

HOW MUCH MAY AN AI AGENT DECIDE ON ITS OWN?

The control plane instructs, and can say no before the act

An agent that settles a claim, clears a payment, or renews a credit line acts for the firm, and the firm answers for every decision taken on its behalf. For people, the firm already governs that delegation, in policies, systems, and judgement. For agents, the same governance has to exist in a form the agent can be checked against at the moment it acts. Platforms already instruct their agents: identity, routing, guardrails, access. The part that can refuse an act is the mandate and the check.

The map, by altitude and essential

	BEFORE THE ACT: THE AGENT · ITS KNOWLEDGE · ITS MANDATE	AT THE ACT: THE CHECK	AFTER THE ACT: THE RECORD · THE LOOP
Enterprise	the agent portfolio · the knowledge estate · ceilings	the control standard	the evidence base · the investment decision
Workflow	the agent · the sources that apply · the mandate	the check before the act	the decision record · narrow or widen
Platform	harness and identity · index and retrieval · entitlements	policy enforcement	traces and cost · evaluations and drift

What the firm has to build

Most of the boxes on the map can be bought or already exist. Three joins make them a control plane, and they meet at the mandate:

1. a versioned mandate the check can query
 2. a record keyed to the mandate version and the build
 3. a return path from the record to the mandate

Questions for any agent that acts

1. Who signed this agent's mandate, and where is it written?
 2. Is that exact mandate checked before every act?
 3. For any decision, can you show which mandate, wording, and release applied?
 4. When outcomes go wrong, how does the mandate change, and who approves the change?

How a firm starts

With one workflow and half a day with a facilitator from the Hub: business, risk, and technology mark every box of the map for that workflow, from not in place to evidenced, keep their disagreements with a name against each, and leave with a first step, its owner, and its date. A first agent recommends before it decides; it runs under its own identity from the first day, and every proposal is recorded with its versions.

By day ninety, as a target one workflow within agreed limits · one consequential act checked, by the gate or by a person · a referral queue with an owner · one review rule tested from trigger to decision