

# AI Control Architecture

Make AI's authority bounded, provable, and reversible — before you grant it, not after it fails.

## THE GAP

AI is being given deterministic authority — over data, decisions, and actions — faster than control can absorb it. A policy states intent; it cannot, by itself, bound what a specific system can see, decide, or do.

## THREE SURFACES OF AUTHORITY

- See** what data & systems it can access
- Decide** what decisions its output influences
- Do** what tools & actions it can run

Risk rises See → Decide → Do, as human oversight falls away.

## WHAT “UNDER CONTROL” MEANS

- We know what it can see, decide, and do.
- We have graded how strongly each boundary is held.
- A named human owns every consequential outcome.
- We can observe and reconstruct what it did.
- We can stop it — and we can recover.

## CONTROL SCALES WITH RISK

- T1** Fast-track — low-risk, public data
- T2** Standard — enterprise data or vendor AI
- T3** Enhanced — influences decisions & records
- T4** Action-capable — calls tools, triggers workflows
- T5** High-impact — autonomous, regulated, hard to reverse

## TEN PILLARS, IN TWO HALVES

07-13 bound what AI is and can do. 14-16 make it provable, observable, reversible.

## ONE ASSESSMENT, MANY OBLIGATIONS

- NIST AI RMF
- ISO 42001
- EU AI Act
- SR 11-7
- NYDFS 500
- OWASP

Open • CC BY 4.0. Start with one use case, end to end.