continuous adaptation.

AI capability is outpacing the system's ability to use it in practice

The primary constraint is implementation, not model performance

Governance, workflow, and accountability have not kept pace with adoption

Incremental adjustments to existing clinical roles will not be enough. Physicians are already working alongside AI systems, whether formally deployed or informally adopted, and patients are increasingly doing the same. As these tools become part of how care is delivered, they become part of the care team. Defining the physician's role, responsibilities, and boundaries in that context is central to ensuring that this transformation improves care rather than fragments it. This is the challenge this convening is designed to address.

3. AI IS ALREADY CHANGING CLINICAL PRACTICE

Inside the clinical workflow

Artificial intelligence is already embedded in clinical workflows across many health systems. It supports documentation, imaging, triage, and elements of clinical decision-making, shaping how work is performed in day-to-day practice.

These tools are changing how work is performed. Ambient documentation systems reduce time spent on note-taking but introduce a new task: reviewing and correcting AI-generated content.

Decision support tools surface recommendations earlier in the clinical process, shaping how clinicians gather information and consider next steps. In imaging and screening, AI systems increasingly influence which findings are prioritized and how quickly they are acted upon.

As systems take on more complex, multi-step tasks, the nature of supervision is changing. In traditional decision support, a clinician reviews a single recommendation. In emerging workflows, AI systems may generate hypotheses, suggest tests, and propose plans in sequence. Errors introduced early in that sequence can propagate forward, producing outputs that appear coherent but are clinically incorrect.

AI is already shaping how clinical work gets done

Supervision is becoming more complex as systems take on multi-step tasks

Human–AI collaboration introduces new sources of error and variability

At the same time, human performance is dynamic. Studies show that AI assistance can improve accuracy in some cases and reduce it in others, with no reliable way to predict which clinicians will benefit.

Exposure to AI can also change unaided performance over time, particularly in perceptual tasks, where reliance on automated detection may reduce the need for independent signal recognition.

The result is a clinical environment in which AI functions as an active participant in how decisions are formed, reviewed, and executed. As such, we must begin to consider these tools part of the multi-disciplinary care team.

Outside the clinical workflow

AI is arriving in healthcare directly via patients and clinicians, as well as through formal deployment in clinical workflows.

Patients are increasingly using general-purpose AI tools to understand symptoms, interpret diagnoses, and consider next steps before seeking care. Many arrive at clinical encounters having already formed a working hypothesis about their condition, often based on AI-generated information that appears tailored and authoritative.

Clinicians are doing the same. In the absence of enterprise solutions that meet their needs, many are turning to consumer tools for administrative support, documentation, and information retrieval. This use often occurs outside institutional oversight, without formal validation for clinical accuracy, privacy, or compliance.

Together, these behaviors are creating a parallel layer of AI-informed decision-making that sits alongside formal clinical systems. Information is being generated, interpreted, and acted upon before it reaches the clinical record, shaping both patient expectations and clinical interactions.

The clinical encounter is being forced to evolve in response. Physicians are increasingly asked to interpret AI-generated inputs brought by patients or used by colleagues in addition to traditional symptoms and data. Managing this information, assessing its reliability, and integrating it into care are additional tasks that physicians must complete.

Patients are already using AI to interpret symptoms and guide decisions

---

Clinicians are using AI tools outside institutional governance

---

A parallel layer of AI-informed decision-making is emerging

4. WHAT THIS MEANS FOR THE PHYSICIAN ROLE

Clinical judgment

AI systems can retrieve information, recognize patterns, and generate plausible clinical outputs. In controlled settings, they can match or exceed human performance on specific tasks. In practice, these outputs become one input among many in clinical decision-making.

The nature of clinical judgment, however, remains unchanged. Diagnosis and treatment decisions are rarely made with complete information. They require weighing uncertainty, interpreting conflicting signals, and adapting to the specific context of the patient in front of the clinician. Current AI systems do not reliably perform these functions, particularly when information is incomplete or conditions deviate from those in the training data.

The physician's role is evolving from sole decision-maker to responsible integrator of multiple inputs, including AI-generated outputs. This requires a different skill set. Clinicians must assess when AI is operating within its validated use case, recognize when outputs are unreliable, and determine how much weight to assign to them in the context of a specific patient.

AI can support clinical reasoning but does not replace it

Uncertainty, ambiguity, and context remain physician-led

The core skill is judgment under AI, not independent of it

Clinical judgment is increasingly exercised in partnership with AI systems.

Accountability

As AI systems become more embedded in clinical workflows, they influence decisions in ways that are not always visible. Recommendations may shape diagnostic pathways, treatment plans, and patient communication before a clinician formally reviews them.

Despite this, accountability has not shifted. Legal and professional standards continue to place responsibility for clinical decisions on the physician, even when those decisions are informed by AI systems.

This creates a structural tension. AI can contribute to decisions without sharing responsibility for their outcomes. At the same time, the complexity of AI-assisted workflows can make it difficult for clinicians to fully evaluate how a recommendation was generated or whether it is appropriate in a given context.

The concept of "human in the loop" describes a workflow position and does not guarantee meaningful oversight.

Effective supervision depends on whether clinicians have the time, information, and authority to question or override system outputs. In many current implementations, those conditions are not consistently met.

Responsibility for clinical decisions remains with the physician

AI introduces new failure modes that are difficult to detect

"Human in the loop" does not guarantee meaningful oversight

Trust & the patient relationship

Patients are already using AI to interpret symptoms, understand diagnoses, and evaluate treatment options. Many arrive at clinical encounters with expectations shaped by AI-generated information that is personalized, detailed, and often presented with confidence.

This changes the nature of the clinical interaction. Physicians are no longer the sole interpreters of medical information. They are increasingly asked to validate, contextualize, or correct AI-generated inputs.

At the same time, the core elements of the patient relationship remain. Trust is built through communication, empathy, and the ability to respond to a patient's concerns in context.

These elements of care require continuous human presence and judgment.

AI can support aspects of this work, including information retrieval and communication, but it does not replace the physician's role in establishing trust or taking responsibility for the patient's care.

As AI becomes more embedded in care delivery, maintaining that trust becomes more, not less, important.

AI changes how patients understand and engage with their care

The physician remains responsible for trust, communication, and context

Certain elements of care remain inherently human

Defining the physician's role in an AI-enabled system is an immediate operational requirement.

5. SYSTEM CONSTRAINTS

Policy & regulation

Regulatory bodies, including the U.S. Food and Drug Administration (FDA), have established and continue to refine pathways for evaluating AI-enabled medical technologies. Recent guidance reflects a shift toward lifecycle-based oversight, including expectations for transparency, performance monitoring, and the management of post-deployment changes. This includes guidance on clinical decision support software, AI-enabled device lifecycle management, and the use of predetermined change control plans to enable iterative updates while maintaining regulatory oversight.

These frameworks represent substantial evolution in regulatory thinking. They acknowledge that AI systems may change over time and begin to define mechanisms for managing that change within a regulated context.

At the same time, regulatory evaluation remains anchored at the point of authorization. Submissions assess safety and performance under defined conditions, typically prior to widespread clinical use.

In practice, AI systems are deployed into environments that differ from those in which they were validated. They are integrated into complex workflows, used by clinicians with varying levels of experience, and applied to patient populations that may not be fully represented in training data. Evidence from post-deployment studies shows that performance can shift under these conditions, particularly when underlying data distributions change.

Regulatory frameworks are advancing to address AI across the product lifecycle

Approval establishes safety and performance at a point in time, not in practice

Oversight models are still evolving to address real-world use and continuous change

Regulatory frameworks are beginning to address aspects of this variability, including expectations for post-market monitoring. However, they do not fully account for how these systems are used in practice, where outcomes depend on workflow integration, clinician interaction, and ongoing operational oversight.

This is reflected in the current evidence base. A systematic review of 519 clinical AI studies found that only 5% included real patient data, 15.8% reported fairness metrics, and 1.2% reported calibration, limiting the ability to assess performance across populations and settings.

In parallel, the Office of the National Coordinator for Health Information Technology (ONC) has advanced requirements through the HTI-1 Final Rule, including provisions related to transparency, risk management, and the use of predictive decision support interventions. These requirements are designed to increase visibility into how AI-enabled tools are developed and used, and to support more consistent oversight across health IT systems.

Payment & incentives

Reimbursement structures have not kept pace with AI-enabled care delivery. In many cases, there are no clear or consistent payment pathways for AI-supported services, even when they improve efficiency or outcomes.

Health systems are expected to absorb the costs of implementation, including integration, training, and ongoing monitoring. These costs are non-trivial. Implementation of enterprise AI systems requires significant investment in IT infrastructure, workflow redesign, and governance processes.

At the same time, financial benefits are often distributed across the system. Improvements in outcomes or reductions in downstream utilization may accrue to payers or other providers rather than the organization making the investment.

Incentives also shape how technology is used. In the absence of alignment to quality and safety, AI may be deployed to optimize throughput or reduce administrative burden without consistent improvement in clinical outcomes.

Payment models do not consistently support AI-enabled care

Incentives are not aligned with safe adoption or sustained use

The costs of implementation are borne locally, while the benefits are often distributed

The result is uneven adoption, with leading organizations investing heavily while others remain constrained by financial uncertainty and unclear return on investment.

Infrastructure & operations

AI systems depend on high-quality data, interoperable systems, and integration into clinical workflows. In many environments, these prerequisites are not consistently in place. Fragmented data infrastructure remains a persistent barrier to scaling AI across health systems.

Operationalizing AI requires more than technical deployment. It includes governance structures, performance monitoring, clinician training, and processes for updating or withdrawing systems when performance changes. These capabilities are still developing across most health systems.

The organizations demonstrating consistent success share common characteristics: strong data infrastructure, dedicated implementation teams, and leadership support for iterative deployment. Evidence from large-scale deployments shows that these enabling conditions are concentrated in well-resourced systems and are not widely distributed.

As a result, the ability to use AI safely and effectively varies significantly across settings, limiting the scalability of current approaches.

Reliable use of AI depends on data, integration, and operational capacity

Most health systems are not equipped to support continuous monitoring and adaptation

Implementation remains resource-intensive and unevenly distributed

6. PRIORITY QUESTIONS TO ADDRESS

The following questions are intended to guide discussion and define direction.

Evaluation & Evidence

- What constitutes sufficient evidence for AI systems used in clinical care, particularly when performance evolves over time?
- How should real-world performance be measured, monitored, and reported across settings?
- What standards are needed to assess safety, fairness, and reliability in practice?

Clinical Responsibility & Accountability

- What decisions must remain physician-led, regardless of AI capability?
- How should responsibility be defined when AI systems influence clinical decisions?
- What does meaningful oversight look like in workflows where AI contributes to multi-step decisions?

Use in Practice

- How should health systems govern the use of AI tools by clinicians outside formal deployment?
- What expectations should be set for the use of AI in patient care, documentation, and communication?
- How should AI-generated inputs from patients be incorporated into clinical decision-making?

Policy, Payment, & Infrastructure

- How should regulatory models evolve to account for continuously learning systems?
- What payment structures are needed to support safe and effective use of AI in care delivery?
- What infrastructure and operational capabilities are required to support sustained use?

Boundaries & the Future of Care

- Which aspects of care should remain human by design, regardless of technological capability?
- How should the physician's role evolve as AI becomes part of the care team?
- What does good look like for an AI-enabled healthcare system from the perspective of patients, clinicians, and society?

Notes