



When banking's marginal cost hits zero

As AI bends the cost curve toward zero, the uneconomic becomes profitable, and the value chain gets rewritten.



AI is becoming the operating layer that global finance runs on. As the marginal cost of banking approaches zero, the economics of the industry flip, and the products banks once ignored become their next source of growth.

Why the cost curve is the story banks are underpricing

- AI is moving from copilots to the substrate beneath origination, servicing, payments, and risk, which changes what banking costs to run.
- When serving one more customer costs almost nothing, the arithmetic that ruled entire product lines stops holding.
- Cloud and foundation models are commoditizing fast, so proprietary data becomes the only durable source of advantage.
- Banks that optimize but defer reinvention hand faster rivals the structural lead, one segment at a time.

The cost curve bends, and the economics invert

Every bank runs on a simple rule. To serve a customer, assess a loan, or advise a client, you need people, and people cost money. That rule set the boundary of who a bank could profitably serve. Below a certain account size, the cost to serve exceeded the revenue, so whole segments were left untouched. AI breaks that rule. As the marginal cost of these tasks approaches zero, underwriting, know your customer (KYC) review, and dispute handling stop scaling with headcount and start scaling with compute [\[McKinsey Global Banking Review 2025\]](#). The boundary moves. A wealth manager who once needed a \$500,000 minimum can now advise a client with a fraction of that, profitably. A lender who skipped thin-file applicants can now assess them at almost no incremental cost. These examples represent not a productivity tale about doing the same work faster but an economics shift about what becomes possible. Micro-loans, thin-file insurance, and advice for the long tail of customers move from uneconomic to feasible. The addressable market widens at the exact moment the cost to serve it collapses.

What the flip unlocks across the value chain

Four shifts compound once the curve bends.

- First, products that were uneconomic become feasible, including credit, advice, and protection priced for a segment of one.
- Second, building gets faster and cheaper, as AI-led development compresses new onboarding journeys, core modernization, and payment corridors from multi-year programs into quarters.
- Third, technology operations runs leaner. AI-led automation across cloud, testing, and support cuts run cost and incident volume in environments where downtime carries regulatory weight.
- Fourth, and most important, data becomes the asset that decides who wins.

Once every bank can access the same cloud and the same models, technology is no longer the differentiator. It is clean, compliant, readily available data about your clients, your processes, and your workflows that makes a difference.

If cost falls toward zero for everyone, then ‘cost advantage’ is not an advantage anymore. It is table stakes. The edge shifts to the two things rivals cannot copy quickly: proprietary data and the willingness to reinvent the process itself, not just automate the old one.

Why winners optimize and reinvent at once

The instinct is to treat AI adoption as a sequence. Optimize the current model first, then reinvent once the savings arrive. That sequence is the costly mistake. Winners run two speeds inside one operating model. Mode one is optimization. Use AI to run the current model faster, cheaper, and increasingly on its own. That means AI-enabled legacy modernization, a scalable data layer, and leaner technology operations. This is the near-term return that funds the journey. Mode two is reinvention. Reimagine processes, products, and business models for what AI now makes possible, from segment-of-one personalization to ambient customer journeys, and new revenue. The structural advantage compounds. The two modes aren’t a choice. They run in parallel, and they play out concretely across the value chain:

- **Origination and onboarding:** Optimize by automating KYC and decisioning to strip cost and delay from the current funnel. Reinvent toward continuous, ambient onboarding, where the customer is recognized and provisioned in the flow rather than pushed through an application.
- **Servicing:** Optimize with agent-assisted, self-healing support that resolves issues at lower cost and higher reliability. Reinvent toward servicing that anticipates, acting on the customer’s need before they raise it.
- **Payments:** Optimize by modernizing corridors and testing to cut run cost and failure rates. Reinvent toward programmable, agent-led money that moves and settles autonomously under set rules.
- **Wealth management:** Optimize by automating research and reporting so advisors spend time with clients, not spreadsheets. Reinvent toward personalized advice at scale, extending tailored guidance profitably to the long tail once priced out.
- **Risk and compliance:** Optimize by automating controls and evidence to reduce manual review and audit load. Reinvent toward continuous assurance everywhere, with risk monitored in real time across every process rather than sampled after the fact.
- **Treasury and markets:** Optimize with faster analysis and forecasting to sharpen decisions on today’s positions. Reinvent toward real-time, agentic liquidity that senses and rebalances continuously.

Beneath all six sit three horizontals that carry the same two-speed logic:

- Technology build optimizes through AI-enabled modernization and delivery, then reinvents as an agentic software factory that ships what was previously infeasible.
- Technology operations optimize through AI-led, self-healing cloud, then reinvents toward autonomous operations that adapt in real time.
- Enterprise functions automate the HR, finance, procurement, and legal service desks, then reinvent them as agents that resolve rather than route.

Most banks fund only the first and defer the second. That is precisely how an institution becomes the one a faster competitor outcompetes.

Efficiency roadmaps will produce a generation of laggards

Many pour AI budget into cost takeout and call it strategy. That produces a cleaner version of the model they already run, on the same terms as every rival doing the same thing. We predict the market will split within three years. The banks that reinvented their processes will hold structural advantages in cost, speed, and reach, and the efficiency-only banks will spend the next decade trying to buy their way back into contention.

Four moves that convert the cost flip into advantage

- **Reframe AI as economics, not efficiency:** Map which products and segments turn profitable once cost to serve approaches zero, and you convert a defensive cost program into a growth thesis, opening the thin-file, micro-loan, and mass-affluent segments that were previously off-limits.
- **Fund reinvention alongside optimization:** Run both as one operating model and the optimization savings become the self-funding engine for reinvention, so you compress modernization from years to quarters without waiting on a fresh budget cycle.
- **Make data the priority investment:** Get clean, compliant, accessible data in place and ensure every downstream agent inherits it, which turns data quality into the one moat rivals cannot buy off the shelf even when they run the same cloud and models.
- **Reimagine the value chain area by area:** Name the efficiency play and the reinvention play for each of the six processes, and you replace a vague AI ambition with a concrete portfolio of moves, each with an owner, a near-term return, and a structural payoff.

About Brillio

Brillio is The Enterprise AI Accelerator helping Fortune 1000 companies move from AI ambition to scaled impact, faster. Powered by our AI accelerator platform – Agentic Data and Application Management (ADAM), Brillio is one of the fastest-growing digital technology service providers, delivering transformation across five core workstreams: business-led transformation, customer experience transformation, AI and data engineering, digital engineering, and infrastructure engineering.

With 14 delivery locations across North America, Europe, and Asia and a team of over 6,000 customer-obsessed professionals, Brillio combines deep industry expertise, modern engineering, and accelerators to deliver measurable outcomes.

Headquartered in Dallas, Texas, Brillio serves clients globally with a commitment to speed, scale, and measurable impact.



<https://www.brillio.com/>

Contact Us: info@brillio.com

