ASPEN POLICY ACADEMY
aspen institute

Policy Brief

Monitoring AI tools in Healthcare for Developers and Implementers

By Arfa Rehman

January 2026

This policy project was developed by an Aspen Policy Academy fellow while participating in the Science and Technology Policy Fellowship. This policy brief is intended for developers and implementers of AI-enabled clinical decision support tools, including health systems, vendors, and technology partners involved in deploying and managing these tools in clinical settings. The full project is available here. Please note that the author's opinions published here are their own. This publication does not reflect the views of the Aspen Policy Academy or the Aspen Institute.

Executive Summary

Developers and implementers of AI-enabled clinical decision support (AI-CDS) tools should adopt a rigorous approach to monitoring these tools after they are deployed. This project developed and recommends the adoption of the **TRACE monitoring framework, which provides practical guidance for tracking real-world performance, safety, and equity of AI-CDS tools**. While much of today's guidance focuses on AI governance, risk assessments, and evaluating models before deployment, TRACE fills a critical gap by offering actionable guidance for what happens after these tools go live. Without post-deployment monitoring, AI-CDS tools may be used in unintended ways, fall out of sync with clinical guidelines, or produce uneven results across patient populations — undermining safety, trust, and the value these tools are meant to deliver. TRACE gives organizations a clear roadmap for what to monitor and how — helping them anticipate potential issues, detect emerging risks, and establish processes to respond effectively when problems arise. Ultimately, this reduces safety and performance risks, supports stronger clinician and leadership trust, and helps meet expectations from payers, accreditors, and regulators.

Background

AI-powered clinical decision support (AI-CDS) tools are increasingly being integrated into clinical workflows for tasks such as suggesting diagnoses and recommending next steps for patients. There is limited post-deployment monitoring of these tools, however. Many of these tools are not subject to regulation by the US Food and Drug Administration (FDA)

because they do not meet the definition of a regulated medical device. Even among the tools that are FDA regulated, few have formal mechanisms for postmarket surveillance. One recent analysis found that only 9% of FDA-cleared AI-enabled devices have a postmarket surveillance plan.

> *"Developers and implementers should integrate post-deployment monitoring into their tools from the start."*

This lack of monitoring presents significant risks for AI-CDS tools based on large language models (LLMs). Unlike traditional software, which is deterministic and produces the same output for a given input, LLMs are probabilistic — meaning their outputs can vary based on changes in input, context, and updates to the underlying model. As a result, LLM-based tools require continuous monitoring to ensure that their outputs remain clinically appropriate. Without monitoring, developers and implementers of AI tools may be unable to detect when performance degrades, when clinicians begin to overrely on a tool, or when a tool produces uneven results across patient populations.

Multiple studies highlight risks that post-deployment monitoring could help address. For example, one case involving an AI tool for sepsis detection found that the system often failed to align with clinician workflows or needs, resulting in low adoption and trust over time. Without ongoing monitoring of tool usage and override behavior, organizations may struggle to identify and act on such misalignments early. In another study, researchers observed automation bias in AI-assisted tasks, where clinicians, under time pressure, were more likely to accept incorrect model outputs. Monitoring clinician interactions with AI tools could provide early signals of inappropriate overreliance.

While there are a number of established frameworks for AI risk and safety, few have addressed the specific challenges of post-deployment monitoring, particularly in clinical settings. There remains a gap in operational guidance detailing what should be monitored and how monitoring should be implemented. Recent studies and expert consensus highlight the need for robust monitoring protocols that provide oversight of clinical AI tools in real-world environments.

Recommendations

To ensure the safe, effective, and equitable use of AI-CDS tools in clinical settings, developers and implementers should integrate post-deployment monitoring into their tools from the start. This means designing tools with real-world oversight in mind — defining what will be monitored, how performance will be measured over time, and how issues will be managed when they arise.

This project recommends applying the TRACE framework as a practical foundation for building this monitoring approach. TRACE outlines 5 essential domains that, together, ensure comprehensive oversight:

1. **Technical Integrity:** Is the tool functioning reliably and consistently across time and settings?
2. **Real-World Use:** How are clinicians actually using the tool, and is that use meeting expectations?
3. **Alignment and Accuracy:** Are the tool's outputs clinically sound and consistent with current guidelines?
4. **Clinical Fairness:** Does the tool work equally well across patient subgroups, or are disparities emerging?
5. **Escalation and Safety Response:** When problems arise, is there a process in place to detect, investigate, and respond?

Developers and implementers should adopt the TRACE monitoring framework because it is specifically designed to be actionable and adaptable to different clinical environments. Rather than requiring teams to design monitoring protocols from scratch, TRACE offers a clear, implementation-ready framework that can be integrated into existing governance and quality improvement processes. It is grounded in real-world challenges and provides concrete metrics, practices, and governance considerations that can be tailored to the tool and context.

Other high-stakes fields have long relied on post-deployment monitoring frameworks, such as aviation's Safety Management System, good pharmacovigilance practices in drug safety, and site reliability engineering practices in software engineering, to ensure that technological solutions remain safe and effective in real-world conditions. TRACE would bring a structured approach to AI-CDS, filling a critical gap in healthcare AI governance.

Specifically, AI-CDS developers and implementers should adopt the following practices:

1. Develop a TRACE-aligned monitoring plan before deployment

Post-deployment monitoring should be incorporated as a defined workstream within the overall development life cycle of an AI-CDS tool. In other safety-critical industries such as aviation and pharmaceuticals, postmarket monitoring is standard practice. In software engineering, observability and error tracking are at the core of maintaining systems in production. AI tools deployed in clinical environments should follow the same logic. Planning for monitoring in advance enables organizations to specify the scope, methodology, data sources, and governance processes that will be used once the tool is operational. It ensures that the necessary infrastructure and institutional roles are in place to evaluate performance over time, rather than reacting only after problems surface.

The TRACE framework provides a practical guide for this planning, outlining key monitoring domains that help ensure all relevant dimensions of real-world tool behavior are considered. Establishing a monitoring plan early also facilitates alignment with internal leadership and compliance teams, integration into workflows, and preparation for regulatory, accreditation, or payer requirements.

"Developers and implementers should adopt the TRACE monitoring framework because it is specifically designed to be actionable and adaptable to different clinical environments."

2. Define monitoring metrics as key performance indicators

Monitoring plans should specify a set of key performance indicators (KPIs) that will be used to evaluate the AI-CDS tool in real-world settings. Establishing KPIs is a widely accepted practice across industries, including software development, healthcare quality improvement, and safety-critical systems, where measurement is essential for ensuring performance and accountability over time. Applying this principle to AI-CDS tools is essential given their complexity, context sensitivity, and potential to change behavior over time.

The TRACE framework includes example KPIs that can be adapted to different organizational contexts. These metrics correspond to common implementation risks and are feasible to measure in most clinical environments. Making them part of routine performance review processes enables early detection of degradation, inequitable

outcomes, or changes in utilization. By tracking these metrics, developers and implementers will have necessary evidence to support decisions on tool maintenance, modification, or retirement.

3. Establish clear accountability and governance for monitoring

A monitoring framework is only effective if roles and responsibilities for its execution are clearly defined and consistently followed. Organizations should assign accountability for each stage of the process, including data collection, metric calculation, review of findings, and identified through monitoring may be overlooked, delayed, or inconsistently addressed — undermining both patient safety and clinician trust. Responsibilities should be documented in internal governance policies and aligned with existing protocols for escalating clinical or safety concerns. The TRACE framework supports this process by helping organizations clarify who is responsible for monitoring across each domain and by encouraging inclusion of both technical and clinical stakeholders in decisionmaking. Establishing clear accountability structures reduces the risk of inaction when problems are detected and enables more coordinated and transparent oversight over time.

If you'd like to learn more, see the full project, including the TRACE framework and an FAQ guide, at aspenpolicyacademy.org/project/ai-cds-monitoring-2025.

aukidphumsirichat via Canva

Sezeryadigar from Getty Images Signature via Canva

ASPEN POLICY ACADEMY
aspen institute

aspen institute

About the Aspen Policy Academy

The Aspen Institute's Policy Academy helps community leaders and experts across the political spectrum elevate their voices, influence key decisions, and strengthen democracy from the ground up. Our innovative training programs and resources equip people across sectors – from tech to the environment, science to civic engagement – with the skills to shape critical policy efforts. Learn more at aspenpolicyacademy.org.