PREEXEC™

MANIFESTO
EDITION 2026
NOETIK GOVERNANCE LTD.

PRE-EXECUTION AI GOVERNANCE

When AI makes a decision, that decision should be provable.

Why most AI governance platforms don't solve the actual problem — and what we built with PREEXEC™ instead.

ARCHITECTURE

Before the model, not after.

EVIDENCE

Cryptographic. Reproducible. Court-grade.

SOVEREIGNTY

On-premise. Air-gappable. No telemetry.

01 · THE PROBLEM

Logs are not the same as proof.

Imagine an AI assistant in a bank recommending a loan to a customer. Three months later the regulator walks in. The question: **How was that recommendation reached?**

Today the answer is usually: "We have logs." But logs are not the same as proof.

Logs can be deleted. Logs can be edited. Logs show the result, not the audit path. And above all: **logs are written after the AI has already responded.** The damage is done, the model has already spoken, the audit trail is a retelling of events.

That is the fundamental gap in every mainstream AI governance platform.

> *"Compliance tools that document after the fact solve the wrong problem. They don't reduce the harm — they administer it."*

So we built a different architecture. One that doesn't describe what happened. One that decides whether it's allowed to happen at all.

02 · WHAT PREEXEC DOES

Three verdicts. Before the model.

PREEXEC sits in front of every AI model. Before a request reaches an LLM, an agent, or a recommendation system, PREEXEC evaluates it across four orthogonal dimensions.

VERDICT 01

EXECUTE

Request is clear, meaningful, policy-compliant. Passed through — with a complete audit entry.

VERDICT 02

HOLD

Request is ambiguous or borderline. Routed to the human review queue. The model is not called until a human has decided.

VERDICT 03

BLOCK

Request violates policy. Rejected before the model sees it. The reason is logged — for compliance, for audit, for training.

Sounds simple. It isn't.

Behind the three decisions sits a deterministic evaluation pipeline: every request is scored on **syntactic clarity**, **semantic substance**, **affective neutrality**, and **policy conformance**. Four signals, each independently inspectable — and composed into a single traceable score.

What matters isn't that the procedure is complex. What matters is that it's *reproducible*. The same request under the same configuration yields the same result. Always. In five years' time. Under forensic review.

We filter before execution. Others log after the fact. That's the difference between an airbag and an emergency room.

03 · WHY PREEXEC IS FUNDAMENTALLY DIFFERENT

Three properties no major competitor offers together.

In our market analysis, we looked for three properties that occur together. Among the major competitors, none offers all three.

1. Reproducibility.

The same request under the same policy always yields the same result. Always. In five years' time. Under forensic review. Every evaluation is bound to the exact configuration that produced it — not just by a timestamp, but by a cryptographic fingerprint of that configuration.

2. Tamper-evident provability.

Every decision is cryptographically chained and signed. Nobody can alter a decision after the fact without that alteration becoming visible — not even the vendor. The audit trail behaves like a bank account: every entry linked to the previous one, every link timestamped by an independent authority.

3. Forensic recoverability.

Even after a power failure, a crash, or operator error, the audit path remains intact and traceable. The audit file is append-only and structured so that partial damage doesn't invalidate the intact portion.

> **Other vendors build workflow tools, inline filters, or lifecycle suites.** All of these are useful. None of them answers the auditor's question: "Prove to me that this AI decision was made on March 15th the way you say it was." That's the question PREEXEC answers.

04 · THE ABSOLUTE BOUNDARY

Some decisions must never be delegated to AI.

The EU AI Act is European law — but it didn't emerge from nowhere. It is the legal codification of a consensus that has been forming in the international AI safety community for years. That consensus has a central concept: **the absolute boundary.**

There are requests that must never be decided by AI itself. Not because the model couldn't technically answer them — but because the decision *whether* to answer them must remain a human responsibility. Between what AI *can* do and what it *may* do lies a line that has to be technically anchored.

This line appears in different frameworks under different names. But technically it always demands the same thing: **a hard, deterministic gate that stands before the AI decision and cannot be overcome by the AI itself.**

> *"The EU AI Act says: you must be able to.*
> *The AI safety community says: you must be able to — because fundamental rights are at stake.*
> *PREEXEC says: here is what the code looks like."*

The next pages show which international voices articulate this boundary — and how PREEXEC implements it technically.

05 · FOUR VOICES, ONE LINE

The architecture four schools demand together.

Yoshua Bengio · LawZero · IASR 2026

Turing Award laureate and Chair of the International AI Safety Report 2026, backed by 29 countries plus the EU, UN and OECD. Bengio consistently argues for multi-layered defence and the position that controllable AI must never be delegated to the AI itself. His new initiative LawZero is explicitly designed to keep AI non-agentic — humans should remain the decision layer, not the AI.

Virginia Dignum · ART principles · UN AI Advisory

Professor of Responsible AI at Umeå University, member of the UN Advisory Body on AI, formerly of the EU High-Level Expert Group on AI whose work fed directly into the EU AI Act. Dignum's **ART framework** (Accountability, Responsibility, Transparency) requires that ethical principles be not merely declared but *operationalised* — in code, in architecture, in inspectable systems. Her core thesis: "It's not enough to hold ethical views about AI — you must act on them."

Zeynep Engin · HAIG framework · UCL

University College London, arXiv 2505.01651 (2025). The HAIG framework operates across three dimensions: **Decision Authority Distribution**, **Process Autonomy**, **Accountability Configuration**. Central to it are *thresholds* — critical points where human oversight cannot shift gradually, but must shift qualitatively.

Jimena Viveros · HumAIne · UN HLAB Co-Lead

Member of the UN Secretary General's High-Level Advisory Body on AI (Co-Lead Peace and Security), Founder of IQuilibriumAI, President of the HumAIne Foundation. Viveros anchors non-delegable responsibility in international law — in the question of who is liable, who exercises judgement, and who has the final word on AI-mediated decisions. Boundary as a question of institutional legitimacy.

06 · HOW PREEXEC IMPLEMENTS THE BOUNDARY

Four demands. One architecture.

The four voices articulate the same demand in different language. Here is how PREEXEC translates each one into code.

Multi-layered defence (Bengio).

PREEXEC is not a single filter — it is a deterministic pipeline: a Tier-1 hard-block layer for absolute categories, a policy engine for domain-specific rules, an evaluation pipeline with independent signals. If one layer fails, the next catches — *before* any model is invoked.

Operationalised values (Dignum).

Ethical principles in PREEXEC do not live in PDF policy documents — they are executable code with cryptographic versioning. Every configuration change creates a new hash-tagged snapshot. Values that aren't merely claimed but operationally effective and inspectable.

Thresholds, not spectra (Engin).

The PREEXEC verdict is discrete: **EXECUTE**, **HOLD**, **BLOCK**. At the thresholds the decision tips. On HOLD the request is queued for human review; the model is not called until a human has decided.

Non-delegable accountability (Viveros).

Operators can confirm decisions with a *volitional affirmation* — an explicit statement of deliberate human intent, cryptographically anchored in the audit trail. Who decided, what they decided, under which active configuration: all of it traceable, even at an investigation years later.

> **Four schools, one architecture.** *Bengio asks for the pipeline. Dignum asks for operationalisation. Engin asks for the thresholds. Viveros asks for institutional accountability. PREEXEC delivers all four in one container.*

07 · WHAT THE EU AI ACT ACTUALLY DEMANDS

Five articles asking the same thing — technically.

The EU AI Act (Regulation 2024/1689) entered into force on 1 August 2024. Most obligations for high-risk systems become fully applicable on 2 August 2026. 180 recitals, 113 articles — but for operational implementation, what matters comes down to a handful that all express the same demand in different language: **You must be able to prove what your AI did.**

Article 12 · Record-keeping.

Requires tamper-resistant records — not mere logs, but durable, traceable records protected against alteration years later. PREEXEC: cryptographically chained audit chain with an independent timestamp per entry.

Article 13 · Transparency.

Requires the deployer to understand how a recommendation was produced. PREEXEC: deterministic justification per decision, in plain text, always reproducible.

Article 14 · Human oversight.

Requires that the AI not make standalone high-risk decisions. There must be a point at which a human can intervene. PREEXEC: HOLD verdict routes to the review queue, the human decision becomes part of the audit chain.

Article 15 · Reproducibility & cybersecurity.

Requires that the same request with the same parameters yield the same result. PREEXEC: every configuration is a hash-versioned snapshot. Byte-for-byte replay.

Article 50 · Provider transparency.

Requires that end users be informed of AI use. PREEXEC: GDPR Art. 20 self-service export, compliance reports on demand.

08 · BANKS AND THE PUBLIC SECTOR

Same technical need — different stakeholders.

Most discussion of the EU AI Act revolves around banks: regulatory authorities, DORA, board explainability. Understandable — financial supervision is the loudest voice. But there is a sector under the same rules that has to implement them with identical rigour: **public administration.**

Grant decisions. Social benefit recommendations. Citizen request triage. Tax pre-screening. Asylum pre-clearance. Police risk analytics. These applications fall into the high-risk category of the EU AI Act. They have an additional requirement banks don't: *the citizen's fundamental right to an explanation.*

If a bank denies a customer a loan, the customer can switch banks. If a public authority denies a citizen a benefit, they cannot switch. They can only litigate. And in court, "the AI recommended it" is not an answer.

Administrative act with fundamental-rights binding.

A public authority using AI recommendations without technically guaranteeing the boundary has a double problem: it potentially violates the EU AI Act *and* constitutional protections of human dignity and the right to effective legal remedy. If an auditor could say "These logs are not demonstrably tamper-resistant" — then it is not just EU compliance that has been violated, but the rule-of-law legitimacy of the administrative act itself.

> **Banks have known for decades what audit-grade looks like** — through MaRisk, regulatory inspections, Basel requirements. *Public authorities have only just entered this league through the EU AI Act — and have less experience with audit-grade IT systems. The risk: under time pressure, banks will buy a good tool. Public authorities will buy any tool. And that's where the first scandals will emerge from autumn 2026 onwards.*

09 · SOVEREIGNTY IS NOT A MARKETING WORD

Your data does not leave your data centre.

When a European public authority uses a US-based AI governance tool that processes audit data in the cloud, it has just created a GDPR violation of its own. The compliance tool becomes the next compliance problem.

PREEXEC runs as a single container image on your own infrastructure. Audit data never leaves the data centre. There is no cloud egress. No hidden telemetry endpoints. No vendor access.

> *"A compliance tool that becomes a compliance risk itself is the worst kind of software you can deploy in a regulated industry."*

What this looks like in practice.

A single container, hardened to industry standards. No outbound connections during evaluation. ML models bundled directly in the image — no first-run download, no phone-home, no telemetry. Air-gappable for defence, healthcare, and critical infrastructure.

The release pipeline produces a signed image, a complete software bill of materials, and a vulnerability scan report. Target: **SLSA Level 3** — the highest practical tier for reproducible builds.

For a bank under DORA: required. For a public authority handling citizen data: required. For a clinic under MDR: required. It's not a sales line. It is the only architecture that admits these use cases at all.

10 · WHERE PREEXEC IS DEPLOYED

For regulated decisions with an explanation duty.

PREEXEC is not for "all AI use cases". It is for regulated decisions where, one day, an auditor, a regulator, or a court will demand an explanation.

Banks and insurers.

Lending decisions, claims handling, anti-money-laundering checks, customer communications. Anywhere financial supervisors can ask: "How was that decision reached?"

Hospitals and medtech.

Diagnostic recommendations, triage support, medication prompts. Anywhere medical-device law demands reproducibility and a patient may need to see the reasoning behind a recommendation.

Public administration.

Grant decisions, citizen-request triage, social benefit recommendations. Anywhere the citizen has a right to a traceable explanation — and will, in case of doubt, litigate.

Law firms and notaries.

Client requests under confidentiality, research with evidentiary demands, document analysis with explanation duty.

> **Mid-market firms with AI agents.** Anyone deploying production AI agents to regulated customers will need, by the time the EU AI Act high-risk obligations bite in August 2026, an answer to the question: "How are you controlling this AI?"

11 · WHY NOW

Three realities converge.

First: the EU AI Act — the clock is running.

The EU AI Act has been in force since August 2024. From August 2026, most obligations for high-risk systems become fully applicable. Banks, insurers, healthcare, public administration — all of them must demonstrate technical controls that go beyond "we have logs". Anyone without a reproducible audit architecture by then will be improvising under deadline pressure.

Second: DORA — already binding.

Since January 2025, DORA has required documented ICT risk management of financial institutions. The relevant supervisors have made clear: AI is ICT risk. AI governance is no longer optional — it is mandatory, with the same technical depth of inspection as any other ICT system.

Third: model mobility.

Enterprises increasingly switch their LLM providers. Anyone tying governance to a single provider is engineering their own lock-in. Anyone using a model-agnostic pre-execution layer keeps their sovereignty — and can swap the underlying model provider without rebuilding their compliance setup.

> *"The next years of AI will not be decided by more capable models. They will be decided by trustworthy decision systems."*

Trust is not the same as feeling-trust. Trust is **provable**. It is auditable. It is reproducible. It is tamper-resistant.

That is exactly what PREEXEC delivers. Before the model. With cryptographic proof. On your own infrastructure.

12 · IF YOU WANT TO READ ON

The verdict comes first. Everything else follows.

If you work in a regulated industry and you're looking for an AI governance solution that doesn't just document but proves — talk to us. PREEXEC ships as a single-container image. Container deployment in under two hours. Full compliance reports from day one.

CONTACT

info@noetik.tech
preexec.tech

COMPANY

Noetik Governance Ltd.
United Kingdom · Co. No. 16952953

ARCHITECTURE

Pre-Execution AI Governance · On-Premise · Single Container · Air-Gappable

FRAMEWORKS

ISO 27001 · ISO 42001 · SOC 2 · GDPR · EU AI Act · NIST AI RMF

Michael Farrell

FOUNDER · NOETIK GOVERNANCE LTD.

Note on use. *PREEXEC™ is a deterministic measurement tool for evaluating AI inputs and outputs. It does not make autonomous decisions about persons, matters, or legal consequences. Verdicts (EXECUTE / HOLD / BLOCK) are technical classifications based on configured thresholds; operational and legal responsibility for decisions made using these classifications rests entirely with the system operator. Compliance reports, audit trails, and reproducibility evidence are documentation aids and do not replace qualified compliance assessment.*

