

Q3 2026 · SESSION FOUR

You've Hired the Team. Now What?

Managing AI beyond the pilot: what a year of running it for real has taught us.

Rob Williams · Chief AI Officer · SoftHouse Advisory
PEO Leadership | Innovators Alliance · AI Peer Group · July 28, 2026

02

Rob Williams

Chief AI Officer

SoftHouse Advisory

- 30 years in technology, across many industries
- CPO / CTO / CAIO
- AI consultant in Insurance, Agriculture, VC, Government, Entertainment and SaaS
- Running a live, full-time multi-agent AI harness at home and in his business

NEWS

03

Four model releases worth knowing

What landed since the last session

- **Claude Fable 5:** Anthropic's most capable model, returned July 1 after a brief pause
- **Claude Opus 5:** near-Fable performance at half the cost, built for agents, debugging and enterprise workflows
- **GPT-5.6 (Sol, Terra, Luna):** OpenAI's new tiered family, generally available July 9, now ChatGPT's default
- **Gemini 3.5 Pro delayed:** Google found enterprise testers were consuming too many tokens in agentic tasks and pushed the release from June into July

NEWS

04

Costs fell 67%. Bills went up.

Evidence

The blended cost of an enterprise AI API call fell 67% year over year, from \$18.40 to \$6.07 per million tokens between Q1 2025 and Q1 2026. Inference now costs roughly what bandwidth did in 2003.

73%

of enterprises

exceeded their AI budget despite falling per-token costs

10x

annual decline

in inference cost per comparable capability tier, 2022 to 2026

4x

more likely

to see revenue growth if AI is embedded broadly vs. still piloting

Sources: Blended enterprise API cost Q1 2025 to Q1 2026, multiple inference-cost reports. BCG AI Radar 2026.

NEWS

05

The EU AI Act becomes enforceable in five days

August 2, 2026

- On August 2, the EU AI Act becomes fully enforceable. The world's first comprehensive, binding AI regulation. Penalties exceed GDPR: up to **EUR 35 million or 7% of global annual turnover**.
- Transparency obligations under Article 50 kick in on that date, as do enforcement powers over general-purpose AI models and the full fine regime.
- Multi-agent deployments in high-impact sectors (finance, insurance, HR, infrastructure) are classified as **high-risk**: mandatory human-in-the-loop oversight, immutable audit trails, and persistent identity management for every agent in the chain.

The Brussels Effect is real: even if you don't operate in Europe, your enterprise customers there do. The standard they are setting will travel.

Source: EU AI Act official enforcement timeline. European AI Office, 2026.

NEWS

06

The infrastructure for AI teams is here

Tools

- **Claude Code Dynamic Workflows:** hundreds of parallel subagents in a single session; codebase-scale migrations across hundreds of thousands of lines in one run
- **Claude Managed Agents:** a cloud service handling sandboxing, orchestration and governance so you can ship production agents without building the plumbing yourself
- **OpenClaw:** 382,000 GitHub stars, the most-starred software project in GitHub history, overtaking React's 10-year record in 60 days. Not a model; a general-purpose AI assistant layer on top of any model.
- **Microsoft Copilot Agents 365:** every M365 tool now has an agent governance layer. Not just AI in the product; AI reporting to an audit trail.

The tools have moved from giving you an agent to giving you a control plane for many agents.

TOPIC

07

You've Hired the Team. Now Who's Managing Them?

In May we talked about how to put your first agent to work. This session is about what happens next: when you go from one agent to many, and from a pilot to something you're accountable for.

DEMO

DEMO

What is a harness?

DEMO

DEMO

This is Home

My harness. Running full time, for over a year.

TOPIC

08

One agent is a tool. Five is a team.

- A single agent automates a task. Multiple agents start to coordinate, hand off, and make decisions that depend on each other. That is qualitatively different from what you managed before.
- What changes at scale: task routing decisions, conflict between agents reaching contradictory conclusions, and the cost of a silent failure multiplying across the whole chain.
- Gartner: inquiries about multi-agent systems rose **1,445%** between Q1 2024 and Q2 2025. The question is no longer whether to build one. It is how to run it once it is running.
- Tim Cook, Apple Q2 earnings call (April 30, 2026): Mac mini and Mac Studio "may take several months to reach supply/demand balance" and are "amazing platforms for AI and agentic tools." Apple's CEO is on the record that demand for this hardware is outstripping supply because people are building agentic AI on it.

KPMG 2026: 75% of large enterprises now name security, compliance and auditability as their primary requirement for agent deployment, ahead of capability.

Sources: Gartner inquiry growth Q1 2024 to Q2 2025. Tim Cook, Apple Q2 earnings call, April 30, 2026. KPMG 2026 Enterprise AI Survey.

TOPIC

The green light problem

- This morning my alarm did not go off. The system that wakes me reported itself perfectly healthy: scheduled, armed, lights configured, no issues. Every check was green. There was simply no check anywhere for whether a sound had actually been made.
- Nothing broke. Everything carried on working. That is exactly what makes it worth your attention: the failure was silent, and the only thing that eventually noticed was a human being wondering why it was quiet.
- This is the risk nobody warns you about: not capability failure, but **supervision failure**. Nothing was broken. It was that the check had been thought about once, assumed done, and lost from view. The agent tells you it is fine, you believe it, and there is no alert when it is not.
- At pilot scale you notice. At production scale, with five or ten agents running concurrently, you rely on the dashboard. And the dashboard is only as honest as the agent reporting into it.

The executives who are winning ask not just "is the agent running?" but "can I verify what it actually did?"

HOW TO FIX IT

- **Deterministic programmatic checks:** after each AI action, verify it independently. Confirm the file was actually written, the message was actually sent, the record was actually updated. Do not take the agent's word for it.
- **A smaller, gated AI checking the bigger one:** before the output propagates, a second agent verifies the first. One to do; one to confirm. The checker has no incentive to cover for the doer.

TOPIC

10

The 12% who achieved both

- BCG surveyed over 1,500 executives in January 2026. **56% of CEOs reported neither increased revenue nor reduced costs** from AI in the past twelve months. Not less-than-expected returns: zero measurable return on either dimension.
- A further group saw one or the other. **12% achieved both**, and that 12% is the most important number in this room.
- What they have in common: they embedded AI across decision-making and operations, not just distributed licences. They did not buy a tool; they **rewired a workflow**.
- Organizations with fully integrated AI are **four times** more likely to report AI-driven revenue growth than those still running pilots (58% vs 15%). The gap is widening, not narrowing.

The question is not whether you have AI. It is whether your AI is woven into how decisions get made, or sitting beside the workflow waiting to be asked.

Source: BCG AI Radar, January 2026. Survey of 1,500+ executives.

POLL

AI LIFEBLOOD IN BUSINESS

AI Teams

Have you moved from one AI tool to multiple AI tools working together?

Think: more than one AI assistant, agent, or automated workflow running at the same time on interconnected tasks.

The series so far: *transcription* → *markdown* → *connected data* → *AI teams*

DISCUSS

12

Over to the room

- Introduce yourself
- What does your AI team look like right now?
- Who in your organisation is actually checking that it is working?