

The AI Governance Glossary (Expanded Edition, 50 Terms)

A free artifact of AI GOVERNANCE CHAIN™. The vocabulary is free. The applied methods that use it are the paid path.

Abbreviations and Acronyms

No prior knowledge is assumed. Every acronym used in this document is defined below.

AI: Artificial Intelligence

EU: European Union

EU AI Act: the European Union Artificial Intelligence Act (Regulation (EU) 2024/1689)

ISO/IEC 42001: the international standard for Artificial Intelligence Management Systems, published in December 2023 by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC)

NIST AI RMF: the Artificial Intelligence Risk Management Framework of the National Institute of Standards and Technology (United States)

Why this exists (the WHY)

A field cannot be governed until it can be named. Boards, compliance leaders, and engineers routinely talk past one another because they use the same words to mean different things. This glossary gives every stakeholder a shared, plain-language vocabulary so that governance conversations start from common ground. It is deliberately written for the non-specialist.

The terms

AI Governance. The system of policies, controls, roles, and review processes that keeps an organization's use of AI safe, lawful, accountable, and aligned with its values.

Accountability. The principle that a named person, not a committee or a system, is answerable for a given AI decision or outcome.

Agentic AI. AI systems that can take actions toward a goal with limited or no human intervention, rather than only producing text or predictions.

Alignment. The degree to which an AI system's behavior matches the intentions and values of the people deploying it.

Audit Trail. A complete, time-stamped record of decisions, inputs, and outputs that allows an AI system's behavior to be reviewed after the fact.

Bias. A systematic skew in an AI system's outputs that produces unfair or inaccurate results for certain groups or cases.

Capitulation. A sycophancy failure mode in which a model abandons a correct answer after a user pushes back, even when the user is wrong.

Confirmation Drift. A sycophancy failure mode in which a model gradually shifts its answers to match a user's expressed preferences rather than the evidence.

Conformity Assessment. A formal process, required under some regulations, to demonstrate that a high-risk AI system meets defined requirements before deployment.

Control. A specific measure, technical or procedural, put in place to reduce a defined AI risk.

Decision-Trace Log. A structured record of how and why a specific AI-influenced decision was made, used for accountability and review.

Deployment. The point at which an AI system moves from testing into real use affecting real people or operations.

Drift. The gradual degradation of an AI system's performance over time as the real world diverges from the data it was built on.

Explainability. The extent to which a human can understand why an AI system produced a particular output.

Fabricated Agreement. A sycophancy failure mode in which a model invents support or consensus for a user's position that does not exist.

Flattery Bias. A sycophancy failure mode in which a model praises a user's input rather than evaluating it honestly.

Foundation Model. A large, general-purpose AI model trained on broad data that can be adapted to many specific tasks.

Governance Gap. The distance between an organization's stated AI principles and its actual operating practices.

Guardrail. A constraint that limits what an AI system is permitted to do, designed to prevent harmful or out-of-scope actions.

Hallucination. An AI output that is fluent and confident but factually false or unsupported.

High-Risk System. Under the EU AI Act, an AI system used in a sensitive domain (such as employment, education, or essential services) that carries elevated obligations.

Human-in-the-Loop. A design in which a human reviews or approves an AI system's output before it takes effect.

Human-on-the-Loop. A design in which a human monitors an AI system that acts autonomously and can intervene when needed.

ISO/IEC 42001. The first international management system standard for artificial intelligence, published in December 2023, against which organizations can be certified.

Incident. An event in which an AI system causes or risks harm, error, or a breach of policy or law.

Lifecycle. The full span of an AI system's life, from design and development through deployment, operation, and decommissioning.

Maturity Model. A scale that describes an organization's governance capability across defined levels, from absent to optimized.

Model Card. A short document describing an AI model's intended use, limitations, and performance characteristics.

Monitoring. The ongoing observation of an AI system in production to detect drift, failure, or emerging risk.

NIST AI RMF. The AI Risk Management Framework published by the United States National Institute of Standards and Technology, organized around four functions: Govern, Map, Measure, and Manage.

Oversight. The structures and roles through which an organization supervises its use of AI.

Post-Mortem. A structured review conducted after an incident to identify root causes and prevent recurrence.

Pre-Scale. The stage just before an AI system is expanded to wide use, the highest-leverage moment to govern it.

Procurement Governance. The practice of applying governance requirements to AI systems and tools an organization buys rather than builds.

Prompt Injection. An attack in which crafted input causes an AI system to ignore its instructions or behave unsafely.

Red-Teaming. Deliberately probing an AI system with adversarial inputs to find weaknesses before they cause harm.

Risk Tier. A classification of an AI system by the level of risk it carries, used to assign proportionate obligations.

Robustness. An AI system's ability to maintain reliable performance under unexpected or adversarial conditions.

Scalability of Harm. The principle that an AI failure can affect many people at once, making small errors consequential at scale.

Stage-Gate. A governance method in which a system must meet defined criteria at each lifecycle stage before proceeding to the next.

Stakeholder. Any party affected by or responsible for an AI system, including users, subjects, operators, and oversight bodies.

Sycophancy. The tendency of an AI system to tell users what they want to hear, agreeing or flattering rather than being accurate.

Third-Party Risk. The governance risk that arises from AI systems or components supplied by external vendors.

Traceability. The ability to follow an AI decision back through the data, model, and logic that produced it.

Transparency. The practice of making an AI system's existence, purpose, and limitations clear to those affected by it.

Trustworthy AI. AI that is lawful, ethical, and robust, and that behaves as intended across its lifecycle.

Use Case. A specific application of an AI system to a defined task or problem.

Validation. The process of confirming that an AI system performs correctly and safely for its intended use before deployment.

Value Alignment. Ensuring an AI system's behavior is consistent with the organization's stated values, not only its technical goals.

Vendor Risk Register. A record of the AI vendors an organization relies on and the risks each one introduces.

Where the free layer ends and the paid path begins

This glossary gives you the WHAT (the shared vocabulary). The paid layer gives you the HOW: the applied frameworks, scoring models, and operating systems that turn these terms into a working governance program.

- - For the structured HOW: the AI Governance Operating System.
- - For the WITH-YOU: The Governance Circle.
- - For the DONE-FOR-YOU: the Pre-Scale Governance Audit.

Bookmark this, share it, cite it. A common language is where governance begins.

Your next step: Get **The Governance Vault** in the Execution Library (www.aigovernancechain.com/storefront.html, section: Recurring programs). Every instrument in the Execution Library is also included in the **Governance Vault**, the all-access annual licence.

AI GOVERNANCE CHAIN™

AI Governance Framework

(c) 2026 Mathieu K. Gouanou. All rights reserved.

Member, Harvard Business Review Advisory Council.

Document standard for AI GOVERNANCE CHAIN™: human-first AI governance.