Pritect

# Managing AI Risks: Inventory, Assessments and Evaluations

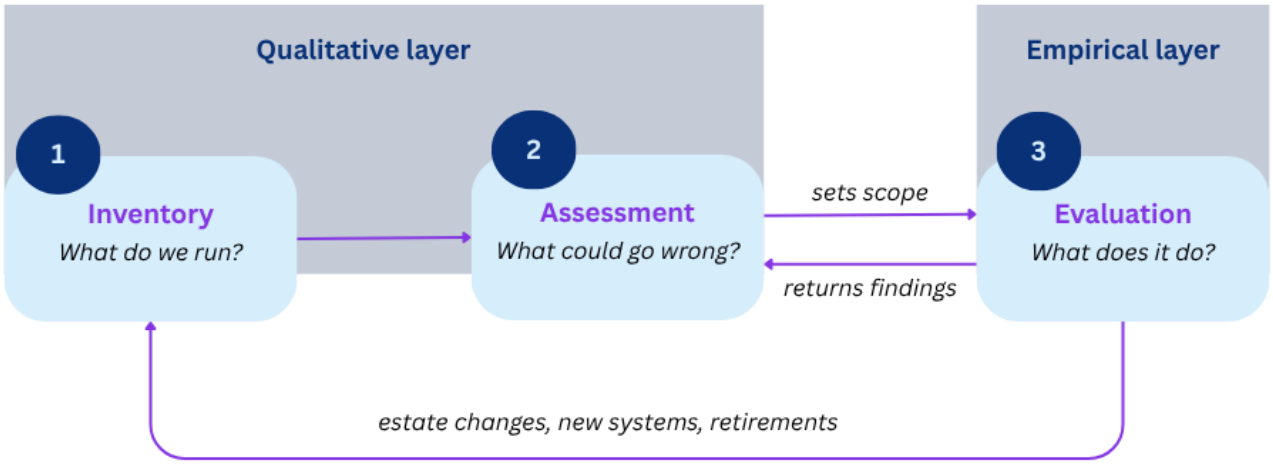
August 2026

# Executive Summary

Three pillars make an AI governance programme defensible, and the relationship between them matters as much as each one alone. Most organisations have some version of the first, a partial version of the second, and nothing resembling the third.

- The **AI inventory** is the register of every AI system the organisation builds, buys or uses, including the shadow estate arriving through everyday SaaS subscriptions. It is rarely a legal obligation, but it is the index everything else attaches to. A system that is not on the inventory cannot be assessed, evaluated or controlled.
- **AI assessments** answer the qualitative question: what is this system for, who could it harm, and what controls make that acceptable? They are increasingly required by law and converged upon by the major standards. What they cannot do is tell you how the system actually behaves.
- **AI evaluations** answer the empirical question: run against curated inputs with known expected behaviours, does the system do what we said it would? This layer is less optional than it looks, and it is the one most often skipped.

The pillars are sequential to build and circular to run. Assessments set what is worth evaluating; evaluations return evidence that confirms or unsettles the assessment; and the inventory keeps moving underneath both.



## Introduction

AI governance has a discovery problem before it has a control problem. Most organisations cannot list, with confidence, the AI systems deployed across their IT environments, the suppliers behind them, or the data those systems process.

Until that baseline exists, every downstream control, every policy and every assurance statement is partial. But establishing the baseline is only the beginning. Two questions follow.

The first is qualitative: *does this system, as designed and as deployed, do what we think it does, and what could go wrong with it?*

The second is empirical: *when we run it on inputs that resemble the real world, does it behave the way we said it would?*

The two questions sound similar, but they are not. Answering them well requires two different disciplines that AI governance programmes routinely conflate or skip entirely.

This piece sets out the three core pillars of a defensible AI governance programme:

1. **AI inventory**,
2. **AI assessments** and
3. **AI evaluations**.

It looks at the role each one plays and at the relationship between them.

## Pillar 1: AI Inventory

**Before an organisation can govern its AI, it needs to know what it has, who supplied it, and what data it touches, and keep that register current as the estate changes.**

An AI inventory is the foundation. It is the register of every AI system the organisation builds, buys or uses, together with the suppliers, models and datasets behind them.

Unlike other well-known registers, such as the Records of Processing Activities (RoPA) required by many data protection laws (Art. 30 GDPR; Art. 31 Saudi PDPL; s. 28 ADGM Data Protection Regulations 2021; Art. 37 LGPD), the AI inventory is generally not a statutory obligation.

And while we can find some regimes that require organisations to build and publish their AI inventories (e.g., Executive Order 13960, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, and Office of Management and Budget Memo M-25-21, Accelerating Federal Use of AI through Innovation, Governance, and Public Trust), building the AI inventory is in general regarded as a governance and risk management measure (see for instance NIST AI RMF: Govern 1.6).

A noticeable exception is DIFC Data Protection Regulations, whose Regulation 10.2.2(g) requires deployers and operators of autonomous and semi-autonomous systems processing personal data to provide, on request, a register listing the use cases in which those systems are used, the necessity and proportionality of the processing, whether the system will be used solely to make automated decisions, and which third parties process the personal data involved and where they are located.

A useful inventory captures more than names. It records what each system does, who owns it, its lifecycle stage, the data it processes, the suppliers involved and the underlying architecture. It distinguishes systems built in house or integrated through APIs from those embedded in third party tools. The honest version of any organisation's AI footprint is larger than the sanctioned version, so the inventory must also capture the shadow estate: the long tail of AI that creeps in through everyday SaaS subscriptions and through employees using tools of their own choosing.

Two further properties matter:

1. The inventory is a living artefact. New AI features are rolled into existing software every week, and product teams ship and modify AI systems faster than ever. Discovery is therefore a continuous exercise, not a one off one.
2. The inventory is the index to which every other governance activity attaches. If a system is not on the inventory, it cannot be assessed, evaluated, monitored or controlled. The inventory is the spine of the AI governance programme.

The common failure mode is to treat it as a static compliance artefact, a register populated once to satisfy the board or another stakeholder and then left to decay. That gives false confidence that the organisation knows what it has, when in fact it knows only what it had some time ago.

## Pillar 2: AI Assessments

**The AI inventory tells you what exists. AI assessments enable you to know what the risks are and what mitigations should be implemented to reduce risks to acceptable levels.**

Assessments are the qualitative, structured judgement layer of an AI governance programme. They are typically questionnaires, and they aim to provide a holistic and dynamic view of a given AI system: its purpose, its data flows, its decision logic, the populations it affects, the legal bases it relies on, the harms it could cause, and what controls would need to be implemented, and by whom.

Some assessments are required by law. A Data Protection Impact Assessment (DPIA) is mandatory where processing is likely to result in a high risk to individuals. The conditions are set out in Art. 35 GDPR and its equivalents in the UK, Switzerland, and other jurisdictions that have modelled their data protection frameworks on the same structure. The California Consumer Privacy Act (as amended by the CPRA) now requires businesses to conduct cybersecurity audits and privacy risk assessments in connection with certain uses of automated decision-making technology and sensitive personal information, following CPPA regulations approved by the Office of Administrative Law on 22 September 2025 and effective from 1 January 2026, with deadlines phased through 2030. Notably, the obligation reaches backwards: processing that began before the regulations took effect and continues after it must also be assessed, by 31 December 2027. The duty is not limited to what an organisation launches next; it extends to what it is already running.

Colorado offers a cautionary counterexample. SB 24-205 would have required developers and deployers of high-risk AI systems to conduct impact assessments and maintain risk management programmes. It never took effect: the state repealed and replaced it in May 2026 with SB 26-189, effective 1 January 2027, which drops the impact assessment mandate, the risk management programme and the duty of care in favour of notice, adverse outcome disclosure and a right to meaningful human review. The assessment work does not disappear, though. Developers must supply deployers with formal documentation about their systems, and the law voids any clause attempting to shift liability for a party's own discriminatory use of automated decision-making technology onto another party. Neither party can identify what falls in scope without first inventorying the tools that materially influence consequential decisions.

Within the EU, the AI Act requires providers of high-risk AI systems to maintain a formal risk management system under Art. 9, and requires providers' quality management systems to cover that risk management system under Art. 17(1)(g). These provisions establish a continuous, documented, and auditable risk management obligation that a well-structured AI assessment process is designed to satisfy.

Beyond statutory obligations, leading standards converge on the same requirement. ISO/IEC 42001:2023 at clause 6.1 requires organisations to determine the risks and opportunities that need to be addressed in their AI management system and to plan actions to address them. The NIST AI Risk Management Framework calls for AI risks and benefits to be mapped across the lifecycle, with the basis for risk prioritisation documented. The European standard EN 18286:2026, at clause 8.1, sets out requirements for AI risk assessment as part of an AI management system. These instruments make the point that structured AI risk assessment is, increasingly, a legal and standards-based requirement whose absence will be difficult to defend to a regulator or an auditor.

Assessments cover the territory where automation can assist but not replace, because the questions require human oversight and considered judgement. *Is the use case proportionate to the harm it could cause? Is the population affected adequately represented in the training and evaluation data? Is there a meaningful human in the loop, and does that human have the information, time and authority to override the system?* These are judgement calls. They need to be documented as such, by people accountable for the answer, with evidence that supports potential audits.

A good assessment programme, or risk management programme more broadly, also has a review cadence. Systems and other architectural details change. Suppliers change models under the same product name. The regulatory environment evolves. A defensible assessment can stop reflecting reality within months, hence the need to reassess on triggers and on schedule.

However, questionnaire type assessments have limitations. What assessments do not do, and cannot do, is tell you how the system is actually behaving in development or production. They tell you what the organisation thinks the system should do, and what it has done or committed to do in order to keep risk inside tolerance. For the empirical question, you need the third pillar: AI evaluations.

## Pillar 3: AI Evaluations

**Assessments tell you what the organisation believes about a system. Evaluations tell you what the system actually does when you run it.**

Evaluations are empirical. They run a system, model or agent against a golden dataset, a curated set of inputs with known expected behaviours or properties, and score the outputs. Scoring is often done by a second model acting as a judge (LLM-as-a-judge), sometimes by deterministic checkers, sometimes by both, and in the highest stakes cases by human reviewers or structured red teaming. The result is a quantitative picture of how the system performs across the dimensions that matter to the use case.

Those dimensions vary with the use case. For a customer-facing assistant, they typically include:

- factual accuracy against a domain-specific corpus (product information, for instance),
- role violation (whether the system stays within its defined role and boundaries),
- toxicity,
- resistance to prompt injection, and
- adherence to brand and policy constraints.

For a retrieval-augmented system, they include:

- correctness (whether an answer semantically matches a known-correct reference, with no factual errors), and
- hallucination rates against the retrieved context.

For an agent, they extend to:

- tool-selection accuracy,
- plan adherence and
- task completion.

Evaluations are not purely a matter of good practice. For instance, the EU AI Act requires that high-risk AI systems be tested to identify the most appropriate and targeted risk management measures, and that testing ensure the system performs consistently for its intended purpose, with testing performed at any time throughout the development process and, in any event, before the system is placed on the market or put into service, against prior defined metrics and probabilistic thresholds appropriate to the intended purpose. That last phrase is the statutory description of an evaluation harness. Art. 17(1)(d) then requires the examination, test and validation procedures, and the frequency with which they are carried out, to be documented within the provider's quality management system. The NIST AI Risk Management Framework makes the same move architecturally, separating Map from Measure and devoting the latter to test, evaluation, verification and validation.

Evaluations run repeatedly: on every model change, and on a schedule even when nothing has changed, because the model behind a product name may change without notice. They produce signals rather than verdicts: toxic output where none was expected, a rise in hallucination rate, a tool-use error rate above tolerance. These signals feed back into the assessment layer, where someone accountable decides what they mean and what to do about them.

Evaluations should not be interpreted in isolation. They should sit within a wider AI risk assessment, where the findings are translated into risks, and mitigations are proposed to mitigate them. In a sense, they should be treated as evidence that the belief recorded in the assessment is either supported or not.

## The Role of Assessments and Evaluations in AI Governance Programmes

The most common mistake in AI governance programmes is treating assessments and evaluations as alternatives. The more common one is not considering them at all, particularly evaluations. The two are complementary, not substitutes. They sit at different layers, run on different cadences, and have different objectives.

Assessments are deliberative and produce documented judgements. They record what an organisation understands about a system, identify

the risks it presents, agree the controls that reduce those risks, and assign each control to a named owner. Evaluations are mechanical, repeatable and narrowly scoped to observed behaviour of a system, model or agent. They show how the organisation tests that behaviour against bias, toxicity, hallucination, task completion and whatever other metrics the use case demands.

The link between them runs in one direction, with a feedback loop. AI evaluations surface findings: a refusal failure on a category of sensitive prompts, a hallucination rate above tolerance on a particular query class, an uncompleted task in an agentic workflow. Each finding is a candidate risk. It may have already been captured in the relevant assessment, in which case the evaluation provides live evidence of the status of a known control and a basis for confirming, tightening, or relaxing the mitigation. Or it may be new, in which case it should be considered into the assessment, scored for likelihood and severity using the same scale the organisation already uses, and matched to a mitigation, either from the existing library or, where no clean fit exists, as a custom risk paired with a custom mitigation. Either way, the trail from automated finding to documented decision is the loop that makes the programme real.

Treated this way, the evaluation layer becomes a continuous early-warning system feeding the assessment layer, and the assessment layer remains the place where the organisation reasons about whether, on balance, it is comfortable running the system, in this form, for this purpose, with this set of mitigations.

## How Pritect Supports This

Pritect organises AI governance around these three pillars. The AI inventory captures systems, models, datasets, and the suppliers behind them, with a path for shadow AI discovered across the estate. Assessment templates cover the main system archetypes, including standard AI systems and agentic AI, each with its own risk library and mitigation catalogue, and room for custom risks and custom mitigations where the standard set does not fit. Evaluations run against golden datasets and produce findings with confidence levels, a recommended risk framing, and suggested likelihood and severity inputs that practitioners can carry into the relevant assessment.

The point is that inventory, assessment, and evaluation are three different jobs, and a programme that conflates them, or skips any of them, will struggle to give a credible answer the first time a regulator, a customer, or a board member asks one.

## Closing

AI risk management is knowing what you run, forming a defensible view of each system, and continuously testing that view against the system's actual behaviour.

The order matters: inventory before assessment, assessment setting what is worth evaluating, evaluation returning evidence that confirms or unsettles the view, and the whole loop feeding back into the inventory as systems change, arrive and are retired.

Get the three pillars right and the rest of the programme, the policies, the training, the supplier oversight, the board reporting, has something solid to attach to.

Pritect

# Contact

hello@Pritect.com
+45 71 74 74 54
www.pritect.ai

NORWAY

## Main Office Oslo

+47 4141 2168

Fjordalléen 16
0250 Oslo