Proof, not measurement

EMILIA and EU AI Act Article 14 — what counts as evidence of human oversight ·

Annex III · 2 Aug 2026

> *Decision logs are testimony. Scores are opinion. Receipts are evidence.*

THE ARTICLE 14 QUESTION

Article 14 requires high-risk AI systems to be under **effective human oversight** — a person must be able to decide not to use an output, **override** it, and **intervene or stop** the system. When a regulator, court, or insurer later asks the one question that matters —

> *"Show me that a specific authorized human approved **this** irreversible action **before** it ran."*

— most "human oversight" tooling produces something that **cannot answer it**: a **decision log** (self-reported, mutable — testimony, not evidence), an **oversight score / dashboard** (a heuristic rating of how attentive the reviewer *probably* was, computed over self-reported telemetry — an opinion *about* the human, not a binding *of* the human to the act), or a **"human-in-the-loop" toggle** (proof a step existed, not that a named person authorized *this* action).

THE BRIGHT LINE: MEASUREMENT VS. PROOF

Evidence of human oversight has to do two things a score cannot: **(1) bind** a named human to an exact action before it executed, and **(2) verify without trust** — let an outside party confirm it offline, with a public key, without trusting the vendor or the log.

	OVERSIGHT SCORE / DASHBOARD / LOG	EMILIA AUTHORIZATION RECEIPT
Artifact	A number or narrative *about* the review	A cryptographic **proof** of the authorization itself
Evidence	Statistical / behavioral — an indicator	Mathematical — Ed25519 over RFC 8785 canonical JSON
Binds	How attentive the reviewer probably was	A named human → an exact action, pre-execution
Outsider-verifiable?	Only by trusting the issuer	Yes — offline, with a public key. No backend, no vendor trust
Enforcement	Advisory — scores after the fact	Fail-closed — `no receipt, no execution`
Gameable?	Yes — telemetry is self-reported	No — a forged or altered authorization fails verification

A score tells you the oversight was *probably* real. A receipt **proves** the authorization happened — and refuses to let the action run without it. For irreversible-action accountability, the binding is load-bearing; the score is, at most, an annotation on top of it.

WHAT EMILIA PROVIDES FOR ARTICLE 14

- **The audit artifact a regulator can actually check** — deterministic verification with a public key; the receipt *is* the evidence, reproducible by anyone, years later.
- **Override / intervention you can prove** — approve, decline, or stop captured as signed, tamper-evident, per-action proof, not a log entry.

- **Enforcement, not exhortation** — `428 — no receipt, no execution`. Oversight that can be bypassed isn't oversight.
- **Two-person control where stakes demand it** — quorum receipts (a cryptographic two-person rule) + scoped delegation with verify-time constraint enforcement.
- **Built on accepted standards** — Ed25519 (RFC 8032), JCS (RFC 8785); expresses as JWS (RFC 7515) / COSE for interop; loggable to a transparency service (SCITT). Open IETF Internet-Draft, Apache-2.0, verifiers in JS/Python/Go.

HONEST SCOPE

A receipt is **necessary, not sufficient**. It proves a named human authorized the exact action; it does not prove the decision was wise, lawful, or fully informed — and a quality signal (e.g. a third-party judgment score) can sit *inside* a receipt as one more claim. EMILIA supplies the load-bearing, verifiable binding the rest of an oversight program builds on.

EMILIA Protocol · authorization receipts for irreversible AI-agent actions · `draft-schrock-ep-authorization-receipts` · Apache-2.0 · team@emiliaprotocol.ai