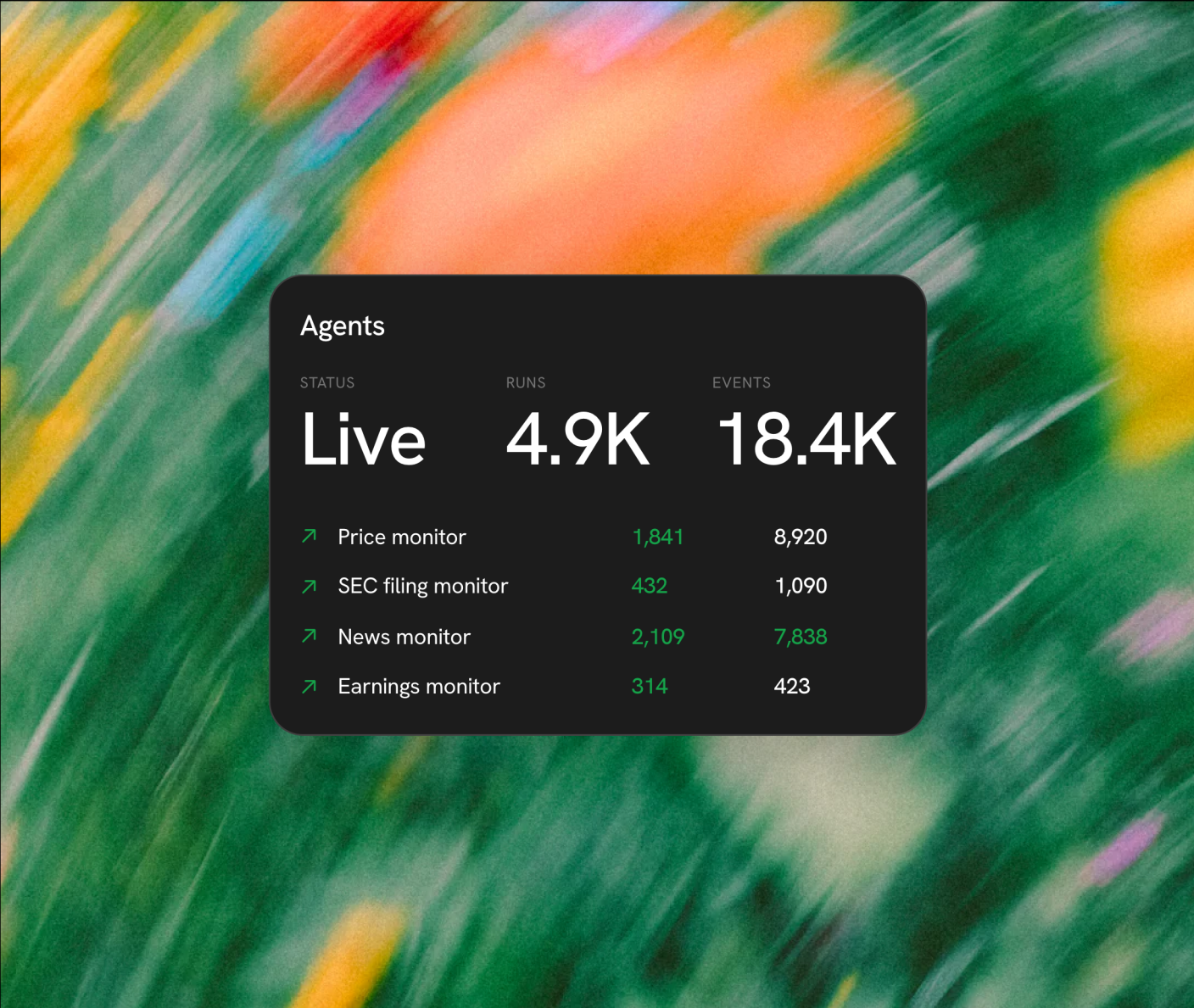




An Intro to Multi-Agent AI for Asset Managers

EXECUTIVE SUMMARY

Most funds already use AI; few can show a return on it. Partly because of undefined benchmarks and partly because they rely on one general model that doesn't have context for how the firm invests. A solution to the "one general model" problem is a committee of models that argue a name out: their level of disagreement signals conviction, running the debate across different models reduces any single model's blind spots, and the result is a cited view a fund can defend. Building that in-house is an ongoing cost most teams underestimate, so buying infrastructure that already exists is usually the better call.



Why widespread AI use hasn't produced widespread returns

At most funds, AI is already embedded in the daily work of research. It almost always operates in a single form: one general chatbot, reacting to a query.

There is no question that the industry will adopt AI. EY's 2025 survey of wealth and asset managers found that 95 percent had scaled generative AI across multiple use cases and 78 percent were exploring agentic systems.

Whether that use is paying off is harder to answer, partly because many firms have never landed on what "paying off" would mean. MIT's Project NANDA, in its July 2025 report *The GenAI Divide*, found that only about 5 percent of integrated AI pilots were generating rapid revenue impact, while the vast majority showed little or no measurable effect on profit and loss.

Gartner, from another direction, projected in June 2025 that more than 40 percent of agentic AI projects would be abandoned by the end of 2027, citing escalating costs, unclear business value, and inadequate risk controls. The operative word in both findings is *measurable*. A firm that never defines what a return would look like will never record one, so there is no score to work toward.

Poor scorekeeping is only part of it. Even where the benchmark is clear, the single-model setup most firms begin with has a ceiling. The MIT authors reach the same conclusion from the data, placing the failure in how firms absorb these tools: a general model serves individual people competently for drafting or first-pass analysis, yet stalls once inside a firm because it does not learn the way that firm works.

That limit is what a committee of models is built to clear, and it sets the two questions a research team now has to settle. The question used to be whether AI belonged in research at all. Now it is how far beyond a single chatbot a firm should go, and whether to build the resulting system in-house or buy one already built.

What a single model can and cannot do for research

Consider the off-the-shelf AI most analysts use. A single general-purpose model is useful, but it labors under four constraints that bear more heavily on investment research than on almost any other application.

01

It supplies only one reading. A position put to it returns a single synthesized view, with nothing to indicate what the model discounted or set aside, and no dissent anywhere in the exchange.

02

It inclines toward consensus rather than a house view. Asked to assess Microsoft, a general model gravitates to the market's median position, something near the conclusion that the stock looks cheap; to a fund whose thesis treats Microsoft as an AI-risk name, that answer is close to worthless, since the two readings imply entirely different trades and the model cannot know which lens the firm uses.

03

It changes under the people relying on it. Firms built on general tools report the same recurring complaint: a product changes without notice, and behavior a team had learned to trust is suddenly altered. In a process that depends on consistency, that instability is a cost in its own right.

04

It retains nothing about how the fund invests, holding neither the methodology nor the book nor the screening criteria, so each session resumes from close to zero.

These constraints confirm MIT's findings: a general model stalls inside a firm because each session is isolated and the model never adapts to the firm's workflow.

What "multi-agent" actually means, once the hype is stripped out

The term doing the most work in this discussion is also the most debased, since "agent" now can mean nearly anything. Gartner has a label for the practice, "agent washing," meaning the rebranding of ordinary chatbots and robotic process automation as agents, and estimates that only around 130 of the thousands of vendors advertising agentic capability are building the real thing.

Think of agents as individual workers, sharing the work among themselves, each with a single job. Consider the way a newsroom puts out a story: a reporter gathers the facts, a writer shapes them into a piece, an editor challenges the weak spots, and a copy desk checks everything before it runs. A multi-agent system works the same way. One model gathers the evidence, others build the case, argue against it, and review what the rest produced.

In investment management, this is a process all funds already run themselves. An investment committee stages a bull case against a bear case and charges a moderator with forcing the two to reconcile. Analysts and portfolio managers perform a compressed version of it whenever they weigh a name. A multi-agent system does the same thing, then runs it across different model families so the exchange is something other than one model in conversation with itself.

The most consequential feature is the disagreement. When a bull and bear case are set against one another, conviction can be inferred from how hard they were to reconcile. Rapid, clean agreement marks a strong case, while a protracted and contested exchange is itself a finding,

evidence that a name carries more uncertainty than any lone confident answer would concede. More controversy, lower conviction.

Distributing the work across model families carries a second advantage. Every model has characteristic blind spots, and compelling several independent ones to defend a position lowers a fund's exposure to the weakness of any single one. The claim is not that multi-model methods produce unbiased answers — nothing does — but that they reduce single-model bias.

What it really costs to build this in-house

Asset managers we speak to agree that the multi-model approach is a compelling one, and they understand why a fund would want a committee of arguing agents. And for many teams with capable engineers (or ambitious enough laypeople), the instinct is to build it. That instinct is where the MIT failure figures enter the picture. When these builds fall short, the model at the center is rarely the reason. The AI models themselves are now the easy part, much the same from one vendor to the next. What sinks a build is everything around the model: the data, the integrations, the controls, and that surrounding work keeps demanding attention long after the system goes live.

DATA	A research system is only as trustworthy as the data beneath it, and it requires verified, point-in-time inputs before any output merits action. A model fed the revised historical record rather than the version available on the date in question learns from a hindsight no investor actually had, and rewards a pattern that will not survive contact with live capital. The effort this demands is what teams most dependably underestimate.
COMPLIANCE	The audit trails, source citations, and governance controls an investment committee already expects have to be designed into the first prototype. Appended late, they seldom fit the system they are fastened to, and the seams tend to part at the least forgiving moment.
MAINTENANCE	Ongoing support falls hard even on the strongest engineering teams, because competence cannot avert it: the field advances faster than any roadmap. A system that takes a year to build is designed around what AI models could do at the start. Models improve so fast that those assumptions are stale by roughly the halfway point, so the finished system can already look behind on the day it launches. Keeping up means maintenance that never really stops.

None of this makes building the wrong decision. Funds with real data infrastructure, and the staff and

money to sustain it, have produced strong systems, and the control such ownership confers is worth a great deal to certain firms. For others, though, the true cost is almost never the figure from the kickoff slide, and it keeps accruing once the system is live. Among the funds that chose to build, a recognizable fatigue has set in: an acknowledgment that the engineering never quite reaches a state that can be left alone. A firm that has not started should price that recurring cost before it commits. A firm already underway faces a different question: not whether the system can be finished but whether finishing and maintaining it is the best use of a team that might instead spend its hours on the portfolio.

Build it, or buy it already built

That is the decision most research teams are now circling, and it comes down to two paths. A fund can build the committee in-house, clear-eyed about the data, the controls, and the maintenance the choice entails. Or it can license a version already built and tested, and reserve its own people for the one input no vendor can supply: investment judgment.

Boosted.ai has recently launched Alfa Prime, the culmination of eight years of creating artificial intelligence systems for asset managers. Allowing funds to put their expertise into an AI system is something we've had customers ask for since the beginning. Prime is an AI investment committee that adapts to a firm's own methodology, returning a citable memo with source-linked reasoning alongside a conviction score. It is modular, so a firm activates only the portion it is ready for: research and vetting at the outset for most, the fuller pipeline later, if at all.

Prime's financial-model output is directional. It frames a bull, base, and bear case and should not be read as a forecast of where a stock will trade. It gives an analyst a running start on diligence, and leaves the judgment to the analyst.

In our experience, most funds have bought rather than built. The logic is simple: a single model has a ceiling, a committee of models that argue clears part of it, and building that committee in-house is a serious and ongoing engineering job. For most firms, buying the version that already works is the better call, and it gets the team back to investing sooner. The quickest way to judge it is to turn the committee loose on a name the firm knows cold and weigh its reasoning against the desk's own.

Sources

EY, GenAI in Wealth & Asset Management Survey 2025.

MIT Project NANDA, *The GenAI Divide: State of AI in Business 2025* (July 2025).

Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027" (June 25, 2025).



www.boosted.ai