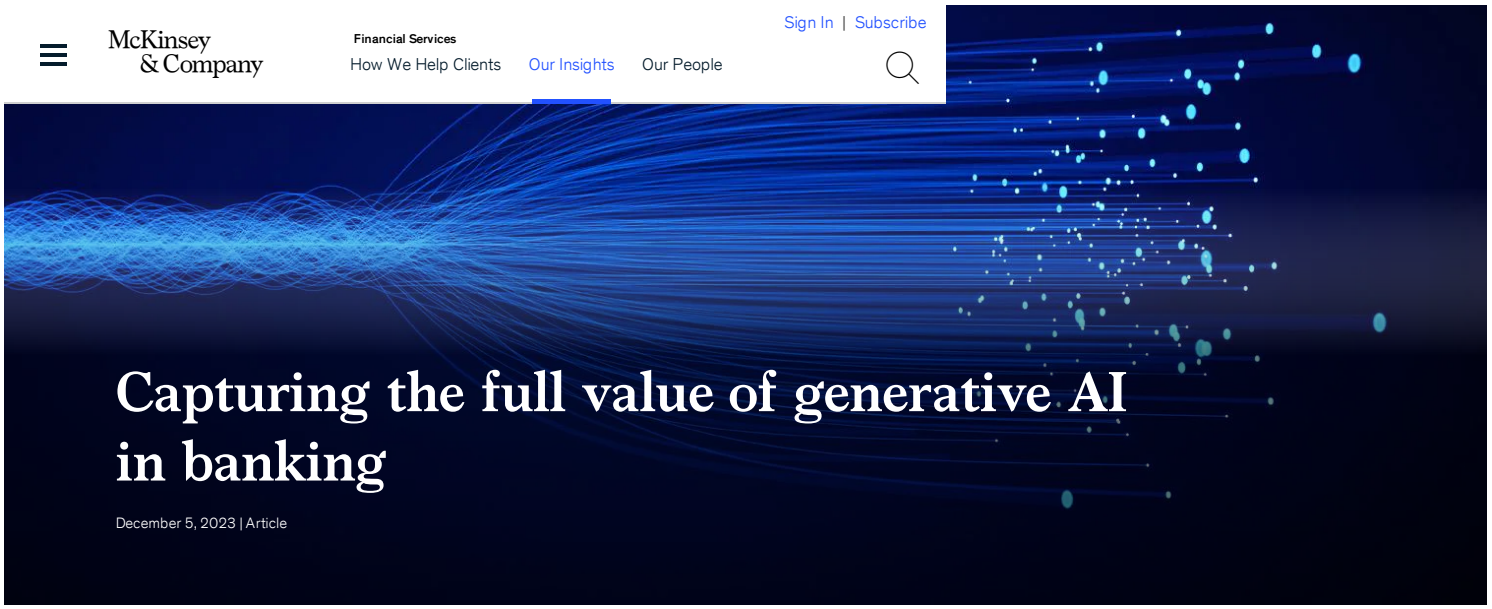


Capturing the full value of generative AI in banking

December 5, 2023 | Article

Setting up gen AI pilots is easy; scaling them to capture material value is hard. A recipe for success is emerging.

Generative AI burst onto the scene in early 2023 and is showing clearly positive results—and raising new potential risks—for organizations worldwide. Banking leaders appear to be on board, even with the possible complications. Two-thirds of senior digital and analytics leaders attending a recent McKinsey forum on gen AI[1] said they believed that the technology will fundamentally change the way they do business. The pressing questions for banking institutions are how and where to use gen AI most effectively, and how to ensure the applications are fully adopted and scaled within their organizations.

Sidebar

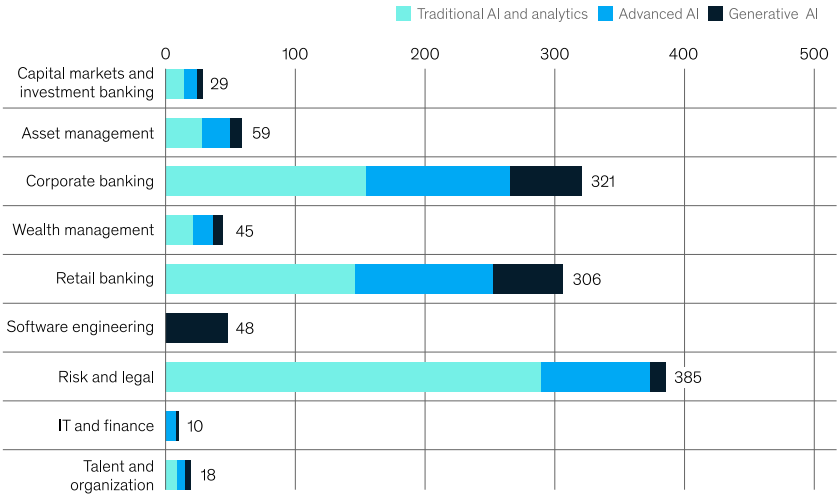
About the authors

The McKinsey Global Institute estimates that among industries globally, gen AI could add the equivalent of $2.6 trillion to $4.4 trillion annually in value across the 63 use cases it analyzed.[2] Among industry sectors, banking is expected to have one of the largest opportunities: an annual potential of $200 billion to $340 billion (equivalent to 9 to 15 percent of operating profits), largely from increased productivity (exhibit).[3] The economic impact will likely benefit all banking segments and functions, with the greatest absolute gains in the corporate and retail sectors ($56 billion and $54 billion, respectively; see sidebar "How banks are using gen AI"). (Notably, while banks have rightly focused on productivity in their initial gen AI pilots due to the broader pressure on banking economics,[4] the technology could greatly alter how some jobs are done and how customers interact with banks. It might even lead to entirely new business models.)

Exhibit

Generative AI has the potential to deliver significant new value to banks—between $200 billion and $340 billion.

Value created by AI at stake by segment and function,[1] $ billion



[1]Assumes 0% overlap of traditional AI and generative AI (generative AI assumes the lower end of value at stake), top-down estimation based on projected growth and value pools.
Source: *The economic potential of generative AI: The next productivity frontier*, McKinsey Global Institute, June 2023; QuantumBlack, AI by McKinsey traditional advanced analytics and AI analysis

McKinsey & Company

Sidebar

Share

How banks are using gen AI

For banks seeking to tap this valuable technology, a gen AI scale-up is in some ways like any other—it requires old-school change management skills, upfront senior leadership alignment and sponsorship, business-unit accountability for results, value-centered use cases, clear targets, and so on. In other ways, a gen AI scale-up is like nothing most leaders have ever seen.

Several factors explain why scaling gen AI is different. The first is the scope of the task and related implications. Just as the smartphone catalyzed an entire ecosystem of businesses and business models, gen AI is making relevant the full range of advanced analytics capabilities and applications. Executive teams are suddenly awakening to the power of AI. Almost overnight, banking leaders are having to pick their way through a thicket of once obscure terms such as *reinforcement learning* and *convolutional neural networks*.[5] But scaling gen AI will demand more than learning new terminology—management teams will need to decipher and consider the several potential pathways gen AI could create, and to adapt strategically and position themselves for optionality.

The second factor is that scaling gen AI complicates an operating dynamic that had been nearly resolved for most financial institutions. Just as banks could believe they were finally bridging the infamous divide between business and technology (for example, with agile, cloud, and product operating model changes), analytics and data rose to prominence and created a critical third node of coordination. While analytics at banks have been relatively focused, and often governed centrally, gen AI has revealed that data and analytics will need to enable every step in the value chain to a much greater extent. Business leaders will have to interact more deeply with analytics colleagues and synchronize often-differing priorities. In our experience, this transition is a work in progress for most banks, and operating models are still evolving.

Third, the pace of change has never been faster. While smartphones took many years to move banking to a more digital destination—consider that

MOST POPULAR INSIGHTS

1. Rethinking the role of the middle manager
2. The State of Fashion 2024: Finding pockets of growth as uncertainty reigns
3. How is the CHRO role changing?
4. An affordable, reliable, competitive path to net zero
5. The secrets of outperforming family-owned businesses: How they create value—and how you can become one

mobile banking only recently overtook the web as the primary customer engagement channel in the United States[6]—adoption of gen AI tools is happening in a fraction of that time. Goldman Sachs, for example, is reportedly using an AI-based tool to automate test generation, which had been a manual, highly labor-intensive process.[7] And Citigroup recently used gen AI to assess the impact of new US capital rules.[8] For slower-moving organizations, such rapid change could stress their operating models.

Finally, scaling up gen AI has unique talent-related challenges, whose magnitude will depend greatly on a bank's talent base. Leading corporate and investment banks, for example, have built up expert teams of quants, modelers, translators, and others who often have AI expertise and could add gen AI skills, such as prompt engineering and database curation, to their capability set. Banks with fewer AI experts on staff will need to enhance their capabilities through some mix of training and recruiting—not a small task.

Successful gen AI scale-up—in seven dimensions

While implementing and scaling up gen AI capabilities can present complex challenges in areas including model tuning and data quality, the process can be easier and more straightforward than a traditional AI project of similar scope. High-quality use cases can be launched in a matter of days or weeks. From our early involvement in gen AI, both for internal use (check out McKinsey's gen AI insights expert) and in our work with banks that are successfully scaling gen AI across the enterprise, we have found that delivering sustained value, beyond initial proofs of concept, requires strong capabilities across seven dimensions.

1. Strategic road map

Management teams with early success in scaling gen AI have started with a strategic view of where gen AI, AI, and advanced analytics more broadly could play a role in their business. This view can cover everything from highly transformative business model changes to more tactical economic improvements based on niche productivity initiatives. For example, leaders at a wealth management firm recognized the potential for gen AI to change how to deliver advice to clients, and how it could influence the wider industry ecosystem of operating platforms, relationships, partnerships, and economics. As a result, the institution is taking a more adaptive view of where to place its AI bets and how much to invest.

This kind of senior leadership alignment can generate strong business-level sponsorship for use-case domains. An effective strategic road map for a gen AI scale-up may also include:

- Vision, alignment, and commitment from senior leadership and business-unit-level accountability for delivering results
- A list of priority domains (functions or business units) where several related use cases can be built—each with a clear business case based on value potential and delivery feasibility (gen AI is not always the right solution; sometimes traditional analytical AI is better)
- Clear "from/to goals" that reimagine priority domains
- Assessment of enabling capabilities, including talent, agile operating model, technology, and data
- A thorough scale-up plan that sequences when and how to tackle each domain and build enabling capabilities
- A detailed partnership plan, as necessary, to potentially augment existing capabilities or acquire new ones

2. Talent

The speed of gen AI's emergence as a critical capability has left banking leaders little time to prepare for the effects on their people—and for how to upskill employees or attract the talent they'll need to keep pace.[9]

The solution starts at the top. Leaders must acquire a deep personal understanding of gen AI, if they haven't already. Investments in executive education will equip them to show employees precisely how the technology and the bank's operations connect, thereby generating excitement and overcoming trepidation.

To further demystify the new technology, two or three high-profile, high-impact value-generating lighthouses within priority domains can build consensus regarding the value of gen AI. They can also explain to employees in practical terms how gen AI will enhance their jobs.

There's also an elephant in the room: Much of the discussion on gen AI centers on the potential for automation and job losses. McKinsey's own projections see the technology enabling automation of up to 70 percent of business activities. Leaders must address these employee concerns head-on; transparency should be a priority. They can also provide clear messaging about how gen AI can automate certain tasks and manual work, improving overall productivity and employee experience.

Gen AI is also giving rise to new talent profiles. Prompt engineering and model fine-tuning were not skills on the radar of most banks' talent leaders before gen AI emerged. Few companies will have the right mix of talent out of the gate, so they need to commit to building the required roles, skills, and capabilities for the long term. The process must be continual: Some gen AI initiatives may be up and running in the near term; others may not bear fruit for a few years. Upskilling employees therefore requires a sustained approach that accounts for an evolving set of required skills and capabilities.

Banks also need to evaluate their talent acquisition strategy regularly, to align with changing priorities. They should approach skill-based hiring, resource allocation, and upskilling programs comprehensively; many roles will need skills in AI, cloud engineering, data engineering, and other areas. And as always, retaining talent means more than offering competitive pay. Clear career development and advancement opportunities—and work that has meaning and value—matter a lot to the average tech practitioner.

3. Operating model

Too often, banking leaders call for new operating models to support new technologies. But we believe that "gen AI operating model" is a misnomer. Successful institutions' models already enable flexibility and scalability to support new capabilities. An operating model that is fit for scale-up is cross-functional and aligns accountabilities and responsibilities between delivery and business teams. Cross-functional teams bring coherence and transparency to implementation, by putting product teams closer to businesses and ensuring that use cases meet specific business outcomes. Processes such as funding, staffing, procurement, and risk management get rewired to facilitate speed, scale, and flexibility.

Given how nascent gen AI is, many banks have centralized how they design and implement execution standards, allocate resources, provide access to foundation models, steer research and development, create reusable components, manage risk, and ensure alignment with overall digital and AI strategy.[10] More than 50 percent of the banks in a recent McKinsey gen AI maturity benchmark survey of US and European banks[11] had adopted a "more centralized" gen AI organization, even in cases where their usual setup for data and analytics is relatively decentralized. Whatever the degree of centralization, close and early collaboration with business teams is critical when identifying, prototyping, and deploying gen AI applications, and while integrating the models into the business flow. Involving business early in evaluating use cases can yield operational insights on high-impact opportunities, data availability, and implementation requirements. And throughout the prototyping and deployment phases, continual cross-functional dialogue ensures that models encounter and

learn from real business scenarios and uncover potential risks while unlocking the art of the possible. Soliciting continuous user feedback helps teams deliver and refine gen AI solutions that get truly embedded in decisions and workflows. Banks that foster integration between technical talent and business leaders are more likely to develop scalable gen AI solutions that create measurable value.

As the technology advances, banks might find it beneficial to adopt a more federated approach for specific functions, allowing individual domains to identify and prioritize activities according to their needs. Institutions must reflect on why their current operational structure struggles to seamlessly integrate such innovative capabilities and why the task requires exceptional effort. The most successful banks have thrived not by launching isolated initiatives, but by equipping their existing teams with the required resources and embracing the necessary skills, talent, and processes that gen AI demands.

4. Technology

Early successes in scaling gen AI occurred when banks carefully weighed the "build versus buy versus partner" options—that is, when they compared the competitive advantages of developing solutions internally with using market-proven solutions from ecosystem partnerships. Capabilities such as foundation models, cloud infrastructure, and MLOps platforms are at risk of becoming commoditized, given how rapidly open-source alternatives are developing. Making purposeful decisions with an explicit strategy (for example, about where value will really be created) is a hallmark of successful scale efforts.

For banks, navigating this maze is intricately challenging. Their history of procuring third-party IT solutions, such as databases and cloud services, has familiarized them with associated risks, but the inherent uncertainty of gen AI models presents a novel challenge. Adopting those models demands a heightened trust in vendors that might surpass banks' established risk or regulatory guardrails, potentially making them favor gen AI applications that maintain risk levels beneath a specific threshold. This limitation is something banks must carefully consider in their application and use-case decisions.

Equally, having an integrated view of the architecture that supports gen AI is critical. The new gen AI stack must be mutually reinforcing and internally consistent, not just with its different components but also with the existing legacy stack. Most banks are likely to deploy a wide range of gen AI models, each to be integrated with their existing systems, workflows, enterprise applications, and data sources. This is a critical, complex task. Effective integration and model maintenance will depend on multiple architecture components: context management and caching, policy management, a model hub, a prompt library, an MLOps platform, a risk-management engine, large language model (LLM) Ops and so on.[12]

5. Data

Gen AI's heavy reliance on unstructured data adds another layer of data-related complexity, and banks' current data strategies and architectures may not be up to the task. For example, some data migrations to cloud or third-party platforms create both constraints and degrees of freedom that must be understood clearly.[13] And while most banks have developed strong capabilities in using structured data, many have struggled to leverage the unstructured kind, largely because they lack the capabilities (such as natural language processing techniques) and infrastructure (especially computing power) to deploy the significantly more sophisticated AI models. Gen AI itself may provide a solution. Gen AI's natural language capabilities can extract insights from unstructured data like historical service interactions, social posts, news, and web pages and provide frontline bank employees with prompts that enhance their engagement with customers. The strategic deployment of tailored gen AI solutions enables financial institutions to profoundly enhance their service operations and improve the overall customer experience. Simultaneously, it facilitates the democratization of data access, unlocking the full value of unstructured data for the entire organization. Similarly, with regard to data architecture, the focus should be on developing capabilities to support the broadest set of high-value applications. Relevant capabilities, such as vector databases and data pre- and post-processing pipelines, must be built in.

Data quality—always important—becomes even more crucial in the context of gen AI. Again, the unstructured nature of much of the data and the size of the data sets add complexity to pinpointing quality issues. Leading banks are using a combination of human talent and automation, intervening at multiple points in the data life cycle to ensure quality of all data. Data leaders also must consider the implications of security risks with the new technology—and be prepared to move quickly in response to regulations.

6. Risk and controls

Gen AI, along with its boost to productivity, also presents new risks (see sidebar "A unique set of risks"). Risk management for gen AI remains in the early stages for financial institutions—we have seen little consistency in how most are approaching the issue. Sooner rather than later, however, banks will need to redesign their risk- and model-governance frameworks and develop new sets of controls.

Sidebar

Share

A unique set of risks

Responsible use of gen AI must be baked into the scale-up road map from day one. Naturally, banks encounter distinct regulatory oversight, concerning issues such as model interpretability and unbiased decision-making, that must be comprehensively tackled before scaling any application.

To reduce the risks associated with gen AI "hallucinations," which occur when models produce answers or outputs that are illogical or not based on actual data, the current approach is to loop in subject matter experts to validate model outputs. However, this process may not be scalable across all potential use cases with material value. To help subject matter experts focus their time and effort, banks are developing automation, validation methodologies, and playbooks. For example, hallucinations can be controlled in practical ways: adjusting LLM parameters' settings, such as temperature setting, which controls the randomness of the output; or setting up a post-processing first line of defense, such as automated content moderation to flag toxicity in the output.

7. Adoption and change management

How a bank manages change can make or break a scale-up, particularly when it comes to ensuring adoption. The most well-thought-out application can stall if it isn't carefully designed to encourage employees and customers to use it. Employees will not fully leverage a tool if they're not comfortable with the technology and don't understand its limitations. Similarly, transformative technology can create turf wars among even the best-intentioned executives. At one institution, a cutting-edge AI tool did not achieve its full potential with the sales force because executives couldn't decide whether it was a "product" or a "capability" and, therefore, did not put their shoulders behind the rollout.

In today's rapidly evolving landscape, the successful deployment of gen AI solutions demands a shift in perspective—that is, starting with the end user experience and working backward. This approach entails a rethinking of processes and the creation of AI agents that are not only user-centric but also capable of adapting through reinforcement learning from human feedback. This ensures that gen AI-enabled capabilities evolve in a way that is aligned with human input.

A successful gen AI scale-up also requires a comprehensive change management plan. Such a plan keeps teams engaged by including user-centered change management; incorporates training for senior leadership and employees; involves role-modeling by leaders and influencers; articulates a clear-eyed view of the expected priorities, investments, and outcomes; cogently delineates how to change mindsets and culture; and defines explicit and implicit incentives for people to use the capability. Most important, the change management process must be transparent and pragmatic.

Gen AI certainly has the potential to create significant value for banks and other financial institutions by improving their productivity. Indeed, new examples emerge weekly. But scaling up is always hard, and it's still unclear how effectively banks will bring gen AI solutions to market and persuade employees and customers to fully embrace them. Only by following a plan that engages all of the relevant hurdles, complications, and opportunities will banks tap the enormous promise of gen AI long into the future.

ABOUT THE AUTHOR(S)

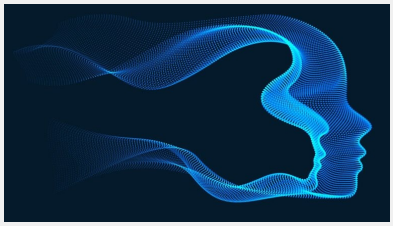
Vishnu Kamalnath is a partner in McKinsey's Boston office; **Larry Lerner** is a partner in the Washington, DC, office; **Jared Moon** and **Gökhan Sari** are senior partners in the London office; **Vik Sohoni** is a senior partner in the Chicago office; and **Shuo Zhang** is an expert in the New York office.

Talk to us

EXPLORE A CAREER WITH US

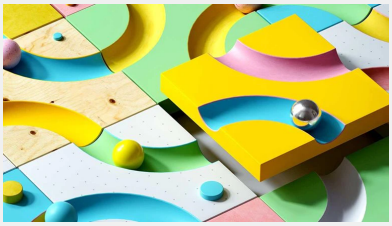
Search Openings

RELATED ARTICLES



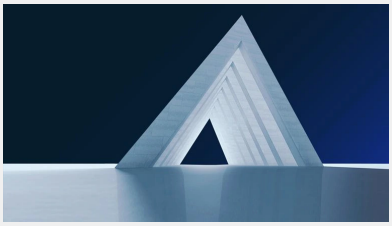
Article

Been there, doing that: How corporate and investment banks are tackling gen AI



Article - *McKinsey Quarterly*

The future of banks: A $20 trillion breakup opportunity



Article

Introduction: Building the AI bank of the future

McKinsey & Company

Subscribe

Select topics and stay current with our latest insights

Email address

Submit

McKinsey Insights - Get our latest thinking on your iPhone, iPad, or Android device.

Download on the App Store

GET IT ON Google Play