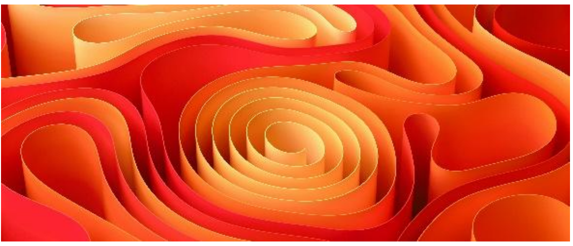
Pills

Regulatory/Supervisory Pills | N.45 March 2026

Author: Dario Esposito



TOPICS:

Technology

SOURCE:

National Institute of Standards and Technology

NIST: Challenges to the Monitoring of Deployed AI Systems

- The report examines **the growing need for systematic monitoring of artificial intelligence after deployment** and identifies key barriers, gaps and open questions in implementing effective monitoring frameworks.
- A central premise of the report is that **pre-deployment testing alone is insufficient** to ensure AI reliability and safety. Because AI systems operate in dynamic real-world environments and often produce non-deterministic outputs, monitoring after deployment is necessary to validate system performance, detect unexpected behaviors, and identify downstream impacts arising from changing inputs or contexts. Post-deployment monitoring also enables organizations to feed operational insights back into system development and risk management processes.
- The report introduces **six primary categories of monitoring** designed to structure the evaluation of deployed AI systems. These include: **Functionality monitoring**, ensuring the system continues to perform as intended; **Operational monitoring**, assessing infrastructure reliability and service consistency; **Human factors monitoring**, evaluating transparency, user interaction, and output quality; **Security monitoring**, identifying vulnerabilities, adversarial attacks, or misuse; **Compliance monitoring**, ensuring adherence to regulatory and governance requirements; **Large-scale impacts monitoring**, assessing broader societal or systemic consequences.
- Across these categories, the report highlights **cross-cutting challenges** that impede effective monitoring. These include the lack of standardized monitoring methods and trusted tools, limited visibility into model internals and training data, rapid technological change, and organizational barriers such as resource constraints and misaligned incentives. Financial costs, workforce expertise shortages, and privacy considerations further complicate monitoring efforts.
- The study also identifies **category-specific difficulties**, such as detecting model drift, collecting reliable ground-truth data, managing fragmented logging across distributed systems, understanding human-AI feedback loops, and navigating complex regulatory environments. Monitoring large-scale societal impacts is particularly challenging due to difficulties in defining meaningful metrics and tracking downstream effects of widely distributed models.