A LANGHAM GUIDE

Governing AI in Production

A short, practical guide for leaders putting AI to work in the real business — the risks that appear once it's in production, and the questions worth answering before they do.

LANGHAM

2026 EDITION

01 — THE SHIFT

The risk doesn't begin when you adopt AI. It begins when AI starts doing the work.

Most companies are past the question of whether to use AI. It is already in inboxes, spreadsheets, support queues, and codebases. The line that matters is no longer adoption — it is the quiet moment a model stops *assisting* a person and starts *acting* inside a workflow. The draft that gets sent. The invoices it flags. The policy it retrieves for a representative to read aloud. The action it takes through a connected tool.

That shift — from experiment to production — is where governance starts to matter, and it is also where most organizations have the least visibility. An experiment has a human reading every output. A production system often does not. The very qualities that make AI useful — speed, autonomy, reach — are what make an ungoverned production system a liability.

This guide is deliberately short. It will not turn you into an AI risk function, and it is not meant to. It offers a way to think about production AI, the questions worth asking, and the patterns that tend to cause trouble — enough to judge, fairly quickly, whether a given system is in capable hands.

> An experiment has a human reading every output. A production system often does not.

Seven questions every production AI system should be able to answer

It isn't a policy — it's a test. If a system is live and no one can answer these crisply, that gap *is* the finding.

1 **Who owns it?**

A single named person accountable for what it does — not a team, not the vendor.

2 **What data can it reach?**

Everything it can read, store, or send — including the sensitive things no one mentioned.

3 **What can it actually do?**

The gap between drafting a reply and sending one is the gap between low risk and high.

4 **How was it tested?**

Not "did it demo well," but how it behaves on the hard, adversarial, and unhappy cases.

5 **Who approved it for production?**

A deliberate decision by someone with the authority to accept the risk on the company's behalf.

6 **How is it monitored?**

What is logged, what is watched, and who notices when its behavior drifts.

7 **What happens when it fails?**

Because it will. Is there a way to catch it, stop it, and undo what it did?

The value isn't the list. It's how confidently your organization can answer it for the AI already running.

03 — A SENSE OF PROPORTION

Govern AI by what it can do — not by what it is

The instinct is to govern AI by category: "we have an AI policy." But a meeting-notes summarizer and an agent that can move money are both "AI," and holding them to the same standard guarantees one of two failures — you smother the harmless one, or you under-govern the dangerous one.

The axis that matters is **capability and consequence**. Most systems fall into one of three bands:

Low	Internal and assistive, with a human reading every output. A document summarizer; drafting internal text. Light-touch.
Moderate	Touches sensitive data or shapes a real decision, with a person still in the loop. Drafting customer replies from CRM data; an internal knowledge assistant. Worth documenting and watching.
High	Makes or strongly shapes consequential decisions, acts on its own, or speaks to customers in sensitive contexts. An autonomous outreach or transaction agent; AI in hiring, lending, or care. Earns real scrutiny, before and after launch.

The principle is simply: **more capability and consequence, more structure; less, less.** A surprising amount of what gets called "AI governance" is just the cost of ignoring this — applying one heavy standard to everything, until people route around it.

THE SHAPE OF AN OPERATING MODEL

Govern — decide how you'll decide · **Map** — know every system you have · **Assess** — judge each one's risk · **Control** — put proportionate guardrails on it · **Operate** — launch through a gate, then watch and review

Five moves, run as a loop rather than a one-time project. The artifacts that make them real — a living inventory, a one-page record per system, a pre-launch gate, a review rhythm — are where the actual work lives.

Where production AI actually goes wrong

Across real deployments, the trouble clusters. Two kinds of system account for most of it: the ones that *act*, and the ones that *retrieve*.

Systems that act

AGENTS & AUTOMATIONS

Too much access

An agent handed broad credentials "to be useful" can do broad damage. The remedy is least privilege — narrow, explicit permissions — which almost nothing has by default.

Steered by untrusted input

Text in an email, a web page, or a document can carry instructions the agent quietly follows. If outside content can change what it does, that is a way in.

No brakes

No human gate on irreversible actions, no spend or rate limits, no way to switch it off. Speed stops being a feature and becomes the hazard.

Systems that retrieve

KNOWLEDGE & RAG

Permission leakage

The system surfaces a document the user was never allowed to see — because retrieval ignored the permissions the rest of the business carefully enforces.

Confident wrongness

A fluent, certain-sounding answer drawn from stale or contradictory sources — trusted precisely because it sounds sure.

No provenance

When you can't see where an answer came from, you can't catch it when it's wrong — and neither can the person relying on it.

None of these are exotic. They are the default state of a system built for a demo and shipped without governance.

05 — A TEN-MINUTE CHECK

Questions to ask about your own AI

Honest answers, for the systems already running in your business.

- ☐ Could you list every AI system in production right now — including the ones a team stood up without telling anyone?
- ☐ Does each one have a single, named owner?
- ☐ For each, do you know exactly what data it can reach and what actions it can take?
- ☐ Have the high-risk ones been tested against adversarial and failure cases — not just the happy path?
- ☐ Did someone with real authority approve the consequential ones for production?
- ☐ If one began behaving badly right now, would you know — and how quickly?
- ☐ Could you stop it, and undo what it had already done?
- ☐ Are the systems that touch customers, money, or sensitive data held to a higher bar than the harmless ones?

If a question is hard to answer, that isn't a failure — it's a starting point. Most companies are in exactly that position.

ABOUT LANGHAM

Langham is an AI strategy and build partner for growing companies. We help leadership teams put AI into production responsibly — setting the strategy, building the systems, and running them alongside the team. This guide reflects how we work: **practical, proportionate, and grounded in the established standards** for AI governance and security.

langham.ai

Informed by ISO/IEC 42001 · ISO/IEC 23894 & 42005 · NIST AI RMF · OWASP Top 10 for LLM & Agentic Applications · MITRE ATLAS · Five Eyes agentic-AI guidance