Hitachi Vantara

White Paper

Architecting AI for Financial Services

Aligning Infrastructure and Strategy to Realize AI Value

© Hitachi Vantara LLC 2025. All Rights Reserved.

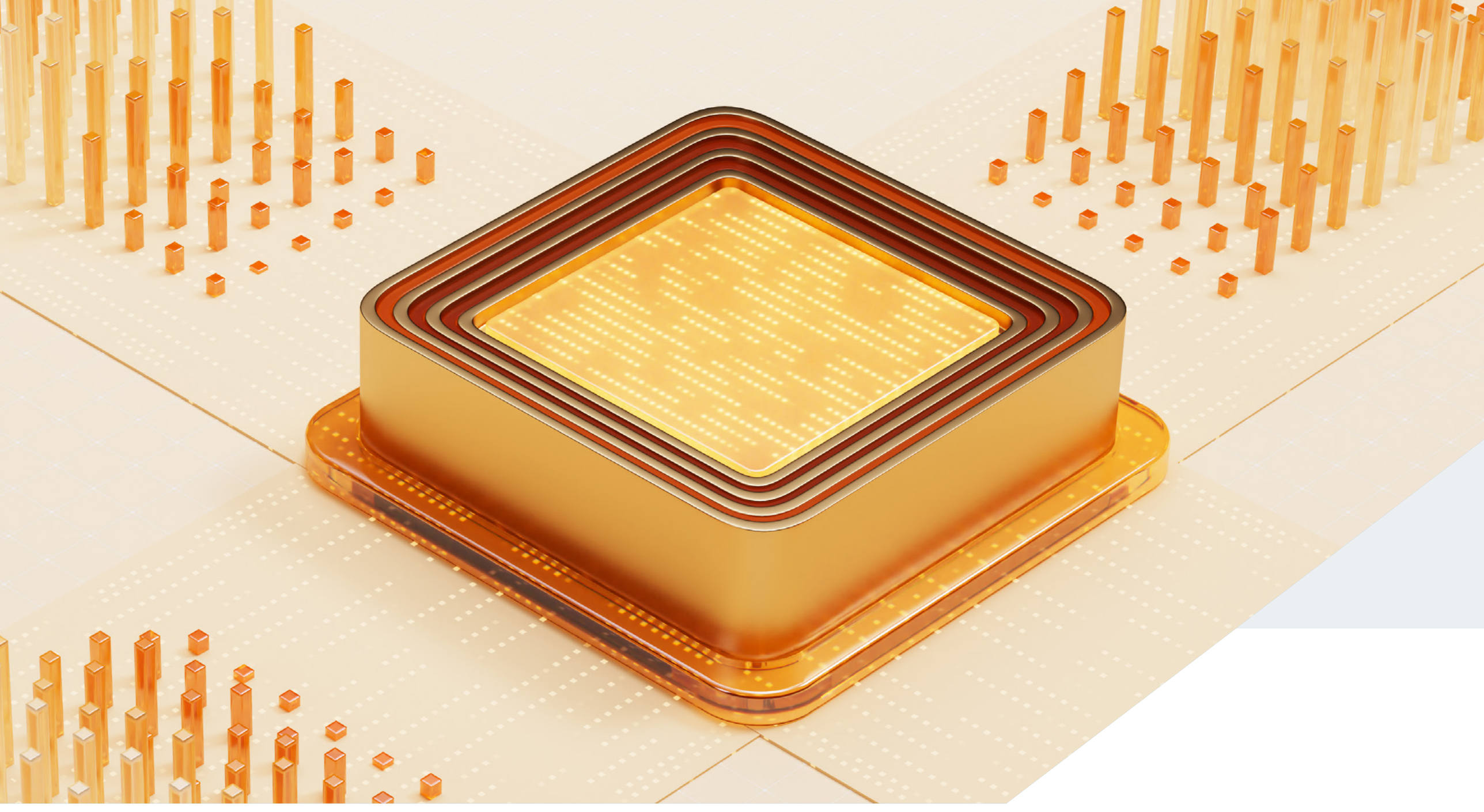


Table of Contents

Introduction: AI's Fork in the Road 3

1. **From AI Experimentation to Execution: The State of Play** 4
2. **Defining Business-Ready AI Infrastructure** 5
3. **Blueprinting to Accelerate Time to Value** 8
4. **Why Governance Is More Important Than Ever** 10
5. **Building Executive Alignment and External Partnerships** 12
6. **Use Cases That Deliver ROI—Fast** 14
7. **Practical Steps to Architect for AI** 16

Conclusion: From Vision to Value 18



Introduction:

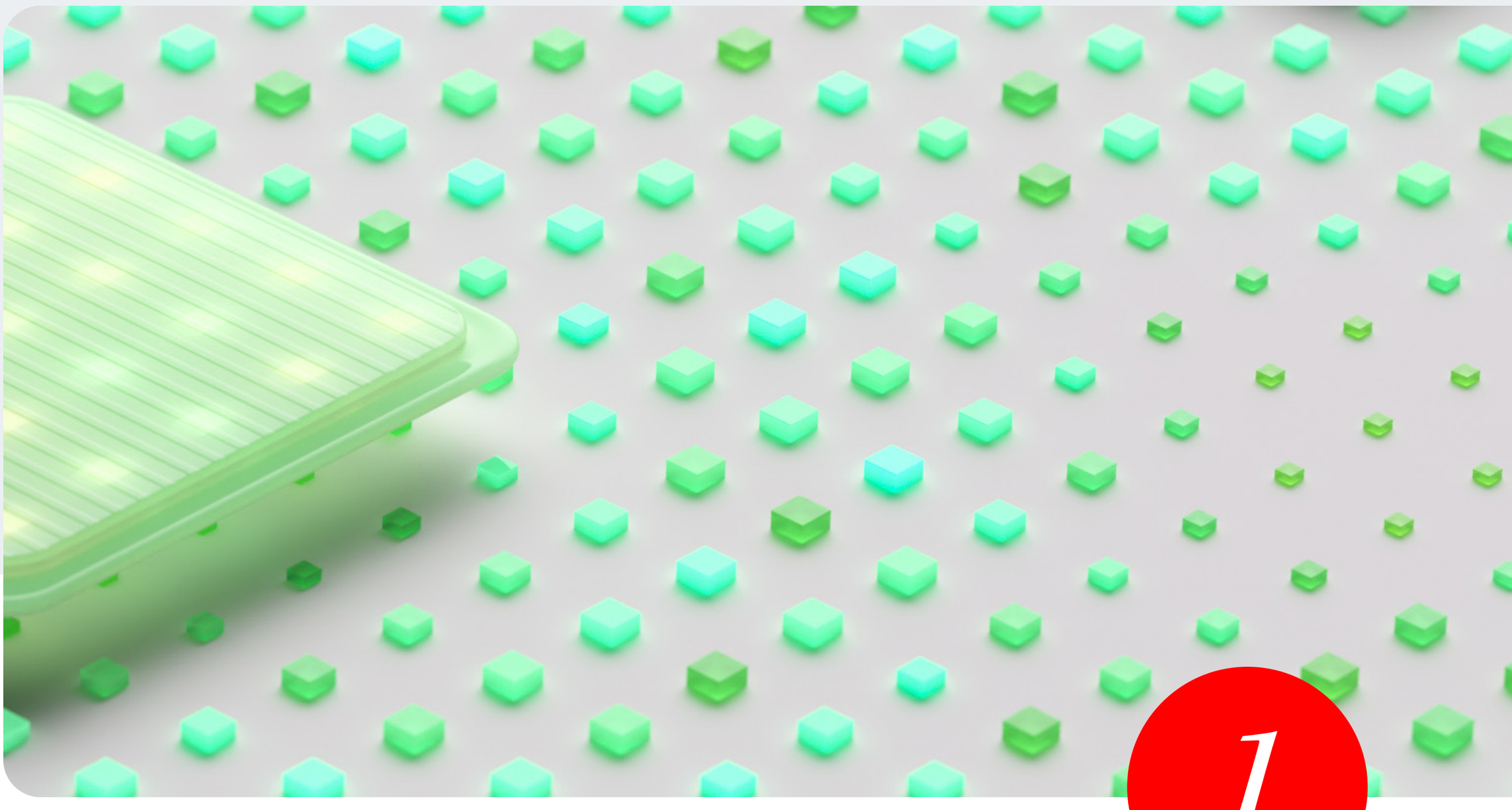
AI's Fork in the Road

The financial services industry is at a pivotal juncture in its AI evolution. While firms are increasing their AI IT budgets, many financial institutions are still struggling to translate ambition into value. According to BCG, 60% of banks haven't defined financial performance indicators to track impact. Even more telling, a global S&P Intelligence survey found that 42% of enterprises have scrapped most AI projects—a sharp rise from 17% the year prior.

As Mark Katz, CTO for Financial Services at Hitachi Vantara, puts it: "We're at an inflection point. The gold rush is over. Value has proved elusive, and a new risk landscape has emerged."

"The biggest challenge is biting off more than one can chew without a real plan," explains Jason Hardy, CTO of Artificial Intelligence at Hitachi Vantara. "Organizations jump into AI wanting to show value, but they waffle in science experiment land because they don't have a real strategy beyond 'we want AI.'"

This paper explores why many AI initiatives in financial services fall short of value and how institutions can overcome those challenges by aligning infrastructure, governance, leadership, and external expertise. It offers a practical roadmap for building scalable, business-ready AI systems that deliver measurable outcomes.



1

From AI Experimentation to Execution: The State of Play

*Financial services firms spent **$35 billion** on AI in 2023 and are projected to invest **$97 billion by 2027**, according to Statista Research Department.*

Much of this is driven by line-of-business (LOB) demand to embed intelligence into customer-facing and risk-reduction applications. But many of these initiatives have failed to move past the pilot phase. The reasons: limited alignment with business goals, lack of scalability, and infrastructure that could not support real-world demands.

Hardy identifies a critical organizational challenge: “AI projects often land on the CIO’s desk, but this is truly a cross-business collaboration effort. Without the right perspective and disciplines involved, organizations lose sight of what they’re really trying to accomplish.”

Despite these challenges, the ambition of an AI-supported industry hasn’t faded. AI remains a funded and strategic priority. What’s changed is the recognition that success will require a more disciplined, integrated approach that goes beyond box-buying and model tuning.



2

Defining Business-Ready AI Infrastructure

"We need to move AI out of the science experiment phase and into the production-ready phase. That means time to value must drop—and infrastructure must be able to deliver services with SLAs the business can depend on," says Katz.

AI initiatives in financial services demand more than raw compute. They require an infrastructure stack that is:

- **High-performance:** GPU-accelerated, low-latency, and capable of handling large data streams
- **Hybrid:** Balancing cloud with on-prem control
- **Secure and governed:** Capable of managing risks around data privacy, model hallucination, and regulatory exposure
- **Business-aligned:** Tuned to deliver specific use cases with measurable outcomes

Success isn't only about buying more hardware. It's about engineering systems that align with enterprise priorities in the BFS landscape: SLAs, resilience, explainability, and operational scale.

The Cloud Cost Reality

Hardy also flags a shift in cloud economics: “70% of AI workloads are repatriating right now, and cost is a big piece of that.”

Many institutions begin in the cloud for flexibility. But when scaled to production—especially for inference and training-intensive use cases—cloud costs can rapidly exceed forecasted ROI. As a result, BFSI organizations are increasingly moving mission-critical workloads back on-prem, especially for model training and sensitive data classification.

And this shift is not only about cost—it’s also about control and predictability. For inference-heavy workloads processing thousands of transactions per hour, on-prem infrastructure can provide the data residency and latency control that regulatory frameworks increasingly demand.

Scalable AI Infrastructure

Hardy identifies a critical infrastructure oversight: “Organizations start small and want to scale, but they forget they need to scale the infrastructure as much as the solution.”

The problem occurs when teams build for the proof-of-concept phase where a few terabytes of data are handled manually or semi-automatically. But in production, everything changes.

“Say you’re bringing in a TB of data for a proof of concept. When you move to pilot or production and your data increases tenfold, or you need real-time processing, or you can’t have a developer manually coerce the data—suddenly all those tools built for small scale become inadequate,” Hardy explains.

Defining Business-Ready AI Infrastructure *cont.*

The solution to scaling isn't just better hardware—it's better architecture. To scale successfully, organizations must plan for production from the start. This means building modular infrastructure with automation, observability, and workload readiness already in place. By testing systems under real conditions early—while also ensuring alignment across teams—enterprises can move beyond prototypes and scale AI with speed and reliability.

Leading institutions are adopting modular infrastructure blueprints built around:

- **High-speed object and file storage**
- **Federated data access**
- **Integrated orchestration and observability**
- **Support for streaming ingestion, model training, RAG, and inferencing**

These requirements translate into specific architectural patterns: a hub-and-spoke data lake architecture with zone-based security, containerized model serving with auto-scaling capabilities, and event-driven data pipelines that can handle both batch and streaming workloads.



Read the Report →

The State of BFSI Data Infrastructure



3

Blueprinting to Accelerate Time to Value

Hardy advocates for a systematic approach to AI development: "The biggest thing is you want to blueprint your successes. Take the success you've learned from and your failures to improve on that blueprint so you're not reinventing the wheel every time."

This blueprinting approach creates:

Standardization

Consistent execution frameworks across initiatives

Acceleration

Faster startup for new projects leveraging previous work

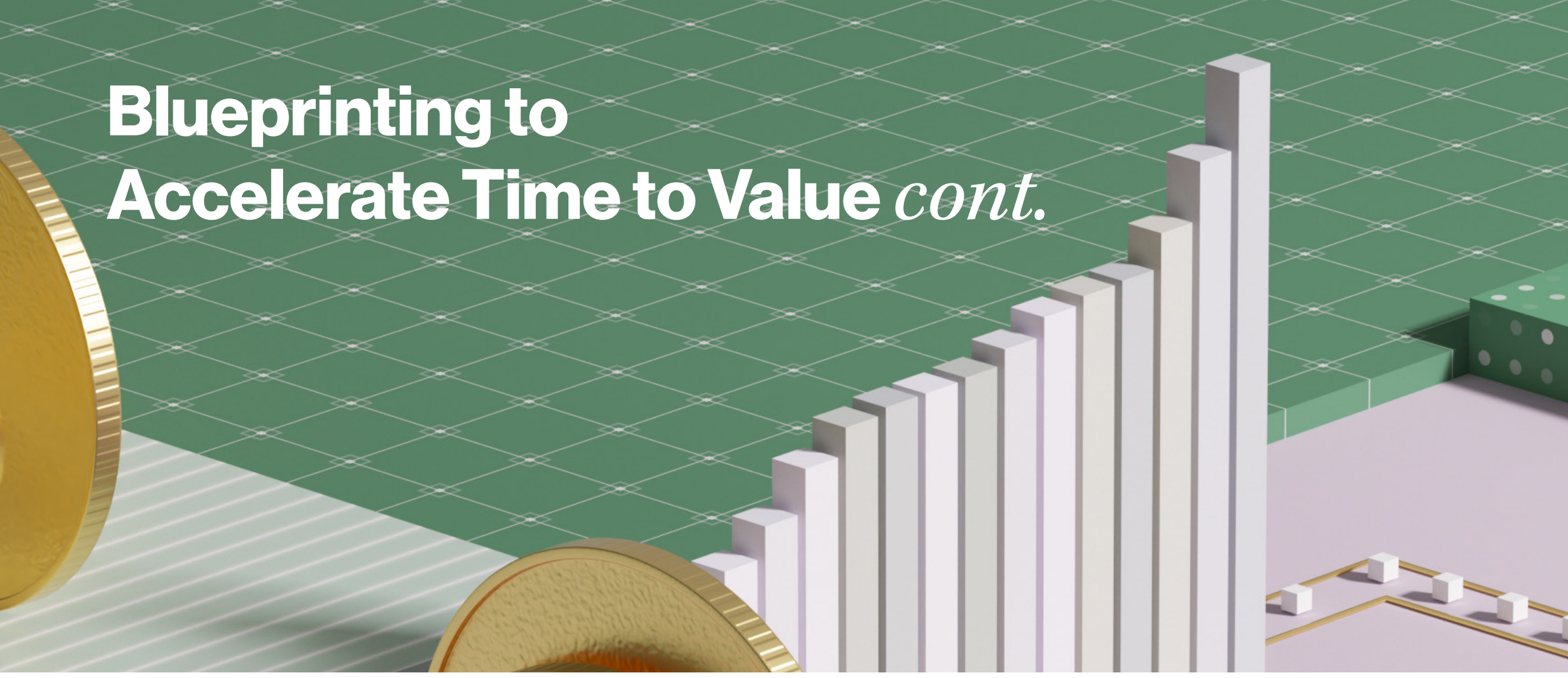
Reusability

Enterprise-wide adoption through proven patterns

Continuous improvement

Each project benefits from accumulated knowledge

Blueprinting to Accelerate Time to Value *cont.*



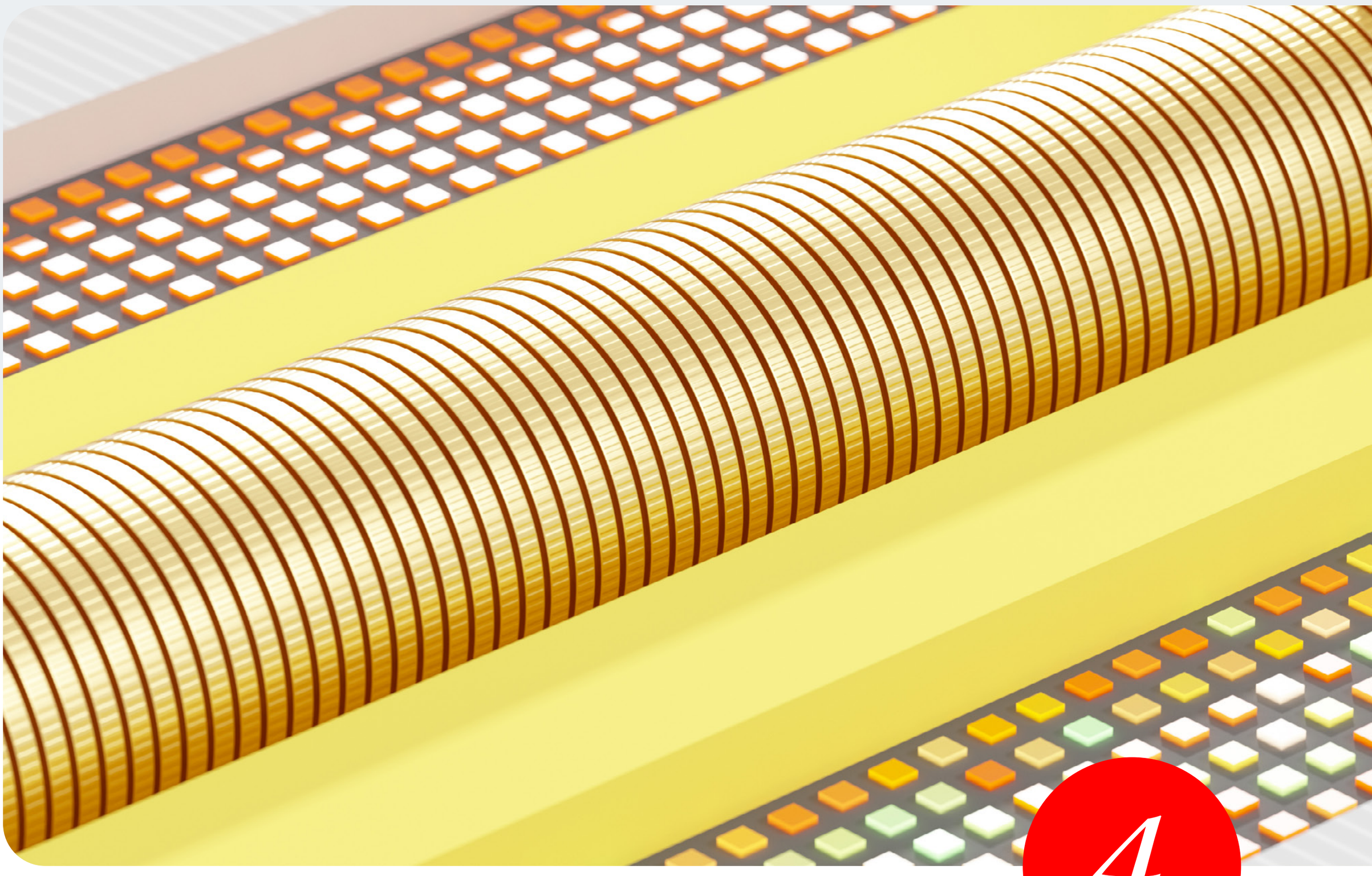
When you build this blueprint, it helps you start initiatives in a more accelerated manner, gives you standardization to execute from, and creates reuse so your next project can benefit from all the previous work," Hardy explains.

Many financial institutions are building internal AI platforms with service catalogs, curated data pipelines, and containerized model orchestration. This approach allows teams to accelerate development, enforce governance, and reduce redundant infrastructure spend.

JP Morgan's COiN platform, for example, is an early success story—saving an estimated 360,000 lawyer-hours per year by automating contract review. That efficiency is now shaping other document-based AI services across the firm.

Both Katz and Hardy emphasize that successful AI requires diverse perspectives from the outset. "You want to have a diversity of people to understand what problem you're trying to solve and ensure it's being applied to the business appropriately," says Hardy.

This cross-functional approach prevents the tunnel vision that leads to technically sound but business-irrelevant solutions. As both Hardy and Katz note, it's important to remember the user. For example, if an internal HR person is going to use a new AI tool, involve internal HR in the development of the tool.



4

Why Governance Is More Important Than Ever

Even the best models fail on poor data. Data quality, provenance, and regulatory exposure are critical concerns for all AI projects, but especially those in financial services.

Regulators are now demanding more accountability in how models are trained, audited, and explained. Yet, according to a 2024 FStech and Hitachi Vantara survey, only 32-percent of BFSI institutions report having comprehensive regulatory compliance frameworks for AI. That leaves many firms exposed.

Hardy identifies the core challenge: “The biggest disconnect is either assuming your data is in a perfect state—and you’re just in denial—or not understanding what outcome you’re trying to create, so you have a misconception of the data needed to drive that outcome.”

Why Governance Is More Important Than Ever *cont.*

Top Data Risks

Hallucination and bias due to unvetted training inputs

Legal liability from training on copyrighted or PII-laden data

Loss of control over data flow and model output without auditing

Mitigation Strategies

- Using Retrieval-Augmented Generation (RAG) to validate outputs against trusted data sets
- Classifying and assessing data with data governance platforms
- Using pre-trained or fine-tuned models to minimize exposure to low-quality data
- Starting with edge cases before moving to core business processes



5

Building Executive Alignment and External Partnerships

Executive Alignment

AI adoption at scale requires more than investment and infrastructure—it requires conviction and clarity from leadership, paired with the humility to seek the right expertise.

Hardy emphasizes that successful AI adoption "starts at the top. You need executive endorsement, which I think most companies are willing to do today." But he also notes that sponsorship alone isn't enough. It must be accompanied by a tolerance for iteration. "If a customer wants 100% accuracy but really only needs 75%, and 100% is impossible—having the right understanding of how good is good enough becomes critical."

Leadership also plays a vital role in setting boundaries and focus. In highly regulated industries like banking and financial services, leaders need to move more carefully. Hardy's guidance: start with low-risk areas like customer service, and ensure that auditability, traceability, and legal review are built in before moving into areas like credit approval or fraud decisions.

Building Executive Alignment and External Partnerships *cont.*

External Partnerships

Leadership, however, does not only involve internal coordination. In a field like AI, leadership also means knowing when to ask for help. As Katz puts it: "This is one of those [situations] where leaders must put the other hat on and say, 'Okay, we're going to accept help here.' The job is to take on the right set of people to help us get to a place where we can have effective knowledge transfer, come up to speed, and figure out what it takes to run an effective service."

Rather than treating external providers as commodity vendors, Katz recommends evaluating partners as long-term allies. True partnership is built on understanding the technology necessary for AI implementation, as well as an institution's particular journey. Questions to consider when choosing a partner: Do they have success stories? Do they understand your use cases? Do they understand where you are in the process and what you've done already? Do they understand where you want to go?

The right partner brings proven methodologies for AI readiness assessment, use case prioritization, and infrastructure blueprinting. Rather than generic consulting, financial institutions need partners who can rapidly assess their data estate, identify high-ROI use cases, and architect the infrastructure needed to support them—all within the context of regulatory requirements and risk tolerance.



Innovate Your AI →

Hitachi iQ: AI Solutions Stack for Enterprise Workloads



6

Use Cases That Deliver ROI—Fast

After many months of AI experimentation, financial institutions are beginning to see where value is—and isn't—being realized. According to Katz, not all AI use cases are created equal. Some are beginning to show repeatable promise, while others have proven underwhelming despite heavy investment.

Katz highlights several high-ROI use cases that are gaining traction in BFSI, such as Know Your Customer (KYC) automation, fixed income and FX trading, and regulatory response automation.

These use cases benefit from large historical datasets and well-defined business processes, making them ideal for AI augmentation. The most promising AI applications in financial services today share three traits: they are narrowly scoped, tied to specific business outcomes, and don't require full LLM generalization. Others include:

- **Intelligent Document Processing for Underwriting and Claims:** For example, HSBC platform processes 1.35B transactions/month, doubling fraud detection while reducing false positives by 60%
- **Risk Modeling and Pricing:** AI is improving risk-based pricing in lending and insurance, analyzing diverse structured and unstructured data inputs. According to Bain, productivity has improved by up to 20% in functions such as underwriting and pricing due to GenAI adoption.
- **Regulatory and Compliance Support:** Leading banks are using AI for KYC, AML, and real-time transaction monitoring.

Use Cases That Deliver ROI—Fast *cont.*

By contrast, credit card fraud detection and equities trading have underperformed, as they are already well-served by existing technology, according to Katz.

The pattern across successful implementations is clear: start with use cases that have defined inputs and outputs, measurable business metrics, and existing process workflows. Document processing and risk modeling excel because they leverage structured decision trees that AI can augment rather than replace entirely.



Read the eBook →

Turning AI into ROI for Banking and Financial Services

Practical Steps to Architect for AI: Jason Hardy's Blueprint



7

1 *Outcome First, AI Second*

- **Clearly define business value and KPIs.**
- **Start with user-driven, measurable goals.**
- **Avoid tech-for-tech's-sake initiatives.**

2 *Think Big, Start Small, Move Fast*

- **Prioritize achievable early wins.**
- **Rapidly prototype, learn, and adapt.**
- **Realism beats grandiosity every time.**

3 *AI is a Team Sport—Build Your Dream Team*

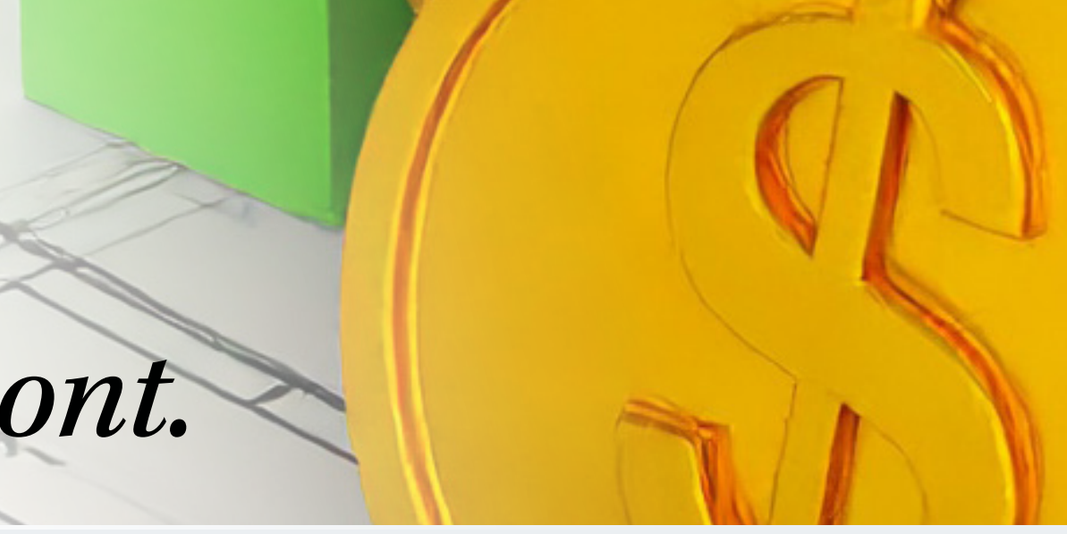
- **Blend technical, business, and domain expertise.**
- **Foster collaboration, not silos.**
- **AI success requires diverse perspectives.**

4 *Data is Your Fuel—Use It Wisely*

- **Identify and prioritize critical data.**
- **Manage quality without over-engineering.**
- **Make systems robust against data noise.**

5 *Infrastructure for the Win*

- **Plan for scale early in your design.**
- **Build flexible, scalable AI infrastructure.**
- **Ensure tech stack readiness for growth.**



Practical Steps to Architect for AI: Jason Hardy's Blueprint *cont.*

6 *Done Beats Perfect*

- **Iterate quickly, avoid perfection paralysis.**
- **Test solutions in the real world rapidly.**
- **Let user feedback guide refinements.**

7 *Users Define Success*

- **Engage end-users throughout development.**
- **Design intuitive, easy-to-adopt solutions.**
- **Invest in user training and change management.**

8 *Trust Matters—Reliable AI Wins*

- **Embed transparency, fairness, and privacy.**
- **Build reliability through rigorous testing.**
- **Treat AI like any critical business system.**

9 *Partners in Power*

- **Identify and fill capability gaps strategically.**
- **Choose partners who complement internal skills.**
- **Own your strategy, leverage external expertise.**

10 *Blueprint Your Success—Scale and Repeat*

- **Standardize and document winning practices.**
- **Create reusable platforms and solutions.**
- **Drive enterprise-wide adoption from pilots.**

Conclusion:

From Vision to Value

AI is no longer a speculative future. It's a strategic imperative. 2025 is the year the financial industry demands results from its AI investments, but getting there requires a shift from opportunistic experiments to aligned execution.

As Hardy concludes, "AI is going to transform how we do business. It's just a matter of how we do that in the regulated space. Step one is not jumping right into the regulated part—it's going after those more open spaces so we can develop and flex that muscle before we step into things that can get us sued out of existence."

That means designing infrastructure around business outcomes, enforcing data discipline, selecting use cases that deliver value fast, and fostering the cross-functional collaboration essential for AI success.

Hitachi Vantara

About Hitachi Vantara

Hitachi Vantara is transforming the way data fuels innovation. A wholly owned subsidiary of Hitachi, Ltd., we're the data foundation the world's leading innovators rely on. Through data storage, infrastructure systems, cloud management and digital expertise, we build the foundation for sustainable business growth.

© Hitachi Vantara LLC 2025. All Rights Reserved. HITACHI is a trademark or registered trademark of Hitachi, Ltd. All other trademarks, service marks and company names are properties of their respective owners.
HV-EB-Architecting-AI-for-Financial-Services-1Aug-A-BNG.pdf

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact