The claim

Somewhere in most AI deployment approvals sits one sentence doing enormous work: *a human reviews every decision.* It appears in vendor contracts, risk registers, board minutes and regulatory filings, converting uncomfortable authorisations into comfortable ones. The decision is whether to accept that sentence as evidence of a control, or to treat it as an unproven claim about a person the board has never met, doing a job it has never costed.

Our position is blunt: a human checkpoint functions as a control only when the human holds four things at the moment of refusal — authority to say no without career consequence, time to actually look, context to see what the machine saw and what it did not, and an incentive structure that does not punish refusal. Remove any one and the checkpoint becomes a costume: it changes who is blamed after the failure without changing its probability. We call that liability theatre; it passes audits until the day it meets a loss.

The evidence

Start with the law, which is more honest than most deployment plans. Article 14 of the EU AI Act does not say "put a human in the loop." It requires that high-risk systems be *designed and developed* so that natural persons can effectively oversee them, and enumerates what the overseer must be enabled to do.

> EVIDENCE
>
> Article 14 of the EU AI Act requires that high-risk AI systems be designed so they can be effectively overseen by natural persons, including the overseer's ability to remain aware of automation bias, to interpret output correctly, to decide not to use the system, and to intervene or stop it; measures must be commensurate with risk, autonomy and context.
>
> Source: Regulation (EU) 2024/1689, Article 14 (EU AI Act), text verified 2026 · Geography: European Union · Method: statutory text · SEG-5 · Caveat: Article 14 imposes a design-capability duty on providers and deployers, not a guarantee that oversight occurs; the application of the high-risk obligations was postponed by the Digital Omnibus on AI, Regulation (EU) 2026/1744, in force 27 July 2026.

Article 14 is a list of the conditions under which a human can refuse a machine: the legislator does not assume the human will function, it orders the system built so that functioning is possible.

Automation bias is one of the most stable findings in human-factors science. Skitka, Mosier and colleagues showed in 1999 two characteristic errors: omission, missing events the aid failed to flag even when other indicators showed them plainly, and commission, following the aid against training and fully valid contrary information. Parasuraman and Manzey's 2010 integration added that complacency emerges under multiple-task load — the normal condition of a production review queue — appears in experts as well as novices, and is not eliminated by practice.

EVIDENCE

Peer-reviewed studies find that humans working with automated aids commit systematic omission and commission errors, and that automation complacency arises under multi-task load, affects experts as well as novices, and is not overcome by practice alone.

Source: Skitka, Mosier & Burdick, International Journal of Human-Computer Studies 51(5), 1999; Parasuraman & Manzey, Human Factors 52(3), 2010 · Geography: laboratory and simulator studies, primarily US and Europe · Method: controlled experiments and integrative literature review · SEG-4 · Caveat: findings are strongest in laboratory and simulator settings; effect sizes in specific production deployments vary and should be measured, not assumed.

NIST's AI Risk Management Framework — voluntary, but the reference vocabulary of most AI governance programmes — places oversight under GOVERN as an organisational design problem and treats the insertion of a human as itself a risk decision: oversight roles inherit the biases the framework catalogues.

The law demands that refusal be possible; the psychology shows it is unlikely under load unless deliberately protected; the framework makes that protection a design task, not a default. What none supplies — because it is firm judgment, not measurement — is our operational test: authority, time, context, incentive, all four, held by the named person on the worst day of the quarter. One consequence is measurable: a checkpoint that has never refused the machine is not evidence that the machine is good, but evidence that the checkpoint is absent. Override rates are a control metric, and zero is a finding.

Evidence cards

CLM-AIACT-ART14-OVERSIGHT-CAPABILITY · SEG-5. Claim: EU AI Act Article 14 requires high-risk AI systems to be designed for effective human oversight, enumerating the overseer's capabilities (awareness of automation bias, correct interpretation, the option not to use, intervention and stop), with measures commensurate with risk, autonomy and context. Context: Regulation (EU) 2024/1689, text verified 2026. Method: statutory text. Contradictory evidence: the article imposes design capability, not observed effectiveness; the Digital Omnibus on AI (Regulation (EU) 2026/1744, in force 27 July 2026) postponed application of the high-risk obligations. Causal confidence: none claimed — legal fact. Transferability: EU-regulated deployments; persuasive elsewhere. Review date: 2026-08-02.

CLM-AUTOMATION-BIAS-PERSISTENCE · SEG-4. Claim: humans working with automated aids commit systematic omission and commission errors; complacency arises under multi-task load, affects experts, and is not eliminated by practice. Context: Skitka, Mosier & Burdick 1999; Parasuraman & Manzey 2010. Method: controlled experiments and integrative review. Contradictory evidence: most findings come from laboratory and simulator settings; magnitude in any specific production deployment must be measured. Causal confidence: experimental for the core effects, in-setting. Transferability: strongest for high-volume review queues under load. Review date: 2026-08-02.

CLM-NIST-RMF-OVERSIGHT-RISK · SEG-5. Claim: the NIST AI Risk Management Framework (AI 100-1, January 2023) treats human oversight as an organisational design task under GOVERN and identifies human-AI configurations, including oversight roles, as carrying their own risks. Context: voluntary US framework, current edition 2026. Method: framework text. Contradictory evidence: the framework is non-binding and does

not prescribe thresholds or staffing ratios. Causal confidence: none claimed — institutional fact. Transferability: any organisation adopting the RMF vocabulary. Review date: 2026-08-02.

CLM-FIRM-OVERSIGHT-FOUR-CONDITIONS · SEG-1 (firm position). Claim: a human checkpoint functions as a control only when the named person holds authority, time, context and incentive to refuse the machine; absent any one, the checkpoint is liability theatre, and a zero override rate is a finding of absence, not of quality. Context: firm operating practice in the corridor. Method: interpretation, not measurement; no deployment statistics cited. Contradictory evidence: in genuinely low-stakes, high-accuracy tasks a lightweight checkpoint may be proportionate; the four conditions are conjunctive by design and unvalidated as a scored instrument. Causal confidence: none claimed. Transferability: bounded — strongest for high-risk decisions reviewed under production load. Review date: 2026-08-02.

The limits

The four conditions fail predictably where our clients operate. AI deployments arrive priced on vendor-country staffing assumptions: a queue sized for a European back office, minutes-per-item ratios validated in the vendor's home market. An Abidjan operating company inherits the queue without the staffing plan behind it, and oversight lands on whoever is available, at the end of a reporting line whose real authority sits in Paris or Geneva. Each condition then degrades independently. Authority: the reviewer who refuses is overruling a system the holding company paid for, from the bottom of the organisation chart. Time: the queue arithmetic was never recomputed for local staffing. Context: the model was trained on markets with formal credit histories the corridor does not have. Incentive: throughput is measured, refusal is friction. There is also a regulatory asymmetry: the EU AI Act reaches the European holding company; no equivalent statute operates in the WAEMU jurisdictions where the system runs. Whatever oversight exists in Abidjan exists because the group built it.

Sources and limitations

Sources and limitations. The legal and framework facts rest on the text of Regulation (EU) 2024/1689 Article 14, its application timeline as amended by Regulation (EU) 2026/1744, and NIST AI 100-1; the behavioural findings rest on the peer-reviewed automation-bias literature cited above, whose laboratory provenance is flagged, not hidden. The four-conditions test is the firm's position, graded as interpretation: it is not validated as a scored instrument, and we publish no client oversight metrics because none has passed our evidence-release gate. The transferability boundary is explicit: strongest for high-risk, high-volume decisions reviewed under load; weakest for low-stakes tasks where a light checkpoint may be proportionate. This note is not valid as legal advice on AI Act compliance, nor as a claim that human oversight is useless — only that it must be built, resourced and measured before it is counted as a control.

What to do on Monday

1. The chief risk officer: publish the override rate of every checkpoint counted as a control. A rate of zero is not comfort; it is the strongest available signal that the checkpoint is a costume.

2. The operations director: recompute the queue arithmetic — items per reviewer per day against minutes of genuine review per item — on the staffing that exists, not the vendor's deck. If it does not close, the time condition has already failed.

3. The head of internal audit: trace one real refusal end to end. Who was questioned, whose metrics suffered, what changed. If none has ever occurred, return to action one.

4. The deployment sponsor: name the checkpoints that exist solely so that a person is blamable, and decide by month-end to remove or resource them. An honest answer names at least one.