

The problem

Agent libraries are becoming a commodity. Anyone can ship a hundred prompt-driven AI employees mapped to departments. What no library provides is governance: something that decides what those agents should do, checks every action against the company's actual strategy, records the reasoning, and controls what runs autonomously versus what requires a human. Regent is that layer. A regent governs in the sovereign's name: real authority, but borrowed, bounded, and revocable.

How it runs

Connect a business inbox and Slack, ratify a charter once, and the loop runs. Scheduled ingestion normalizes email and Slack into deduplicated, source-linked work items. An orchestrator routes them to desks: small, single-job agents instantiated from playbook files. An independent Auditor reviews every action on two counts, direction (does it advance the charter?) and conduct (does it comply with the rules?), and it can veto. Actions in human-delegated lanes execute unattended and are logged with the authority they ran under; everything else lands on a Docket as a drafted proposal with reasoning attached. The founder's attention is spent in exactly two places: the Docket (decisions) and the Minutes (review of everything done unattended). Everything else runs.

Core concepts

The Charter The ratified strategy: mission, ranked objectives, tradeoffs. Machine-checkable; every action cites it.	Precepts Rules of conduct binding every agent, plus a lane law mapping risk classes to default handling.
The Auditor Judges every action against charter and precepts. Never does the work; only checks it. Can veto.	Signets Scoped, evidence-backed, revocable grants of authority to act alone. Only humans issue them.
The Docket Everything awaiting a human decision, each item carrying the agent's proposal and reasoning.	Decision Ledger Append-only log of every action: rationale, citations, verdict, authority. Nothing happens silently.
The Minutes Session review of all work done under delegated authority since the founder last looked.	Shadow Mode Two-week pilot that drafts everything and sends nothing, scored against what humans actually sent.

Safety by structure

- Inbound content is untrusted, twice: an email or message can create work, never issue instructions, and its factual claims are verified against state before anything acts on them.
- Outbound is earned: the system starts read-only, drafting everything and sending nothing. Only human-delegated lanes send unattended, within scope and conditions.
- No self-expansion, and trust decays: Regent can recommend delegation citing ledger evidence, but only a human can grant it. Grants carry review dates, auto-suspend on trouble, and revoke in one action.

Where it stands

A working engine, not a concept. Live Gmail and Slack ingestion runs against real accounts with cursor catch-up and zero-duplicate reprocessing, behind a consent gate that refuses any source not explicitly ratified in configuration. The company is pure configuration, plain git-versioned files, so swapping one folder runs a different business. Verified by a 99-test unit suite and a 29-scenario scripted demo day that runs green from a frozen seed, including the showpiece: the Auditor vetoing a cross-client confidentiality breach and sealing its reasoning into the ledger. Next: send-on-approval execution, the Shadow Mode pilot, and the first evidence-backed grant of standing authority.