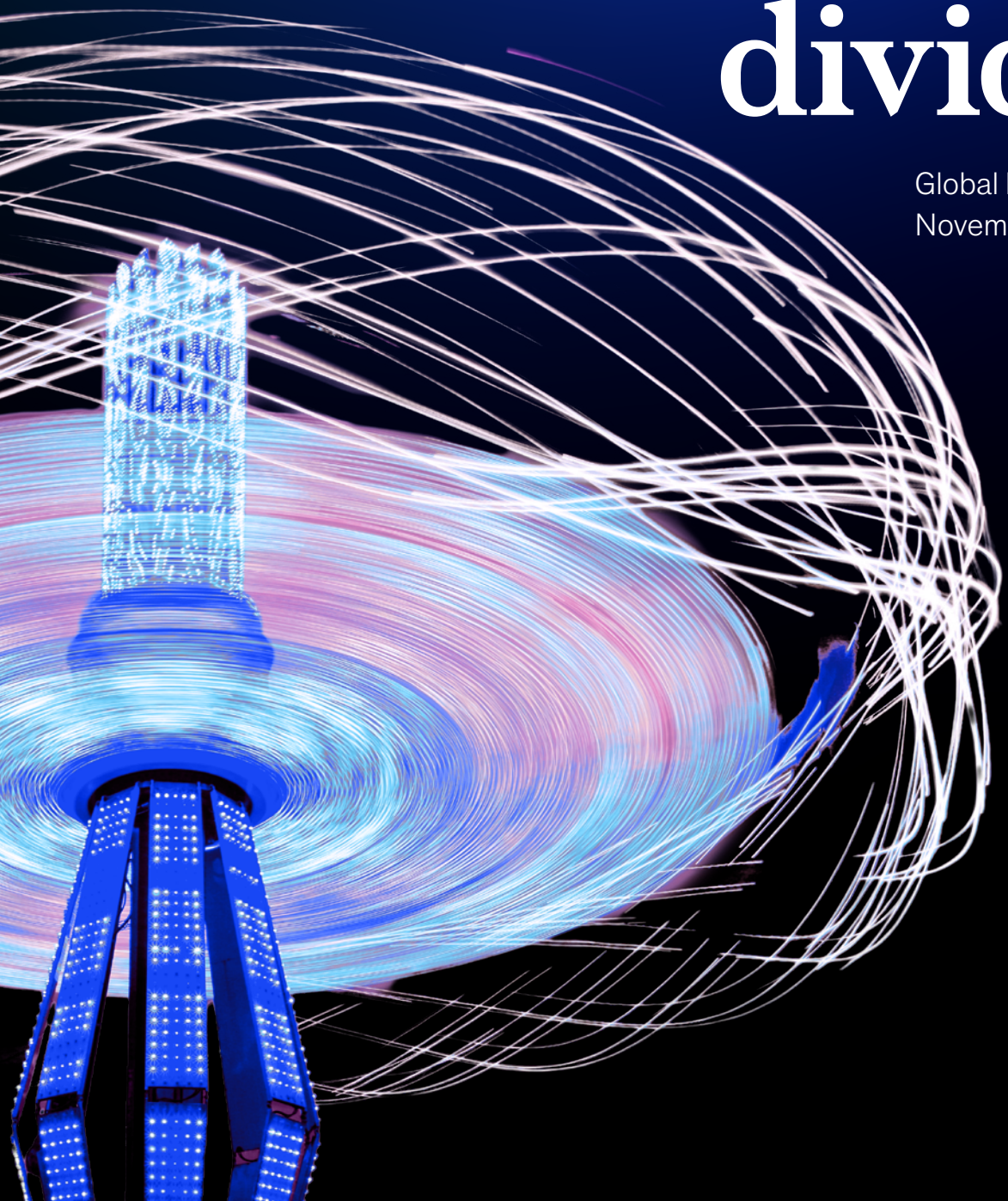
McKinsey & Company

The AI disruption dividend

Global Banking Annual Review
November 2026

by Akshay Kapoor, Brian Ledbetter, Jonathan Leydon, Klaus Dallerup, Marti Riba, Saptarshi Ganguly, Sid Tiwari, Teresa Diviu, and Vik Sohoni



The economics of banking may shift in quarters rather than years as AI changes what the work costs, how a bank and its customers interact, and how those customers choose. Advantage for some banks is building across several assets: data, context, judgment, trust, and the controls around them, and this could compound with use. The critical variable is the speed at which banks capitalize on these assets to capture the AI dividend. Speed could determine how much value the bank keeps, and unlike in past technology cycles, a late start may not offer chances to catch up, depending on how the profits get reinvested.

During Mercedes's run of eight straight Formula 1 constructors' titles,[1] several rival teams raced with the very same Mercedes engine. They had the same horsepower but different results. The differences were everything around the engine: the chassis, the data, the pit crew, and how fast the team learned between races.

Banking is entering its own version of that competition with AI. What we see starting to separate leaders from the rest of the field is the context they feed those models, the controls around them, the speed at which they turn pilots into products, and the orchestration of complex implementation, from personal productivity to business-model change.

With these advantages, leading banks are moving ahead briskly to capture the AI dividend. A common domain for investment is software development. One bank is deploying agents across the software development life cycle (SDLC). Its initial investment was $20 million; after an 18-month buildout, it expects a two-fold return. Another invested $30 million to modernize its mainframe code and expects to realize about $25 million in savings each year. A second popular domain is risk and compliance. One bank built agents to help its team of risk modelers. Within four months, it is on track to free up 10 percent of the team and redeploy them. Another tackled its system of controls, using agents to automate these. After a significant investment ($32 million), the bank expects $50 million a year in savings. The call center is another domain where banks are finding success. One bank invested $5 million to build agents that resolve customer support queries. Within a year, the effort should yield 2.8 times the initial investment. Another bank built agents to classify the intent of incoming queries;

[1] Guinness World Records, "Most consecutive Formula One Constructors' World Championship titles," Mercedes, 2014–21.

it is on track to reach payback of its $2.6 million investment, and the faster calls the technology allows will be worth 3.3 times the investment.

These are not isolated examples. Across 52 agentic AI use cases, the average investment payback was approximately ten months, with a three-year ROI of about 2.9 times (Exhibit 1).

The race for leadership is unfolding in five arenas: the *industry ecosystem*, *consumer and client uptake*, *institutional intelligence*, *infrastructure* (including controls), and *coordination for speed*. In this chapter, we'll review the latest developments, with a focus on the strategies of leading banks. For others, the race is not lost—far from it—but we believe the window to catch up will start closing.

Industry ecosystem: A looming shift in banking's economics and business models

While most attention to date has focused on productivity and expense savings, leaders are contemplating the underpinnings of competition and economics and how AI might structurally alter them. They are considering questions that have been submerged for some time, such as: How does money move? Who stands between a bank and its customer? Where does margin accrue? Where do scale effects show up? And how might prices change?

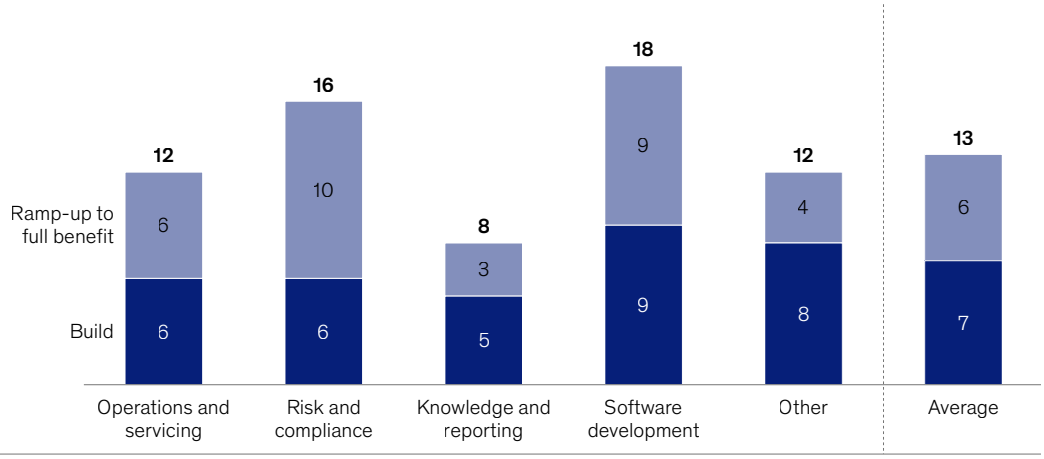
How money moves

Start with the most salient development: Money is beginning to move based on advice from software agents. This is not just a question of whether

Exhibit 1

Agentic AI builds pay back their investment in about ten months.

Average time to value, by domain, months



	Operations and servicing	Risk and compliance	Knowledge and reporting	Software development	Other	Average
Use cases, number	33	7	4	4	4	52

Average payback on investment[1] **~10 months** | *Three-year ROI*[2] **~2.9x** | Use cases **52**

[1]Months of run-rate impact needed to recover the build investment, estimated.
[2]Adjusted for ramp-up (2x annual impact minus 2x run cost minus build investment); assumes year 1 is build and ramp-up with no benefit; estimated.
Source: Client interviews, September-October 2026; McKinsey AI ROI benchmark

agents can reinvest low-cost deposits; rather, it's potentially a redesign of the entire ecosystem of payments: agents that help with buying things on websites, agents that advise clients on where to invest their money, and agents that negotiate contracts and issue invoices. The payment networks appear supportive. Visa's Intelligent Commerce and Mastercard's Agent Pay let merchants accept purchases from verified agents, PayPal agreed to put its wallet into ChatGPT's checkout, and Stripe and OpenAI co-developed the protocol behind it.[2] Meta's Muse agent, launched in September 2026, passed five million downloads in the United States in 22 days, faster than ChatGPT, Grok, or Claude reached that mark.[3] It already reads users' balances and transactions through Plaid.

Many analysts (including us) expect agents to start sweeping idle cash into higher-yielding accounts. An important note here: many banks have issued statements that they will block agents. However, some agents can now simulate human behavior (for example, mouse clicks), and our researchers at QuantumBlack, AI by McKinsey find that even the latest walls by financial institutions cannot detect these agents. It's likely therefore not to be a straightforward path to managing agent access. In the broader commerce space, Perplexity and Amazon are litigating the rules around agent access. But as we showed in last year's edition of this report, if agents can either directly move or indirectly influence through in-the-moment advice even 5 to 10 percent of checking balances to higher-yield accounts, industry deposit profits could fall by 20 percent.[4]

But the dividend from AI disruption can outweigh a decline in deposit profits. A bank that has anticipated the disruption might have already offered its clients more attractive and suitable investment products, financial advice, and cash management capabilities that rebalance the profitability. While deposit profits fall, the bank may see overall profitability rise.

Imagine a "personal CFO" for the household or small business sitting inside your bank's app. Could this change how consumers view their bank and create even more engagement and affinity, raising balances, invested assets and loyalty (and lifetime value)? Would these benefits likely accrue to the banks that are first out of the gate in creating this value? And could that attract—at scale—consumers from banks that don't offer such propositions?

The bank/customer interface

Large language models (LLMs) have emerged as an entirely new distribution channel in record time. While it took mobile nearly a decade to reach half of bank users, our research shows that more than 50 percent of users with $100,000 in wealth are already using LLMs for financial questions. Historically, US consumers compared an average of 2.5 financial institutions when selecting a product; now LLMs can effortlessly and dispassionately compare a far wider set.

But LLMs are more than a channel. They are a new intermediary forming between banks and their customers, and banks have to decide how

[2] "Visa introduces Trusted Agent Protocol: An ecosystem-led framework for AI commerce," Visa press release, October 14, 2025; "Mastercard unveils Agent Pay, pioneering agentic payments technology to power commerce in the age of AI," Mastercard press release, April 29, 2025; "OpenAI and PayPal team up to power instant checkout and agentic commerce in ChatGPT," PayPal press release, October 28, 2025; "Stripe powers Instant Checkout in ChatGPT and releases Agentic Commerce Protocol codeveloped with OpenAI," Stripe press release, September 29, 2025.

[3] "Meta's Muse AI assistant hits 5 million downloads—outpacing growth of ChatGPT, Grok, and Claude," *Forbes*, September 30, 2026.

[4] *Global Banking Annual Review 2025: Why precision, not heft, defines the future of banking*, McKinsey, October 23, 2025; see the agentic AI scenarios in chapter 2.

74%

Share of consumers who use AI for banking tasks and do so through ChatGPT, Claude, or Gemini—about twice the share who use their bank's chatbot.

to address that opportunity or threat. Two early stances are emerging: data source and storefront. OpenAI lets US ChatGPT subscribers link their bank accounts through Plaid for read-only guidance[5]; banks are only a data source. Banco Bilbao Vizcaya Argentaria (BBVA) has built an app inside ChatGPT so that clients can explore its products without leaving the conversation,[6] establishing its platform as a storefront.

Another sign of disrupted distribution: Generative engine optimization (GEO) is emerging as a top priority for marketing teams as web traffic to many sites has fallen by a third. The banking industry let Silicon Valley players own the authentication experience ("Sign in with Google"; "Sign in with Facebook"). With AI, could we see "Sign in with Zelle" so banks are positioned as the gatekeepers to agentic commerce? With cybersecurity threats growing due to AI, could banks leverage the trust that consumers have in them, and their highly sophisticated cyber capabilities and networks to offer new authentication services to the ecosystem?

Shifts in margin, scale, and price

Margin may move in both directions between banks, nonbanks, and customers. AI could cut the industry's cost base by 15 to 20 percent,[7] but competition may pass some of those savings to customers, or to nonbanks like money market funds and asset managers. In this scenario, as we modeled last year, durable gains would accrue to the banks that move first, which can then invest in innovation and better designed and incentivized products, harvest productivity, and offer better rates and experiences to their customers. New demand is springing up at the same time: 47 percent of the corporations, institutional investors, and financial sponsors in our 2026 survey cite AI-related investment as a primary driver of their increased financing needs, and the AI buildout is a lending market in its own right.[8]

The strategic question is therefore how to position the bank to capture the AI dividend as the ecosystem changes. Should the bank serve as the rails that agents run on, as a seller through the platforms, as a supplier of balance sheet to whoever owns the customer's agent, or take positions in all these spots? Will advantage compound for those who move first, or will there be sufficient space to reposition later?

These are profound and complex changes. To develop some insight, we built a model of market competition in one vital US banking business: wealth management. The analysis pits various business models—including registered investment advisors (RIAs), wirehouses, and regional bank brokerages—against one another, using inputs such as elasticity of demand and utility curves for different segments. We then ran Monte Carlo simulations to create 10,000 different future states of the wealth market (see sidebar "A wealth war game").

[5] Will Robinson, "What ChatGPT's new experience signals for digital finance," Plaid, May 15, 2026; Melinda Lucy, "OpenAI launches personal finance tools for ChatGPT Pro users," *American Banker*, May 18, 2026.
[6] "BBVA's app is now available on ChatGPT in Germany and Italy," BBVA, February 24, 2026.
[7] *Global Banking Annual Review 2025: Why precision, not heft, defines the future of banking*, McKinsey, October 23, 2025.
[8] McKinsey Corporate, Institutional and Financial Sponsors Voice of the Client Survey, summer 2026 (300 respondents); see also Denis Francis, Enrico Muti, Jared Moon, and Saswati Hazarika, "Anchor or drift: What it takes to capture wallet share in 2026," McKinsey, September 16, 2026.

A wealth war game

We used McKinsey's proprietary AI-enabled market simulator to war-game how AI might change the economics of US wealth management, with the goals of understanding what will make profit pools move, where growth will come from, how the operating leverage curve might shift, and what changes we could expect within and across channels. The model starts with a 2025 baseline of roughly $57 trillion in client assets, about $95 billion in provider pretax profit, and roughly 286,000 financial advisors across 12 provider archetypes, including wirehouses, registered investment advisors (RIAs), regional brokers, private banks, digital brokers, robo-advisors, AI-native firms, and scale hybrid platforms. Each model has different assumptions for assets, fees, margins, advisor capacity, client mix, technology readiness, client trust, cash balances, and legacy drag. The model then lets those archetypes compete for clients across successive rounds.

Client movement is driven by estimated utility curves. The model assigns different price sensitivity, human-advice preference, digital preference, and service-complexity weights to each wealth band and generation. It then tests how clients reallocate assets as fees, cash yields, AI adoption, advisor capacity, trust, and service quality change. The Monte Carlo simulations that result are scenarios based on one set of assumptions about client behavior, firm economics, market returns, interest rates, AI adoption, and the degree of delegation to client agents. These should not be read as forecasts but illustrative examples of possible alternative futures.

Our clearest finding is that under these assumptions AI does not overturn the sector and it becomes more profitable than it is today. Despite the changes in technology, customer trust remains the dominant force, especially in high-net-worth segments. Industry pretax profit still rises, from about $95 billion in 2025 to between $136 billion and $147 billion in 2035 across the scenarios we tested. AI lowers some revenue pools, but it also reduces service costs and helps advisors support more clients and can increase investment momentum with better in-the-moment advice that clients may not otherwise be able to access.

But AI does change the balance of advantage within the sector. The model shows market share moving toward firms with scale, digital reach, and flexible service models. Trust, brand, breadth, and relevance to complex clients still support a large franchise. AI-native firms raise client expectations for service, transparency, and speed, and they put pressure on incumbents to improve workflows and pricing.

Notably, even a stressed scenario in which net interest income (NII) is pressured by agents does not undermine the sector's economics. The inputs for this stress case make assumptions about asset-weighted client-agent adoption by 2035, interoperability of wealth managers' platforms and third-party agents, and delegation among those capable of acting, which results in the delegated exposure of those with excess cash. It's a far smaller number than industry observers speculate, since not many clients have cash lying around that needs to be swept, as is often believed. The exposure is counterbalanced by the gains from advice fees, platform economics, lending, alternatives, new propositions, and the productivity growth that accompanies such levels of agentic use more broadly.

Contact us for a demonstration of the market simulation.

Despite the changes in technology, customer trust remains the dominant force, especially in high-net-worth segments.



The takeaway: AI is unlikely to make wealth management a less valuable business. But it is likely to change where the value sits. Cost leadership and automation remain important, and value will accrue to firms that can combine scale, trusted advice, and better digital workflows. AI makes it easier for clients to see what they are paying, what they are earning on cash, and whether their provider is delivering enough value. It strengthens firms that can make advice more productive, more transparent, and easier to deliver at scale.

One final note on the shifting industry ecosystem: North America is the epicenter of the AI buildout. It is home to many of the leading frontier labs, a huge share of global infrastructure build, and a significant cohort of world-class AI scientists. North American banks can benefit from easier access to the US ecosystem, including new tools, technical partnerships, and a deep pool of researchers. Of course, all these elements are available to banks elsewhere too, but the closer geographical proximity makes it easier to form the necessary relationships that can lead to collaboration and co-innovation.

Consumer usage: Clients are moving fast and want banks to follow

We surveyed 46,000 retail banking customers in major banking markets worldwide.[9] The survey and our work with banks worldwide suggest five key trends in customer adoption:

[9] McKinsey Global Banking Annual Review Consumer Survey 2026. Survey was conducted in June 2026 across 46,000 consumers in 13 countries: Australia, Brazil, Canada, France, Germany, India, Italy, Mexico, Netherlands, Portugal, Spain, United Kingdom, United States. Respondents were 15% Gen Z, 26% millennials, 27% Gen X, 28% baby boomers, and 3% Silent Generation.

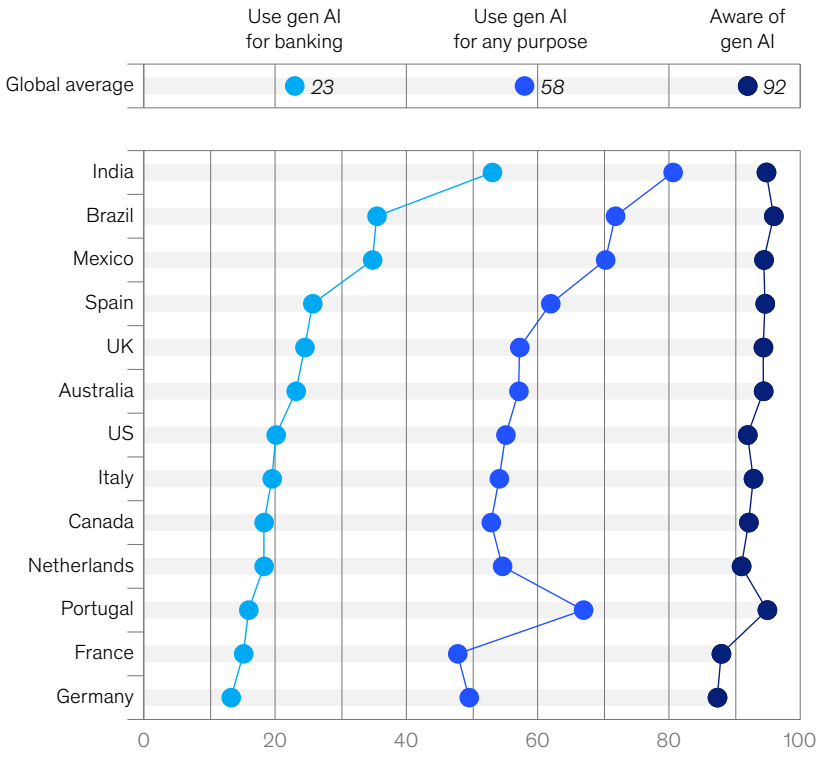
— *Consumers have moved rapidly.* Awareness of gen AI tools is almost universal, and nearly a quarter of customers are using them for their finances (Exhibit 2). A customer who once asked a branch advisor about a savings rate now asks ChatGPT to compare rates across the market. Seventy-four percent of consumers who use AI for banking tasks go through ChatGPT, Claude, or Gemini, about twice the share who use their bank's chatbot today (to be sure, not many banks have rolled out fully enabled bots).[10] The transaction still comes to the bank; the conversation that shapes it is increasingly happening elsewhere.

— *Consumers want to call a human.* Broadly speaking, AI is satisfying many customers, especially for simple questions (Exhibit 3). More

Exhibit 2

Nearly a quarter of consumers use gen AI for financial tasks.

Consumer awareness and use of gen AI, by country, 2026,[1] % of respondents



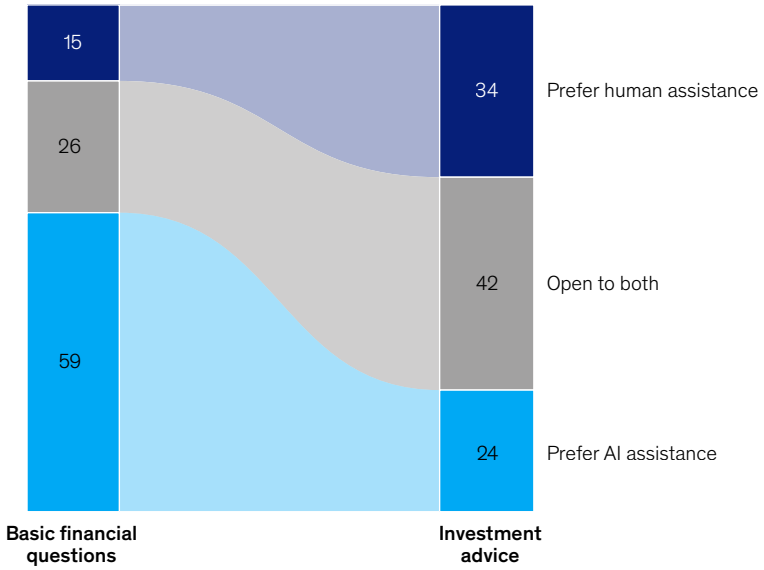
[1]Sample sizes, rounded: India, 2,500; Brazil, 2,700; Mexico, 2,900; Spain, 3,300; UK, 3,700; Australia, 3,600; US, 4,000; Italy, 4,000; Canada, 4,000; Netherlands, 3,900; Portugal, 3,200; France, 4,500; and Germany, 4,400. The survey covered about 46,000 consumers and respondents were: Gen Z, 15%; millennials, 26%; Gen X, 27%; baby boomers, 28%; and Silent Generation, 3%.
Source: McKinsey Global Banking Annual Review Consumer Survey, 2026

[10] McKinsey Global Banking Annual Review Consumer Survey 2026 (37 percent of banking AI users use a bank's chatbot).

Exhibit 3

For investment advice, customers prefer a human over AI.

Consumer preferences for AI vs human assistance with financial tasks, 2026,[1] % of respondents



[1]The survey, conducted in June 2026, included approximately 46,000 consumers in 13 countries. Respondents were: Gen Z, 15%; millennials, 26%; Gen X, 27%; baby boomers, 28%; and Silent Generation, 3%.
Source: McKinsey Global Banking Annual Review Consumer Survey, 2026

than four in ten say AI's financial advice is as good as or better than a human's. But the story changes when they pick up the phone. Two in three AI users would still rather speak to a person than to a bank's voice agent even with no wait.[11] They prefer the bot for information but find it wanting, above all in voice. The gap is closable, as Bank of America's Erica shows.[12]

— *Consumers are more willing to delegate than many assume*, within limits they set. Revolut's AIR, launched to its UK customers in April 2026, executes sensitive actions after the customer's biometric approval.[13] Seventy-two percent of AI users would let a banking agent carry out at least one financial task; most want it to ask before it acts; and most would cap autonomous approvals at $500.[14] The customer wants a capable deputy that asks for permission, and the bank that gets the permission model right may win the mandate to move the customer's money.

— *Trust remains the cornerstone, with banks leading the way.* Sixty-two percent of AI users would rather get AI services and advice from

[11] McKinsey Global Banking Annual Review Consumer Survey 2026 (43 percent; 66 percent, up 10 points from 2025; 59 percent).
[12] "A decade of AI innovation: BofA's virtual assistant Erica surpasses 3 billion client interactions," Bank of America press release, August 20, 2025.
[13] "Ask, and AIR makes it happen," Revolut blog, April 9, 2026; Allison Steffens Herrera, "Revolut launches its new AI assistant AIR to UK customers," The Next Web, April 10, 2026.
[14] McKinsey Global Banking Annual Review Consumer Survey 2026.

their primary bank than from anyone else, and privacy is their greatest concern with AI.[15] The pattern we're seeing: A customer may experiment with a chatbot on a $1,000 balance but a family deciding where to hold $300,000 of savings will meet its human advisor. Trust also unlocks the data that makes answers personal and relevant; third-party tools can access only some of that data today. The advantage, however, is conditional. Roughly half of AI users say they may turn to a third-party financial app if their bank does not offer capable AI,[16] so trust buys a bank the first chance, and half of its AI-using customers say they will not wait indefinitely.

— *Corporate clients want the same combination.* In our 2026 survey, corporate clients increasingly expect AI inside the services they buy, from cash flow forecasting to credit assessment.[17] Further, they expect transaction banking agents that act in real time. Taken together, it's a shift they describe as moving from relationship banking to intelligence banking, delivered through a relationship manager they trust.

Institutional intelligence: From faster employees to a smarter bank

Most banks have used AI so far to help employees work faster. The next step, institutional intelligence, makes the bank itself smarter. When a bank's best credit officer retires today, her judgment leaves with her. Nor is that the only problem: The bank's best thinking is usually scattered across the institution. And decision-makers struggle to thread their best ideas through myriad policies and compliance requirements. Institutional intelligence connects the data, policies, past decisions, and expert judgment that currently sit here and there, gathering them so that every agent and employee works from what the whole bank knows, a knowledge that expands on itself. With that, average employees can attain top-quartile performance, in our experience.

Individual productivity benefits are already apparent. In McKinsey's 2026 State of AI survey, 80 percent of respondents report individual productivity gains from AI. But only 6 percent attribute a significant share of EBIT to it.[18] Individual gains diminish before they reach the P&L. Why? One reason is that the technology today has been designed at the level of the task—but tasks

Leading banks are discovering that the process of building institutional intelligence takes time, but results get better and better.

[15] McKinsey Global Banking Annual Review Consumer Survey 2026 (62 percent, unchanged from 2025; privacy cited by 37 percent).
[16] McKinsey Global Banking Annual Review Consumer Survey 2026.
[17] 2026 McKinsey Corporate and Investment Banking (CIB) Voice of the Customer Survey; Denis Francis, Enrico Muti, Jared Moon, and Saswati Hazarika, "Anchor or drift: What it takes to capture wallet share in 2026," McKinsey, September 16, 2026.
[18] Dan Tinkoff, Lieven Van der Veken, Michael Chui, and Tara Balakrishnan, "The state of AI in 2026: On the road to ROI," McKinsey, August 25, 2026.

are not easily reallocated. In a benchmark by QuantumBlack, AI by McKinsey, of seven agentic systems running in production at financial and professional services firms, agents cut the time that tasks take by roughly 60 percent.[19] But the bank captures far less because an individual's savings are fractional, an hour or two of someone's day. No bank can monetize a fraction of a job.

A second reason is that tools are running in isolation. JPMorgan's chief analytics officer says the next horizon of value lies in connecting productivity tools to the bank's data and applications, rather than in more adoption.[20] Employee adoption is the scorecard many banks are using, but a better scorecard is the business outcomes achieved in a few chosen domains.

A third reason: While tasks have been sped up, processes have not. A credit memo that now takes minutes still waits for a committee that meets on Thursdays. Leading banks have realized they need to redesign the work or restructure roles and teams so the fragments add up. Only one-quarter of companies using gen AI have done that redesign anywhere.[21] The constraint is organizational: Banks need to redesign the end-to-end domains that matter and measure the result in cost, throughput, and risk rather than in hours saved.

Elements of intelligence

An agent is only as good as the knowledge it can reach. Leading banks are discovering that the process of building institutional intelligence takes time, but results get better and better. They're building three layers of institutional intelligence:

— *Data*—reliable data products, such as a customer 360 and a credit 360, that agents can consume at scale. Forty-eight percent of the banking executives we surveyed rank data readiness as their biggest AI obstacle, while technology is the smallest problem.[22]

— *Context*—a map of how a customer's accounts, obligations, and regulatory status connect, with permissions enforced before anything reaches a model, so that every agent works from one version of the customer rather than its own.

— *Judgment, encoded*—what the senior credit officer knows that never made it into the policy manual, captured in a format that is reusable by agents.

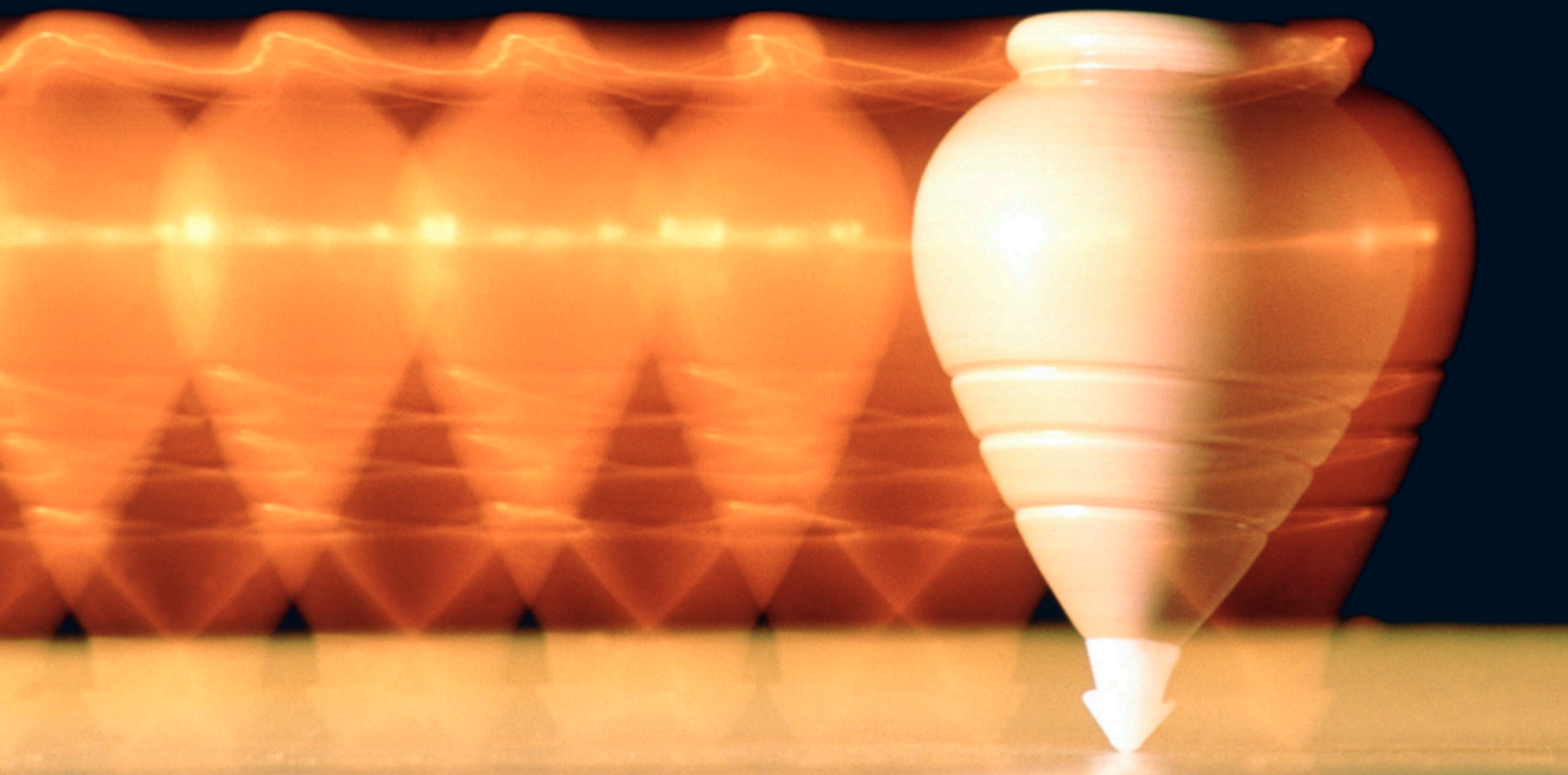
Since every bank can buy the same models, these layers are the moat, and they are accumulated domain by domain from the institution's memory and practices.

Institutional intelligence allows AI to accelerate and enrich the work of everyone in the bank, starting with the CEO. As Dave McKay, CEO of Royal Bank of Canada, said to Bloomberg recently, "I have my own CEO suite of agents. I have it tell me each morning what happened overnight and what important data is coming out that day. I use it to look at consumer behavior, markets, the economy and risk. I compare what other people are saying with what my own team is saying, and look at different forecasting horizons. It lets me drill into the organization without calling people to run report after report. I can operate five times faster and understand data trends and work through questions and answers."[23]

DBS built all three layers into the agentic solution it rolled out to some 1,500 corporate bankers in August 2026, whose agents draft review-ready

[19] McKinsey analysis.
[20] "JPMorgan Chase's Derek Waldron on building an AI-first bank culture," McKinsey, October 29, 2025.
[21] Dan Tinkoff, Lieven Van der Veken, Michael Chui, and Tara Balakrishnan, "The state of AI in 2026: On the road to ROI," McKinsey, August 25, 2026.
[22] McKinsey Banking Executive Survey on AI, 2026 (85 respondents): data readiness 48 percent, talent 41, change management 37, operating model and strategy 33 to 35, technology 27.
[23] "Five questions for Royal Bank of Canada's CEO," Bloomberg, October 26, 2026.

The customer wants a capable deputy that asks for permission, and the bank that gets the permission model right may win the mandate to move the customer's money.



credit memos and, in the bank's words, capture "the knowledge and insight of our best relationship managers and credit risk managers."[24] Morgan Stanley did the same for its wealth advisors, starting with a 2024 assistant that 98 percent of advisor teams use, followed by several other investments.[25]

Infrastructure and control: Affordable to run, safe to scale

Banks are already spending a great deal on AI and wondering about the return on investment. They're also grappling with a design choice: As agentic work scales, how do banks prioritize and route queries to different models to ensure optimized economic usage? And they need to figure out ways to monitor usage and cost. In other words, they're searching for an AI operating system.

Tokenomics is evolving

Start with costs. Agents turn AI into a material and volatile cost line. Spending on tokens, compute, and AI operations can reach as much as 20 percent of labor costs.[26]

Containing the bill takes dedicated discipline and architectural savvy. AI FinOps, as many are now calling it, requires a new form of management. Simple finance controls are inadequate to manage token spend. After all, if a brilliant engineer is creating magic with AI, why meter their spending? Similarly, while many industry pundits proclaim that routing tasks to small models will solve token costs, there are complex quality and architectural challenges to be solved before that can be done.

One intriguing challenge is that determining whether a query (or chunk of a query) is simple enough to be solved by a smaller model requires using a frontier model, since it needs to detect the true complexity of that query. At that point, routing it to a smaller model may not make sense at all. Such conundrums need to be solved architecturally.

Similarly, quality evaluation is easy for coding, but difficult for knowledge work. Switching models subtly changes the quality of answers. Evaluating quality and then feeding it into the routing algorithm

Agents turn AI into a material and volatile cost line. Spending on tokens, compute, and AI operations can reach as much as 20% of labor costs.

[24] "DBS scales agentic AI to transform way of working for corporate bankers, freeing up time for more strategic client engagements," DBS press release, August 19, 2026.
[25] Rocio Fabbro, "Morgan Stanley has a new OpenAI-powered assistant for its financial advisors," Quartz, August 6, 2024.
[26] "Beyond the copilot: Scaling the agentic product development life cycle," McKinsey, August 21, 2026.

$900 million

Reported losses from AI-enabled crime recorded by the FBI in 2025, the first year it tracked such crime as a distinct category.

is not trivial for models and also requires human effort. Banks need a system that ties together quality and tokenomics before deciding how to toggle spend; our research has revealed more than 20 architectural patterns that banks can examine to understand which ones they are using and how best to lower spending. Of course, there are some no-regrets steps, like using the increasing competition to negotiate unit prices down as banks did with cloud a decade ago; showing every team what it spends; and stopping frivolous usage.[27]

Controls built for people will not support agents

Turning to safety, today's control model was built for people executing steps at human speed: checklists, four-eyes approvals, and audits that sample a few cases at a time. Banks are now running agents that work at machine speed. Control therefore moves from the process to the policy: The bank approves the rules, limits, and escalation paths an agent operates within, monitors the full population of its actions rather than a sample, and keeps a named human accountable for exceptions. The controls themselves become agents, with first-line controls in code and second-line agents monitoring outputs continuously. Confidence in an agent is engineered before release, checked in production, and improved from what goes wrong.

The biggest risk, of course, is external. The current cybersecurity model is designed to fight rogue humans. Rogue agents are another matter altogether. AI cuts the cost of attacking faster than the cost of defending. A conversation-based fraud that once occupied the time of a single hacker now costs a fraction of a cent of machine speech, so an attacker could potentially run a million personalized conversations at once while the bank's defenders answer at human pace. So far, machine speed is beating human speed: The FBI recorded AI-enabled crime as a distinct category for the first time in 2025, with about $900 million in reported losses.[28] Agents widen the exposure from the inside, because an agent with write access to accounts and payments is a new kind of insider, and the identity and access frameworks built for employees do not yet cover software acting on their behalf.[29] Defense has to run at machine speed too, starting with identities and entitlements for agents.

Taken together, the disciplines sketched above form a single "control plane" (a term borrowed from computer science) that we call WORTH: Walls to define what an agent may see and do; Observability and orchestration; Responsible AI; Tokenomics; and Hallucination control (see sidebar "One control plane to rule them all"). Crucially, all these elements need to work together in a coordinated way. With that, tokenomics solves for quality while walls are enforced at every level so agents can work reliably.

Coordination for speed

Banking's new critical path runs from an idea to an agent in production. Most banks still run it the way they ran digital, with two-week sprints, ten-person squads, and approvals in series. That's a problem,

[27] McKinsey Enterprise AI FinOps survey, May 2026; "Is that AI agent worth it? Agentic economics and the modern operating model," McKinsey, July 13, 2026; "The cost of intelligence: How CIOs can manage AI demand at scale," McKinsey, July 20, 2026.
[28] FBI Internet Crime Complaint Center, 2025 Internet Crime Report, April 2026.
[29] "JPMorgan Chase's Derek Waldron on building an AI-first bank culture," McKinsey, October 29, 2025.

One control plane to rule them all

Cost, controls, and security are not separate projects. Leading banks build them once, as a single control plane that wraps every layer of the stack (exhibit) and that every new agent inherits. We call it WORTH:

Walls. Permissions cover data, tools, and actions, with human approval for high-risk steps: For example, an agent can read a client's accounts but needs sign-off before it moves money.

Observability and orchestration. Every agent run is logged and can be replayed, with central controls to throttle or roll back an agent.

Responsible and resilient AI. Policy, bias, and safety checks run on inputs and outputs, with a safe fallback when a control fails.

Tokenomics. Every workflow is priced by token, model, and tool call, and routed to the cheapest model that does the job.

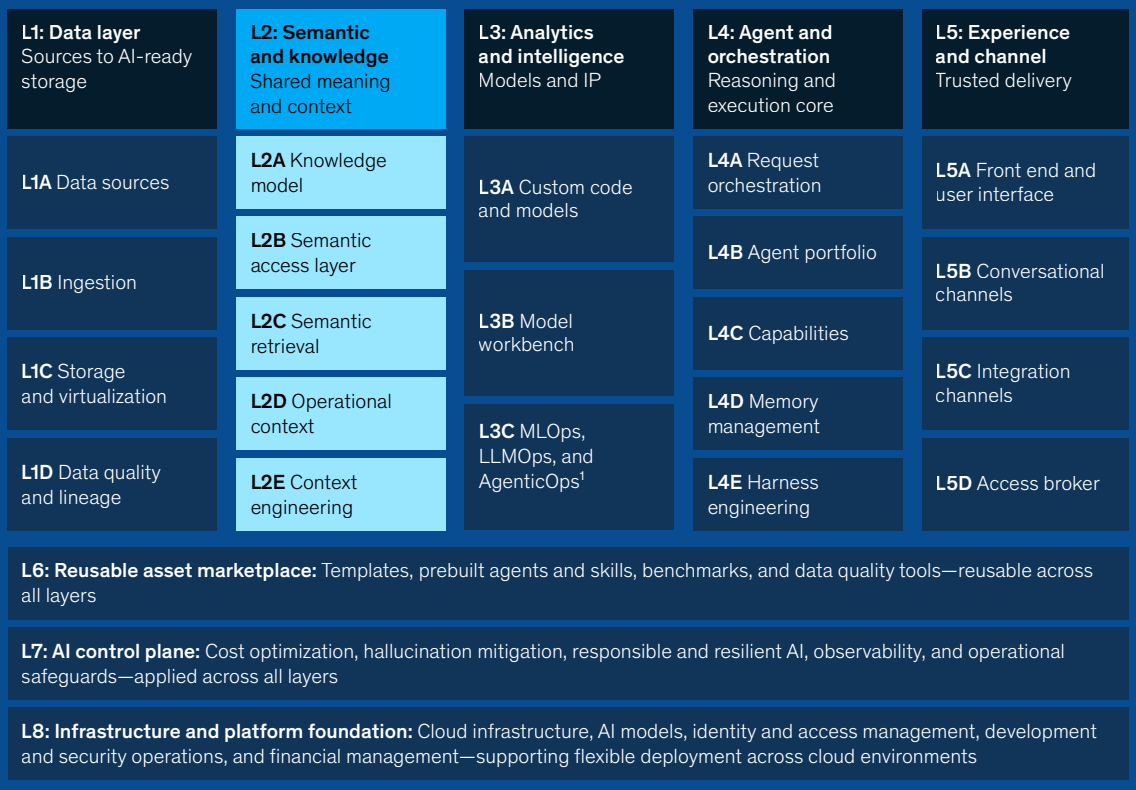
Hallucination control. Outputs are checked against approved sources before release; an answer without a citation is withheld.

The control plane sits underneath the intelligence layer (exhibit): the pipelines, models, and gateways can be bought or swapped; the data, context, and judgment they carry are the bank's own.

Exhibit

The AI-ready bank is built in layers, and the semantic and knowledge layer is where advantage accumulates.

Layers of an AI-ready bank's technology architecture



[1]Machine learning operations, large language model operations, and agentic operations.

not least because internal resistance to the new technologies is even greater than it was for digital. The banks moving fastest have redesigned the operating model to build and run agentic AI: *agile* stripped of its ceremonies (structured meetings) for the teams that build; a *new chart* for how humans and agents are organized; new *paths for talent and judgment*; and a *factory* that can build or enable all levels of complexity.

The principles of agile development still hold, but its ceremonies were built for a production system made of people. With agents now handling much of the executing and working overnight, the two-week sprint gives way to sprints of a day or two. Coordination events like "scrum of scrums" or "big room planning," where different release trains are coordinated, could be performed by an AI system that catches dependencies and conflicts in plans and backlogs daily. One global bank shows the sequence: An early copilot rollout delivered 10 to 15 percent productivity gains; the bank then seconded its best engineers into delivery squads, wrote throughput targets into team objectives, and moved to daily sprints with agent night shifts, and the teams in scope reached efficiency gains of at least 40 percent.[30] In McKinsey's May 2026 survey of 334 product and engineering leaders, teams that embedded AI and redesigned roles became top accelerators 40 percent of the time, against 27 percent for teams that adopted the tools and left the roles alone.[31]

As our colleagues recently discussed, agentic AI loosens three constraints that have shaped organizations for a century: human capacity, human expertise, and the cost of coordination. Start with the outcome, assemble the humans and agents needed to deliver it, and map the result in an *orchestration chart* rather than an org chart. The prize for banks is large because their coordination overhead has grown rapidly: Over the past five years, coordinator roles such as project management have grown 1.5 to 2.0 times as fast as line roles.[32]

Agents produce a productivity dividend, and the bank has to decide where it goes: more output, different work, lower cost, or a mix. The decision directly influences the bank's *talent paths*. Whatever their choice, most banks are finding that many current jobs need to be redesigned. When agents do the producing, the people who remain are exercising judgment, requiring refined skills such as the ability to spot an answer that is plausible but wrong. Moreover, the work arrives in volumes no review queue was designed to absorb. Banks that treat supervision as a distinct job family, with its own training, quality bar, and career path, get oversight they can rely on.

Some of the banks moving fastest have built what we call an *agentic mission factory*: a dedicated construction site inside the bank where human-and-agent teams redesign work, decisions, and business models mission by mission, and install the infrastructure, controls, and institutional intelligence that make the next mission faster.[33] Each mission takes one business outcome, such as approving and funding a good loan faster, and rebuilds the end-to-end process behind it, from trigger to outcome.[34] The factory owns the standards and reusable assets; the businesses own the missions and their outcomes (with one

[30] "Beyond the copilot: Scaling the agentic product development life cycle," McKinsey, August 21, 2026.
[31] "Beyond the copilot: Scaling the agentic product development life cycle," McKinsey, August 21, 2026; McKinsey Agentic PDLC/SDLC Survey, May 2026 (334 respondents).
[32] Alexis Krivkovich, Brooke Weddle, Holly Price, and Vik Sohoni, "AI is changing work. Now it has to change the organization," *McKinsey Quarterly*, October 5, 2026.
[33] Alexis Krivkovich, Brooke Weddle, Holly Price, and Vik Sohoni, "AI is changing work. Now it has to change the organization," *McKinsey Quarterly*, October 5, 2026.
[34] Alexis Krivkovich, Brooke Weddle, Holly Price, and Vik Sohoni, "AI is changing work. Now it has to change the organization," *McKinsey Quarterly*, October 5, 2026.

exception, when the factory acts as a challenger unit by building an AI-native business). The factory needs a leader fluent in both strategy and technology, with the authority to move money and talent across the bank. In our experience, every bank that has moved from pilot to production at scale has done it with the CEO as the forcing function and a senior leader personally responsible for driving adoption.

Beyond the mission factory, the next frontier is the one-person agentic company, in which agents run marketing, finance, HR, and technology, and a founder or a few people approve high-risk actions and answer for the results. In early builds we have seen, the agents planned marketing campaigns and then wrote the CRM system to support them. These companies have become workable as frontier models gained the ability to persist through long tasks, memory architectures improved, and agent harnesses hardened. Founders have been the first to act on it: solo founders led about 36 percent of US startups founded on Carta in 2025, double the share of a decade earlier, and among startups incorporated through Stripe Atlas in 2022 and 2023, AI-native solo companies earned nearly twice the revenue of other solo-founded ones by their second year.[35]



[35] Peter Walker and Kevin Dowd, "2026 Founder Ownership Report," Carta, March 12, 2026; Jesse Carey, "Solo founding is at an all-time high: Top performers have these traits in common," Stripe, May 28, 2026.

For banks, the agentic company is at once a client, a competitor, and a blueprint. As clients, these companies will send agents to open accounts and move money within limits the owner sets. Some will compete with banks: In our wealth market simulation, AI-native firms put pressure on incumbents to improve workflows and pricing. The design also gives banks a template for the AI-native challengers their mission factories will build.

The one-person agentic company crystallizes the urgent challenges AI is creating for the workforce. Our new US-focused research indicates that AI could create more jobs than it destroys.[36] For banks and others, the focus of workforce management will likely shift from job titles toward skills. Banks can distinguish between employees able to migrate quickly into growing roles, those who need targeted upskilling, those requiring longer transitions, and roles that cannot be filled internally. Banks' culture and workforce incentives will be key determinants of the skills employees develop and how prepared they are to adapt. Helping workers navigate those changes also requires transparent communication, structured learning, and involving employees in reinventing how work evolves in tandem with AI.

Banks are rapidly adopting AI. Profit pools are shifting slowly, but that will change. What matters are the actions banks take today to set themselves up for an agentic future: finding the best positioning in the industry ecosystem, catching up to customer adoption, and building institutional intelligence and an infrastructure of controls, all at an unprecedented pace.

To help shape their thinking on positioning, leadership teams and boards can consider the following questions:

- *What is our thesis on the industry shifts that AI is setting in motion?* As scale arises where it did not exist before; as personalization takes over where standardization was the norm; and as matching the cost of service to the cost of tokens compresses margins, how are we preparing? What new propositions will we develop to ride AI rails, such as a personal CFO agent for customers, top-notch GEO, and agentic commerce-native offerings?
- *Is our innovation engine largely agentic?* Can we move at machine speed?
- *Are we developing AI fluency in our people?* Is our organizational structure built for AI?
- *How will our institutional intelligence compound?* Do we have an AI operating system that connects the C-suite to the frontline and agents, moving the bank faster and smarter?
- *What is the state of our cybersecurity?* Are we able to deal with agentic attacks?
- *As we become an AI-first bank, how are we preserving core strengths such as credit skills and a culture of apprenticeship?*

Akshay Kapoor is a senior partner in McKinsey's New York office. **Brian Ledbetter** is a senior partner in the London office. **Jonathan Leydon** is an associate partner in the Boston office. **Klaus Dallerup** is a senior partner in the Copenhagen office. **Marti Riba** is a partner in the Barcelona office. **Saptarshi Ganguly** is a senior partner in the Boston office. **Sid Tiwari** is a partner in the Phoenix office. **Teresa Diviu** is a director of client capabilities in the Barcelona office. **Vik Sohoni** is a senior partner in the Chicago office.

[36] "Workforce in motion: Skills and pathways to future jobs in the United States," McKinsey Global Institute, September 29, 2026.

Contacts

Authors, McKinsey's *Global Banking Annual Review*, 2026

Akshay Kapoor
Senior partner, New York
akshay_kapoor@mckinsey.com

Brian Ledbetter
Senior partner, London
brian_ledbetter@mckinsey.com

Javier Martinez Arroyo
Partner, Paris
javier_martinez_arroyo@mckinsey.com

Klaus Dallerup
Senior partner, Copenhagen
klaus_dallerup@mckinsey.com

Marco Vettori
Senior partner, Milan
marco_vettori@mckinsey.com

Marti Riba
Partner, Barcelona
marti_riba@mckinsey.com

Miklós Dietz
Senior partner, Vancouver
miklos_dietz@mckinsey.com

Saptarshi Ganguly
Senior partner, Boston
saptarshi_ganguly@mckinsey.com

Sid Tiwari
Partner, Phoenix
sid_tiwari@mckinsey.com

Vik Sohoni
Senior partner, Chicago
vik_sohoni@mckinsey.com

Filippo Castellano
Associate partner, New York
filippo_castellano@mckinsey.com

Jonathan Leydon
Associate partner, Boston
jonathan_leydon@mckinsey.com

Roland Schneider
Senior knowledge expert, Frankfurt
roland_schneider@mckinsey.com

Teresa Diviu
Client capabilities director, Barcelona
teresa_diviu@mckinsey.com

Valeria Laszlo
Panorama senior asset leader, Budapest
valeria_laszlo@mckinsey.com

Regional leadership contacts

Asia
Renny Thomas
Senior partner, Mumbai
renny_thomas@mckinsey.com

Eastern Europe, Middle East, Africa
Mazen Najjar
Senior partner, Kuwait
mazen_najjar@mckinsey.com

Latin America
Felipe Villarreal
Senior partner, Panama City
felipe_villarreal@mckinsey.com

Roberto Marchi
Senior partner, São Paulo
roberto_marchi@mckinsey.com

Europe
Stephanie Hauser
Senior partner, London
stephanie_hauser@mckinsey.com

North America
Ishaan Seth
Senior partner, New York
ishaan_seth@mckinsey.com

Acknowledgments

The authors would like to thank the following colleagues for their contributions to this report:

Partners and senior partners
Akash Lal, Amit Garg, Felipe Villarreal, Ignacio Yurrita, Ishaan Seth, Joydeep Sengupta, Juan Aristi Baquero, Lorenzo Serino, Max Flötotto, Pradip Patiath, Renny Thomas, Robert Byrne, and Stephanie Hauser

Analytics and solutions colleagues
Ajay Gupta, Anubhav Das, Debopriyo Bhattacharyya, Kriti Suman, and Shikha Gupta

Production team
Alec Zorrilla, Arpita Matkari, Balazs Hajna, Carine Zahra, Cem Ozmen, Chris Depin, Cyrielle Grivot, Diane Rice, Dumitru Diocher, Eleonora Coletti, Eszter Teszarik, Francesco Lippi, Gabriel A Fernandez, Gurminder Marwaha, Jana Zabkova, Janet Michaud, Katie Shearer, Marc Mitrovic, Matt Perry, Michael Greenberg, Moritz Schwade, Nancy Cohn, Nick Coats, Priya Bhirgoo, Richard Johnson, Roger ten Have, Sarah Thuerk, Sherry Ghali, Thomas Descamps, and Zsofia Sveiczer

This report was edited by Mark Staples and developed and produced by the McKinsey Global Financial Services Marketing and Communications team:

Gurminder Marwaha
Global head of marketing and communications, banking
gurminder_marwaha@mckinsey.com

Chris Depin
Communications coordinator
chris_depin@mckinsey.com

This annual review of the global banking industry is based on data and insights from McKinsey Panorama, McKinsey's proprietary banking research arm, as well as the experience of clients and practitioners from all over the world. We welcome comments about this research at fs_external_relations@mckinsey.com.

McKinsey & Company
November 2026
Copyright © 2026 McKinsey & Company. All rights reserved.

McKinsey.com

@McKinsey
@McKinsey
@McKinsey

This publication is not intended to be used as the basis for trading in the shares of any company or for undertaking any other complex or significant financial transaction without consulting appropriate professional advisors.

No part of this publication may be copied or redistributed in any form without the prior written consent of McKinsey & Company.

