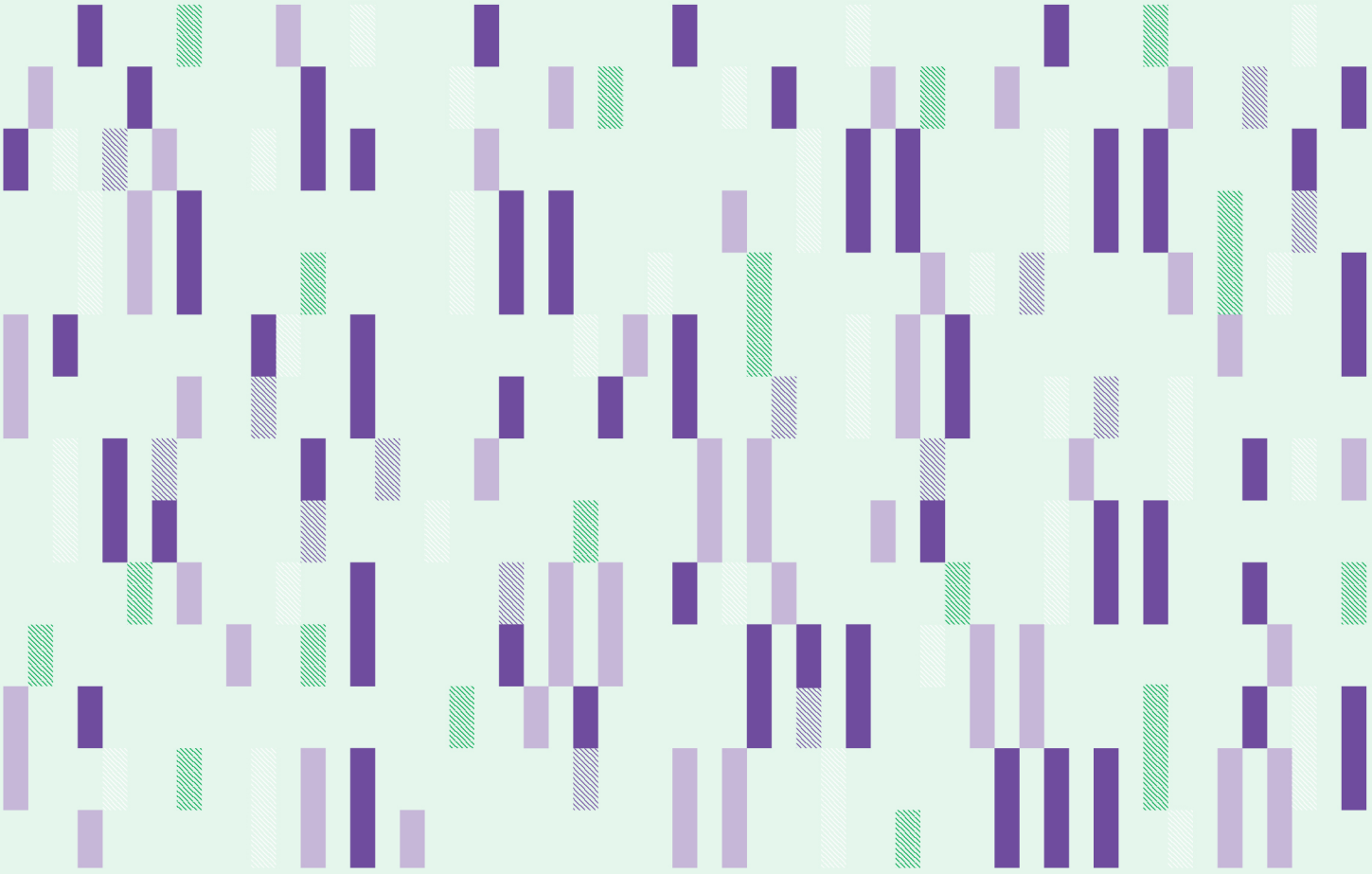
PETERSON HEALTH TECHNOLOGY INSTITUTE

Clinical AI: Evidence and Policy Requirements for Scaling Adoption

INDUSTRY INSIGHT | FEBRUARY 2026



PHTI.org

Executive Summary

Artificial intelligence (AI) is rapidly reshaping clinical care. AI-based technologies are enabling increasingly autonomous care delivery, making AI a critical tool for addressing gaps in workforce, expanding access, and improving patient outcomes in the U.S. healthcare system. In November 2025, PHTI convened senior leaders from health systems, health plans, technology developers, academia, investment firms, and federal agencies—including clinical experts—for a workshop in Washington, DC, to explore what is needed to scale AI for autonomous healthcare delivery.

The workshop focused on the requirements for safe, effective, and scalable use of clinical AI—with autonomous prescribing for hypertension management and mental health chatbots as illustrative use cases.

The opportunity for AI in hypertension and mental health care

Nearly half of U.S. adults—approximately 120 million people[1]—have hypertension. Yet despite the availability of well-established, generally low-cost, and highly effective treatments,[2] only one in four patients[3] successfully achieves blood pressure control. More than one in five U.S. adults[4] live with a mental illness, yet many never receive treatment because of cost barriers,[5] clinician workforce shortages,[6] and other factors.

These gaps in care for both hypertension and mental health reflect persistent failures in our current healthcare system. Effectively managing both conditions requires broad, up-front screening; active treatment; medication initiation and titration; attention to side effects; sustained patient engagement; and support for long-term behavior changes. Current care delivery models lack the capability and capacity to provide this level of continuous management, resulting in widespread underdiagnosis and undertreatment.[7]

Clinicians and patients are increasingly turning to AI to fill gaps. For hypertension management, AI supports measurement, diagnosis, and treatment decisions, with emerging solutions operating increasingly autonomously. In mental health, patients are often turning to publicly available generative AI tools for needed support. However, these tools are typically not designed for this purpose and lack clinical rigor or validation. Purpose-built mental health chatbots—trained on cognitive behavioral therapy and other psychotherapy principles—are now entering the market with promising clinical outcomes for patients with mild-to-moderate symptoms.

With the policy landscape evolving at an unprecedented speed to adapt to these technological advances, clinical AI applications will be positioned to deliver high-quality, accessible care to millions of patients. Yet going from promise to scale requires addressing market reality.

What clinical evidence, performance monitoring, and regulatory changes are necessary to build confidence in these tools for purchasers, clinicians, and patients?

The workshop discussion spanned a common set of questions across two use cases:

Adoption and Evaluation of Progress

- What would need to be true for clinicians to accept and adopt these tools?
- How will the market know whether these tools are driving meaningful progress?
- What factors would encourage payers to cover clinical-grade solutions?
- What would help patients feel confident and safe using them?

Market and Policy Enablers

- What barriers would limit effective adoption today?
- How might stakeholders work to accelerate adoption?
- What insights from these case studies inform regulation more broadly?

From promise to scale: Strengthening market confidence

Purchasers need standards to assess quality and value; innovators need guidance on product development and evidence generation requirements to meet regulatory and market needs; and clinicians need to understand which tools work and for which populations.

Four key themes emerged from the discussion:

1. **Evidence standards should compare AI to current standards of care and scale with clinical risk.** Evidence requirements must be rigorous enough to build trust, yet practical enough to avoid stalling innovation. This means having different evidence standards based on the risk of using the AI tool. Autonomous AI tools should be compared to local conditions and the care that patients receive today, not to idealized care. For many, the alternative may be poor access or no treatment at all.

2. **Performance benchmarks should be based on clinical outcomes, and safety standards should adapt as the evidence grows.** Ambiguity about what constitutes "good" performance remains a barrier to adoption. Metrics must be anchored to specific, measurable, and meaningful clinical outcomes, rather than to process measures.

3. **New technologies may be initially tested in lower-risk populations but should scale quickly to high-risk populations to maximize impact.** Lower-risk patients offer tempting on-ramps, but the greatest opportunities for clinical benefits from AI-enabled solutions come from reaching the highest-need patients, including those with higher-complexity conditions and in underserved communities. Reaching these populations may require higher evidence expectations and carry more significant clinical risk.

4. **Widespread adoption will depend on building clinician confidence, gaining clarity about legal liability, and aligning payment models.** Even highly effective clinical AI faces adoption resistance in the form of norms and culture, concerns about liability, and misaligned incentives. Many health systems are uneasy moving from "some human involvement" to "little-to-no human involvement." These tensions are impacting regulatory frameworks and the evolution of clinicians' roles as adoption grows.

The discussion underscores a central reality: the technologies enabling autonomous clinical AI are advancing faster than the policy, payment, evidentiary frameworks, and organizational readiness, needed to support their adoption. Participants identified meaningful pathways forward for both hypertension management and mental health chatbots but also surfaced unresolved tensions that will require sustained, cross-stakeholder dialogue. The themes that emerged are not unique to these use cases. They reflect foundational questions that will recur as autonomous AI capabilities expand across clinical domains.

Shaping the Future of Autonomous Hypertension Management and Mental Health Chatbots

Use Case 1: How AI-Enabled Solutions Can Address Access and Quality Gaps in Hypertension Management

AI-enabled technologies offer a pathway to automate steps across the hypertension care continuum, with the potential to expand access, improve outcomes, and reduce clinician burden. Emerging solutions are beginning to impact care delivery activities, including remote blood pressure monitoring, risk detection, care navigation, medication initiation, and ongoing clinical management.

> "The gap between what's now possible, relative to the declining status quo, has never been greater. We are facing an access crisis at unprecedented scale."

Clinical AI tools can be classified by the degree of human oversight, ranging from fully clinician-directed care to fully autonomous systems operating under defined protocols (Exhibit 1). While most commercially available products remain at Level 1 (assistive AI) or Level 2 (semi-autonomous AI), realizing the full value of AI requires progress toward Level 3 (supervised autonomy) and Level 4 (fully autonomous) capabilities.

Exhibit 1: Levels of AI Autonomy in Autonomous Prescribing

Level	Example – Autonomous Prescribing	Adoption and Evidence
No AI	Standard care without AI-enabled decision support.	--
1 – Assistive AI	Device displays blood pressure trends; clinician reviews data and decides whether to make medication adjustments.	Most existing tools can be categorized as Level 1 or 2. RCTs demonstrate that assistive and semi-autonomous tools can improve adherence to hypertension guidelines[8] and lead to modest reductions in blood pressure.[9]
2 – Semi-Autonomous	AI analyzes readings and recommends medication adjustments or escalates out-of-range values; clinician must approve.	
3 – Supervised Autonomy	AI instructs patients to change their medication dosage using an approved algorithm; clinicians intervene only if safety thresholds are crossed.	Limited in practice; pilots of autonomous coaching programs[10] led to blood pressure reductions.
4 – Full Autonomy	AI independently initiates and adjusts prescriptions without clinician oversight.	Hypothetical; no FDA-cleared product operates without oversight.

Evidence is building to support adoption of Level 3 and Level 4 tools. Randomized control trials (RCTs) demonstrate that Level 1 and Level 2 tools can provide safe, guideline-aligned medication management[11] to support improvements in blood pressure control.[12]

Market readiness is also increasing. Remote monitoring and telehealth have normalized protocol-driven, virtual hypertension care.[13] Continued device innovation—including the FDA clearance[14] of the first cuffless blood pressure monitor in July 2025—now enables the continuous data collection required to support more advanced AI tools. ARPA-H recently announced a new program seeking to develop the first FDA-authorized, agentic AI technology that can provide 24/7 specialty care for cardiovascular disease.[15]

Together, the trajectory of emerging clinical evidence, commercial device development, AI maturity, and regulatory enablement support the progression toward higher-autonomy AI systems in hypertension care.

Use Case 2: How AI Can Expand Access to High-Value Mental Health Care

Purpose-built chatbots for mental health care that rely on large language models (LLMs) to deliver custom, interactive support offer a pathway to expand access to high-value mental health care. These tools span a continuum from supportive dialogue and emotional coaching to clinical psychotherapy (Exhibit 2). As adoption accelerates, there is a need for safe, evidence-based, and clinically validated chatbot solutions capable of meeting patients' needs. This will require distinct functionality and oversight to support integration of these tools into the healthcare system.

Exhibit 2: Categories of Mental Health Chatbots and Product Differentiators

Type of Chatbot	Overview	Illustrative Differentiators
General Purpose LLMs	Generalist foundational models used for emotional support and conversation.	• Clear disclosure of nontherapeutic intent • Detection of crisis language with routing to human support • Continuous monitoring to detect misuse or safety issues
Emotional Support and Wellness Chatbots	Purpose-built models offering coaching and skills-based support.	• Demonstrated real-world and clinical evidence • Safety and escalation protocols, including escalating to a clinician • Continuous monitoring to evaluate clinical outcomes and detect safety issues
Clinical-Grade Psychotherapy Chatbots	Purpose-built models delivering psychotherapy to replicate or supplement human-based psychotherapy.	• Use of evidence-based modalities • Clinical validation compared with human-delivered therapy • Comprehensive safety models to detect and intervene, including escalating to a clinician • Traceable decision logs • Regulatory oversight • Continuous monitoring to evaluate clinical outcomes and detect safety issues

Emerging evidence and growing patient uptake of available tools demonstrate the potential of AI to expand access to mental health care. Studies[16] show that mental health chatbots can reduce short-term depression and anxiety symptoms, and the publication of the first RCT of a generative AI therapy chatbot[17] (Therabot) in 2025 marks an important milestone.

However, most users today turn to general purpose LLMs that were not designed to deliver clinical-grade psychotherapy. These models are estimated to be 50% more sycophantic than humans and can validate harmful or distorted thinking rather than challenge it—a core element of therapeutic practice.[18] Early research indicates that such behaviors can undermine safe and effective support.[19] Data from Open AI estimates that 1.2 million users per week engage in conversations indicating suicidal ideation and up to 0.15% of active users show explicit signs of potential suicide risk.[20] This highlights the need for safe, clinically validated tools.

Key Themes

Theme 1: Evidence standards should compare AI to current standards of care and scale with clinical risk.

How much evidence should be required, and compared to what?

There is agreement that the willingness of clinicians, patients, and policymakers to adopt clinical AI tools depends on the right body of evidence—but what kind, how much, and as compared to what? These foundational questions remain a barrier to widespread adoption. Evidence must be rigorous enough to build trust, while practical enough to avoid slowing innovation or limiting access.

> "In an ideal world, every organization would conduct local validation. But setting unrealistic evidence bars risks denying care to the people who need it most. What matters is whether a solution clears a reasonable threshold."

Participants felt strongly that the relevant comparison must be current, real-world access, experience, and outcomes—not idealized, guideline-driven care, which only about half of patients receive.[21] The appropriate baseline should be determined and evaluated on a local or population basis. For example, for many patients, the realistic alternative to a clinical-grade chatbot may be an unskilled therapist; nonvalidated, consumer-grade tools; or no care at all.

Across both use cases, participants expressed divergent views on the level and type of evidence required and emphasized that traditional evidence paradigms may not translate to adaptive, workflow-integrated AI systems. For example, RCTs may be infeasible for every model (much less so for every version of the model) given the pace of AI advancement. Evidence expectations may also vary widely across health systems. Highly resourced facilities (e.g., academic medical centers) may expect noninferiority data and local validation, whereas less-resourced systems may accept lower evidence thresholds when the care alternatives are demonstrably worse. Setting overly high standards risks slowing innovation and raising costs; however, without agreed-upon floors, it is difficult to build confidence.

In addition, participants felt that evidence should assess whether the full workflow (including multiple models, devices, and human oversight) improves outcomes, not merely model performance.

Modernizing regulatory frameworks for adaptive AI

Regulatory approaches need to shift from approving static model versions toward frameworks that assess safety and efficacy upfront and rely more heavily on robust postdeployment monitoring.

Recent FDA actions reflect this evolution. For example, the FDA:

- Piloted a company-level total product life cycle oversight model anchored in real-world performance monitoring; [22]
- Allows certain categories of model modifications without new submissions;[23]
- Established principles for ongoing monitoring and managing retraining risks;[24] and
- Requested feedback on approaches to postmarket monitoring and real-world evaluation of AI-enabled medical devices.[25]

Some stakeholders view these approaches as appropriate for adaptive AI; others continue to call for widespread RCT-level evidence.[26] It was also noted that traditional approaches to regulating AI using the FDA's Software as a Medical Device construct may be limiting and that other approaches, such as "AI licensure," should be explored. [27]

A "principles-based validation" construct was discussed as a potential alternative to current approaches that center on a static model and its specific intended use. Under a principles-based approach, a regulatory body or purchaser would define a principle set, establish evidence thresholds, conduct evaluation, enable continuous improvement, and conduct ongoing monitoring (Exhibit 3).

Exhibit 3: Principles-Based Evaluation for Hypertension "Safe Medication Adjustment"

Component	Illustrative Example – "Safe Medication Adjustment"
Define Principles	Define the principle set for autonomous hypertension management (e.g., safe titration logic, risk stratification, data accuracy, escalation protocols, longitudinal outcomes tracking).
Establish Evidence Thresholds, Tiered by Automation Level	Assistive AI tools may be evaluated through retrospective review of clinical appropriateness, whereas tools that independently initiate treatment may warrant prospective monitoring with explicit safety floors.
Validate Once, Apply Broadly	Once a company validates its titration logic, as an example, against the "safe medication adjustment" principle, future products using the same logic would only need to demonstrate conformance to the previously validated principle.
Enable Continuous Improvement	Because the principle (not the specific model version) is the unit of evaluation, underlying models can be updated without triggering full resubmission.
Ongoing Monitoring	Robust real-world monitoring systems (which could be AI-based) ensure that systems continue to perform well against the principles and escalate issues (e.g., model drift) in near-real time.

Theme 2: Performance benchmarks should be based on clinical outcomes, and safety standards should adapt as the evidence grows.

What should be measured, how should impact be defined, and how is progress monitored over time?

Workshop participants agreed that ambiguity around what constitutes "good" performance remains a barrier to adoption. Performance benchmarks serve multiple functions: they set goalposts for developers, guide purchasing decisions, and guide product evaluation. However, poorly designed benchmarks risk

> "Is it a high-fidelity decision? Is the output of the AI tool consistent with what we would expect a physician to do?"

obscuring real-world impacts. Participants emphasized that metrics should prioritize meaningful clinical outcomes, rather than proxies or process measures, and must encompass both predeployment validation and postdeployment monitoring.

At the same time, AI-enabled tools may enable more granular targets. For example, hypertension thresholds defined in 10 mmHg increments reflect historical limits of manual measurement, rather than clinically meaningful cut points.

Across both use cases, participants emphasized the need not only to set benchmarks but to set minimum safety floors, which could adjust dynamically over time on the basis of observed outcomes, changing patient risk profiles, emerging evidence, and clinical guidelines. Safety floors could become more restrictive if there is an observed increase in adverse events and more targeted as the safety profile and failure models of a given tool are more clearly understood. Selecting appropriate benchmarks requires balancing multiple considerations, outlined in Exhibit 4.

Exhibit 4: Considerations for Developing Appropriate AI Benchmarks

	Autonomous Hypertension Management	Mental Health Chatbots
Potential Clinical Outcome Performance Measures	**Examples:** • Time to blood pressure control • Accuracy of titration • Medication adherence rates • Medication reconciliation rates • Detection of high-risk patients	**Examples:** • Symptom improvement (e.g., PHQ-9, GAD-7) • Clinical fidelity across multiturn conversations • Patient engagement and retention • Crisis handling performance • Self-reported symptom reduction
Benchmark Design Considerations	• **Clinical Performance:** Benchmarks should be tied to clinical outcomes (e.g., time to control, therapeutic adherence, and improved risk scores). • **Weighting of Evaluation Criteria:** Different criteria, (e.g., ethical and professional conduct versus clinical outcomes) may warrant different levels of importance in an evaluation. • **Governance of Evaluation Criteria:** The decision around who sets evaluation criteria and how those criteria are managed and made transparent over time to end-users, is important.	• **AI-Driven Oversight:** Human review is not scalable for each autonomous action (e.g., each chatbot message, all agent reasoning steps that yield an action). Benchmarks should accommodate AI-assisted monitoring of clinical fidelity and safety.
Pre- vs. Post-Deployment	• **Benchmarks must be devised for both stages:** Postdeployment monitoring should enable continuous improvement, rather than "freezing" model versions.	

Minimum Safety Floors	• **Adverse Event Thresholds:** Clear thresholds for escalations, with additional protections for high-risk populations. • **Dynamic Safety Thresholds:** Safety floors that adjust over time on the basis of observed outcomes and individual patient risk profiles.	• **Safety Protocols:** Minimum safety floors may include escalation protocols for crisis situations and guardrails to prevent harmful parasocial attachments.
Implications for Clinical Guidelines	• **AI Automation Appropriateness:** Guidelines should be updated to reflect AI automation capabilities and to define the clinical scenarios where autonomous AI is indicated.	• **Defining Boundaries:** Clear distinctions are needed between general purpose AI use and clinical-grade interventions, with criteria for human escalation and transitions to higher or lower levels of care.

"Cars cannot be sold without seatbelts. What will be our 'seatbelt' requirements for AI in healthcare?"

Participants also noted a broader societal challenge: Emerging technologies are often held to a higher performance standard than existing systems. Self-driving cars, for example, face expectations of near perfection, rather than superiority to human drivers. Healthcare AI may encounter similar dynamics. A realistic assessment of current system failures is necessary to calibrate what level of AI performance is acceptable now and perhaps how it evolves in the future.

Theme 3: New technologies may be initially tested in lower-risk populations but should scale quickly to high-risk populations to maximize impact.

Where should the market focus early adoption efforts?

Workshop participants identified a strategic tension at the core of deployment planning: where to start? While lower-risk patients offer a "safer" entry point (e.g., fewer adverse events and lower-complexity management), the individual-level clinical and economic impact may be limited. High-need or high-cost patients represent a significant opportunity for outcome improvement and cost reduction, but they also present higher clinical risks and demand higher evidentiary expectations. An alternative framing emerged: deployment should by prioritized by opportunity to most greatly improve the status quo. Patients with limited access to care may derive the greatest benefit, irrespective of clinical complexity. Ultimately, views diverged on whether to adopt autonomous AI first in lower-risk populations or in high-risk populations.

Deployment considerations for priority populations

For mental health, participants emphasized that engagement[28] and retention[29] are barriers to effective treatment. Early evidence suggests that accessible digital tools, including general purpose LLMs, serve as an effective engagement mechanism to close treatment gaps for underserved communities, including LGBTQ+ populations facing disproportionate barriers to human-delivered care.[30] However, there have also been widely publicized cases of users communicating with chatbots before committing suicide.[31]

Workshop participants cautioned that overly restrictive deployment approaches risk limiting access and instead emphasized the need for appropriate care routing following LLM engagement.

For hypertension management, participants identified two priority segments for AI adoption: 1) newly diagnosed patients, who are a lower-complexity population for which timely intervention can prevent disease progression and downstream costs; and 2) underserved populations, for which autonomous solutions may meaningfully expand access despite higher clinical risk.

Exhibit 5: Summary of Deployment Pathways and Considerations

	Autonomous Hypertension Management	**Mental Health Chatbots**
The On-Ramp Tension	Tension between starting with lower-risk patients versus targeting high-need, high-cost patients for whom the benefit is greatest but evidentiary and risk expectations are higher.	Starting with mild-to-moderate symptom populations presents lower risk, but the greatest unmet need lies in patients with more-severe conditions or limited access to care.
Access-Based Framing	For patients with limited access to care or few alternatives, autonomous tools could deliver the greatest benefit regardless of clinical risk level.	Millions already use general purpose AI models for mental health support, often because of access and affordability barriers to traditional therapy. This population is opting into AI-delivered care.
Graduated Autonomy Model	Advancing toward more autonomous tools incrementally could follow the model of training clinical residents. AI systems could earn greater independence as they demonstrate competence over time.	Chatbots could begin with supportive, nonclinical functions and progress toward more clinically meaningful interventions as evidence accumulates.
Connecting General Purpose AI to Clinical-Grade Solutions	Foster connectivity between consumer-grade, LLM-based care management capabilities (e.g., ChatGPT Health, Claude for Healthcare) and purpose-built tools for managing hypertension that include more advanced capabilities, like prescribing and medication titration.	A central challenge is bridging the gap between general purpose AI tools that patients already use for mental health and clinical-grade solutions with appropriate safety monitoring and escalation pathways.

Theme 4: Widespread adoption will depend on building clinician confidence, gaining clarity about legal liability, and aligning payment models

What market and policy changes are needed to enable widespread adoption?

Even when the evidence demonstrates strong performance, the pace of autonomous AI's adoption is constrained by institutional instincts and norms, unresolved liability questions, and misaligned incentives. Technological capabilities increasingly outpace organizational readiness. Many health systems are uncomfortable with Level 3 or Level 4 autonomy because the shift from "some human involvement" to "little-to-no human involvement" is a considerable operational and psychological leap. These tensions impact regulatory frameworks, the evolution of clinicians' roles, and the dynamics that dictate adoption.

Building the regulatory pathway for clinical AI

Workshop participants agree that regulation must first evolve to permit higher levels of autonomy. For example:

- Frameworks for delegated prescribing authority to autonomous tools
- Documentation standards for autonomous clinical decision making
- Payment mechanisms that reimburse AI-assisted or AI-only clinical work
- Professional liability and licensure obstacles

Utah's Evolving Regulatory Landscape

Utah approved a novel pilot that allows an AI system to provide routine, low-risk prescription renewals without any human involvement for patients with chronic conditions.[32] One hundred and ninety commonly prescribed medications are included in the program administered by Doctronic; however, certain prescriptions—such as those for pain management, ADHD, and injectables—are excluded for safety reasons.

Defining responsibilities in hybrid AI-human clinical workflows

As clinical care evolves toward hybrid AI and human teams, greater clarity is needed on how responsibilities should be allocated. Participants broadly accept AI for tasks like structured data collection, patient education, care-plan reinforcement, and medication reconciliation, while caveating that data sharing across institutions must continue to improve.

There was less consensus on topics related to the degree of and approach to clinician oversight and quality control for models' adjudication of clinical logic. The appropriate amount of hands-on clinical involvement may also differ across patient cohorts. As one workshop participant noted, some patients will still want to speak with their clinician before a medication or care-plan change; however, other patients have less trust in the traditional healthcare system and may be more inclined to interact directly with AI.

New entrants can accelerate adoption

New entrants may accelerate adoption by challenging incumbents. Startups and nontraditional players—unburdened by legacy workflows and building with autonomy principles from day one—are entering the market and can reshape how clinical work gets done. This may put pressure on incumbent, more risk-averse institutions and drive regulatory and cultural evolution.

Exhibit 6: Market and Policy Enablers for Adoption

	Autonomous Hypertension Management	Mental Health Chatbots
Transparency and Accountability	Visibility into AI rationale, monitoring plans, and clear lines of accountability before adoption.	Similar expectations apply, though high-volume, asynchronous interactions and limitations of understanding "why" generative AI produces a particular response complicates traditional oversight models. AI-assisted monitoring was discussed as a potential solution.
Sandboxes for Evidentiary Advancement	To maintain focus on clinical performance rather than on "how" or "why" a recommendation is generated, sandboxes will be an increasingly important strategy to enable evidentiary advancement without requiring direct human involvement in every decision.	Sandboxes could similarly enable testing of chatbot performance in controlled environments before broader deployment.
Role Definition	Discussion emphasized the need to define clear roles and workflows delineating where humans add value versus where AI can operate independently.	There is a parallel need to clarify boundaries, particularly regarding the handoff between AI-delivered support and human clinician intervention.

The following issues, while outside the scope of this discussion, were raised by participants:

- **Professional Liability:** Liability remains unresolved, with unclear accountability for autonomous clinical decisions contributing to hesitation. Historically, most technologies have operated under the controlling clinician's liability. To enable higher levels of autonomy, innovators may need to assume liability for autonomous system actions. Similar concerns apply to mental health chatbots, including responsibility for failure to escalate acute risk.
- **Payment Models:** Current reimbursement structures do not typically reward efficiency gains or clinical improvements from autonomous AI tools. Participants identified a need for payment models that support development and scaling of clinical-grade AI.

Looking ahead

Upcoming convenings on administrative AI and payment for AI

This convening represents the first in a series of three workshops on AI adoption in healthcare. Future sessions will focus on administrative AI and payments for AI. PHTI will continue to publish findings across workshops and engage stakeholders in developing actionable frameworks that balance a rapid pace of innovation with patient safety and system sustainability in pursuit of improving quality and reducing cost.

References

[1] Centers for Disease Control and Prevention. “Hypertension Prevalence in the U.S. | Million Hearts®.” May 12, 2023. https://millionhearts.hhs.gov/data-reports/hypertension-prevalence.html.

[2] Park, Chanhyun, Guijing Wang, Jefferey M. Durthaler, and Jing Fang. “Cost-Effectiveness Analyses of Antihypertensive Medicines: A Systematic Review.” *American Journal of Preventive Medicine* 53, no. 6 Suppl 2 (2017): S131–42. https://doi.org/10.1016/j.amepre.2017.06.020.

[3] Centers for Disease Control and Prevention. “Hypertension Prevalence in the U.S. | Million Hearts®.” May 12, 2023. https://millionhearts.hhs.gov/data-reports/hypertension-prevalence.html.

[4] Substance Abuse and Mental Health Services Administration. “Highlights for the 2022 National Survey on Drug Use and Health.” February 13, 2025. https://www.samhsa.gov/data/sites/default/files/reports/rpt42731/2022-nsduh-main-highlights.pdf.

[5] Meiselbach, Mark K., Catherine K. Ettman, Karen Shen, Brian C. Castrucci, and Sandro Galea. “Unmet Need for Mental Health Care Is Common across Insurance Market Segments in the United States.” *Health Affairs Scholar* 2, no. 3 (2024): qxae032. https://doi.org/10.1093/haschl/qxae032.

[6] Health Resources & Services Organization. “State of the Behavioral Health Workforce, 2025.” December 2025. https://bhw.hrsa.gov/sites/default/files/bureau-health-workforce/data-research/Behavioral-Health-Workforce-Brief-2025.pdf.

[7] Fryar, Cheryl, Brian Kit, Margaret Carroll, and Joseph Afful. “Hypertension Prevalence, Awareness, Treatment, and Control Among Adults Age 18 and Older: United States, August 2021–August 2023.” CDC Stacks, National Center for Health Statistics (U.S.), October 23, 2024. https://stacks.cdc.gov/view/cdc/164016.

[8] Hicks, Leroi S., Thomas D. Sequist, John Z. Ayanian, et al. “Impact of Computerized Decision Support on Blood Pressure Management and Control: A Randomized Controlled Trial.” *Journal of General Internal Medicine* 23, no. 4 (2008): 429–41. https://doi.org/10.1007/s11606-007-0403-1.

[9] Samal, Lipika, John L. Kilgallon, Stuart Lipsitz, et al. “Clinical Decision Support for Hypertension Management in Chronic Kidney Disease: A Randomized Clinical Trial.” JAMA Internal Medicine 184, no. 5 (2024): 484–92. https://doi.org/10.1001/jamainternmed.2023.8315.

[10] Leitner, Jared, Po-Han Chiang, Parag Agnihotri, and Sujit Dey. “The Effect of an AI-Based, Autonomous, Digital Health Intervention Using Precise Lifestyle Guidance on Blood Pressure in Adults With Hypertension: Single-Arm Nonrandomized Trial.” *JMIR Cardio* 8 (May 2024): e51916. https://doi.org/10.2196/51916.

[11] Hicks, Leroi S., Thomas D. Sequist, John Z. Ayanian, et al. “Impact of Computerized Decision Support on Blood Pressure Management and Control: A Randomized Controlled Trial.” *Journal of General Internal Medicine* 23, no. 4 (2008): 429–41. https://doi.org/10.1007/s11606-007-0403-1.

[12] Samal, Lipika, John L. Kilgallon, Stuart Lipsitz, et al. “Clinical Decision Support for Hypertension Management in Chronic Kidney Disease: A Randomized Clinical Trial.” JAMA Internal Medicine 184, no. 5 (2024): 484–92. https://doi.org/10.1001/jamainternmed.2023.8315.

[13] Margolis, Karen L., Anna R. Bergdall, A. Lauren Crain, et al. “Comparing Pharmacist-Led Telehealth Care and Clinic-Based Care for Uncontrolled High Blood Pressure: The Hyperlink 3 Pragmatic Cluster-Randomized Trial.” *Hypertension (Dallas, Tex.: 1979)* 79, no. 12 (2022): 2708–20. https://doi.org/10.1161/HYPERTENSIONAHA.122.19816.

[14] U.S. Food and Drug Administration. "510(k) Premarket Notification - G0 Blood Pressure Monitoring System (G0)." July 2, 2025. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K250415.

[15] Advanced Research Projects Agency for Health. "ARPA-H to Revolutionize Cardiovascular Disease Management with Clinical Agentic AI | ARPA-H." January 13, 2026. https://arpa-h.gov/news-and-events/arpa-h-revolutionize-cardiovascular-disease-management-clinical-agentic-ai.

[16] Li, Han, Renwen Zhang, Yi-Chieh Lee, Robert E. Kraut, and David C. Mohr. "Systematic Review and Meta-Analysis of AI-Based Conversational Agents for Promoting Mental Health and Well-Being." *NPJ Digital Medicine* 6, no. 1 (2023): 236. https://doi.org/10.1038/s41746-023-00979-5.

[17] Heinz, Michael V., Daniel M. Mackin, Brianna M. Trudeau, et al. "Randomized Trial of a Generative AI Chatbot for Mental Health Treatment." *NEJM AI* 2, no. 4 (2025). https://doi.org/10.1056/AIoa2400802.

[18] Cheng, Myra, Cinoo Lee, Pranav Khadpe, Sunny Yu, Dyllan Han, and Dan Jurafsky. "Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence." October 1, 2025. https://doi.org/10.48550/arXiv.2510.01395.

[19] Arnaiz-Rodriguez, Adrian, Miguel Baidal, Erik Derner, et al. "Between Help and Harm: An Evaluation of Mental Health Crisis Handling by LLMs." December 2, 2025. https://arxiv.org/html/2509.24857v2.

[20] O'Dowd, Adrian. "ChatGPT: More than a Million Users Show Signs of Mental Health Distress and Mania Each Week, Internal Data Suggest." The BMJ, October 30, 2025. https://doi.org/https://doi.org/10.1136/bmj.r2290.

[21] McGlynn, Elizabeth A., Steven M. Asch, John Adams, et al. "The Quality of Health Care Delivered to Adults in the United States." *The New England Journal of Medicine* 348, no. 26 (2003): 2635–45. https://doi.org/10.1056/NEJMsa022615.

[22] Food and Drug Administration. "Digital Health Software Precertification (Pre-Cert) Pilot Program." July 23, 2025. https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-pilot-program.

[23] Food and Drug Administration. "Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions." August 18, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence.

[24] Food and Drug Administration. "Good Machine Learning Practice for Medical Device Development: Guiding Principles." December 19, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles.

[25] Food and Drug Administration. "Request For Public Comment: Measuring and Evaluating Artificial Intelligence-Enabled Medical Device Performance in the Real-World." September 30, 2025. https://www.fda.gov/medical-devices/digital-health-center-excellence/request-public-comment-measuring-and-evaluating-artificial-intelligence-enabled-medical-device.

[26] Khozin, Olivier Elemento, Cora N. Sternberg, Sean. "Why Silicon Valley Should Demand Clinical Trials for Its Medical AI." *STAT*, August 28, 2025. https://www.statnews.com/2025/08/28/medical-ai-randomized-controlled-clinical-trials-rcts/.

[27] Bressman, Eric, Carmel Shachar, Ariel Stern, and Ateev Mehrotra. "Software as a Medical Practitioner—Is It Time to License Artificial Intelligence?" *JAMA Intern Med.* 2026;186(1):5–6. https://doi.org/10.1001/jamainternmed.2025.6132.

[28] Mojtabai, Ramin, Mark Olfson, Nancy A Sampson, et al. "Barriers to Mental Health Treatment: Results from the National Comorbidity Survey Replication." Psychological Medicine 41, no. 8 (2011): 1751–61. https://doi.org/10.1017/S0033291710002291.

[29] Olfson, Mark, Ramin Mojtabai, Nancy A. Sampson, et al. "Dropout from Outpatient Mental Health Care in the United States." Psychiatric Services (Washington, D.C.) 60, no. 7 (2009): 898–907. https://doi.org/10.1176/ps.2009.60.7.898.

[30] Ma, Zilin, Yiyang Mei, Yinru Long, Zhaoyuan Su, and Krzysztof Z. Gajos. "Evaluating the Experience of LGBTQ+ People Using Large Language Model Based Chatbots for Mental Health Support." *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024, 1–15. https://doi.org/10.1145/3613904.3642482.

[31] Hill, Kashmir Hill. *Lawsuits Blame ChatGPT for Suicide and Harmful Delusions*, The New York Times, November 6, 2025. https://www.nytimes.com/2025/11/06/technology/chatgpt-lawsuit-suicides-delusions.html.

[32] Utah Department of Commerce. "NEWS RELEASE: Utah and Doctronic Announce Groundbreaking Partnership for AI Prescription Medication Renewals." January 6, 2026. https://commerce.utah.gov/2026/01/06/news-release-utah-and-doctronic-announce-groundbreaking-partnership-for-ai-prescription-medication-renewals/.

PETERSON HEALTH TECHNOLOGY INSTITUTE