PREEXEC™

MANIFEST
AUSGABE 2026
NOETIK GOVERNANCE LTD.

PRE-EXECUTION AI GOVERNANCE

Wenn AI eine Entscheidung trifft, sollte sie beweisbar sein.

Warum die meisten AI-Governance-Plattformen das eigentliche Problem nicht lösen — und was wir mit PREEXEC™ stattdessen gebaut haben.

ARCHITEKTUR

Vor dem Modell, nicht danach.

BEWEISFÜHRUNG

Kryptographisch. Reproduzierbar. Vor Gericht haltbar.

SOUVERÄNITÄT

On-Premise. Air-gappable. Keine Telemetrie.

01 · DAS PROBLEM

Logs sind nicht dasselbe wie Beweise.

Stellen Sie sich vor, ein KI-Assistent in einer Bank empfiehlt einem Kunden einen Kredit. Drei Monate später steht die Aufsicht im Haus. Die Frage: **Wie kam diese Empfehlung zustande?**

Heute ist die Antwort meistens: „Es gibt Logs." Aber Logs sind nicht dasselbe wie Beweise.

Logs können gelöscht werden. Logs können verändert werden. Logs zeigen das Ergebnis, aber nicht den Prüfweg. Und vor allem: **Logs werden geschrieben, nachdem die KI bereits geantwortet hat.** Der Schaden ist da, das Modell hat schon gesprochen, der Audit-Trail ist eine Nacherzählung.

Das ist die fundamentale Lücke jeder gängigen AI-Governance-Plattform.

> *„Compliance-Tools, die nach der Tatsache dokumentieren, lösen das falsche Problem. Sie reduzieren den Schaden nicht — sie verwalten ihn nur."*

Wir haben deshalb eine andere Architektur gebaut. Eine, die nicht beschreibt, was passiert ist. Sondern eine, die entscheidet, ob es überhaupt passieren darf.

02 · WAS PREEXEC MACHT

Drei Verdicts. Vor dem Modell.

PREEXEC sitzt vor jedem AI-Modell. Bevor eine Anfrage an ein LLM, einen Agenten oder ein Empfehlungssystem geht, prüft PREEXEC sie in vier orthogonalen Dimensionen.

VERDICT 01

EXECUTE

Anfrage ist klar, sinnvoll, regelkonform. Wird ausgeführt — mit vollständigem Audit-Eintrag.

VERDICT 02

HOLD

Anfrage ist mehrdeutig oder grenzwertig. Geht in die menschliche Review-Queue. Modell wird nicht gerufen, bis ein Mensch entschieden hat.

VERDICT 03

BLOCK

Anfrage verstößt gegen eine Policy. Wird abgewiesen, bevor das Modell sie sieht. Begründung wird geloggt — für Compliance, für Audit, für Schulung.

Klingt simpel. Ist es nicht.

Hinter den drei Entscheidungen steht eine deterministische Bewertungs-Pipeline: jede Anfrage wird auf **syntaktische Klarheit**, **semantische Sinnhaftigkeit**, **emotionale Unauffälligkeit** und **Regelkonformität** geprüft. Vier Signale, jedes für sich auswertbar — und gemeinsam zu einem nachvollziehbaren Score komponiert.

Wichtig ist nicht, dass das Verfahren komplex ist. Wichtig ist, dass es *reproduzierbar* ist. Dieselbe Anfrage unter derselben Konfiguration ergibt dasselbe Ergebnis. Immer. Auch in fünf Jahren. Auch unter forensischer Prüfung.

Wir filtern vor der Ausführung. Andere protokollieren danach. Das ist der Unterschied zwischen einem Air-Bag und einem Notarzt.

03 · WARUM PREEXEC FUNDAMENTAL ANDERS IST

Drei Eigenschaften, die zusammen nirgendwo sonst vorkommen.

In der Marktanalyse haben wir nach drei Eigenschaften gesucht, die zusammen vorkommen. Bei keinem der großen Wettbewerber findet man alle drei.

1. Reproduzierbarkeit.

Dieselbe Anfrage unter derselben Policy ergibt immer dasselbe Ergebnis. Immer. Auch in fünf Jahren. Auch unter forensischer Prüfung. Jede Bewertung wird mit der exakten Konfiguration verknüpft, die sie produziert hat — nicht nur mit einem Zeitstempel, sondern mit einem kryptographischen Fingerprint dieser Konfiguration.

2. Fälschungssichere Beweisbarkeit.

Jede Entscheidung wird kryptographisch verkettet und signiert. Niemand kann eine Entscheidung nachträglich ändern, ohne dass es auffällt — auch nicht der Hersteller selbst. Der Audit-Trail funktioniert wie ein Bankkonto: jede Buchung mit der vorhergehenden verknüpft, jede Verkettung mit einem unabhängigen Zeitstempel versehen.

3. Forensische Wiederherstellbarkeit.

Selbst nach Stromausfall, Crash oder Bedienerfehler bleibt der Prüf-Verlauf intakt und nachvollziehbar. Die Audit-Datei ist append-only und so geschrieben, dass eine teilweise Beschädigung den intakten Teil nicht entwertet.

> **Andere Anbieter machen Workflow-Tools, Inline-Filter oder Lifecycle-Suiten.** Alles davon ist nützlich. Aber keines beantwortet die Frage des Auditors: „Beweisen Sie mir, dass diese AI-Entscheidung am 15. März so getroffen wurde, wie Sie behaupten." Genau diese Frage beantwortet PREEXEC.

04 · DIE ABSOLUTE BOUNDARY

Es gibt Entscheidungen, die niemals an eine KI delegiert werden dürfen.

Der EU AI Act ist europäisches Recht — aber er ist nicht im luftleeren Raum entstanden. Er ist die juristische Kodifizierung eines Konsenses, der sich seit Jahren in der internationalen AI-Safety-Community herausgebildet hat. Dieser Konsens hat ein zentrales Konzept: **die Absolute Boundary.**

Es gibt Anfragen, die niemals von einer KI selbst entschieden werden dürfen. Nicht, weil das Modell sie technisch nicht beantworten könnte — sondern weil die Entscheidung, ob sie überhaupt beantwortet werden darf, eine menschliche Verantwortung bleiben muss. Zwischen dem, was eine KI *kann*, und dem, was sie *darf*, liegt eine Linie, die technisch verankert sein muss.

Diese Linie taucht in unterschiedlichen Frameworks unter unterschiedlichen Namen auf. Aber technisch verlangt sie immer dasselbe: **Eine harte, deterministische Sperre, die vor der KI-Entscheidung steht und nicht von der KI selbst überwunden werden kann.**

> *„Der EU AI Act sagt: Du musst es können.*
> *Die AI-Safety-Community sagt: Du musst es können — weil es um Grundrechte geht.*
> *PREEXEC sagt: So sieht der Code aus."*

Auf den nächsten Seiten zeigen wir, welche internationalen Stimmen diese Boundary formulieren — und wie PREEXEC sie technisch umsetzt.

05 · VIER STIMMEN, EINE LINIE

Die Architektur, die vier Schulen gemeinsam fordern.

Yoshua Bengio · LawZero · IASR 2026

Turing-Preisträger und Vorsitzender des International AI Safety Report 2026, getragen von 29 Staaten plus EU, UN und OECD. Bengio plädiert konsequent für mehrschichtige Verteidigung — und für die Position, dass kontrollierbare KI niemals an die KI selbst delegiert werden darf. Sein Stiftungsprojekt LawZero ist explizit darauf ausgerichtet, Menschen als Entscheidungsschicht zu erhalten.

Virginia Dignum · ART-Prinzipien · UN AI Advisory

Professorin für Responsible AI (Universität Umeå), Mitglied der UN Advisory Body on AI, ehemals EU High-Level Expert Group on AI. Dignums **ART-Framework** (Accountability, Responsibility, Transparency) verlangt, dass ethische Prinzipien *operationalisiert* werden — in Code, Architektur, nachprüfbaren Systemen. Kernthese: „Es reicht nicht, ethische Ansichten zu haben — man muss nach ihnen handeln."

Zeynep Engin · HAIG-Framework · UCL

University College London, arXiv 2505.01651 (2025). Das HAIG-Framework operiert auf drei Dimensionen: **Decision Authority Distribution**, **Process Autonomy**, **Accountability Configuration**. Zentral sind *Thresholds* — kritische Punkte, an denen menschliche Aufsicht nicht graduell, sondern qualitativ kippen muss.

Jimena Viveros · HumAIne · UN HLAB Co-Lead

Mitglied des UN Secretary General's High-Level Advisory Body on AI (Co-Lead Peace and Security), Gründerin von IQuilibriumAI, Präsidentin der HumAIne Foundation. Viveros verankert die nicht-delegierbare Verantwortung im Völkerrecht — wer haftet, wer Recht spricht, wer das letzte Wort behält. Boundary als Frage der institutionellen Legitimation.

06 · WIE PREEXEC DIE BOUNDARY TECHNISCH UMSETZT

Vier Forderungen. Eine Architektur.

Die vier Stimmen formulieren denselben Anspruch in unterschiedlicher Sprache. Hier ist, wie PREEXEC jede einzelne Forderung in Code überführt.

Mehrschichtige Verteidigung (Bengio).

PREEXEC ist nicht ein einzelner Filter, sondern eine deterministische Pipeline: Tier-1-Hardblock-Schicht für absolute Kategorien, Policy-Engine für Domänenregeln, Bewertungspipeline mit unabhängigen Signalen. Versagt eine Schicht, fängt die nächste auf — *bevor* ein Modell angesprochen wird.

Operationalisierte Werte (Dignum).

Ethische Prinzipien stehen in PREEXEC nicht in PDF-Policies — sie sind ausführbarer Code mit kryptographischer Versionierung. Jede Konfigurationsänderung erzeugt einen neuen Snapshot mit eigenem Hash. Werte, die nicht behauptet, sondern operativ wirksam und nachweisbar sind.

Schwellen statt Spektren (Engin).

Der PREEXEC-Verdict ist diskret: **EXECUTE**, **HOLD**, **BLOCK**. An den Schwellen kippt die Entscheidung. Bei HOLD wird die Anfrage in eine menschliche Review-Queue gestellt; das Modell wird nicht gerufen, bis ein Mensch entschieden hat.

Nicht-delegierbare Verantwortung (Viveros).

Operatoren bestätigen Entscheidungen mit einer *Volitional Affirmation* — eine kryptographisch im Audit-Trail verankerte Erklärung des deliberaten menschlichen Willens. Wer entschieden hat, was, unter welcher Konfiguration: alles bleibt nachvollziehbar.

> **Vier Schulen, eine Architektur.** *Bengio fordert die Pipeline. Dignum fordert die Operationalisierung. Engin fordert die Schwellen. Viveros fordert die institutionelle Verantwortung. PREEXEC liefert alle vier in einem Container.*

07 · WAS DER EU AI ACT TECHNISCH VERLANGT

Fünf Artikel, die das gleiche technisch verlangen.

Der EU AI Act (Verordnung 2024/1689) ist seit dem 1. August 2024 in Kraft. Die meisten Pflichten für High-Risk-Systeme werden ab dem 2. August 2026 voll anwendbar. 180 Erwägungsgründe, 113 Artikel — für die operative Umsetzung kommt es auf eine Handvoll an: **Sie müssen beweisen können, was Ihre KI getan hat.**

Artikel 12 · Aufzeichnungspflicht.

Verlangt manipulationssichere Protokolle — keine bloßen Logs, sondern dauerhafte, nachvollziehbare Aufzeichnungen, die auch Jahre später gegen Veränderung geschützt sind. PREEXEC: kryptographisch verkettete Audit-Chain mit unabhängigem Zeitstempel pro Eintrag.

Artikel 13 · Transparenz.

Verlangt, dass der Einsetzer der KI versteht, wie eine Empfehlung zustande kam. PREEXEC: deterministische Begründung pro Entscheidung, in Klartext, immer reproduzierbar.

Artikel 14 · Menschliche Aufsicht.

Verlangt, dass die KI keine eigenständigen Hochrisiko-Entscheidungen trifft. PREEXEC: HOLD-Verdict führt zur Review-Queue, die menschliche Entscheidung wird Teil der Audit-Chain.

Artikel 15 · Reproduzierbarkeit & Cybersicherheit.

Verlangt, dass dieselbe Anfrage mit denselben Parametern dasselbe Ergebnis liefert. PREEXEC: jede Konfiguration ist ein hashversionierter Snapshot. Replay byte-genau möglich.

Artikel 50 · Anbieter-Transparenz.

Verlangt, dass Endnutzer über die KI-Nutzung informiert werden. PREEXEC: GDPR-Art-20-Self-Service-Export, Compliance-Reports auf Knopfdruck.

08 · BANKEN UND BEHÖRDEN

Derselbe technische Bedarf — unterschiedliche Stakeholder.

Die meisten Diskussionen zum EU AI Act drehen sich um Banken: BaFin, DORA, MaRisk, Vorstandserklärbarkeit. Verständlich — die Finanzaufsicht ist am lautesten. Aber es gibt einen Sektor, der unter denselben Regeln steht und ihn mit identischer Strenge umsetzen muss: **die öffentliche Verwaltung.**

Förderbescheide. Sozialleistungs-Empfehlungen. Bürgeranfragen-Klassifizierung. Steuer-Vorprüfungen. Asyl-Vorklärungen. Polizei-Risikoanalysen. Diese Anwendungen fallen in die High-Risk-Kategorie des EU AI Act. Sie haben zusätzlich noch eine Anforderung, die Banken nicht haben: *das Grundrecht des Bürgers auf Erklärung.*

Wenn eine Bank einem Kunden einen Kredit verweigert, kann der Kunde wechseln. Wenn eine Behörde einem Bürger einen Zuschuss verweigert, kann er nicht wechseln. Er kann nur klagen. Und vor Gericht reicht „die KI hat das so empfohlen" nicht aus.

Verwaltungsakt mit Grundrechtsbindung.

Eine Behörde, die KI-Empfehlungen nutzt, ohne die Boundary technisch zu garantieren, hat ein doppeltes Problem: Sie verstößt potenziell gegen den EU AI Act *und* gegen Artikel 1 GG (Würde des Menschen) sowie Artikel 19 Absatz 4 (Rechtsschutzgarantie). Wenn ein Auditor sagen könnte: „Diese Logs sind nicht nachweisbar manipulationssicher" — dann ist nicht nur EU-Compliance verletzt, sondern die rechtsstaatliche Legitimation des Verwaltungshandelns selbst in Frage gestellt.

> **Banken wissen seit Jahrzehnten, was Audit-tauglich heißt** — durch MaRisk, BaFin-Prüfungen, Basel-Vorgaben. *Behörden sind durch den EU AI Act erst in dieser Liga angekommen — und haben weniger Erfahrung mit Audit-tauglichen IT-Systemen. Die Gefahr: Banken werden mit knapper Zeit ein gutes Tool kaufen. Behörden werden mit knapper Zeit irgendein Tool kaufen. Und genau dort entstehen die ersten Skandale ab Herbst 2026.*

09 · SOUVERÄNITÄT IST KEIN MARKETING-WORT

Ihre Daten verlassen Ihr Rechenzentrum nicht.

Wenn eine deutsche Behörde ein US-amerikanisches AI-Governance-Tool einsetzt, das Audit-Daten in der Cloud verarbeitet, hat sie selbst einen GDPR-Verstoß erzeugt. Das Compliance-Tool ist dann das eigene Compliance-Problem.

PREEXEC läuft als ein einzelnes Container-Image auf der eigenen Infrastruktur. Audit-Daten verlassen nie das Rechenzentrum. Es gibt keinen Cloud-Egress. Keine versteckten Telemetrie-Endpunkte. Keinen Vendor-Zugriff.

> *„Ein Compliance-Tool, das selbst zu einem Compliance-Risiko wird, ist die schlimmste Sorte Software, die man in einer regulierten Industrie deployen kann."*

Was das praktisch heißt.

Ein einziger Container, gehärtet nach Industrie-Standard. Keine ausgehenden Verbindungen während der Bewertung. ML-Modelle direkt im Image gebündelt — kein Download beim ersten Start, kein Phone-Home, keine Telemetrie. Air-gappable für Verteidigung, Gesundheitswesen und kritische Infrastruktur.

Die Bereitstellungs-Pipeline produziert ein signiertes Image, eine vollständige Software Bill of Materials und einen Schwachstellen-Scan-Report. Ziel: **SLSA Level 3** — die in der Branche höchste praktisch erreichbare Stufe für reproduzierbare Builds.

Für eine Bank unter DORA: Pflicht. Für eine Behörde mit Bürgerdaten: Pflicht. Für eine Klinik unter MDR: Pflicht. Es ist keine Sales-Linie. Es ist die einzige Architektur, die diese Anwendungsfälle überhaupt zulässt.

10 · WO PREEXEC EINGESETZT WIRD

Für regulierte Entscheidungen mit Erklärungspflicht.

PREEXEC ist nicht für „alle AI-Use-Cases". Es ist für regulierte Entscheidungen, bei denen ein Auditor, ein Regulator oder ein Gericht eines Tages eine Erklärung verlangen wird.

Banken und Versicherungen.

Kreditentscheidungen, Schadenregulierung, Anti-Geldwäsche-Prüfungen, Kundenkommunikation. Überall dort, wo BaFin, FINMA oder DORA fragen können: „Wie kam diese Entscheidung zustande?"

Krankenhäuser und MedTech.

Diagnostische Empfehlungen, Triage-Unterstützung, Medikamenten-Hinweise. Überall dort, wo das Medizinprodukte-Recht Reproduzierbarkeit verlangt und ein Patient die Begründung einer Empfehlung einsehen können muss.

Öffentliche Verwaltung.

Förderbescheide, Bürgeranfragen-Klassifizierung, Sozialleistungs-Empfehlungen. Überall dort, wo der Bürger das Recht auf eine nachvollziehbare Begründung hat — und im Zweifelsfall klagen wird.

Anwaltskanzleien und Notare.

Mandantenanfragen mit Vertraulichkeitsanspruch, Recherchen mit Beweispflicht, Dokumentenanalyse mit Erklärungspflicht.

> **Mittelständler mit AI-Agenten.** Wer KI-Agenten produktiv einsetzt und regulierte Kunden hat, braucht spätestens mit dem Wirksamwerden der EU-AI-Act-High-Risk-Pflichten im August 2026 eine Antwort auf die Frage: „Wie kontrollieren Sie diese KI?"

11 · WARUM JETZT

Drei Realitäten treffen aufeinander.

Erstens: EU AI Act — die Frist läuft.

Der EU AI Act ist seit August 2024 in Kraft. Ab August 2026 werden die meisten Pflichten für High-Risk-Systeme voll anwendbar. Banken, Versicherungen, Gesundheitswesen, öffentliche Verwaltung — sie alle müssen technische Kontrollen nachweisen können, die über „wir haben Logs" hinausgehen. Wer bis dahin keine reproduzierbare Audit-Architektur hat, wird sie unter Zeitdruck improvisieren müssen.

Zweitens: DORA — bereits bindend.

DORA verlangt seit Januar 2025 von Finanzinstituten ein dokumentiertes ICT-Risikomanagement. Die zuständigen Aufsichtsbehörden haben klargestellt: AI ist ICT-Risiko. Damit ist AI-Governance nicht mehr Kür, sondern Pflicht — und zwar mit derselben technischen Prüf-Tiefe wie jedes andere ICT-System.

Drittens: Modell-Beweglichkeit.

Unternehmen wechseln zunehmend ihre LLM-Anbieter. Wer seine Governance an einen einzelnen Anbieter bindet, baut sich ein Lock-in. Wer einen modell-agnostischen Pre-Execution-Layer nutzt, behält seine Souveränität — und kann den darunter liegenden Modellanbieter wechseln, ohne sein Compliance-Setup neu aufzubauen.

> *„Die nächsten Jahre der KI werden nicht durch leistungsfähigere Modelle entschieden. Sie werden durch vertrauenswürdige Entscheidungssysteme entschieden."*

Vertrauen ist nicht das gleiche wie Vertrauen-haben. Vertrauen ist **beweisbar**. Es ist auditierbar. Es ist reproduzierbar. Es ist fälschungssicher.

Genau das liefert PREEXEC. Vor dem Modell. Mit kryptographischem Beweis. Auf Ihrer eigenen Infrastruktur.

12 · WENN SIE WEITER LESEN WOLLEN

Das Verdict kommt zuerst. Alles andere folgt daraus.

Wenn Sie in einer regulierten Industrie arbeiten und nach einer AI-Governance-Lösung suchen, die nicht nur dokumentiert, sondern beweist — sprechen Sie uns an. PREEXEC ist als Single-Container-Image lieferbar. Container-Deployment in unter zwei Stunden. Vollständige Compliance-Berichte ab Tag eins.

KONTAKT

info@noetik.tech
preexec.tech

UNTERNEHMEN

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

ARCHITEKTUR

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.

Hinweis zur Verwendung. *PREEXEC™ ist ein deterministisches Messwerkzeug zur Bewertung von KI-Anfragen und -Ausgaben. Es trifft keine eigenständigen Entscheidungen über Personen, Sachverhalte oder Rechtsfolgen. Verdicts (EXECUTE / HOLD / BLOCK) sind technische Klassifikationen auf Basis konfigurierter Schwellenwerte; die operative und rechtliche Verantwortung für Entscheidungen, die unter Verwendung dieser Klassifikationen getroffen werden, verbleibt vollständig beim Betreiber des Systems. Compliance-Berichte, Audit-Trails und Reproduzierbarkeits-Nachweise sind Hilfsmittel zur Dokumentation, ersetzen aber keine Compliance-Bewertung durch qualifizierte Fachleute.*

