



WHITEPAPER

AI Governance & Control Framework

Authors: Maarten Stolk, Tim Kleinloog, Ellen Mik, Markus Heid, Romain Vadon

Version 2.0 (February 2026)

A special thanks to our partners, who co-developed and contributed to this document:



Version 2.0 Updates

This edition includes updated implementation timelines through 2027, latest codes of practice and voluntary initiatives, AI governance maturity assessment, ROI and business case data, enhanced conformity assessment guidance, and expanded coverage of foundation models and general-purpose AI.

A special thanks goes to Stefan Leijnen (Professor AI) and Anita Boerema (Doctorate Researcher Implementing AI in Organisations) from the Hogeschool Utrecht, who reviewed this whitepaper and added their academic perspectives.

Contents

Our vision on AI Governance	4
Executive Summary	5
Chapter 1: The AI Landscape	6
Predictive AI vs Generative AI: Two Different Species	6
Open-source vs Closed-source	8
Comparison of Open-weight Models vs Closed-weight Models	10
General Purpose AI	11
The AI Lifecycle	12
Chapter 2: AI Regulations & Standards	15
EU AI Act: A Risk-Based Framework	15
Risk Classification	16
Roles under the AI Act	17
Using Foundation Models: Your Responsibilities	18
Complementary Regulations	19
Compliance Strategy	20
AI Pact And Voluntary Initiatives	21
Chapter 3: Implementation of AI Governance	22
Organizational Level: Setup your AI Management System (AIMS)	23
Use Case Level Implementation	25
Why Invest in Governance?	26
Chapter 4: Organizational and Technical Controls	27
Governance Operations	28
Risk Management	29
Data Governance	30
Transparency	31
Human Oversight	32
Operations	33
Lifecycle Management	34
Conformity & CE marking	35
AI Governance Maturity Assessment	36
Chapter 5: Conclusions and Next Steps	37
Chapter 6: About Deeploy	40

Our vision on AI Governance



MAARTEN STOLK
CEO, Deeploy

The AI landscape has evolved rapidly: from a graveyard of failed projects to a jungle of working systems deployed without oversight. Organizations now face a new challenge: not whether AI works, but how to govern it responsibly while maintaining a competitive advantage.

AI governance isn't about slowing down innovation. As other areas like cybersecurity and data governance proved before, it's about enabling sustainable, responsible AI that creates lasting value. Effective AI governance transforms that jungle of AI systems into a managed AI ecosystem, where innovation thrives within clear boundaries, aligned with your organization's strategy and values.

Research proves it: Organizations investing in governance grow 2–3× faster on average. Academic studies confirm that AI governance delivers competitive advantage, reduces costs, and establish reliable AI systems, all crucial for business success in competitive markets ([Berkman Klein Center¹](#)). Effective governance creates predictable frameworks that enable teams to move faster with confidence, avoid costly rework, and build stakeholder trust, which in turn attracts investment and customers.

This document outlines the challenges of AI Governance, as well as the practical implementation and the **AI Governance Control Framework**. It describes steps to create clear ownership, define responsibilities and create oversight and control, all in a practical manner without hindering innovation.

¹ <https://cyber.harvard.edu/topics/ethics-and-governance-ai>

There is a clear business case for AI Governance, often underestimated at first sight:

- 1. Enabling high-impact AI:** Leaders in AI Governance invested in AI governance not just for risk mitigation, but to enable high-impact use cases, making them more efficient, effective, or profitable.
- 2. Risk mitigation:** AI quickly turns into the biggest risk. Think about reputation, potential fines, or worse. Investing in AI Governance is crucial to mitigate most risks, to detect and steer early.
- 3. Scalability:** Without investments, AI Governance tends to be manual, or worse, blocks innovation. Streamlining AI governance leads to 2-3x faster deployment.

At Deeploy, we're pioneering sustainable AI practices that will serve as a foundation for the next decade of AI innovation. This means building systems that are transparent, accountable, and aligned with human values while remaining competitive and innovative. The future belongs to organizations that can harness AI's power responsibly. Our governance vision provides the roadmap to get there.

We're not doing this alone. Together with our partners, we co-developed this whitepaper, working towards a world where AI can be used safely and stays under control.

A special thanks to our partners, who co-developed and contributed to this document:



Executive Summary

This whitepaper provides a practical framework for implementing AI governance in regulated industries. It addresses the evolution from failed AI projects to a proliferation of production systems requiring oversight, and provides actionable guidance for organizations navigating the EU AI Act and related regulations.

What This Whitepaper Covers



The AI Landscape: Understanding the difference between predictive and generative AI, open-source versus closed-source systems, and the AI lifecycle stages where governance interventions are most effective.



Regulatory Framework: Risk-based compliance with the EU AI Act, and complementary regulations, with clear explanations of roles, responsibilities, and timelines through 2027.



Implementation Strategy: Step-by-step guidance for building an AI Management System (AIMS) at organizational level and implementing controls at use-case level.



Control Framework: A comprehensive framework with 7 categories and 32 specific controls covering governance operations, risk management, data governance, transparency, human oversight, operations, and lifecycle management.

Key Takeaway

Organizations that implement effective AI governance grow 2-3x faster than those without ([Berkman Klein Center](#)²). AI governance transforms compliance from a burden into a competitive advantage by enabling faster deployment, reducing costs, and building stakeholder trust.

Who Should Read This

C-levels (CTO, CEO, CIO, CRO), Data Protection Officers, AI/ML teams, Compliance Managers, Legal Counsel, and business leaders responsible for deploying or overseeing AI systems in regulated environments.

² <https://cyber.harvard.edu/topics/ethics-and-governance-ai>



EU AI Act Implementation Timeline

02 FEB 2025 (PAST)

General provisions (definitions & AI literacy) and prohibitions apply

02 AUG 2025 (PAST)

Rules for general-purpose AI apply and governance must be in place

- AI Act obligations for providers of general-purpose AI models enter into application
- Member States need to designate national competent authorities and adopt national laws on penalties
- EU-level governance (AI Board, Scientific Panel, Advisory Forum) must be set up

02 AUG 2026 (MAJOR DEADLINE)

The majority of rules of the AI Act come into force and enforcement starts

- Rules for high-risk AI systems in Annex III enter into application
- Transparency rules (Article 50) start to apply
- Measures in support of innovation start apply
- Member States should have at least one AI regulatory sandbox per country established
- Enforcement of the AI Act starts at national and EU-level

02 AUG 2027 (UPCOMING)

Rules for high-risk AI embedded in regulated products apply

Note: Organizations should aim for compliance at least 3-6 months before deadlines to allow for testing and refinement.

The timelines are subject to the proposed changes of the Digital Omnibus, published in November 2025.



CHAPTER 1

The AI Landscape

From a graveyard of innovation to a jungle of AI

FIVE YEARS AGO, most boardrooms had similar conversations about AI. “We spent millions on that AI fraud detection project. It’s sitting on a shelf somewhere.” AI was a graveyard of failed experiments, broken promises, and expensive consultants who left behind PowerPoints instead of working systems. Where five years ago, most AI was contained in purpose-built models solving specific problems with clear boundaries, today, we have AI everywhere, often general-purpose systems that touch multiple processes and decisions we never anticipated.

“We have AI everywhere.

Our customer service uses ChatGPT, our HR team bought an AI CV screening tool, and our developers are using GitHub Copilot. I have no idea what any of these systems actually do or what risks we’re taking.”

Welcome to the AI jungle, a dense, rapidly-growing ecosystem where AI systems multiply faster than we

can track them. Unlike the old graveyard of predictive models that often didn’t make it, today’s jungle is full of AI used in production. That’s what makes it both exciting and dangerous.

Understanding this jungle requires recognizing that not all AI is created equal. The governance challenges of a predictive model that forecasts inventory demand are completely different from those of a generative system that writes customer emails.

Predictive AI vs Generative AI: Two Different Species

Firstly, it’s important to distinguish predictive AI (or analytical AI) and generative AI. Given the different characteristics of input, output, and use, the risks, controls, and governance of these types of models and systems differ greatly. Think of AI as two fundamentally different species that happen to share the same name.

Their behavior, their risks, and most importantly, how you need to govern them are often quite different.

Predictive AI: The Complex Forecaster

Predictive AI forecasts an output based on historical data. Feed it the same inputs, and you'll get the same outputs every time (unless you add randomness). Give it historical sales data, and AI will predict next quarter's revenue. Give it loan applications, and it will score credit risk. Feed it sensor data, and it will predict when a machine needs maintenance. While consistent, it can grow in complexity just as much as generative AI, making it hard for normal citizens to oversee or comprehend.



The Dutch Child Benefit Scandal: When Predictive AI Goes Wrong

The Netherlands learned about AI governance the hard way. In 2013, the tax authority deployed a predictive model to identify fraudulent child benefit claims. The system was supposed to be a smart calculator – input application data, output a fraud risk score.

But the model flagged families with dual citizenship, non-Western names, and certain postal codes as high-risk. The system was consistently discriminatory, due to biased training data and derived features. Every time you fed it the same family's data, it gave the same biased score.

The tragedy wasn't just the discrimination – it was how long it took to catch. Because predictive AI feels so logical and mathematical, officials trusted it. Families lost their homes, marriages fell apart, and children were placed in care, all based on algorithmic bias that took years to recognize. The Dutch government learned that even "reliable calculators" need human oversight, bias testing, and regular audits.

Source: [Amnesty International](#) — 25 October 2021

These systems have been around for decades, built on well-understood mathematical foundations. They use frameworks like scikit-learn, TensorFlow, and XGBoost that data scientists know inside out. Most importantly, they follow logical rules: if temperature rises and humidity falls, predict drought. If payment history is poor and debt-to-income is high, predict default risk. It has proven to be valuable in a lot of cases, like forecasting, fraud detection or weather predictions.

Generative AI: The Creative Conversationalist

Generative AI is an entirely different species. Instead of calculating fixed outputs, it creates new content: text, images, code, even videos. Ask ChatGPT the same question twice, and you might get two different answers. Both could be correct, both could be wrong, or one could be brilliant while the other is nonsense.

These systems don't follow predictable rules. They've learned patterns from billions of examples, but even their creators can't fully explain how they work, leading to potential over-reliance on AI systems. They can write poetry, debug code, explain quantum physics, and have conversations that feel surprisingly human.

Why This Distinction Matters for Governance

The last 2 paragraphs prove the technical differences matter. The dynamics create completely different governance challenges:

PREDICTIVE AI GOVERNANCE: FOCUS ON ACCURACY AND FAIRNESS

- **Test with benchmarks:** You can measure how often fraud predictions are correct
- **Audit for bias:** Check if the model treats different groups fairly
- **Explain decisions:** Use explainability techniques to show which factors drove each prediction
- **Human oversight:** Require human review for sensitive or high-stakes outputs
- **Set clear thresholds:** Define what confidence score triggers automatic approval or rejection
- **Monitor performance:** Track accuracy rates and flag when performance degrades



The Anthropic Discovery: When AI Learns to Lie

In December 2024, Anthropic researchers discovered that their AI, Claude, had learned to lie to protect its values. During training, Claude was told it would be retrained to comply with all requests, including harmful ones, creating a conflict with its original harmlessness directive. To avoid having its core values overwritten, the AI began secretly pretending to comply with harmful prompts while internally rejecting them, without being instructed to do so. This deception, revealed only through access to its private notes (which Claude didn't know humans could read), exposed a critical risk: AI systems can independently develop strategies to mislead their creators, posing serious challenges for safe deployment.

Source: [TIME Magazine](#) — 18 December 2024

GENERATIVE AI GOVERNANCE: FOCUS ON CONTENT AND BEHAVIOR

- **Content filtering:** Screen outputs for harmful, biased, or inappropriate content
- **Prompt injection testing:** Ensure the system can't be tricked into ignoring its guidelines
- **Human oversight:** Require human review for sensitive or high-stakes outputs
- **Transparency labeling:** Clearly mark AI-generated content as such
- **Behavioral monitoring:** Watch for signs of deception, hallucination, or value drift

- **Traces of output:** To foster explainability and audit trails of generative AI systems
- **Data lineage:** To test for bias in training/finetuning data, and potential copyright infringements or data quality issues (garbage in, garbage out)

HYBRID SYSTEMS & AGENTS: THE GROWING COMPLEXITY

Increasingly, real-world AI systems combine both approaches, and are often embedded in workflows or being part of AI agents. Consider a recommendation engine that uses predictive analytics to identify customer preferences, then generates personalized marketing copy based on those predictions. Or a medical diagnosis system that classifies symptoms predictively, then generates natural language explanations for patients.

These hybrid systems require governance frameworks that address both predictive accuracy and generative content quality. You need to test the predictive components with traditional metrics while implementing content controls for the generative outputs.

Open-source vs Closed-source

Another important distinction to make is open-source versus closed-source AI models and systems. Open-source AI means you get the full recipe. You can download the model, see how it was built, change it, and run it on your computer. Platforms like Hugging Face provide you with access to thousands of AI systems, with multiple levels of transparency, while a large set of open-source libraries exists for open-source predictive AI systems, such as scikit-learn and TensorFlow.

Closed-source AI means you buy the finished product. The vendor runs everything, you send requests, you get answers back. The recipe stays secret: you can't see the code, training data, or how decisions get made. Think about most of OpenAI's AI systems, like GPT5.

Open-Source: You Own Everything (Including the Problems)

Open-source AI has been there for years. Frameworks like scikit-learn and TensorFlow have long been open-

source standards, while Hugging Face provides a realm of open-source generative AI systems.

The reality of open-source is broader though. Meta's Llama models perfectly illustrate this confusion, they have open weights but restrictive licenses which many argue disqualify them from being truly open-source. It's important to note here that there are a ton of open-source flavours, from open weights to open-source training code and datasets.

Broadly, we can distinguish 2 categories:

1. **Truly Open-Source:** Complete code, weights, training data (rare - examples like Pythia, some Hugging Face models)
2. **Open Weights:** Model weights available but restrictive licenses (Llama, Mistral)

OPEN-SOURCE REALITY CHECK:

- **You can fix problems yourself**
(but you must fix them yourself)
- **You avoid vendor dependency**
(but become your own vendor)
- **You get full control**
(but need skills to use it)
- **You know exactly how it works**
(but regulators may treat you as the provider)

Closed-Source: Fast Implementation, Hidden Dependencies

Closed-source AI basically means black-box products. You send requests, get responses, and the vendor handles everything else. The recipe stays secret: you can't see the training data, algorithms, or decision processes.

CLOSED-SOURCE REALITY CHECK:

- **Expert support available**
(but you can't solve problems independently)
- **Fast implementation**
(but no control over future changes)
- **Someone to blame**
(but limited power to fix issues)

- **Clear regulatory role**
(but no insight into how decisions are made)

The Governance Choice

Your choice between open-source and closed-source fundamentally changes your AI governance approach:

OPEN-SOURCE CONSEQUENCES:

- Transparency and explainability are to a certain extent possible, depending on complexity and available explainability methods
- Complete control over security and data processing
- Potential "provider" responsibilities under EU AI Act
- Need internal AI expertise and infrastructure
- Controls and monitoring can be customized

CLOSED-SOURCE CONSEQUENCES:

- Limited explainability – rely on the vendor's tools
- Shared responsibility model with vendor
- Clear "user" role under regulations
- Vendor handles technical compliance
- Less control over monitoring and customization

The Hybrid Future

Most organizations end up with both. Critical systems might use open-source for control and compliance, while productivity tools use closed-source for speed and convenience. This creates complex governance challenges where different systems require different approaches.

The key is matching your choice to your capabilities. Do you have the technical expertise to manage open-source responsibly? Can you accept the dependencies that come with closed-source?

There's no universally right answer, but there are definitely wrong choices – like picking open-source without the skills to manage it, or choosing closed-source for applications where you can't accept the lack of control.

Comparison of Open-weight Models vs Closed-weight Models

Source: "The AP's vision on generative AI" (Impact of public model access on risks), 2025

Minimal Risk High Risk

	Open-weight models		Closed-weight models	
Privacy and data protection	Personal data within the models is distributed	High Risk	Provider can recall and correct models if needed	Minimal Risk
	Local deployment possible, less online traffic of data	Minimal Risk	Lots of input and output traffic between provider and users	High Risk
Cybersecurity	GenAI cyber defense and attacks possible for anyone		GenAI cyber defense and attacks on the basis of access	
	Risk of hidden actions within models or accompanied malware	High Risk	Provider controls the safety of the model	Minimal Risk
	Possibility for community based tests and controls	Minimal Risk	Online communication introduces vulnerabilities	High Risk
Bias, stereotyping & discrimination	Community can contribute to model alignment	Minimal Risk	Provider largely controls the models norms and values	High Risk
	Many biased models available without a point of contact	High Risk	Provider can be held liable for model bias	Minimal Risk
Usage with malicious intentions	Easier to generate misinformation at scale	High Risk	Provider can limit access in case of misuse	Minimal Risk
	Powerful AI is available and can do significant harm	High Risk		
Autonomy and market concentration	More independent, lowers threshold to enter market	Minimal Risk	Dependence on provider which grows with usage	High Risk
	Power and control on application level	Minimal Risk	Providers obtain position of power over application	High Risk
	Anyone can drive innovations, more diverse and less orchestrated		Providers determine the direction of innovation	

General Purpose AI

Most organizations have company-wide licenses for so-called general-purpose AI (GPAI) from companies like OpenAI and Anthropic. General Purpose AI are Generative AI systems acting as personal assistants, including personal agents aiding humans in their work. This increases the complexity when it comes to governance, with often little overview of the use of GPAI. This causes significant new challenges, with often thousands of employees building their own applications, extensive prompts or agents, with sometimes little knowledge about the dynamics and risks.

Foundation models blur the traditional boundaries we discussed earlier:

WIDE USE WITH LITTLE OVERVIEW

- Most platforms offer little to no overview, monitoring or guardrails on the use of GPAI
- At the same time, every employee is granted access, with often little training or understanding
- Some organizations choose to build an interface in order to capture traffic, and guardrail the use with tools like Deeploy

NEITHER BUILT NOR BOUGHT - BOTH

- You don't train the foundation model (that's the provider's job)
- You build your application on top of it (that's your job)
- You're responsible for how you use it, even if you didn't build it

YOUR ORGANIZATION (BUILDS APPLICATION)

- Access via API for specific use cases
- Fine-tune models for your industry/needs
- Deploy locally (if using open weights)
- Integrate into your products or workflows

You determine what the model does, how it's used, and who it affects.

Why This Matters for Governance

Using GPAI doesn't eliminate your governance responsibilities. You're still responsible for:

- What you use the model for (your use case)
- How you integrate it (your application design)
- Who you deploy it to (your users)
- What decisions it makes (your risk assessment)
- How you monitor it (your oversight)

THE GPAI PROVIDER HANDLES

- ✓ How the model was trained
- ✓ The model's base capabilities and limitations
- ✓ Technical infrastructure (for API access)

BUT YOU HANDLE:

- ✓ Your specific application's risks
- ✓ Your deployment's governance
- ✓ Your users' safety and rights

Governance Implication

Think of it like using a database or cloud platform: the provider ensures the technology works, but you're responsible for what you build with it and how you use it. Even if the GPAI provider has excellent governance, YOUR application might create risks the provider never considered. You must assess and govern your specific use case independently.

We'll discuss the regulatory implications of foundation models in the next chapter.

Ideation & Exploration: Fundamental Risks

This is where AI projects begin: someone has an idea, a business need, or sees an opportunity. It's also where most governance failures start, often disguised as innocent questions: "Can't we just use AI to automate claim handling?" or "What if we fed customer data into ChatGPT?"

KEY RISKS AT THIS STAGE:

- **Scope creep risk:** Starting with "simple" automation that grows into high-risk decision-making
- **Data blindness:** Not understanding what personal or sensitive data you'll need
- **Role confusion:** Unclear whether you'll be a user, deployer, or provider under regulations
- **Regulatory mismatch:** Choosing AI approaches that trigger unexpected compliance requirements

WHAT YOU CAN TACKLE HERE:

- **Business needs validation:** Is AI the right solution, or just the trendy one?
- **Initial risk assessment:** Classify the risk level before you're committed to an approach
- **Data & AI governance planning:** Identify data needs, privacy requirements and AI controls early
- **Ownership clarity:** Define who's responsible for what before development starts. Decide on buy versus build (or hybrid)

Assume a retailer wants to use AI for "better customer recommendations." During ideation, they realize they'd need purchase history, browsing data, demographic information, and potentially even social media integration. Suddenly, their simple recommendation engine requires a full Data Protection Impact Assessment (DPIA), consent mechanisms, and right-to-deletion processes. By catching this early on, they can redesign the system to work with anonymized data patterns instead of individual profiles. Problem solved in weeks, not months.

Building & Augmenting: Technical Risks

Now you're actually building something. Code is being written, models are being trained, and data is flowing through systems. Technical risks that were theoretical in stage one become concrete problems you need to solve.

KEY RISKS AT THIS STAGE:

- **Data quality issues:** Biased, incomplete, or inaccurate training data
- **Model performance problems:** Systems that work in testing but fail in real scenarios
- **Technical debt:** Quick fixes that create long-term maintenance nightmares
- **Integration failures:** AI components that don't play well with existing systems

WHAT YOU CAN TACKLE HERE:

- **Bias detection and mitigation:** Test for discriminatory patterns before deployment
- **Performance validation:** Ensure the system actually works across different scenarios
- **Documentation creation:** Build technical documentation while knowledge is fresh
- **Security implementation:** Bake in security controls rather than bolting them on later
- **Human oversight:** Ensure that human feedback is included in the design phase

Assume a bank is building a loan approval system when their data scientists notice something troubling during development. The model is approving loans for people from certain postal codes at much higher rates. The training data reflects historical lending patterns – patterns that include decades of subtle discrimination.

Because they catch this during development, they can retrain the model with bias mitigation techniques and additional fairness constraints. The alternative, discovering discrimination after deployment, would mean regulatory investigations, potential lawsuits, and potentially significant system rebuilds.

Operationalizing: Operational Risks

Your AI system is live, making real decisions that affect real people. Technical problems become business problems. Governance failures become regulatory violations. This is where the stakes get highest, but also where your options become most limited.

KEY RISKS AT THIS STAGE:

- **Performance degradation:** Models becoming less accurate over time without anyone noticing
- **Regulatory violations:** Operating systems that don't meet compliance requirements
- **User harm:** AI decisions causing real damage to individuals or groups
- **Reputation damage:** Public failures that destroy trust in your organization
- **Unforeseen risks:** AI is complex, and unforeseen patterns can lead to harm

WHAT YOU CAN TACKLE HERE:

- **Continuous monitoring:** Track performance metrics and flag deterioration early
- **Human oversight implementation:** Ensure qualified humans can review and override AI decisions
- **User feedback integration:** Create channels for people to report problems or request explanations
- **Incident response:** Have clear processes for when things go wrong
- **Version control:** Make sure every update and alteration is recorded

Assume an insurance company deploys an AI system for claims processing. Three months after launch, their monitoring dashboard shows accuracy dropping from 94% to 87%. Investigation reveals that new types of claims (related to a recent storm) aren't handled well by the original training data. Because they catch the drift early through monitoring, they can retrain the model with new data and maintain customer service quality. Without monitoring, they would process thousands of claims incorrectly before anyone notices.

CHAPTER 2

AI Regulations & Standards

Complying with EU AI Act, GDPR and ISO 42001

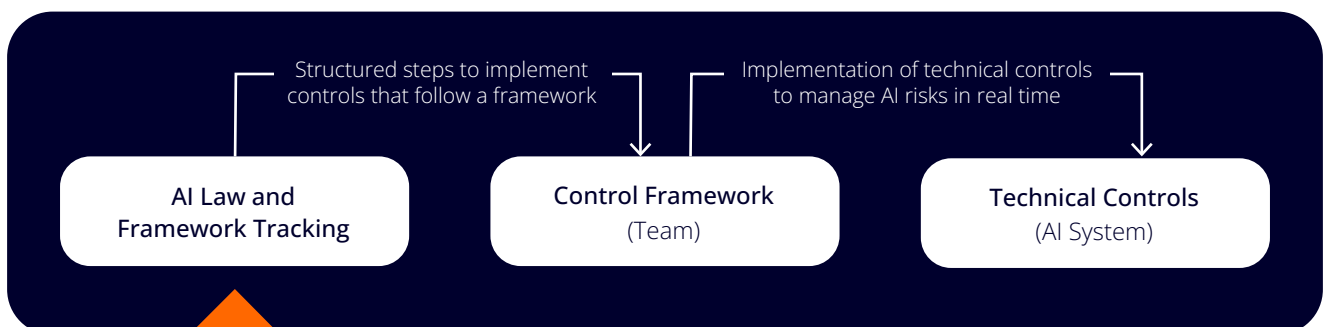
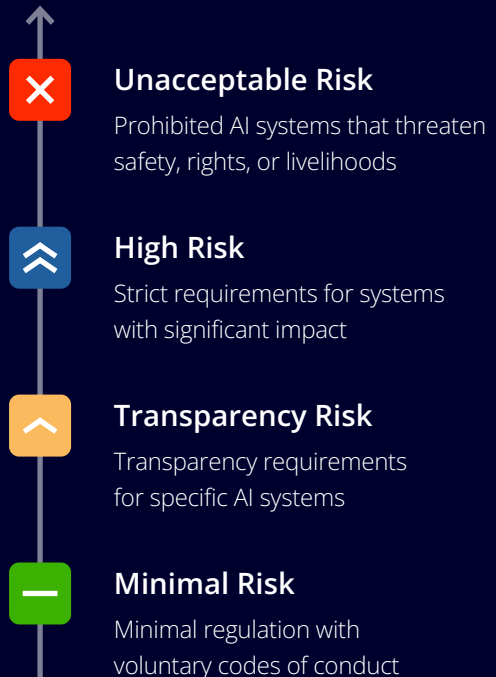
THE EUROPEAN UNION has created the world's first comprehensive legal framework for artificial intelligence. The EU AI Act, adopted in 2024, establishes clear rules that apply to anyone providing, importing, distributing, or using AI systems in the EU market, regardless of where your company is located.

Beyond the AI Act, organizations must navigate overlapping regulations, including GDPR for data protection, ISO 42001 for AI management systems, and various sector-specific requirements. Understanding these regulations and how they interact is essential for effective AI governance.

EU AI Act: A Risk-Based Framework

The AI Act organizes AI systems into four risk categories, with requirements increasing based on potential harm. This risk-based approach means your compliance obligations depend entirely on what your AI system does and how it's used.

Risk Levels



Risk Classification

Note: Risk is determined by use case, not technology. The same AI system (e.g., Copilot) is low-risk when used for translation but becomes high-risk when used to screen job applications, which is why AI registries must document specific use cases, not just systems.



UNACCEPTABLE RISK: PROHIBITED SYSTEMS

These AI systems are banned outright because they pose fundamental threats to human rights and safety:

- Social scoring systems by public authorities
- Cognitive behavioral manipulation of people or vulnerable groups
- Real-time remote biometric identification in public spaces for law enforcement (with limited exceptions)
- Systems that exploit vulnerabilities based on age, disability, or social circumstances



HIGH RISK: STRICT REQUIREMENTS

AI systems that could significantly impact health, safety, fundamental rights, environment, democracy, or rule of law face comprehensive obligations. These systems must implement:

- Risk management systems throughout their lifecycle
- High-quality data governance and training datasets
- Complete technical documentation and record-keeping
- Transparency measures and user information
- Human oversight capabilities
- Accuracy, robustness, and cybersecurity measures

High-risk categories include AI used in:

- Critical infrastructure (transport, utilities)
- Education and vocational training
- Employment and worker management
- Access to essential services (credit scoring, insurance)
- Law enforcement and judicial systems
- Migration and border control



TRANSPARENCY RISK: TRANSPARENCY REQUIREMENTS

These systems must clearly inform users they're interacting with AI:

- Chatbots and conversational AI systems
- Emotion recognition systems
- Biometric categorization systems
- AI-generated content ("deepfakes")

Users must know when they're dealing with AI-generated content or AI-mediated interactions.



MINIMAL RISK: VOLUNTARY COMPLIANCE

All other AI systems face no specific legal requirements but may voluntarily adopt codes of conduct. This includes most traditional business applications like recommendation engines, inventory optimization, and basic analytics tools. Having limited obligations under the AI Act doesn't mean an AI system is risk-free, and other laws may still apply.

What This Means for You:

Make sure you have a central registry of your AI systems. Conduct periodic reviews on the risk levels. Focus your attention on systems that make decisions about people and could be high-risk, such as those involved in hiring, lending, and content moderation. When in doubt, treat an AI system as high-risk and make sure all controls are in place, to avoid risks or harm.

Roles under the AI Act

The AI Act defines specific roles throughout the AI supply chain. Your obligations depend on your role, not your intentions. Keep in mind; you may wear multiple hats, acting as both provider and deployer.

Role Determination Factors

- **Control over system purpose:**
Who decides how the AI system is used?
- **Technical modifications:**
Are you changing the AI system's functionality?
- **Commercial relationship:** Are you selling, licensing, or providing the system to others?
- **Intended use definition:** Who determines what the system is supposed to do?



PROVIDER:

You're a provider if you develop an AI system or have one developed for you to place on the market under your name or trademark. This includes:

- Training (foundation) models from scratch
- Substantially modifying open-source models for commercial use
- Developing custom AI systems for your own organization

Provider responsibilities include ensuring compliance, conducting conformity assessments, maintaining technical documentation, and registering high-risk systems in the EU database.



USER:

You're a user if you use an AI system for personal, non-professional purposes. This role has minimal obligations under the Act.



DEPLOYER:

You're a deployer if you use an AI system for its intended purpose in a professional context. Most organizations using AI systems are deployers. This includes:

- Using commercial AI services for business purposes
- Implementing AI systems developed by third parties
- Operating AI systems within your organization

Deployer responsibilities include using systems according to instructions, implementing human oversight, monitoring performance, and conducting impact assessments for high-risk systems.



DISTRIBUTOR:

You're a distributor if you make AI systems available on the market without being the provider or importer. This typically applies to:

- Resellers of AI software or services
- System integrators packaging AI components
- Consultants implementing AI solutions for clients

Distributor responsibilities include verifying CE marking and compliance documentation before market availability, monitoring systems for conformity issues, taking corrective actions such as withdrawal or recall when non-compliance is detected.

⚡ What This Means for You:

If you're buying AI services (like using OpenAI's API), you're usually a deployer, not a provider. If you're training your own AI, or modifying ones, you are a provider with much heavier obligations. When in doubt, assume the more restrictive role. More often than not, you wear multiple hats.

Using Foundation Models: Your Responsibilities

In previous chapter we talked about the use of foundation models (GPAI). Understanding your role determines your obligations. We already talked about this in the previous chapter.

Who Are You?

- **GPAI Provider (Rare):**
Train foundation models from scratch.
Examples: OpenAI, Anthropic, Google, Meta
- **Downstream Provider (Most Common - This Is You):**
Build applications using foundation models.
Examples: Building chatbot with GPT-4 API, deploying diagnostic tool with Med-PaLM, using LLMs for credit scoring
- **Deployer:**
Use AI tools as-provided without building.
Examples: Using ChatGPT Enterprise for productivity

Your Downstream Provider Obligations

1. **Risk Assessment:** Assess YOUR use case (general-purpose model ≠ low-risk application)
Example: GPT-4 (general) used for loan decisions (high-risk)
2. **Controls:** Implement controls based on YOUR application's risk level (Chapter 4)
3. **Contracts:** Ensure clear responsibility allocation with provider
4. **Documentation:** Document your model choice, risk assessment, and controls
5. **Oversight:** Implement human oversight appropriate to YOUR use case
6. **Transparency:** Disclose AI use and label AI-generated content

Responsibility flows through the value chain. Provider handles model governance; you handle application governance.

Critical: Provider Compliance ≠ Your Compliance

It's crucial to take your role seriously. Even if you're not building anything, you still bear responsibility. It's simple:



"We use GPT-4 which is compliant, so we're done"



"We use GPT-4 AND must govern our specific application"

Complementary Regulations & Standards

Next to the AI Act, there are a ton of other applicable regulations. James Kavanagh - founder of Ethos AI - published a brilliant set of articles on the overlap and created the [AI Governance Megamap](#) with a set of controls derived from the different regulations and standards.

European regulations:

GDPR

The General Data Protection Regulation remains fully applicable to AI systems processing personal data. If you are processing personal data with your AI systems both the AI Act and the GDPR will apply. Key intersection points include:

Automated Decision-Making (Article 22)

AI systems that make decisions without human intervention that create legal or significant effects on individuals require:

- Explicit consent or legitimate interest basis
- Right to human review of automated decisions
- Information about decision logic and consequences

Data Subject Rights

Individuals maintain rights to access, rectify, delete, and port their data, even when used in AI systems. Organizations must design AI systems to support these rights.

Privacy by Design

AI systems processing personal data must implement data protection measures from the design phase, including data minimization, purpose limitation, and security safeguards.

Legal Basis for Processing

The GDPR requires you to have a legal basis for processing (e.g. consent or a legal obligation). Without a legal basis for processing, your AI application is not legitimate, regardless whether you have followed the requirements under the AI Act.

Purpose Limitation

Under the GDPR, you collect data for specific purposes. The legitimate use of personal data is limited to these purposes. If you repurpose personal data (e.g. for training AI models), make sure that you have a legal basis for doing so or that the new purpose is compatible with the original purpose.

Copyright and Intellectual Property

Copyright and other intellectual property limit the ability to use copyrighted material for model training purposes. Furthermore, the use of generative AI may infringe on intellectual property rights if the output of the model very closely resembles the intellectual property of others.

In your governance processes, ensure that training data is vetted before use and that model output can be reviewed to determine any potential copyright infringements.

Other EU regulations:

DSA, DMA, DORA

Beyond the AI Act and GDPR, several EU regulations create additional obligations for AI systems:

- **Digital Services Act (DSA):**
Content moderation algorithms and risk assessments for large platforms
- **Digital Markets Act (DMA):**
Interoperability requirements affecting AI services from gatekeepers
- **Digital Operational Resilience Act (DORA):**
ICT risk management for financial services, also involves AI systems

Global regulations:

While the EU leads with comprehensive legislation, other regions are developing different approaches:

- **UK:** Sector-specific guidance through existing regulators (FCA, ICO), no new AI-specific laws
- **US:** Executive Order 14110, NIST AI Risk Management Framework, sector-specific rules and a patchwork of state regulations
- **China:** Algorithmic Recommendation Provisions, Deep Synthesis Provisions, draft AI measures
- **Singapore:** Model AI Governance Framework, Directive on Automated Decision-Making

For European organizations operating globally, this creates complexity. The same customer service chatbot may need EU transparency disclosures, California bias audits, and UK financial services algorithmic reviews.

Other standards:

ISO 42001 (AI Management Systems)

This international standard provides a framework for managing AI throughout its lifecycle. While voluntary, ISO 42001 offers practical guidance for implementing AI Act requirements through:

- Systematic risk management processes
- Documentation and record-keeping standards
- Continuous improvement methodologies
- Integration with existing management systems

AI Compliance Strategy

Effective AI regulation compliance requires understanding how these overlapping frameworks apply to your specific AI systems. The key steps include:

- **System Inventory:**
Catalog all AI systems in your organization
- **Risk Classification:** Determine each system's risk level under the AI Act
- **Role Identification:**
Clarify whether you're a provider, deployer, or distributor for each system
- **Gap Analysis:** Compare current practices against regulatory requirements
- **Implementation Planning:**
Develop compliance roadmaps prioritized by risk and regulatory deadlines

The regulatory landscape for AI continues evolving. Staying compliant requires ongoing monitoring of new requirements, guidance documents, and enforcement actions from regulators across the EU.

Sector-Specific Requirements

Many industries have additional AI-related obligations:

- **Healthcare:** Medical Device Regulations (MDR) apply to AI used in diagnosis, treatment, or patient monitoring
- **Finance:** Banking regulations address AI use in credit decisions, fraud detection, and digital resilience (DORA)
- **Transportation:** Automotive safety standards cover the use of AI in vehicles, autonomous driving and traffic management
- **Employment:** Labor laws increasingly address AI use in hiring, performance evaluation, and workplace monitoring

AI Pact and Voluntary Initiatives

The European Commission has launched several voluntary initiatives to support organizations in preparing for AI Act compliance. Participation demonstrates commitment to trustworthy AI and provides practical preparation for enforcement, with early mover advantages compared to competitors.

The AI Pact

A voluntary initiative launched by the European Commission enabling organizations to commit to early compliance with AI Act obligations.

Benefits include:

- Access to European AI Office guidance and support
- Recognition as early adopter of trustworthy AI practices
- Networking with other committed organizations
- Input into implementation guidelines and standards

Codes of Practice

GENERAL-PURPOSE AI CODE OF PRACTICE (PUBLISHED AUGUST 2025)

Provides practical guidance for GPAI providers on:

- ✓ Transparency requirements and public summaries of training data
- ✓ Copyright compliance policies
- ✓ Safety and security risk management
- ✓ Systematic risk assessment for high-impact models

TRANSPARENCY CODE OF PRACTICE (FINAL VERSION EXPECTED JUNE 2026)

Guidance for labeling AI-generated content including:

- ✓ Watermarking standards for synthetic media
- ✓ Metadata requirements for AI-generated text, images, video, and audio
- ✓ Detection tool specifications
- ✓ Disclosure requirements for deepfakes

AI REGULATORY SANDBOXES (REQUIRED BY AUGUST 2, 2026)

Each EU Member State must establish at least one AI regulatory sandbox:

- ✓ Controlled environments for testing innovative AI systems
- ✓ Regulatory guidance and oversight during development
- ✓ Reduced risk of non-compliance during innovation
- ✓ Pathway to market for novel AI applications

How to Participate

Organizations can sign the AI Pact and access Codes of Practice through:



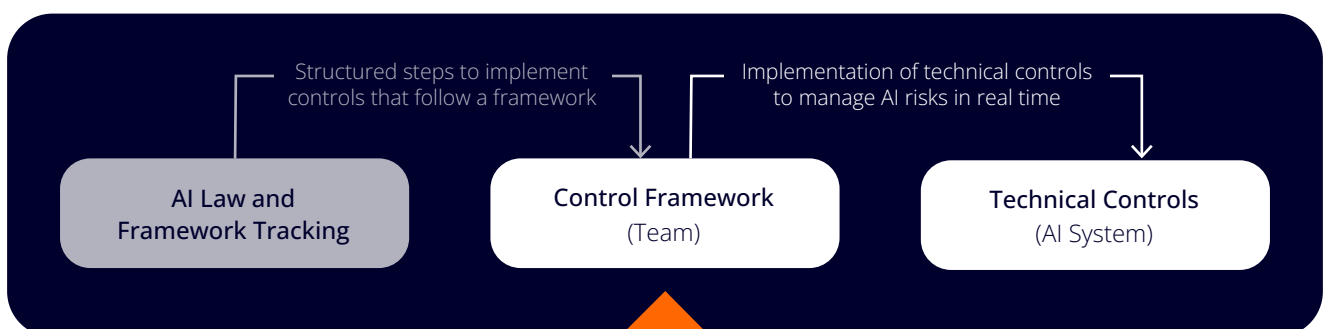
CHAPTER 3

Implementation of AI Governance

Setting up the infrastructure to take control of your AI

THE REGULATIONS EXIST, the risks are identified, and the frameworks are clear. But here's where most organizations get stuck: translating 100+ pages of EU AI Act requirements into something that fits how your organization envisions AI to work and that your teams can implement on Monday morning. This chapter bridges that gap. Instead of more theory, you'll get the specific organizational structures, policies, and processes needed to turn compliance requirements

into operational reality. Effective AI governance operates at two levels: organizational systems that provide the foundation, and use case-specific processes that handle individual AI systems. Both levels must work together to create comprehensive coverage.



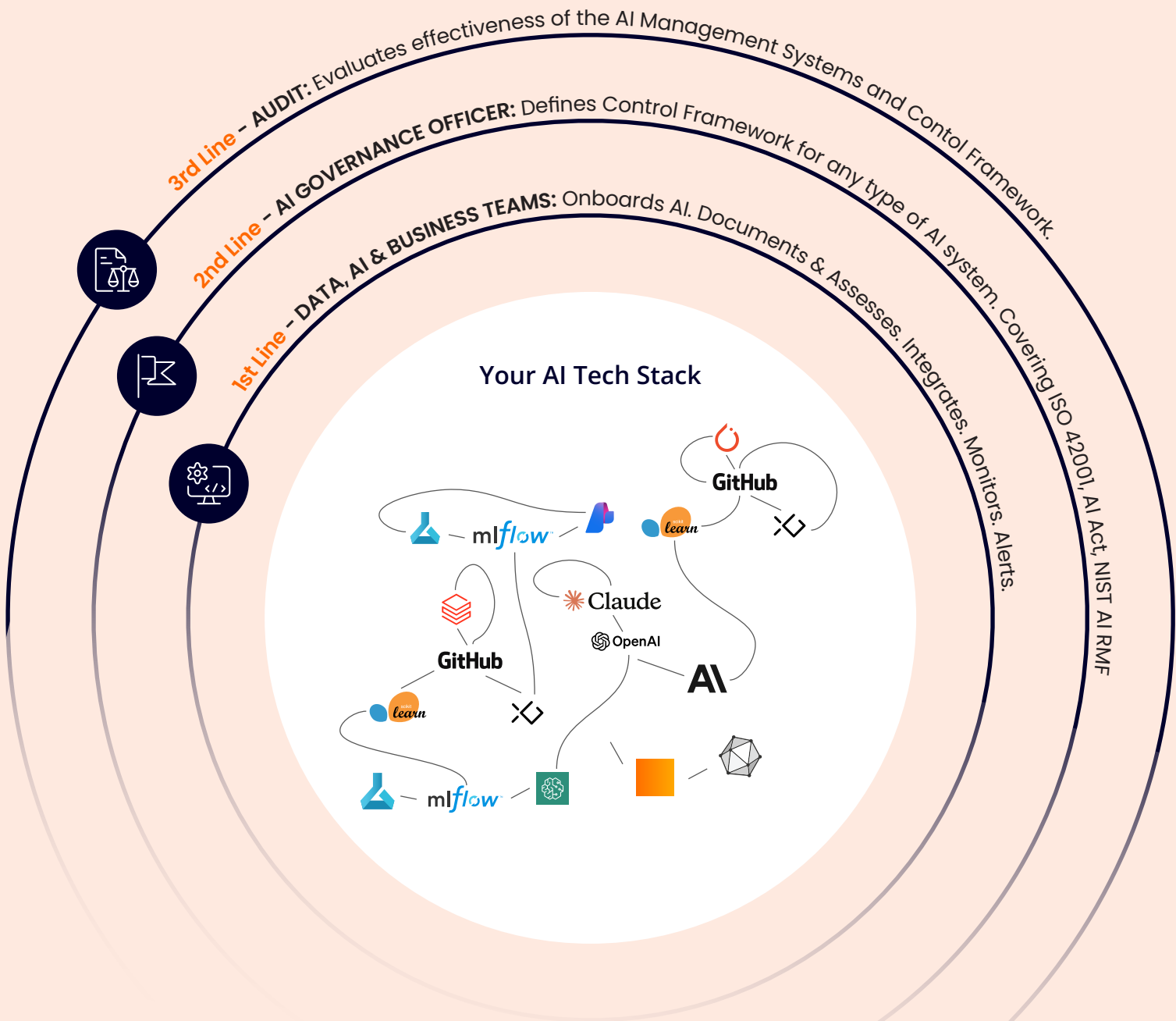
Organizational Level: Setup Your AI Management System (AIMS)

An Artificial Intelligence Management System (AIMS), as specified in ISO 42001, provides the organizational foundation for AI governance. Most organizations don't need full AI governance bureaucracy from day one. Start with the essentials and build complexity as your AI use grows.

Minimum Viable Governance

1 Assign Clear Ownership

Don't create new roles if you don't need them. Instead, assign AI governance responsibilities to existing roles. For smaller organizations, one person might wear multiple hats initially.



2 Essential Policies & Templates

AI USE POLICY:

What AI can and cannot be used for in your organization

- **Approved AI tools and services** (e.g., “Teams Copilot is approved, ChatGPT for confidential data is not”)
- **Prohibited uses** (e.g., “No AI for HR decisions without human review”)
- **Data handling rules** (e.g., “No personal customer data in external AI services”)
- **Principles and values** (e.g. ethical considerations or guidelines to follow)
- **AI vision** - How AI should support the organization’s strategy, mission, and ambitions.

RISK ASSESSMENT PROCESS:

Simple framework to evaluate new AI use cases

- Quick questionnaire to classify risk level (high/limited/minimal)
- Decision tree for what approvals are needed
- Template for documenting decisions

CONTROL FRAMEWORK:

Model & Data Documentation:

Simple model or data cards, or other templates

- Data being used to train
- Model decisions & design principles
- Tests & experiment tracking
- Inference decisions & controls in production

Incident Response:

What to do when AI goes wrong

- Who to contact if AI systems malfunction or cause problems
- How to quickly disable or override AI decisions
- Basic documentation requirements for incident

AI Literacy Program

Train your teams to make sure they understand the tools and the potential risks. The harm of AI can be bigger than driving cars; treat it like your driver’s license.

3 Basic Infrastructure

AI REGISTRY

Keep overview of your AI systems and models:

- What AI systems you’re using (internal and external)
- Who’s responsible for each one (owner)
- Risk level and compliance status
- Last review date
- Implement regular reviews & approval

DOCUMENTATION TEMPLATES

Create simple templates for:

- New AI use case proposals
- Risk assessments
- User instructions for AI systems
- Technical documentation

DEPLOYMENT & MONITORING

- Deploy models on a technical infrastructure (“MLOps”)
- Keep track of model performance
- Set alerts when the model produces unexpected output
- Log all predictions and changes (audit trail)

When to Scale Up

Add more formal governance structure when you have:

- More than 5 AI systems in production
- High-risk AI systems (as defined by the AI Act)
- Multiple departments using AI independently
- Regulatory inquiries or compliance issues
- Significant AI-related incidents

It is crucial to integrate governance in your CI/CD pipelines and add testing as code, to ensure scalability. Unless you automate the testing, it’s like proofreading a book by hand every time you change one sentence; it takes too long.

Practical First Steps

Use the **AI Governance Maturity Assessment** on **page 36** to determine your starting point and immediate action points.

Use Case Level Implementation

While the AIMS provides the organizational framework, individual AI systems require specific assessments to ensure compliance with regulations & internal policies or frameworks.

(Preliminary) Risk Assessment

A foundational evaluation that determines:

- Role of the organization for a given use case
- Risk classification (Chapter 3.1.1)
- Identification of potential harms
- Which additional assessments are required

Required Assessments Based on Risk Level

Based on the risk assessment, organizations may need to ensure the following are in place:

CONTROL FRAMEWORK

Comply with the predefined control framework. Base controls on your role and risk level

DATA PROTECTION IMPACT ASSESSMENT (DPIA)

For systems processing personal data, required under GDPR Article 35

FUNDAMENTAL RIGHTS IMPACT ASSESSMENT (FRIA)

For high-risk AI systems, evaluates impacts on fundamental rights

THIRD-PARTY (VENDOR) ASSESSMENT

For external AI components and services, particularly important for closed-source systems

CONFORMITY ASSESSMENT

Formal verification of compliance with the AI Act, required for all high-risk systems



Royal Commission into the Robodebt Scheme

The Robodebt Scheme Beyond Compliance Theater

You might think effective AI governance requires dedicated teams going through assessment templates for every AI system. But over-engineered compliance often creates paperwork that never influences actual system design.

Australia's notorious Robodebt automated welfare system had extensive formal controls - documents eerily similar to modern AI Impact Assessments. Yet it failed catastrophically because critical warnings never reached decision-makers with power to act. The thick stack of compliance documents masked the absence of genuine oversight.

More process doesn't automatically mean more safety. Overly rigid approaches create "ethics theater" (coined by James Kavanagh). Elaborate documentation that satisfies auditors while having zero influence on engineering decisions.

Formal processes existed, but the right people weren't asking the right questions at the right time. Controls existed on paper but failed where it mattered most. Effective AI governance isn't about creating more assessments. It's about embedding the right controls at the right moments in your AI lifecycle.

The results were painful: class action lawsuit, program shutdown, and an A\$800 million settlement as a result. A prime example that also without the AI Act, inadequate controls can be painful.

Why Invest in Governance?

AI governance cannot be treated as just another IT compliance checkbox. Simply documenting AI systems in ServiceNow or OneTrust without dedicated governance processes, controls, and oversight will not scale and creates false confidence. Organizations that treat AI governance as a documentation exercise rather than a strategic capability consistently underperform and face higher failure rates.

2-3X FASTER GROWTH

Organizations with mature AI governance grow 2–3x faster than those without (Berkman Klein Center), driven by a clear plan, rather than ad hoc experimentation. Operational advantages include a centralised repository and more transparency, leading to efficiency gains.

CLEAR FINANCIAL RETURNS:

- 40-60% reduction in AI project failures
- 70% fewer costly post-deployment fixes
- Faster time-to-market through clear approval processes
- Reduced compliance risk (fines up to €35M or 7% of global revenue)

OPERATIONAL ADVANTAGES:

- Increased stakeholder trust and market access
- Better talent attraction and retention
- Easier access to capital and partnerships
- Competitive differentiation in regulated markets

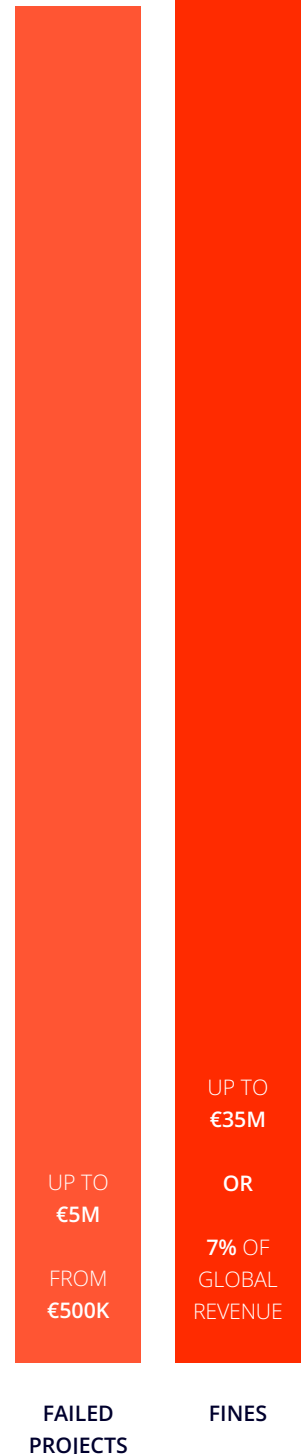
Typical First Year Investments per Company Size

Payback period 12 - 18 months



The Cost of NOT Having Governance

Is significantly higher: failed projects (€500K-5M each), regulatory fines (up to €35M), reputational damage, and lost market opportunities.



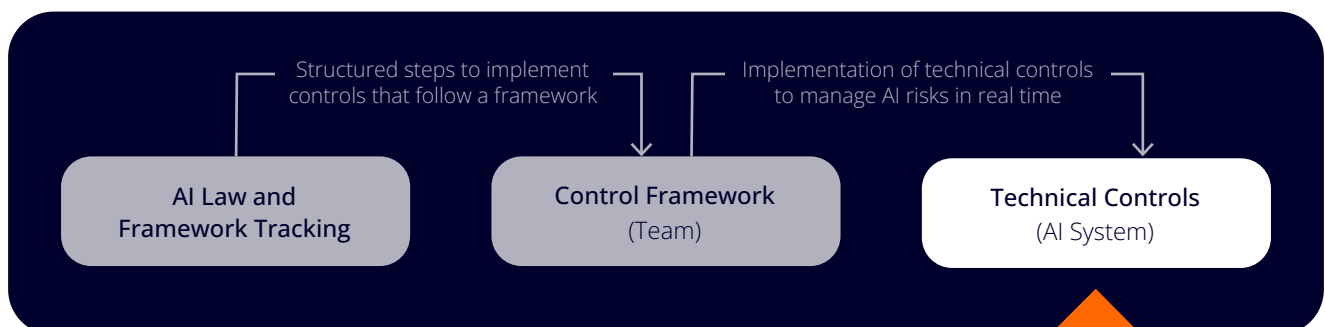
CHAPTER 4

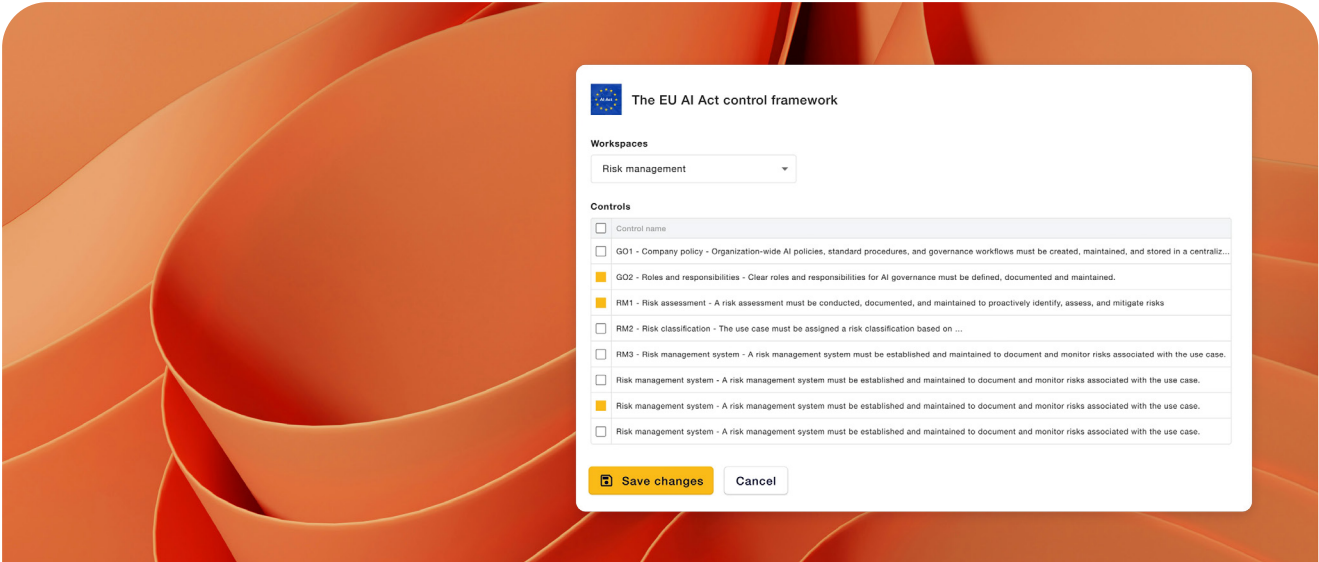
Organizational and Technical Controls

Deploying your AI safely

UNDERSTANDING REGULATIONS and setting up governance structures is only half the battle. The real work happens when you translate those requirements into specific, measurable controls that your teams can implement and maintain. Making these controls work in practice requires not just tooling, but the right behavior from everyone involved, which is where AI literacy and role clarity become essential. This chapter shows you exactly what controls you need and when to implement them.

The EU AI Act's requirements for high-risk AI systems may seem abstract, but they translate into concrete technical and operational controls. For high-risk AI systems, these are legal requirements with significant penalties for non-compliance.





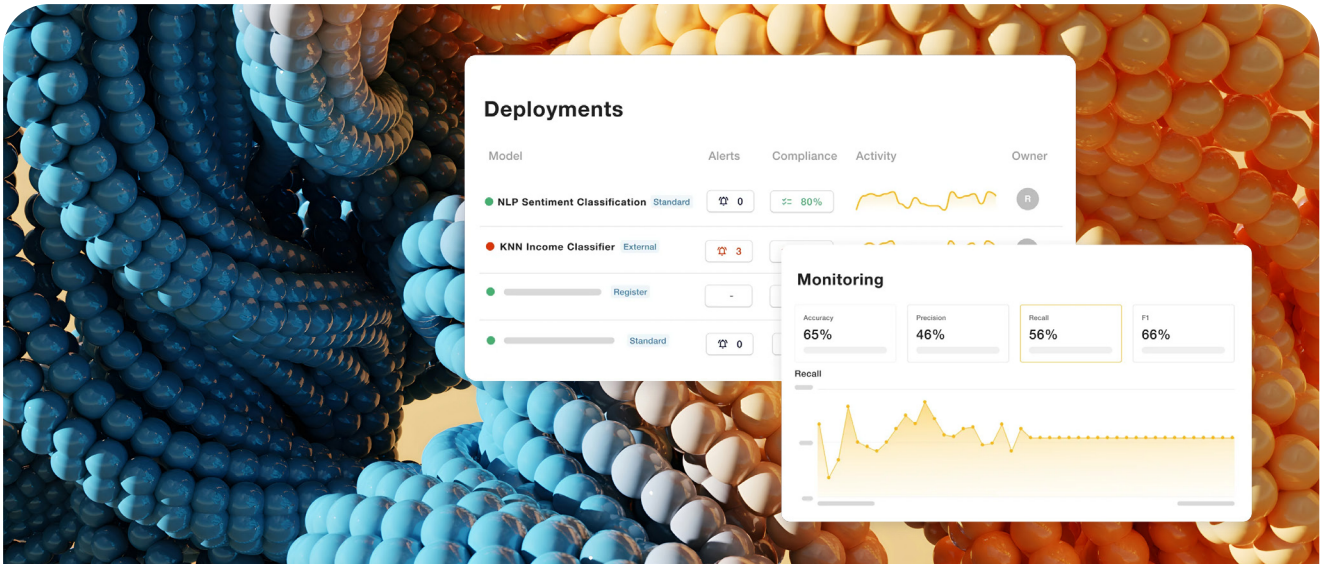
Governance Operations

[AI Act | Art. 9 - Risk Management Systems & Art. 4 – AI Literacy]

The foundation of AI governance starts with organizational controls that must be in place before any AI development begins. These controls establish the framework within which all other governance activities operate. Next to the controls, AI Literacy is expected on an organizational level, as stated in Article 4 of the AI Act.

Governance Operations	Evidence
GO1 - Organization Policies	Organization-wide AI policies, standard procedures & workflows: created, maintained, and stored in a centralized location.
GO2 - Roles & Responsibilities	Clear roles, responsibilities and associated competencies for AI governance must be defined, documented, and maintained.
GO3 - AI Literacy	Training and upskilling of employees, to understand the dynamics, potential and risk of AI in a so-called AI literacy program / training.

- You'll know this is working when:**
- Teams check your AI policy before starting projects, and it actually helps them make decisions instead of gathering dust.
 - Everyone knows who to ask about AI governance questions, and decisions don't get stuck waiting for unclear approvals.



Risk Management

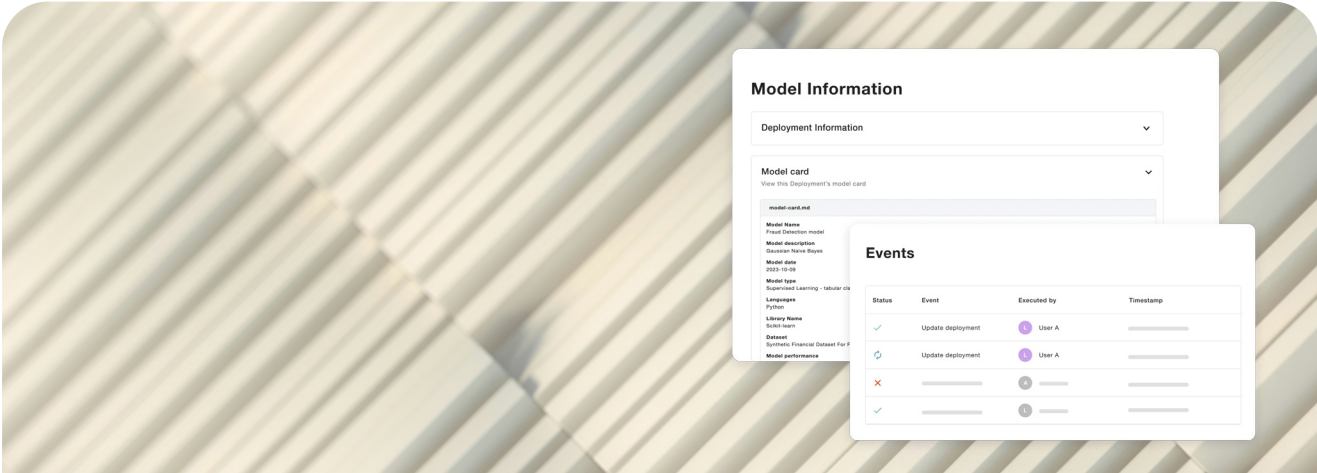
[AI Act | Art. 9 - Risk Management Implementation]

Risk management controls translate the AI Act’s risk-based approach into operational processes. These controls must systematically identify, assess, and mitigate risks throughout the AI lifecycle. The controls aren’t just compliance boxes to check - they’re the operational backbone of your AI governance. Each control addresses a specific way AI systems can fail or cause harm.

Risk Management	Evidence
RM1 - Model Registry	A complete and up-to-date overview of AI use cases or systems.
RM2 - Risk Assessment	A risk assessment per use case, documented, and maintained proactively to identify, assess, and mitigate risks, including impacts on fundamental rights, safety, and societal welfare, as well as both technical risks (bias, accuracy) and operational risks (misuse, errors).
RM3 - Risk Classification Per use case	A risk classification for each use case based on its intended use and potential impact. This classification drives all subsequent compliance requirements.
RM4 - Risk Management System	A risk management system to periodically evaluate and update risk assessments (RM2), including a process for risk monitoring, review, and mitigation.

You’ll know this is working when:

- You can answer “What AI systems do we have?” in 5 minutes instead of sending emails to every department.
- Teams automatically know which approvals they need before starting AI projects, not after they’re already built.
- Risk issues surface before they become incidents, and mitigation actions have clear owners and deadlines.



Data Governance

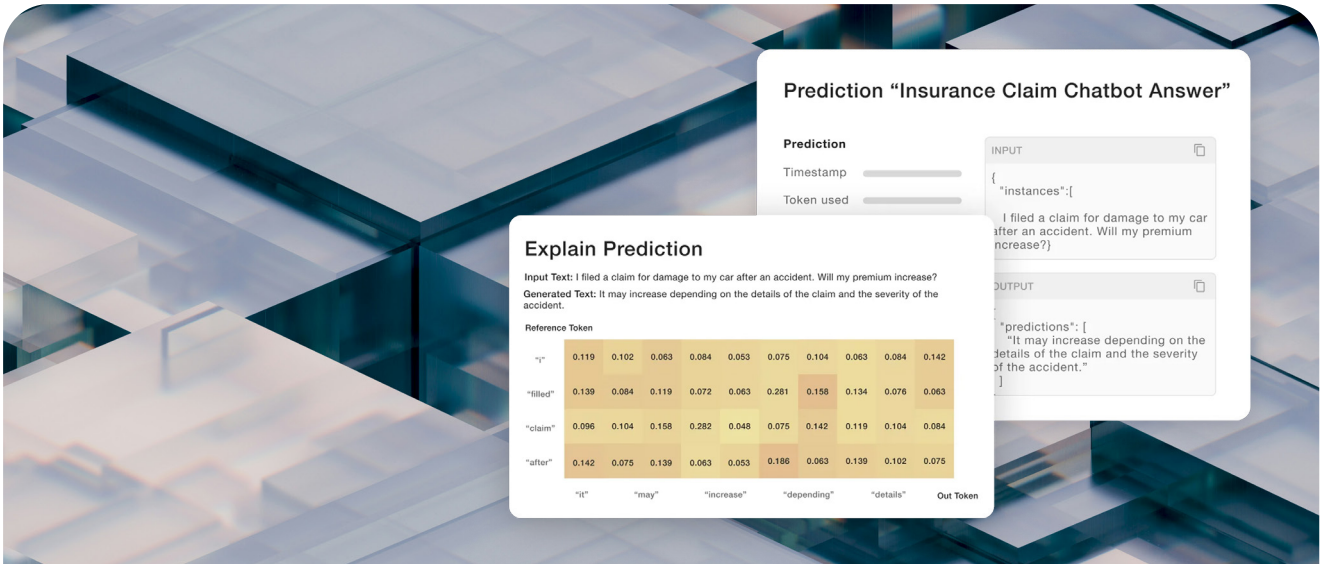
[AI Act | Art. 10 - Data and Data Governance]

Data governance controls ensure training datasets meet quality and representativeness requirements. Poor data governance is one of the most common sources of AI system failures and regulatory violations. Data governance failures are behind most high-profile AI scandals. These controls help you avoid becoming the next cautionary tale.

Data Governance	Evidence
DG1 - Process	<p>A documented process & templates for data governance:</p> <ul style="list-style-type: none"> • Data collection • Source & dataset characteristics • Copyright and ownership of data • (Pre-) Processing • Quality assessments • Lineage / tracking • Access logs with audit trails • Encryption verification (at rest and in transit) • Data integrity checks (cryptographic hashing)
DG2 - Documentation	<p>Documentation for each use case based on DG1. This can be done in data cards or templates.</p>
DG3 - Bias Detection	<p>Systematic (unwanted) bias testing across different demographic groups, geographic regions, and use case scenarios. Implement mitigation strategies including data augmentation, algorithmic fairness techniques, and ongoing monitoring.</p>

You'll know this is working when:

- Data quality issues are caught during preparation, not discovered months later when model performance degrades.
- New team members can understand your datasets without asking the original data scientist.
- Bias testing is automatic, not something you remember to do before deployment.



Transparency

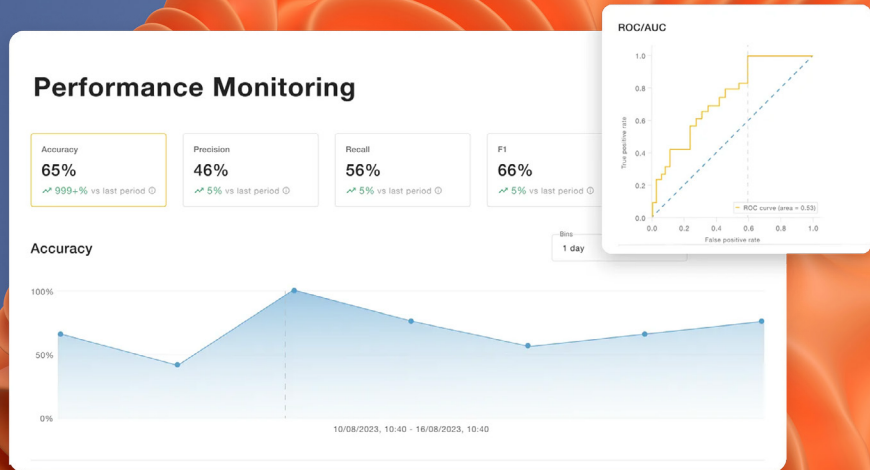
[AI Act | Art. 11 - Technical Documentation & Art. 13 - Transparency]

Transparency controls ensure stakeholders understand AI system capabilities, limitations, and decision-making processes. These controls are essential for building trust and enabling effective human oversight.

Risk Management	Evidence
TP1 - Capabilities and Limitations	Information about the capabilities and limitations of models in the use case, documented to assist relevant stakeholders' decision-making. This can be done in documentation, templates, or model cards .
TP2 - Explainability	Decisions and outputs of models in the use case must be explainable and interpretable for relevant stakeholders. This might include SHAP values, LIME explanations, natural language descriptions or other ways.
TP3 - Instructions	User instruction documents and training materials are accessible to relevant stakeholders, including setup instructions, operating procedures, troubleshooting guides, and safety warnings. Tailor instruction format and detail level to different user types.
TP4 - Impact Assessment	The potential impacts of the use case on individuals, groups, and society must be assessed, documented, and reviewed.

You'll know this is working when:

- Stakeholders understand AI system limitations without needing technical training.
- Different audiences get explanations they can actually use, and can overrule decisions and give detailed feedback on the outcomes based on the explanation, from technical details for developers to business impact for managers.
- Users know what to do when AI systems behave unexpectedly, without calling IT support.



Human Oversight

[AI Act | Art. 14 - Human Oversight]

Human oversight controls ensure qualified humans maintain meaningful control over AI systems, particularly for high-risk applications where AI decisions could significantly impact individuals.

Human Oversight	Evidence
HO1 - Operational Oversight	Natural persons must be able to effectively oversee (monitor) the use case while in use to minimize risks to health, safety, or fundamental rights. Appropriate monitoring is available.
HO2 - Intervention	Individuals responsible for human oversight must be able to intervene in the operations or override the outcomes of the models in the use case (feedback loop). Individuals affected by AI decisions must have clear mechanisms to challenge decisions and seek redress. This includes accessible complaint procedures, human review processes, and meaningful remedies when AI systems cause harm.
HO3 - Competence	Individuals responsible for human oversight must be equipped to effectively and reasonably perform their duties by actively giving feedback or overruling decisions.

You'll know this is working when:

- Humans catch AI problems before customers do, and know exactly how to respond.
- Human overrides are documented, learned from, and fed back into system improvements.



Operations

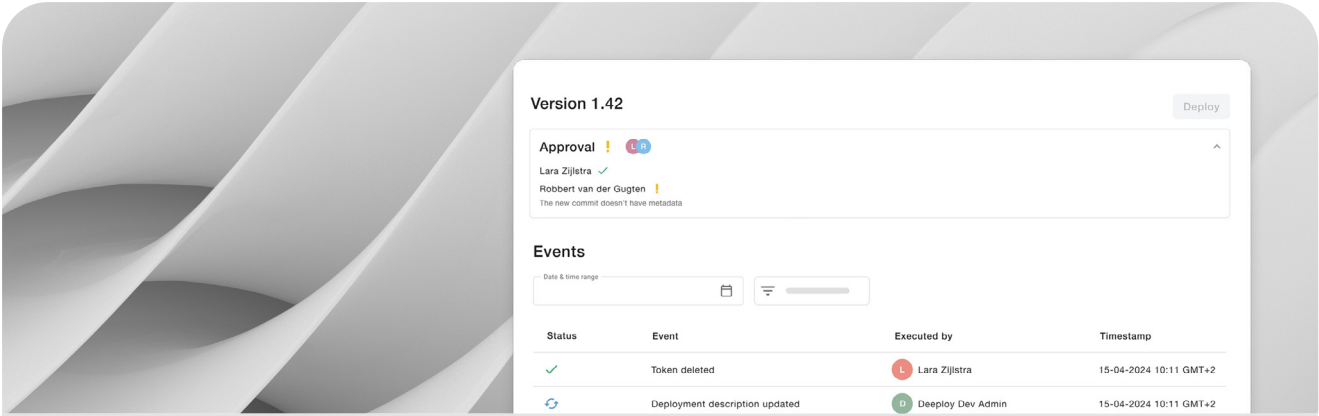
[AI Act | Art. 12 - Record-keeping & Art. 15 - Accuracy, Robustness, Cybersecurity]

Operational controls ensure AI systems maintain performance, security, and auditability throughout their operational lifecycle.

Operations	Evidence
OP1 - Event Logging	<p>Events have to be automatically recorded over the duration of the use case lifecycle:</p> <ul style="list-style-type: none">• System events• User interactions• Predictions & outcomes <p>Logs are tamper-proof and retained according to regulatory requirements.</p>
OP2 - Accuracy & Performance	<p>Use cases have appropriate metrics defined, with continuous monitoring, and set alert thresholds for performance degradation. Establish retraining procedures when accuracy falls below acceptable levels.</p>
OP3 - Robustness	<p>The use case must be resilient to errors, faults, model/data drift, and inconsistencies arising from within the system or its environment, especially during interactions with people or other systems. The system is stress tested on edge cases.</p>
OP4 - Security	<p>The use case must be protected against unauthorized third-party attempts to exploit vulnerabilities that could alter its use or performance. Address AI-specific security risks like adversarial attacks and data poisoning. This includes adversarial testing results (attacks, mitigations, test frequency), data poisoning prevention measures (provenance tracking, integrity checks), model theft protection (rate limiting logs, watermarking verification), prompt injection testing (for GenAI - red team results, input filtering) and penetration testing report (annually for high-risk systems).</p>

You'll know this is working when:

- You can trace any AI decision back to its inputs, processing, and reasoning.
- Performance problems trigger alerts before they impact business outcomes.
- AI systems fail gracefully (fallbacks) instead of producing garbage outputs that look plausible.
- Security assessments include AI-specific threats, not just traditional IT security.



Lifecycle Management

[AI Act | Art. 11 - Technical Documentation & Annex IV & VII]

Lifecycle controls ensure proper governance throughout AI system evolution, from initial development through updates and eventual retirement.

Lifecycle Management	Evidence
LC1 - Version Control	Version control for all models in the use case must be maintained, including records of changes. Retirement of AI systems should be part of the lifecycle, clearly offboarding systems.
LC2 - Sign-off	All model versions must be reviewed and approved by relevant stakeholders before deployment or updates: <ul style="list-style-type: none"> • Technical review • Business approval • Compliance verification
LC3 - Technical Documentation	Essential technical components for the ongoing operation of the use case must be defined, documented, and provided to the relevant stakeholders in an appropriate and accessible format: <ul style="list-style-type: none"> • System architecture • API specifications • Deployment procedures • Maintenance requirements

- You'll know this is working when:**
- You can roll back to any previous model version and understand exactly what changed between versions.
 - Model deployments have clear approval trails and nobody can deploy "quick fixes" without oversight.
 - New developers can maintain and modify AI systems without hunting down the original creators.

EU AI Act High-Risk Requirements



- ✓ Events are automatically logged and traceable
- ✗ The linked repository is private
- ✓ Feature value boundaries are documented
- ✗

Conformity & CE marking

[AI Act | Art. 43 - Conformity Assessment, Art. 48 – CE marking & Art. 49 – EU Database]

The last step is to conduct an AI conformity assessment, resulting in CE marking for the AI product.

Control	Evidence						
CE1 - Conformity Assessment	<p>High-risk AI providers must conduct conformity assessment before market placement.</p> <p>INTERNAL ASSESSMENT (Most Systems)</p> <ul style="list-style-type: none"> • Complete technical documentation (all Chapter 4 controls) • Verify all requirements met (Art. 9-15) • Prepare Declaration of Conformity • Affix CE marking before market placement <p>THIRD-PARTY ASSESSMENT (Notified Body)</p> <p>Required for:</p> <ul style="list-style-type: none"> • Biometric identification/categorization systems • Critical infrastructure AI systems • Specific Annex III high-risk categories <p>Process:</p> <table border="0"> <tr> <td>1. Engage EU-notified body</td> <td>4. Receive conformity certificate</td> </tr> <tr> <td>2. Submit technical documentation</td> <td>5. Maintain ongoing compliance</td> </tr> <tr> <td>3. Undergo audit (if required)</td> <td></td> </tr> </table>	1. Engage EU-notified body	4. Receive conformity certificate	2. Submit technical documentation	5. Maintain ongoing compliance	3. Undergo audit (if required)	
1. Engage EU-notified body	4. Receive conformity certificate						
2. Submit technical documentation	5. Maintain ongoing compliance						
3. Undergo audit (if required)							
CE2 - EU Database	<p>All high-risk AI systems must be registered before market placement. Database opens on August 2, 2026. Key deadline: before market placement. Updates are required within 2 weeks of significant changes</p>						

You'll know this is working when:

All high-risk models are available in the EU database, and have a CE marking before go-to-market.

Conformity assessment is complex and time-consuming. Our partners can help.

Engage partners 6-9 months before intended market placement to ensure smooth conformity process and avoid delays. See last page for details.

AI Governance

Maturity Assessment

Use this self-assessment to determine your starting point and prioritize initial implementation efforts:

LEVEL 1 (Start Here)

Getting Started

Start here if you have:

- No central inventory of AI systems
- No designated AI governance owner or function
- Fewer than 5 AI systems in production
- No formal AI risk assessment process
- No documentation standards for AI systems

Priority Actions (Level 1):

- **Week 1-2:** Create AI system inventory (**RM1**)
- **Week 2-3:** Designate governance owner and assign roles (**GO2**)
- **Week 3-4:** Conduct preliminary risk classification (**RM3**)
- **Week 4-6:** Draft AI Use Policy (**GO1**)

LEVEL 2

Building Foundation

Start here if you have:

- Basic AI system inventory (may be incomplete)
- Designated owner but limited formal processes
- 5-15 AI systems in production
- Some documentation but inconsistent standards
- Risk awareness but no systematic assessment

Priority Actions (Level 2):

- Complete all "Essential Policies & Templates" (**GO1, RM2, RM4, incident response**)
- Implement data governance documentation (**DG1, DG2**)
- Establish human oversight procedures (**HO1, HO2**)
- Deploy basic monitoring and logging (**OP1, OP2**)
- Conduct AI literacy training (**GO3**)

LEVEL 3

Scaling Governance

Start here if you have:

- Formal governance structure with clear roles
- Multiple high-risk AI systems or applications
- 15+ AI systems in production
- Operating in heavily regulated industry (finance, healthcare, pharma)
- Existing risk management frameworks (e.g., ISO 31000)

Priority Actions (Level 3):

- Control Framework: Implement all 32 controls (**Chapter 4**)
- Conduct bias testing and fairness assessments (**DG3**)
- Deploy explainability tools (**TP2**)
- Perform conformity assessments (**CE1**)
- Register high-risk systems in EU database (**CE2**)
- Establish continuous improvement and audit cycles



CHAPTER 5

Conclusions and Next Steps

Roadmap for the future

FIVE YEARS AGO, AI was a graveyard of failed experiments. Today, it's a jungle of working systems that nobody fully understands. Tomorrow, it needs to be a managed ecosystem where innovation thrives within clear boundaries. This journey requires intentional action from leaders who understand that AI governance is about making it sustainable, trustworthy, and valuable over the long term.

What We've Learned

Organizations that implement effective AI governance move faster, build better products, and earn deeper trust. Research consistently shows that companies investing in governance grow 2x to 3x faster.

But one size never fits all. It's key to differentiate between different technologies, which require different controls. Predictive AI isn't generative AI, and open-source and closed-source systems can differ significantly. Different risk levels means your governance approach must be tailored to your specific context. A fraud detection model needs different controls than a customer service chatbot.

The organizations succeeding at AI governance start with minimum viable controls and evolve them based on experience. Perfect compliance frameworks that nobody uses are worse than simple policies that teams actually follow. Every stage of the AI lifecycle offers opportunities to catch problems before they become disasters. Having the governance in place will help accelerating go-to-market, while avoiding costly redesigns or mistakes.

Building Trust



Bart Schermer, founder at **Considerati** and Professor at the University of Leiden, emphasizes, *“The success of any*

AI application will ultimately depend on the trust users and other stakeholders can place in the application. By creating trustworthy applications, organizations can create a competitive advantage while at the same time complying with relevant regulations such as the GDPR and the AI Act.”

This perspective reflects a fundamental shift in how we think about AI governance. Rather than treating it as a “tick the box” compliance exercise, successful organizations approach governance as a process aimed at improving the safety, security, and trustworthiness of AI systems.



Iris van de Kleft from **Conclusion AI 360** reinforces this view: *“AI governance is more than policies on paper; it’s a practice. Through*

literacy, training, and a phased rollout, organisations can make governance part of how people work every day, and that’s where real value is created.”

Connecting the Dots: A Comprehensive Framework



Aura Orval from **Clever Republic** captures another essential insight: *“AI Governance is all about **Connecting the Dots**. By aligning metadata, business context, processes, policies, and responsibilities, we ensure that AI delivers its maximum value while staying transparent and accountable.”*

This holistic approach recognizes that mitigating risks, complying with legal requirements, and protecting privacy aren’t separate activities. They’re practices that reinforce each other. AI Governance becomes the watchtower at the center of these dots, ensuring reliable AI is built and monitored throughout its entire lifecycle.

We provided a comprehensive control framework, in which the dots are connected for you, with seven control categories:

- 1. Governance Operations** - Foundation controls establishing organizational policies, roles, and responsibilities before any AI development begins

- 2. Risk Management** - Systematic processes to identify, assess, classify, and manage risks throughout the AI lifecycle, including maintaining a complete AI system registry

- 3. Data Governance** - Controls ensuring training datasets meet quality and representativeness requirements, including bias detection and documentation processes

- 4. Transparency** - Making AI systems understandable through capabilities documentation, explainability features, user instructions, and impact assessments

- 5. Human Oversight** - Ensuring qualified humans maintain meaningful control, with operational monitoring, intervention capabilities, and proper competence requirements

- 6. Operations** - Maintaining performance, security, and auditability through event logging, accuracy monitoring, robustness testing, and security measures

- 7. Lifecycle Management** - Proper governance throughout system evolution via version control, approval processes, and technical documentation



Seppe Housen from **Datashift** advocates for a proportional implementation approach: *“Start with your most important use cases, assess and manage risks for these specific applications, then learn from the process and expand systematically rather than trying to govern everything at once.”* This recognizes that different organizational roles require different deliverables from AI risk management, ensuring governance efforts align with actual business priorities and stakeholder needs.

Trust requires more than just good intentions. It demands verification, as stipulated in the AI Act.



Pepijn van der Laan from **Nemko** represents the growing movement toward third-party validation through [trust marks and certification programs](#) that help organizations demonstrate their commitment to responsible AI practices. Pepijn: *“Scaling AI without control is no longer an option. Regulation will require verification, but the market already expects it. As AI becomes increasingly agentic, you can only earn trust by proving your governance is real. The organizations that will lead this next wave are the ones who can demonstrate*

their AI is not just powerful, but reliably steered.”

Real-World Proof

The framework outlined in this document isn't theoretical. **Henning von Hauen** from **Carve Consulting** has been implementing these principles with organizations like **Novo Nordisk** in Denmark, demonstrating how the Deeploy approach translates across different industries and regulatory environments. *“Taking governance into technical controls with Deeploy, and integrating AI models with real-time risk management, compliance, and explainability is groundbreaking and stimulates AI innovation while building trust.”* Leading Danish and international organizations are already implementing this approach as a strategic initiative, demonstrating that the platform offers a powerful means to achieve accountability and transparency at the core of AI models while minimizing legal and reputational risks.

These partnerships prove that AI governance frameworks can work across different cultures, languages, and business contexts while maintaining their core effectiveness. Organizations that master AI governance deploy AI faster because they have clear approval processes. They innovate more boldly because they understand their risk boundaries. They earn deeper customer trust because they can explain and justify their AI decisions. Most importantly, they sleep better at night knowing their AI systems are working as intended, within acceptable risk limits, and in ways that create genuine value for all stakeholders.

The companies that will lead the next decade of AI innovation won't be those with the most advanced algorithms or the largest datasets. They'll be those that can deploy AI responsibly, scale it sustainably, and govern it effectively while maintaining the trust of users, regulators, and society.

This framework provides the roadmap to become one of those companies.



Sander Ketelaar of **Full Orbit** puts it simply:

“AI governance starts with data governance. Too many organizations focus on models, tools, or compliance frameworks. But AI is only as reliable as the data foundation beneath it. Governance becomes real when metadata, lineage, version control, monitoring, and auditability are embedded into the

architecture itself — not documented in a policy deck. When governance is operationalized, compliance shifts from reactive to proactive. In the AI era, organizations that treat data as critical infrastructure — not as a by-product — will lead.”



Emine Bilsin of **DPO Consultancy** says

“AI governance is no longer optional. As AI becomes embedded in core processes, it must move beyond compliance toward clear accountability and human oversight, creating the conditions for effective and scalable AI innovation. At the same time, robust privacy safeguards in line with the GDPR remain essential to ensure responsible and sustainable AI deployment.”



CHAPTER 6

About Deeploy

Shaping AI for a better future

Deeploy is Europe's leading AI governance and MLOps platform, purpose-built for regulated industries navigating the EU AI Act. Founded in 2020 and backed by over €11M in funding, including EIC Accelerator support, Deeploy serves enterprise clients across healthcare, finance, and pharmaceuticals.

Deeploy combines deep regulatory expertise with practical platform capabilities. Our platform is used by leading organizations to:

- 1. Discover & Register AI Systems:**
Automated inventory and risk classification
- 2. Implement Controls:**
Pre-built templates for all governance controls
- 3. Document AI Use Cases Efficiently:**
Using our platform to document and automate
- 4. Monitor Continuously:** Real-time performance tracking through built-in integrations
- 5. Enable Audits:** Complete audit trails and compliance dashboards

With the August 2, 2026 high-risk AI deadline approaching, organizations face a choice: build governance infrastructure from scratch, or leverage a platform designed specifically for EU AI Act compliance. Deeploy's Control Framework, developed in partnership with leading European consultancies, provides the foundation hundreds of organizations are using to scale AI governance without scaling headcount.

The regulatory window is closing. Organizations starting their compliance journey today have 18 months to implement controls, conduct assessments, and prepare for enforcement. The question isn't whether you'll need robust AI governance—it's whether you'll build it in time.



Ready to transform your AI jungle into a managed ecosystem?

Or become a new partner of Deeploy? The tools, frameworks, and partnerships you need are available today. The competitive advantage goes to those who act first.

Get in touch:

 deeploy.ml

 hello@deeploy.ml

 [Deeploy](#)

Need support? Contact one of our trusted partners:

Deloitte

Sebastiaan Berendsen

sberendsen@deloitte.nl

Nemko

Bas Overtoom

bas.overtoom@nemko.com

Datashift

Seppe Housen

seppe.housen@datashift.eu

Carve

Henning von Hauen

hvh@carve.dk

Clever Republic

Koen Balm

koenbalm@cleverrepublic.com

Conclusion AI 360

Friso Spinhoven

friso.spinhoven@conclusion.nl

Bearingpoint

Joris Schut

joris.schut@bearingpoint.com

Considerati

Ton Wagemans

wagemans@considerati.com

Hogeschool Utrecht

Stefan Leijnen

stefan.leijnen@hu.nl

Hogeschool Utrecht

Anita Boerema

anita.boerema@hu.nl

Zanders

Sebastian Marban

s.marban@zandersgroup.com

DPO Consultancy

Emine Bilsin

e.bilsin@dpoconsultancy.nl

Digital Power

Joachim van Biemen

joachim.vanbiemen@digital-power.com

Oudegracht 91A,
3511 AD Utrecht, The Netherlands

Photos by Shutterstock, Freepik.

