AI GOVERNANCE IN PRACTICE

How I run a production multi-agent AI system responsibly

Chris Hunter · Marketing Heroes · governance for AI that's live against real customers

AI governance isn't a policy binder nobody reads. It's the set of rules that decide what an agent is allowed to do, who signs off, and what happens when something goes wrong. Here's how I run it on a system that's live every day.

> **The principle: agents do the production, people stay accountable.** Nothing that reaches a customer or the public happens without a human in the loop or a guardrail we put there on purpose.

What's running, and who owns it

AGENT	WHAT IT DOES	AUTONOMY	HUMAN OWNER
Samantha	Answers inbound calls	Within an approved scope; limited on price and commitments	Client + daily review
Chase	Outbound lead follow-up (SDR)	Messaging within an approved scope	Client + daily review
Hope	Executive assistant (internal)	Drafts, triage, tasks; no public output	RMH / client team
Marketing Dept	Blogs, social, SEO, graphics	Drafts on its own; nothing publishes without approval	Content manager
PM-orchestrator	Coordinates agents, briefs humans	Internal coordination only	Ops manager

The controls that actually run

- **Human approval gate.** No public artifact ships on an agent's say-so. It's a hard requirement, not a preference.
- **Daily conversation review.** Every night an agent scrubs the two customer-facing agents' conversations and flags anything off. Catching drift daily, before a complaint, has improved these agents more than anything else I do.
- **Per-client data isolation.** One client's data never gets used to serve another.
- **Scoped knowledge.** Customer-facing agents answer only from client-approved facts. Outside that, they hand off or take a message — they don't guess.
- **Version control and backup.** All code and configuration is tracked in private git repositories, so there's a real audit trail.
- **Vendor-independence by design.** The marketing department has run on both Claude Code and Codex — no single provider can hold the business hostage.

When something breaks

The process is always the same: detect, surface it to the owner, fix it, verify on the next review. Two real ones. Samantha answered some homeowner calls wrong, so we added conversation rails and tightened what she's allowed to answer. Chase gave a prospect a hallucinated answer, so we seeded her with a client-approved FAQ set — accuracy resolved once she had the right facts to work from. The root cause there was missing knowledge, not the model.

Owner: Chris Hunter · Reviewed daily for customer-facing agents · Full framework available on request · chrismhunter98@gmail.com