VLK ADV

Where Law drives Innovation.

ARTIFICIAL INTELLIGENCE
Risks, Incidents, and Governance

FROM REGULATORY UNCERTAINTY TO COMPETITIVE

- 10 reasons to govern AI now
- 10 steps of the AI governance framework

AI INCIDENTS

- What are AI incidents and “hazards”
- Case Study: Strategic Approach

vlklaw.com.br

1 AI GOVERNANCE: FROM REGULATORY UNCERTAINTY TO COMPETITIVE ADVANTAGE

Governing AI does not mean limiting its potential, but creating the conditions to innovate with the protection of rights, efficiency, trust, and security

10 REASONS TO GOVERN AI NOW

1. Protect people and fundamental rights
Prevent discrimination, bias, lack of transparency, errors, inadequate decisions, and privacy violations, ensuring human oversight proportional to risk and protection of rights potentially affected by AI.

2. Comply with rules already in force
Depending on the context of use, AI is already subject to existing legislation (such as the LGPD, the Civil Code, the Consumer Protection Code, the Brazilian Internet Civil Rights Framework, and the Copyright Law) and to sector-specific rules and guidance from authorities and entities (such as ANPD, Anatel, ANS, CFM, and MEC), as well as state legislation (such as those already in force in Goiás and Ceará). The application of these rules already translates into enforcement proceedings and judicial precedents (including at higher courts) directly related to the use of AI.

3. Demonstrate compliance
Turn legal obligations and ethical principles into processes, controls, and verifiable evidence for authorities, clients, investors, and partners.

4. Prevent liability and losses
Mitigate civil, regulatory, labor, consumer, and competition-related risks, as well as reduce the occurrence and impact of incidents, sanctions, financial losses, and reputational damage.

5. Protect business assets
Preserve data, intellectual property, confidential information, and trade secrets throughout the entire life cycle of AI applications.

6. Define responsibilities
Establish roles, competencies, and decision-making authority among administrators, internal areas, developers, suppliers, and users, reducing gray areas in the AI chain.

7. Strengthen cybersecurity and resilience
Protect applications, models, data, and infrastructure against attacks, manipulation, leaks, and misuse, as well as mitigate risks arising from the use of AI to amplify cyber threats.

8. Anticipate future regulation and follow international standards
Bill No. 2,338/2023 proposes a national regulatory framework based on risk management, governance, transparency, human oversight, and accountability of AI agents, in convergence with international parameters such as the EU AI Act, among other rules and best practices. Anticipating these trends reduces compliance costs.

9. Strengthen trust and reputation
Demonstrate responsibility to clients, employees, investors, regulators, and society.

88%
of organizations recognize the benefits of responsible AI for building trust
Cisco 2025 Data Privacy Benchmark Study.

43%
of executives associate responsible AI with brand reputation and reliability
Accenture–Stanford Executive Survey, N=1.000

51%
of leaders associate responsible AI with improved cybersecurity and data protection
PwC, Responsible AI Survey 2025.

10. Turn legal certainty into competitive advantage
Create conditions to test, approve, and scale solutions with greater predictability, speed, and adoption.

49%
of executives consider responsible AI a relevant factor for technology-related revenue growth
Accenture–Stanford Executive Survey, N=1.000

18%
pioneering companies can achieve an average increase of 18% in this revenue
Accenture–Stanford Executive Survey, N=1.000

58%
of executives in the US state that responsible AI initiatives increase ROI and organizational efficiency
PwC, Responsible AI Survey 2025

Organizations with responsible AI policies report better business results, greater customer trust, fewer incidents, and greater operational efficiency.
AI Index Report 2026 (Stanford HAI)

10 STEPS OF THE VLK AI GOVERNANCE FRAMEWORK (VLK)

1

Alignment between governance, innovation strategy, and risk appetite
Align the use of AI with the innovation strategy, corporate objectives, and values, defining principles, priorities, limits, and acceptable risks

2

Governance model and responsibilities
Define responsibilities and commitment from senior management, the central governance body, and other decision-making bodies, distributing responsibilities among business, technology, legal, privacy, risk, compliance, and cybersecurity teams, with decision-making authority, segregation of duties, and escalation

3

Inventory and classification
Map systems, models, agents, use cases, data, integrations, and suppliers, classifying risks (prohibitive, high, or low) and the roles of AI agents in the value chain (developer, deployer, integrator, distributor, or user

4

Risk and impact assessment
Identify the legal requirements applicable to the context of use, and legal, ethical, operational, and cyber risks, with proportional requirements for data quality, representativeness, provenance, and lawful use, and for the protection of models, systems, and infrastructure against attacks, manipulation, and leaks

5

Policies and documentation
Structure policies, procedures, records, technical documentation, and contractual clauses

6

Embedding governance into processes
Integrate AI policies and controls into innovation, development, procurement, approval, and use workflows, with records and technical documentation, such as model cards and system cards, as well as human oversight, where applicable

7

Third-party and AI chain management
Establish criteria for the evaluation, selection, due diligence, contractual requirements, monitoring, and accountability of suppliers and partners

8

AI literacy and culture
Provide AI literacy to leadership, technical teams, and users according to their roles and risk exposure, promoting critical, safe, and responsible use of AI

9

Validation and implementation
Conduct algorithmic impact assessments and performance, accuracy, robustness, bias, security, and human-oversight effectiveness tests before going into production, with acceptance criteria, risk mitigation, and appropriate approval

10

Monitoring, incidents, and continuous improvement
Monitor performance, deviations, changes, incidents, and compliance throughout the entire AI life cycle, with corrective measures, incident response, periodic review, and safe decommissioning

Mature governance is not limited to principles and policies:
it integrates and requires commitment from senior management, culture, responsibilities, classification, impact assessments, evidence, release criteria, production monitoring, change management, and incident response.

2 AI INCIDENTS

WHAT ARE AI INCIDENTS AND "HAZARDS"?

- **Incidents:** events related to the development, use, or malfunction of an AI system that directly or indirectly cause: harm to people's health; disruption of the management and operation of critical infrastructure; violations of human rights or failure to comply with obligations intended to protect fundamental, labor, and intellectual property rights; or damage to property, communities, or the environment.

- **"AI hazard":** is the event that could plausibly cause an AI incident, even if the harm has not materialized.

Font: OCDE.

Mature governance must identify, record, and address both situations

There are **800** to **900** unique incidents involving AI through the first quarter of 2026, growing by roughly **130 to 180 new incidents** per year (AI Incident Database - AIID)

In 2025, there were **362 AI incidents**, growing approximately **55%** compared to the previous year (Stanford AI Index 2026)

HOW DO AI INCIDENTS AND RISKS MANIFEST?

VECTORS AND CAUSES
prompt injection, data poisoning, model manipulation, supplier compromise, and unauthorized use of AI.

SYSTEM OR CONTROL FAILURES
performance degradation over time, unsubstantiated responses, excessive permissions, ineffective human oversight, insufficient segregation between data and instructions, loss of control over agents, or unavailability.

IMPACTS
exposure of personal data and trade secrets, discrimination, inaccurate or incorrect decisions, safety harm, operational disruptions, rights violations, and financial and reputational losses.

PRACTICAL CASE
A company uses an AI agent connected to internal databases to answer questions from employees and clients. Without corporate approval, an external tool is integrated into the agent. When processing content that contained a malicious instruction (prompt injection), the agent performs an unintended action and exposes salary data, performance reviews, and other confidential information. The incident combines unauthorized integration, excessive permissions, and a lack of segregation between data and instructions

HOW TO RESPOND TO THIS INCIDENT?

1. **Detect and classify**
Confirm the systems, tools, and integrations involved. Assess the causal link between the AI and the harm

2. **Contain and preserve evidence**
Isolate or disable the agent, block the integration, revoke access, and preserve prompts, configurations, versions, logs, and actions taken, maintaining the integrity and chain of custody of the evidence

3. **Assess severity and obligations**
Determine the nature, volume, and criticality of the data exposed and operations affected. Assess the severity and any legal obligation or ethical duty to report to regulators, affected individuals, and the market

4. **Investigate the root cause**
Identify, where appropriate with the support of specialized experts, the origin of the incident, the vulnerabilities exploited, and the technical, human, or governance failures that allowed it to occur

5. **Remediate and recover**
Reduce permissions, segregate data and instructions, review input filters, restrict unauthorized tools, roll back versions, and test the system's security before resuming operation

6. **Record**
Maintain a chronological record of the incident, evidence, decisions, communications, and measures taken, formalizing its closure and any residual risks

7. **Learn and monitor**
Update impact assessments, risk indicators, contracts, and release criteria. Reinforce training, improve controls, and adopt intensified monitoring after resuming operations

Mature governance is not synonymous with AI incidents never occurring, but rather the ability to detect them early, contain and remedy their impacts, be accountable, safely resume operations, and turn each occurrence into learning and greater resilience.

Direito,
Inovação
& Tecnologia

ABOUT US

At VLK, **law is never a barrier**. It is the catalyst that empowers **innovation, enables business, and helps build** a more prosperous and equitable society.

We are a boutique Digital Law firm driven by meaningful results that create lasting impact.

We bridge:

» Risk and opportunity;

» Complexity and clarity; and

» Protection and progress.

No matter how ambitious the challenge, we help make it happen—securely, pragmatically, and **without unnecessary bureaucracy**.

» Data Protection
» Ethical and Responsible AI Governance
» Cybersecurity & Incident Response
» Advertising & Marketing Law and Intellectual Property
» Technology Regulation
» Strategic Litigation