Autonomous Finance in Practice: Embedding AI into Core Finance Operations Without Breaking Control

Prithwijit Chaki
Global Finance Advisory Leader
Genpact

Eva Chalioti
Senior Lecturer in Economics (Economics of AI)
Yale University

Sri Ghattamaneni
Field Engineering Leader – Financial Services
Databricks

Vinod Shenoy
Executive Director of Conversational AI
Morgan Stanley

Peter Williams
Head of Banking Solutions Architecture – Global Financial Services
Amazon Web Services

The bottleneck to autonomous finance isn't AI capability — it's decades of data, governance, and organizational debt that agents now ruthlessly expose.

Enterprise leaders from Genpact, Amazon, Morgan Staley, and Databricks engage in a candid, execution-focused conversation on what it really takes to move from automation to autonomous finance — grounded in real workflows, real data challenges, and real governance requirements.

Content produced using generative AI

Key Takeaways

- Four independent vantage points — consulting, cloud, data platform, deploying institution — diagnose the same structural barrier to autonomous finance
- AI initiatives reliably surface a data infrastructure crisis within weeks, regardless of how they were originally scoped
- Governance bolted onto agents after deployment collapses at scale; controls must be codified into agent design from the start
- Most Fortune 500 finance functions remain between first-generation digitization and early AI adoption — nowhere near genuine autonomy
- The post-commoditization competitive moat is institutional data context and semantic architecture, not model access or compute

Introduction

The promise of autonomous finance is real. So is the gap between the boardroom narrative and the actual state of most finance functions. Sri Ghattamaneni, whose firm sits inside the data stacks of major financial institutions, offered the sharpest reframe of the prevailing narrative: "We over-index on the frontier models and agents when it comes to AI and autonomy. But the problem is actually a decades-old problem. You have to get all of your data in one place."

That diagnosis arrived independently from four speakers — a consulting leader advising on finance transformation, a cloud architect tracking deployment patterns across a major bank customer base, a data platform engineer fielding production support requests, and an executive deploying AI inside one of the world's largest wealth management operations. The convergence is the story.

Every AI Initiative Becomes a Data Initiative

"Everyone says AI is finally ready for finance," moderator Eva Chalioti pressed. "So why are most things still stuck at automation and not autonomy?"

The answer from every direction in the room had nothing to do with model capability. Peter Williams described the dominant pattern across Amazon Web Services' financial services customers: "What customers start on as an AI engagement or AI initiative very quickly becomes a data initiative. And once — that's actually one of the biggest challenges — you don't want to boil the ocean." The AI aspiration is functioning as a diagnostic, surfacing data fragmentation that organizations had either normalized or simply never measured. They launched an AI project and discovered a data problem they didn't know they had.

Prithwijit Chaki's maturity framework gives that pattern a structural home. Across Genpact's 25 years in finance operations, he maps five eras: manual-heavy processes, operational streamlining, the first wave of digitization through system-of-engagement vendors, first-generation autonomous finance via RPA and early GenAI pilots, and now the emerging agentic era. Most Fortune 500 finance functions, he argued, sit somewhere between digitally enhanced and data-driven with early AI — and the gap to genuine autonomy is structural, not technological. "Very few organizations have actually made the true pivot to genuine autonomous operations. It's not because AI or organizations are not capable, but because data operating model and the governance underneath has not yet really stepped up to be ready for AI."

The organizations making visible progress are not the ones with the most ambitious roadmaps. Ghattamaneni pointed to expected credit loss reporting as the use case generating the most traction across his firm's banking clients. ECL works precisely because the problem is bounded: data sources are known, the process is repetitive, the governance requirements are well-defined. "We've seen a lot of our financial services customers take those sort of processes, where you know where the data is coming from... put all of it in one place. And then they have been building AI analytics on top of it." Start narrow, govern that scope, prove the model, then extend. The organizations failing are generally the ones that skipped the scoping.

Vinod Shenoy reinforced the point from Morgan Stanley's deployment experience. The temptation, he noted, is to start with the answer rather than the problem — to see that a chat interface worked in one context and assume it works everywhere. "You started with the answer, not the problem. That's usually a recipe for failure pretty quickly." The discipline of defining what success looks like, measuring it through specific KPIs, and treating lessons as non-transferable across contexts is what separates pilots that scale from pilots that stall.

Governance Bolted On Breaks at Agent Speed

The governance problem looks manageable at pilot scale. It becomes structural at production throughput. Williams identified the inflection point: "Where an organization has many batch processes in place, it's then raising the issue of: my agent can make decisions thousands of times a day for me. Let's make sure it has the freshest information." An agent operating at that frequency against stale data or loosely defined controls doesn't just make bad decisions — it makes them faster and more often than any human auditor can track.

Chaki's response to that problem functions as a design principle, not just a compliance footnote. The failure mode is treating governance as a controls overlay applied after the agent is built. His alternative: "If you are doing an automated AP 3-way matching, and let's assume I have 26 validation checks and 26 control checks — those are baked into the agent as part of the agent activity log itself. So you actually have a very fail-safe, auditor-ready, compliant, non-hallucinating transaction getting posted."

Controls as architecture, not audit. The agent cannot operate unless the framework is met. Every output is logged for scrutiny before it ever reaches a human reviewer.

Shenoy's experience deploying production AI at Morgan Stanley illustrates why that design discipline matters at the institutional level. The deployment he described — RAG-based knowledge agents serving financial advisors in wealth management — succeeded because the team built trust asymmetrically: slowly on the way up, catastrophically easy to lose. "It takes a lot to build trust and it takes one small thing to completely erode that trust." Controlling for hallucination wasn't a model tuning problem; it was a system design requirement. Retrieval-augmented generation worked not because it produced more confident outputs but because it grounded responses in proprietary firm knowledge and made the sourcing auditable. The architecture served the governance requirement, not the other way around.

Ghattamaneni pointed to a related problem emerging across his customer base: the question of whether to trust an agent's output isn't answered by accuracy metrics alone. "The ability to measure the answers that you're getting and a way to self-improve your AI models and agents over a period of time has been probably the number one thing a lot of our customers have been asking us." Observability at scale — the capacity for humans to audit edge cases without reviewing every output — is rapidly becoming the category of foundational investment that determines whether a governance framework holds at production volume or collapses under it. Williams connected the measurement problem back to organizational design:

"It's not that the human is going to evaluate every response, but that they have the tools." What that looks like in practice is models and LLMs functioning as independent judges to assess outputs, with humans reviewing flagged edge cases and feeding learnings back into model evolution. The human role shifts from reviewer to architect of the review system.

The Finance Organization That Doesn't Exist Yet

The workforce question is the one most organizations are handling as a footnote. Chaki's framing deserves more weight than it typically gets in finance AI discussions. "Autonomous finance is going to be human-led but AI-operated or AI-run. Humans will still have the critical role on the two edges." His model is precise: strategic judgment and direction-setting at one edge, interpretation of anomalous signals and business partnering at the other. The operational core — transaction processing, controls compliance, routine reporting — becomes AI-run infrastructure.

That is not a reduction in human involvement. It is a wholesale redefinition of what the finance function does. The roles a post-agentic organization needs — people orchestrating bots, exception-case trainers feeding edge cases back to models, governance layer managers reviewing decision matrices, responsible AI monitors watching production behavior — don't exist at scale in most finance functions today. Williams reinforced the point from the AWS customer perspective: "What organizations are starting to realize is we're seeing a real shift... not to reduce the number of people, but actually recognizing that everyone can be 10 times as productive."

The productivity gain is real. So is the requirement to build entirely new role definitions before that gain can be captured.

Shenoy made the organizational readiness argument from the deployment side: "Making sure that your people are ready for the change that is coming with the automation. They are a key part of the transformation here, and a big part of this is making sure that not only are your tools ready, but also your people are ready to support that change." The firms that reach genuine autonomy first will not be the ones with the best models. They will be the ones that built the human infrastructure to operate an agent-driven finance function before they needed it.

Chaki extended that argument into the signal-processing capability that agentic finance makes possible. When sales rises but cash position doesn't move, a transaction-level risk monitoring agent can flag a potential fraud signal in real time rather than at period end. When procurement costs for a material start climbing, an agent with access to external data can link the movement to a geopolitical event and model the margin impact before it becomes a reportable loss. "This is actually the true power of autonomous finance, beyond operations, beyond transactions, to much better business insights." The shift from reporting what happened to recommending what to do next is where the organizational value concentrates — and it requires people who know how to act on those signals, not just read them.

The Moat Is Context, Not Compute

The panel's forward-looking convergence landed on a competitive dynamic that deserves executive attention before models commoditize further. Ghattamaneni was direct: "The companies or the institutions that are going to win are the ones that will have a common data platform and data where the agents can reason and get the context from. So ontology and semantics have been one of the most requested features from our customers." As frontier model performance converges, the differentiator shifts from which model an organization runs to how well that model understands the organization's own data, terminology, and business logic.

Shenoy translated that into operational terms. A well-maintained data dictionary — the kind experienced database administrators have built for decades — provides the semantic grounding that allows agents to reason accurately across business domains. "The power of having a data dictionary is so understated. It explains what's in your data, how to use it." The institutions that invested in data provenance, lineage, and definitional discipline for compliance reasons now hold an AI infrastructure advantage they didn't anticipate. Those that deferred it face a compounding problem: the more capable the agent, the more consequential the semantic gaps.

Ghattamaneni framed it as a pyramid. Business metrics defined by business teams sit at the base, giving agents precise understanding of what a question is actually asking. Semantics, ontology, and knowledge graphs layer on top. The firms that built that pyramid for regulatory reasons are structurally ahead. The firms building it now to serve AI are doing the same work under more pressure and with less time.

Chaki's closing line functions as both diagnostic and mandate: "Autonomous finance is as much a technology journey as it is a people and organization change. So treat that change management as a full-fledged work stream and not as a footnote in your plan." The organizations treating data unification, governance architecture, and role redefinition as preconditions are making measurable progress. The ones treating them as parallel workstreams — things to address while the AI pilots run — are building agents on foundations that will require expensive reconstruction the moment those agents reach production scale. The CFOs who move first on the infrastructure problem, not the model selection problem, are the ones who will still be ahead when the models stop being the differentiator.