

The Quiet Case for Not Building Your Own Agent Platform



You probably know the meeting. The board wants an AI agent strategy by the end of the quarter, a senior leader has read a consulting report, and you have been handed the job of building the platform. The slide says “AI-native,” the success criteria are vague, and someone offers a reassuring “we can just wrap it ourselves.” When you ask what finished looks like, the room goes quiet. That silence is the whole problem. The cost of building gets estimated long before anyone agrees on what is being built, and the work being dropped on internal teams is far larger than the people assigning it understand.

A pendulum that swung in a single year

This is not the first time a new category has tempted teams to build their own. Application servers in the late 1990s, content management systems in the 2000s, container orchestration in the 2010s — the pattern repeats. When something is new, the parts look simple, early adopters build, the market matures, and within about three years those in-house builders are quietly rebuilding on top of the commercial winner that emerged while they were busy.



What is different now is the speed. Menlo Ventures’ enterprise research shows the build-versus-buy ratio flipping in twelve months: in 2024 roughly 47 percent of enterprise AI solutions were built internally, but by the end of 2025 that share had fallen to about 24 percent. The market reached its verdict in a single year, which is unusually fast for a decision this expensive.

For more content like and follow me:

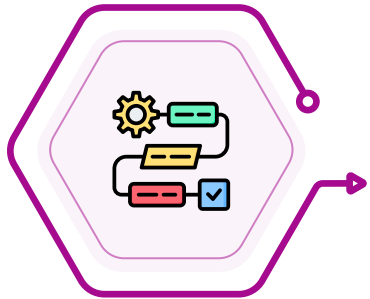


@bertblevins

INCGPT.COM

Most “agent platforms” are workflow systems

Many projects carrying the “agent platform” label are really workflow systems with a language model in the loop. The distinction matters. Workflows orchestrate models and tools along code paths you design in advance; agents decide for themselves how to proceed and which tools to use. Most enterprises today are shipping workflows, which is fine — workflows have bounded requirements, testable behavior, and predictable failure modes, and building one in-house can be reasonable.



The trap springs when a team built for workflows is asked to support real agents and finds the leap is not incremental. Agents need memory that survives across sessions, evaluation that handles nondeterminism, governance that tracks actions rather than outputs, and orchestration that recovers from failures a workflow engine never sees. These four things are not features you add to a workflow engine. Each is a separate product bet with its own maturity curve, its own vendor field, and its own specialists who have worked on nothing else for eighteen months.

The four components teams underestimate

Memory. It is tempting to treat memory as a database problem: pick a vector store, dump in conversation history, retrieve chunks. Production memory is three systems — episodic, semantic, and procedural — each with its own retention and retrieval rules, plus temporal reasoning about when a fact was true, deduplication, and multitenant isolation. The funding tells the story: Mem0 has raised around 24 million dollars, Letta roughly 10 million, and Zep runs as an independent company built on a temporal knowledge-graph engine. On temporal-query benchmarks, these tools differ by double-digit margins, which means they are not interchangeable.

01

Governance. This is not just role-based access control and audit logs. Agent governance has to authorize actions, not only data; reconstruct why an agent acted, not only what it did; and detect behavioral drift. Grant Thornton found that 78 percent of surveyed executives lacked strong confidence they could pass an independent AI governance audit within ninety days. There is a hard deadline attached, too: the EU AI Act becomes fully enforceable for high-risk systems in August 2026, covering credit scoring, hiring, healthcare, and critical infrastructure. If your platform cannot produce conformity assessments, human oversight, and complete audit trails, that is legal exposure, not a second-version feature.

02

Evaluation. Testing agents is not testing software. You evaluate a model's response to a prompt, but you evaluate an agent's entire trajectory — every tool call, state transition, and intermediate decision. Because behavior is nondeterministic, “did it succeed?” stops being a yes-or-no question. Google's Vertex AI has standardized trajectory-based metrics that did not exist eighteen months ago, and a wave of evaluation platforms has grown around them. Gartner projects adoption of AI evaluation and observability tooling rising from 18 percent of engineering teams in 2025 to 60 percent by 2028 — not a one-sprint problem.

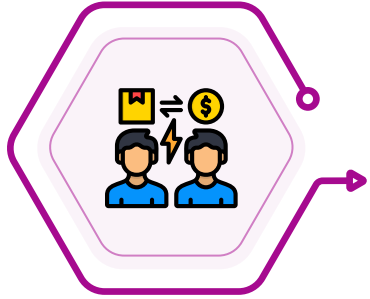
03

Orchestration. The orchestration layer has not converged. Competing frameworks make different bets on state, communication, and coordination, and none are drop-in replacements; switching between them means rewriting most of your agent logic. Underneath, the protocol layer is still being invented, and building on a moving protocol is a cost teams rarely price in. Build your own in 2024 and you are rewriting it now, while the teams that adopted a framework spent two years shipping.



When building is the right call

There are honest reasons to build. Proprietary data is a durable moat — decades of verified records with consistent identifiers is hard and expensive for outsiders to recreate, so if you own data like that, build on it. Regulated industries have real reasons to control the full stack when off-the-shelf tools do not map cleanly to HIPAA, SOX, PCI DSS, and similar frameworks. And lock-in at the AI layer compounds fast across memory, evaluation, and integrations at once.



But these are arguments for building agents on top of platform components, not for building the components themselves. You can own your data, domain logic, evaluation criteria, governance policies, and specific behaviors without owning the memory layer or the orchestration engine beneath them. Build what is specific to your business; buy what is specific to the technology category.

Five questions before you commit

01

Are you building an agent platform or a workflow system?

02

Can you state what “done” means for memory, governance, evaluation, and orchestration in three sentences each?

03

What happens when you need to swap the underlying model — a real risk, given that Anthropic’s share of enterprise model spend climbed from 12 percent in 2023 to 40 percent in 2025 while OpenAI fell from 50 to 27 percent?

04

What happens when the techniques change, as they already have, from flat retrieval to just-in-time context and trajectory-based evaluation?

05

And what happens when the small, mobile team that built your clever platform leaves?

Gartner expects more than 40 percent of agentic AI projects to be canceled by the end of 2027 — mostly internal builds, where the scope estimation error accumulates. The teams that bet on one vendor in 2023 spent 2025 porting; the teams that chose model-agnostic platforms spent it shipping. The only durable bet is the one that assumes the bet will change. The smartest platform decision you make this quarter might be the decision not to build the platform at all.

