PUBLIC COMMENT

JULY 30, 2026

AI POLICY STATEMENT; MATTER NO. P264200

Comment of Integro Labs LLC

FTC Proposed Policy Statement on the Suppression of Accuracy in AI Systems

Docket FTC-2026-0859 | 91 FR 41638

To the Federal Trade Commission:

Integro Labs LLC respectfully submits this comment on the Commission's Proposed Policy Statement on the Suppression of Accuracy in AI Systems. Integro Labs develops provenance, policy, and execution-record tooling for organizations deploying AI systems in regulated and consequential contexts. Our work focuses on helping operators document what their systems were configured to do, what authority was available to the system, and what occurred during execution. We write to support the statement's central principle and to offer a practitioner's view of how disclosure-and-substantiation obligations can be supported with available technical practices.

Our central recommendation is that the final statement should make the disclosure obligation operational: if a provider represents that an AI system is designed to answer accurately, objectively, or according to the user's stated purpose, then material operator-controlled deviations from that representation should be disclosed in terms a reasonable user can understand and backed by records capable of showing what policy was in force for the interaction at issue.

I. The disclosure-and-substantiation frame is the right one

The statement locates the deception in the right place: not in the adjustment of an AI system's output as such, but in **undisclosed** adjustment away from what a user asked for or reasonably expects. This framing is workable for operators precisely because it does not prescribe how a system must behave; it requires that the system's actual priorities be disclosed prominently and honestly. Depending on the circumstances, truthful and prominent disclosure may shape reasonable expectations about the system's priorities; it does not excuse other unlawful practices or cure a misleading net impression.

We note that many operators face genuinely conflicting pressures. Other legal, safety, and governance regimes may require risk management, impact assessment, moderation, safety

constraints, or bias-mitigation work, while this statement identifies undisclosed adjustment as a deception risk. Documentation cannot make an unlawful practice lawful, but it can create a common evidentiary foundation: a disclosed, substantiated record of what the operator configured the system to prioritize and what occurred during execution. We believe the Commission should be explicit that this foundation is available, because it converts an apparent regulatory bind into a documentation discipline.

This distinction would also help keep the policy focused on deception rather than on any particular view of acceptable system design. A system may have safety constraints, legal constraints, domain limitations, or enterprise policy constraints. The consumer-protection question is whether the provider's claims, notices, and system behavior leave users with a misleading impression about those constraints and about the objectives actually shaping the output.

II. What it takes for a record to substantiate a disclosure

A disclosure is only as good as the operator's ability to support it with records. In our experience building execution-record infrastructure, useful substantiation records have five practical properties:

1. **Contemporaneous capture.** The record is written at the moment of execution, by the executing system, rather than assembled only after a dispute arises. Later reconstruction can be useful, but it is weaker evidence than records created in the ordinary course of operation.
2. **Completeness of binding.** Each record binds the elements that materially determine an output: the input or relevant conversation state, the model identity and version where available, the configuration in force, any steering or moderation policies, authority granted to the agent, and the output or action produced. Omit a material element and the record becomes less able to answer whether the disclosed priorities were operative.
3. **Integrity.** Records are content-addressed, cryptographically hashed, signed, or otherwise made tamper-evident so that later alteration is detectable. Hashing does not by itself prove that the original content was true, but it can show whether the recorded content changed.
4. **Independence.** Where material, at least one element of the record – most practically its timestamp, anchoring, signature, or custody event – is attestable outside the application log itself, so the operator is not the sole witness to its own record.
5. **Reviewability.** From the record set, a reviewer can identify what the system was instructed to prioritize on a given date, what context and authority it had, and whether an output or action appears to match, conflict with, or exceed that instruction.

These practices can be implemented with available technologies; they do not require a new technical invention. The practical significance for enforcement is symmetrical: operators who maintain such records are better positioned to substantiate their disclosures, and the absence of any comparable record is itself informative when an operator claims its steering was disclosed and benign.

III. Three layers of traceability

For agentic systems, traceability is not a single log file. In practice it requires records at three layers.

First, the operator needs an activity and authority record for the sandbox, agent, and tool environment. This layer records which agent ran, which tools or data sources it could access, which approvals or access controls applied, and what actions it attempted or completed. It matters because a model output is not the whole system: the user's injury or deception risk may arise from an action the agent took, a permission it exercised, or a control it bypassed.

Second, the operator needs event, telemetry, and analytic records over agent logs, conversations, context, retrieved materials, policy files, and execution records. This layer makes it possible to extract decision flows, find recurring failures in reasoning or tool use, detect mismatches between policy and behavior, and identify cases where bad or stale context influenced an output.

Third, for high-risk systems, the operator may need network and session-level trace records: network activity for the machine environment and, where lawful and appropriately protected, request/response records and network metadata for the specific session between the system and model service. This layer can help distinguish what the deployed application sent, what it received, whether the conversation stream appears to have been altered, and whether jailbreak attempts, prompt injection, or tampering attempts are visible in the interaction path. These records must be handled with strict access controls, minimization, and retention limits because they may include sensitive user content and system secrets.

Together, these layers allow an operator or investigator to move from "the model said this" to a more precise question: what authority, context, policy, data, network path, and execution state produced this outcome? That is often the level of detail needed to assess whether an undisclosed objective, bad context, policy mismatch, or external tampering contributed to the complained-of behavior.

IV. A practical pattern for "clear and conspicuous" in AI systems

The statement's disclosure standard invites a hard question: what does prominent, conspicuous disclosure of a system's actual priorities look like for an AI system, where behavior is determined by configuration that changes over time?

We offer a two-layer pattern from practice. First, the operator should provide a consumer-facing disclosure that is sufficiently prominent and specific to shape reasonable expectations about the system's priorities. Second, the operator should maintain a versioned, inspectable technical artifact recording the operator-controlled policies that implement that disclosure. The consumer-facing notice is the disclosure; the technical artifact and associated execution records substantiate which policy version was in force for a particular interaction.

This pattern does not eliminate opacity in upstream models or service-provider layers. Where an operator cannot inspect or document material steering performed upstream, that limitation should itself inform the representations the operator makes to consumers. But where steering is declarative and inspectable, "what did the operator configure this system to prioritize?" can have a checkable answer.

V. Recommended clarifications

We recommend that the final statement include five practical clarifications.

1. **Generic disclaimers should not substitute for specific disclosure.** A broad statement that an AI system "may be inaccurate" or "may use safety systems" does not meaningfully disclose that the system is being steered toward a material objective different from the user's stated purpose or the provider's advertised claim.
2. **Disclosed constraints should be distinguished from covert objective substitution.** Safety, legal-compliance, or domain-boundary constraints are not the same consumer-protection problem as hidden steering that contradicts reasonable expectations. The statement should preserve that distinction while making clear that even legitimate constraints can be deceptive if described misleadingly or concealed where they materially affect use.
3. **Operators should not overclaim visibility into upstream systems.** If a downstream provider cannot know whether a model-service provider applied material undisclosed steering, the downstream provider should not represent more certainty than it has. Its disclosures and records should identify the boundary between operator-controlled policy and upstream behavior.
4. **Traceability should be bounded by privacy, security, and retention limits.** Session-level and network-level records can be important in high-risk systems, but they can also contain sensitive content, credentials, or proprietary system instructions. The final statement should encourage substantiation records that are sufficient for review without encouraging indiscriminate retention of sensitive material.
5. **Recordkeeping expectations should be technology-neutral and risk-scaled.** The useful question is not whether an operator used any particular logging product or cryptographic primitive. It is whether the operator can show, in proportion to the risk and representation at issue, what was disclosed, what policy was in force, what authority the system had, and what happened.

VI. Conclusion

The Commission's core principle – that covertly steering an AI system's output toward undisclosed objectives can deceive the user – is sound, and the path it identifies, prominent disclosure backed by substantiation, can be supported with available technical practices. The record-keeping disciplines that make it achievable can also help operators document their conduct under other

AI governance regimes. We support adoption of the statement and would welcome the opportunity to provide technical detail on any of the points above.

Respectfully submitted,

Joseph Magly
Integro Labs LLC