THOUGHT LEADERSHIP SERIES

The Deployer's Dilemma: Autonomous Auditing and the Hidden Risks of Third-Party AI Under the EU AI Act

Part 3 of 4: Governing the Users of AI

April 2026

Designed for a procurement world where vendor SOC 2 reports and contractual warranties were sufficient, traditional governance assumes purchasers can rely entirely on the producer's representations of safety and compliance. This assumption collapses when the purchased system is an AI model, which is probabilistic by nature and often continuously evolving in production, thereby carrying legal risk that shifts based entirely on the deployer's specific use case.

For deployers, the structural weakness of traditional auditing is not merely a lack of visibility into the vendor's black-box model. It is the unprecedented regulatory reality that the deployer remains legally accountable for how that model is supervised, monitored, and applied within its own operations, despite having no control over the model's underlying architecture.

These conditions render periodic manual auditing fundamentally inadequate. Deployers require a governance mechanism that operates continuously at the point of use, captures immutable evidence of human oversight as decisions are made, detects use-case drift in real time, and proves to regulators that legal obligations are met in practice, not just promised in policy.

As established in Parts 1 and 2 of this series, the convergence of digital transformation, regulatory explosion, and the unique probabilistic nature of Artificial Intelligence has rendered traditional manual auditing structurally inadequate. As a result, autonomous auditing, which operates in the background and is triggered by operational events, is replacing traditional manual auditing.

Series Context

This third paper shifts the focus from the creators of AI models to the organizations that use them. Where Part 2 examined the governance burden of those who build AI systems, this paper examines the severe, and often underestimated, legal burden of those who procure and deploy them. Procuring a compliant third-party AI model from a major vendor does not absolve the deploying enterprise of regulatory liability. It introduces new categories of risk.

1. The Illusion of Outsourced Accountability

Traditional SaaS procurement operates on a simple, well-understood assumption: if the vendor possesses a SOC 2 report and adheres to security best practices, the purchasing organization is protected. The EU AI Act destroys this assumption.

Under the EU AI Act, a Deployer is defined as any entity using an AI system under its own authority [1]. The regulatory anchor is clear: while the AI Provider is legally responsible for the model's inherent technical safety and initial conformity assessments, the Deployer is legally accountable for how that model behaves in its specific operational context.

The core problem is that deployers inherit the risk profile of a mathematical system they did not train, cannot inspect, and cannot fully control. In essence, deployers suffer from the strict limitations of black-box access [2]. They can only see the outputs generated by their inputs; they do not have white-box access to the model's internal weights or outside-the-box access to the original training data [2]. Compounding this technical limitation is a pervasive liability trap: AI vendors consistently blame deployers for improper use, while deployers blame vendors for flawed models [3]. Furthermore, corporate legal and compliance teams sometimes exhibit systemic willful blindness, operating under the dangerous assumption that avoiding deep audits will protect them from legal discovery [4].

The EU AI Act systematically dismantles this assumption. If a retail bank uses a third-party vendor's API for credit scoring and the system produces discriminatory outcomes, the bank may face regulatory consequences, fines, and reputational damage. The defense that 'our vendor handles compliance' is legally obsolete. The deployer may be held accountable for outcomes generated by a system whose internal operations are fundamentally inscrutable to it [2].

2. Article 26 and the Impossibility of Manually Auditing Human Oversight

The EU AI Act expressly mandates that high-risk AI systems must be subject to genuine human oversight to prevent automation bias and algorithmic harm [5]. Specifically, Article 26 outlines the obligations of deployers, demanding that they assign competent natural persons to oversee the system. These overseers must be equipped with the training, authority, and support necessary to correctly interpret the system's output and, critically, must have the power to decide not to use the system or to disregard its output entirely (Article 14).

As established by Article 26, the nature of human-in-the-loop participation is specific: the human role is to oversee the functioning of the system rather than to interact with it mechanically, for example by continuously providing feedback on the accuracy of the results as

part of an ongoing learning process. At the same time, such supervision must be genuine and not merely formal [9].

The audit challenge this presents is immense. A traditional manual audit might simply verify whether a human-in-the-loop policy exists in the corporate handbook. But how does a periodic auditor prove to a regulator that the human reviewer was exercising independent judgment? The core vulnerability is automation bias: the well-documented psychological tendency of human overseers to blindly trust the machine's recommendations, effectively reducing human-in-the-loop oversight to a rubber-stamping exercise. If an auditor simply verifies that a human clicked approve, they have proven nothing about the cognitive rigor of that approval. Manual oversight becomes an illusion of accountability rather than a genuine safeguard.

> **The Autonomous Auditing Solution**
>
> Autonomous auditing continuously monitors the deployer's operational processes in the background. It generates cryptographic evidence showing exactly when a human reviewed an output, the latency of the review, whether the human altered the decision, and whether oversight matched the provider's instructions for use. It captures the empirical reality of oversight, neutralizing the risks of automation bias by mathematically proving whether the human actively engaged with the decision or merely rubber-stamped it.

3. The Data Black Hole: Feeding the Provider's Post-Market Monitoring

The EU AI Act creates an interdependent compliance relationship between AI creators and AI users. Under Article 72, providers must run a post-market monitoring system to track their model's real-world performance [6]. However, providers rely on deployers to feed them the relevant data necessary to make this monitoring possible.

This creates a profound data black hole and a privacy paradox for the deployer. How does a deployer know precisely what telemetry to send back? More importantly, how does it prove to regulators that it supplied accurate performance data without exposing GDPR-protected customer data, personally identifiable information (PII), or proprietary trade secrets to the third-party AI provider?

Academic legal analysis highlights that this tension between mandatory compliance auditing and the protection of trade secrets represents one of the more significant unresolved conflicts in the AI value chain [4].

> **The Autonomous Auditing Solution**
>
> Autonomous auditing systems sit transparently at the API boundary. They automatically capture, sanitize, and log the input/output telemetry required for Article 72 compliance without human intervention. This creates an immutable audit trail proving the deployer fulfilled its data-sharing obligations securely, without relying on manual data extracts that risk severe privacy or trade-secret violations.

4. Use-Case Creep and Substantial Modifications

An AI system is certified by its provider for a specific intended purpose. Under Article 25 of the EU AI Act, a deployer legally becomes a Provider, inheriting the original vendor's compliance

burden, if it: (1) places its own trademark on an existing high-risk system; (2) makes a substantial modification to a high-risk system, such as changing the method of data processing, processing rules, or rules of human supervision; or (3) changes the intended purpose of a standard AI system so drastically that it becomes high-risk [7, 9].

The audit challenge is that use-case creep happens slowly and informally. Consider an HR department that procures a third-party AI tool to perform basic keyword sorting on CVs. Months later, the department quietly begins using the same tool's outputs to predict employee flight risk. By altering the deployment context, the organization crosses into a high-risk categorization. Deployers often attempt to mitigate this risk by relying on shallow black-box testing, simply prompting the API to see if outputs look acceptable. Academic researchers warn that these inadequate testing norms are becoming dangerously sticky across the industry, creating a false sense of security that will fail under formal regulatory scrutiny [5]. A periodic manual audit will entirely miss this use-case transition until the organization is already in breach [8].

> **The Autonomous Auditing Solution**
> Autonomous auditing continuously maps actual user prompts, API calls, and system outputs against the permitted instructions for use provided by the vendor. If usage drifts outside the certified boundaries, the system flags the violation instantly, preventing the deployer from accidentally reclassifying itself as a Provider.

4.1 Incident Response: Why MLOps Is Insufficient

Deployers frequently assume that if their engineering team has robust MLOps monitoring in place, they are covered for incident reporting. This is a dangerous misconception. MLOps tells you that the model broke; autonomous auditing proves to the regulator that you followed the law after it broke [8].

Crucially, deployers must recognize what triggers this obligation. Under Article 3(49), a serious incident is an event leading to: (a) death or serious harm to health; (b) serious and irreversible disruption of the management or operation of critical infrastructure; (c) a breach of Union law protecting fundamental rights; or (d) serious damage to property or the environment. When these occur, deployers must immediately inform the provider, the importer/distributor, and market surveillance authorities. To ensure investigations into these incidents are possible, deployers are further obligated under Article 26(6) to retain automatically generated event logs for a minimum of six months whenever those logs are under their control, for example when the AI operates within the deployer's infrastructure rather than as an outsourced external service [9].

However, as legal scholars have noted, internal technical and legal teams face an inherent conflict of interest: they may be structurally incentivized to quietly hot-patch a failing model or avoid documenting the severity of an issue to evade liability [4]. Hot-patching a failing system without freezing the system's state risks evidence spoliation and can undermine the logging and incident-reporting obligations that the EU AI Act depends on. Autonomous auditing reduces this conflict of interest by sitting above the MLOps layer. It automatically enforces the incident-reporting timeline, freezes the system state at the moment of failure, and generates the compliance logs required to defend the deployer's actions.

5. Conclusion: Automating the Deployer's Defense

The EU AI Act makes AI procurement a continuous liability rather than a one-off transaction. Manual audits check the vendor contract; autonomous auditing continuously checks the operational reality.

For deployers, autonomous auditing is not just a compliance tool; it is the primary defense mechanism against inheriting regulatory penalties for third-party AI failures. By operating invisibly in the background, it proves the deployer did everything legally required of it when using a system it did not build.

Board Action Items

1. Mandate an API Boundary Audit: Direct the CIO to immediately inventory all third-party AI models currently accessed via API and map their actual internal use cases against the vendors' permitted instructions for use to identify use-case creep.
2. Commission an Article 26 / Article 14 Capability Assessment: Require compliance teams to demonstrate exactly how the organization will prove that human-in-the-loop reviewers are exercising independent judgment rather than suffering from automation bias.
3. Automate Telemetry Sanitization: Implement automated, machine-readable data sanitization at the API boundary to fulfill Article 72 post-market monitoring obligations without leaking GDPR-protected data or corporate trade secrets to upstream vendors.
4. Revise IT Incident Response Frameworks: Formally update the organization's cyber incident response plan to prohibit hot-patching live AI models without first triggering an automated compliance state freeze.
5. Deploy Continuous Auditing (CAAI): Shift vendor risk management from annual security questionnaires to Continuous Auditing of AI Systems, ensuring behavioral drift in upstream foundational models is detected in milliseconds rather than months [8].

Sources

1. EU AI Act (Regulation 2024/1689) - Article 3: Definitions ('Deployer') - https://artificialintelligenceact.eu/article/3/
2. S. Casper, C. Ezell, C. Siegmann, N. Kolt, T. L. Curtis, B. Bucknall, A. Haupt, K. Wei, J. Scheurer, M. Hobbhahn, L. Sharkey, S. Krishna, M. Von Hagen, S. Alberti, A. Chan, Q. Sun, M. Gerovitch, D. Bau, M. Tegmark, D. Krueger and D. Hadfield-Menell, The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 3-6, 2024, Rio de Janeiro, Brazil - https://dl.acm.org/doi/10.1145/3630106.3659037
3. S. Costanza-Chock, E. Harvey, I. D. Raji, M. Czernuszenko and J. Buolamwini, 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22) - https://arxiv.org/abs/2310.02521
4. E. A. Farley and C. R. Lansang, Harvard Journal of Law & Technology 38, Digest Spring 2025 - https://jolt.law.harvard.edu/digest/ai-auditing-first-steps-towards-the-effective-regulation-of-artificial-intelligence-systems
5. EU AI Act (Regulation 2024/1689) - Article 26: Obligations of deployers of high-risk AI systems - https://artificialintelligenceact.eu/article/26/
6. EU AI Act (Regulation 2024/1689) - Article 72: Post-market monitoring by providers - https://artificialintelligenceact.eu/article/72/
7. EU AI Act (Regulation 2024/1689) - Article 25: Responsibilities along the AI value chain - https://artificialintelligenceact.eu/article/25/
8. M. Minkkinen, J. Laine and M. Mäntymäki, DISO 1, 21 (2022) - https://link.springer.com/article/10.1007/s44206-022-00022-2
9. K. Kiejnich-Kruk, Ius Novum 19(4), 121 (2025) - https://reference-global.com/download/article/10.2478/in-2025-0040.pdf