



2024 Global Outlook for Banking and Financial Markets

Regenerate banking with AI

Contents

2	Foreword
3	The game-changer: How generative AI can transform the banking and financial sectors
5	A critical moment
11	Communication: The driving force behind the customer experience
19	A reimagined workforce experience can boost productivity and re-balance costs
27	Risk and compliance: The double-edged sword
35	Manage a trusted AI platform for value, innovation, and risk
39	The 10 guiding actions



Generative AI is more than this year's buzzword. For starters, it can redefine a bank's competitive edge in client relationships, evolve and streamline core banking operations, and bolster cybersecurity.

Key takeaways

- **Almost 8 in 10 institutions (78%) are tactically implementing generative AI for at least one use case.**

Their tactical approaches vary but trend higher in the risk and compliance space, as well as in client engagement. Additionally, 8% of institutions take a broader, more systematic approach by implementing generative AI across a wider set of business domains scaling throughout the bank.

- **AI priorities reflect omnipresent concerns about risk—and client relationships.**

Almost 60% of generative AI decision makers see higher value in risk control, compliance reporting, and client engagement. Keeping data private and earning client trust is essential to winning engagements.

- **AI governance is a must-have.**

Every banker should be an AI risk manager. More than 60% of banking CEOs indicate new vulnerabilities for cybersecurity (76%), legal uncertainty related to operations (72%), difficulties in controlling outcome accuracy (67%), and prejudice from model bias (65%).

Foreword

2023 was a groundbreaking year, with large language models (LLMs) opening unexpected doors for all of us to interact with AI in personal and business life. Banking executives now find themselves both exhilarated and challenged: how can AI innovations alleviate the structural weaknesses in global banking and unlock opportunities for enhanced productivity and profit?

The *2024 Global Outlook for Banking and Financial Markets* explores the potentially far-reaching impacts of generative AI and more, starting from in-depth analysis of financial reports from nearly 2,000 institutions worldwide. This global outlook, created with insights from IBM experts and a survey of 600 forward-thinking executives, delves into the industry landscape and examines the emergence of 19 innovative use cases of generative AI.

Seldom has there been more to discuss.

Generative AI is more than this year's buzzword. For starters, it can help to redefine a bank's competitive edge in client relationships by elevating communication to new levels of personalization and effectiveness. This empowers banks to capitalize on their long-term investments in cloud and AI technologies, paving the way for customized and digitized customer interactions that add more value.

Furthermore, generative AI can assist to evolve core banking operations, streamlining them like never before, beginning with the understanding of code and process complexities. The technology harbors the potential to reshape the workforce experience, boosting productivity and efficiency.

Generative AI can be configured to bolster cybersecurity while alleviating the burden of risk and compliance management. However, it presents challenges for financial organizations as they navigate the balance between value, innovation, and risk on AI platforms that must be both open and trusted. A pragmatic AI governance approach is indispensable, guiding institutions through continuous innovation and application of this new technology. And this approach gets personal: every employee must not only be a risk manager, but an AI risk manager.

How can your organization navigate this level of seismic change? Our comprehensive action guide leads the way. The guide helps you to explore your bank's AI priorities, integrate data and AI into your core operations, and scale AI enterprise-wide.

Generative AI is undoubtedly a hot topic within your C-Suite at this very moment. The *2024 Global Outlook for Banking and Financial Markets* can inform and inspire those discussions, shaping strategies and actions that position you to take advantage of these unprecedented opportunities. Intrigued? Let us embark together on this exciting journey.

Shanker Ramamurthy

Global Managing Partner
Banking and Financial Markets
IBM Consulting

John J. Duigenan

General Manager, Financial Services,
Banking, Financial Markets, and Insurance
IBM Technology

Foreword

The game-changer:
How generative AI can
transform the banking
and financial sectors

A critical moment

Communication:
The driving force behind
the customer experience

A reimagined workforce
experience can boost
productivity and
re-balance costs

Risk and compliance:
The double-edged sword

Manage a trusted AI
platform for value,
innovation, and risk

The 10 guiding actions

The game-changer: How generative AI can transform the banking and financial sectors

Following the astonishing rise of generative AI, artificial intelligence has seized the world's attention. Executives are either dazzled by bright futures or dismayed by dystopian scenarios, and polarizing boardroom discussions proliferate. Many banking executives are brainstorming how to assess and prioritize AI's economic potential, estimate access costs, and manage the risks that come with quickly scaling AI enterprise-wide (see Figure 1).

The most essential question of the moment: how can technological innovation, such as generative AI, help address and course-correct banks' financial performance? This issue requires meticulous assessment. Financial institutions need to reflect on the development and potential of large language models (LLMs) to boost productivity, reduce customer friction, and improve the employee experience. The subsequent insights can help avoid unnecessary hype and assess the real impact of generative AI on bank business models—and define an action plan that mitigates the associated risks.

These considerations are grounded by IBM expertise in providing financial institutions with value-added consulting and breakthrough technology, corroborated by a global survey with 600 executives of financial institutions worldwide. Together, they've shaped the insights revealed in the *2024 Global Outlook for Banking and Financial Markets*.

Our perspective is organized as follows:

- **A critical moment.** Short-term revenue gains can be long-term pains as interest rate hikes factor into business and economic expectations.
- **Communication drives the customer experience.** Generative AI can tackle the core of a bank’s relationship advantage: communication with clients. This elevates decade-long investments in cloud and AI to digitalize and personalize relationships.
- **A reimagined workforce experience can boost productivity and re-balance costs.** Generative AI can facilitate core banking transformation and simplification of operating models.
- **Risk and compliance: The double-edged sword.** Generative AI challenges the risk tolerance of financial institutions and the composition of risk factors, while strengthening cybersecurity and making compliance less onerous.
- **Manage a trusted AI platform for value, innovation, and risk.** Winning the banking future by scaling AI enterprise-wide requires strong technical and cultural foundations. A practical governance of AI guides banks in the continuum of innovation and use cases application.
- **Action guide.** We outline 10 guiding actions across three domains—explore, integrate, and scale—and how to manage them.

FIGURE 1

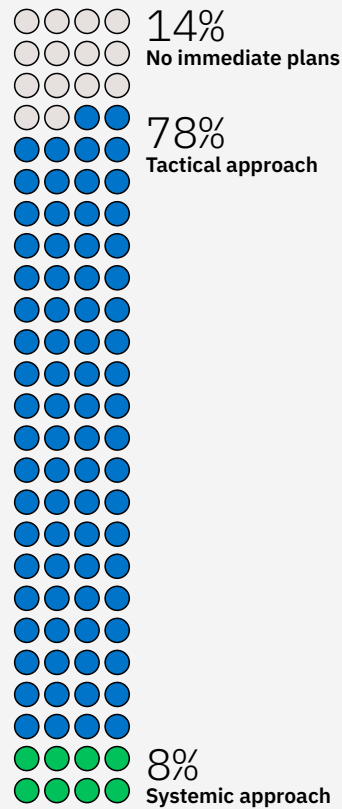
Almost 8 in 10 institutions are tactically implementing generative AI for at least one use case

8% take a more systematic, enterprise-wide approach.

86% of banking organizations are in production or preparing to go live with generative AI use cases. 8% of them are taking a systematic approach, covering all domains with one or more use cases: client engagement, risk and compliance, information technology, and other support functions. 60% of them operate primarily in other advanced and emerging economies.

On the contrary, 14% of organizations have no immediate plans to work with generative AI.

No typical starting point or implementation pattern emerges. The 78% of banking organizations with a tactical approach work on a small number of use cases or domains without significant preference. Yet, they reveal more traction in the risk and compliance space, as well as client engagement.



Q. What is your institution’s approach to implementing generative AI for each use case presented? Note: For a list of domains and use cases, please see page 44.

Foreword

The game-changer:
How generative AI can
transform the banking
and financial sectors

A critical moment

Communication:
The driving force behind
the customer experience

A reimagined workforce
experience can boost
productivity and
re-balance costs

Risk and compliance:
The double-edged sword

Manage a trusted AI
platform for value,
innovation, and risk

The 10 guiding actions

A critical moment

The years 2022 and 2023 were marked by unexpected changes in macroeconomic conditions, characterized by spiking interest rates and sustained inflation. These circumstances were fueled by a post-pandemic reassessment of economic health worldwide and the dramatic resurgence of geopolitical risks.

Deglobalization of the international supply chain ignited inflationary forces that had been repressed for over a decade by accommodative monetary policies. Major central banks, except those in China and Japan, intervened by swiftly raising interest rates to constrain inflation within functional targets. Central bankers started walking a tightrope, balancing the risk of short-term recession with the goal of promoting long-term economic growth.

The financial services industry, which has long been confined to environments of very low-to-negative interest rates, profited from these interest rate hikes. Institutions operating under negative rates saw a robust resurgence of depleted income statements. This was particularly due to the significant gap between the reset of interest earned on lending obligations and the time required to reprice borrowing costs on deposits.

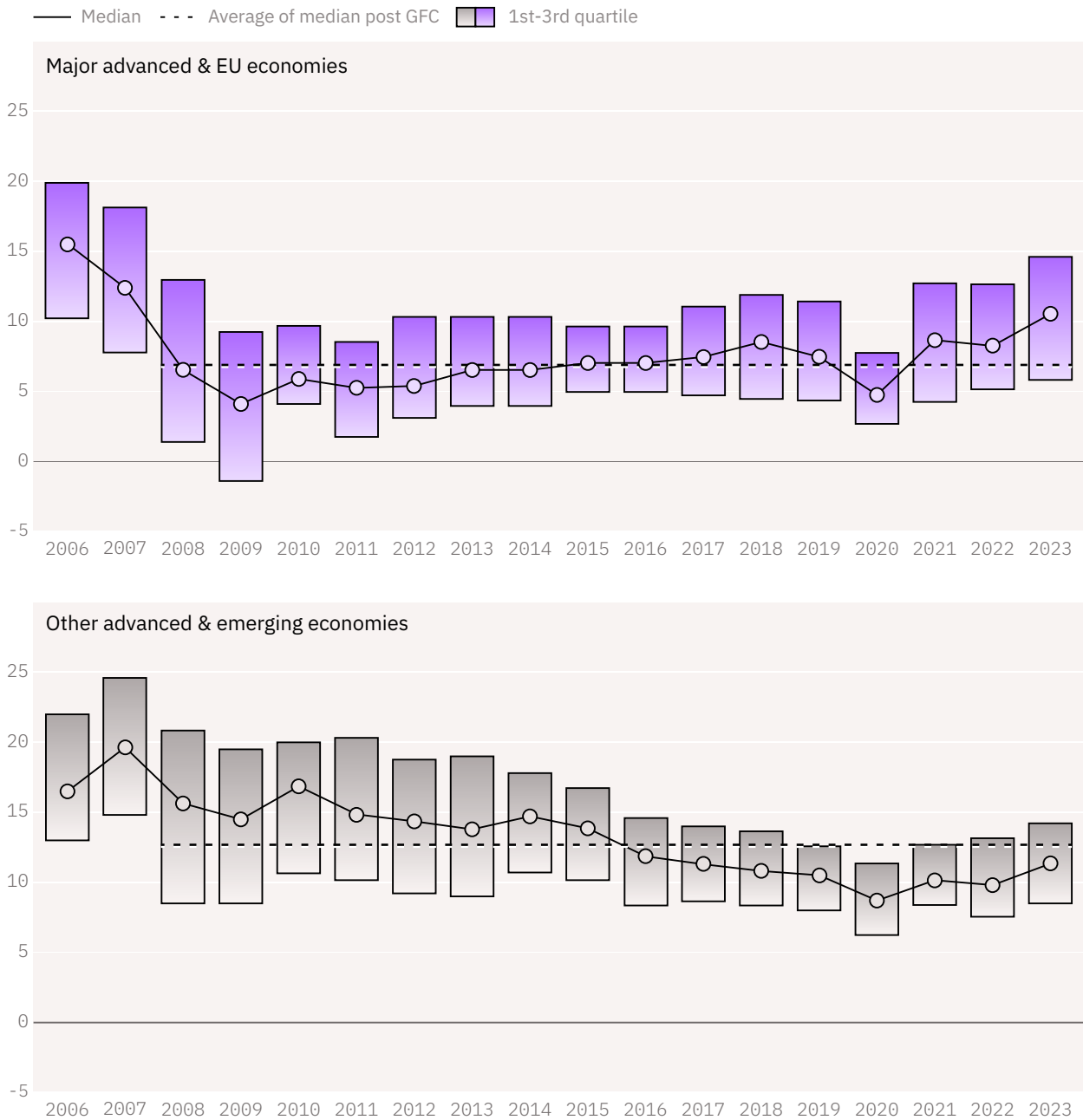
In 2023, most banks in major advanced and EU economies reported Return on Average Equity (ROAE) above the post-global financial crisis average—a key performance metric indicating how much income a bank generates per each dollar of shareholder equity (see Figure 2).¹ Conversely, the ROAE impact on banks in other advanced and emerging economies was muted, with these banks experiencing a smaller percentage increase in interest rates. These economies also emerged weaker from the pandemic: lack of sufficient post-pandemic government guarantees left local banks more exposed to a deteriorating credit cycle.

When adjusting 2023 profit margins for longer-term credit risks, the global outlook may lose its luster. A more rapid-than-expected rise in interest rates can increase profits but also adversely affect the financial position and operations of banks, prompting significant concerns relative to risk management. This new scenario is already testing client appetite for new loans, impacting banks' abilities to service existing debt, and imposing a threat to economic growth, potentially leading to a credit crunch.

FIGURE 2

ROAE: A complicated story

ROAE rises above the global financial crisis average in major advanced and EU economies—but still lags behind in the rest of the world.



Note: Represents banks in each region with total assets of more than USD 50 billion. Source: IBM Institute for Business Value analysis of S&P Global data.

According to the European Central Bank (ECB), Euro-area banks significantly tightened their credit standards (for example, banks' internal guidelines or loan approval criteria) since 2022. Simultaneously, the demand for loans from firms and households has sharply decreased, reaching an all-time low since the ECB launched its bank lending surveys in 2003.²

In the US, average credit card balances per cardholder topped \$6,000 amid persistent inflation, the highest in 10 years, according to the Federal Reserve Bank of New York analysis. The delinquency rate increased to 3% in Q3 2023, correlating with an all-time high credit card balance of \$1.08 trillion in aggregate. The increase was particularly pronounced among millennials and those with auto or student loans.³

As discussed in our *2023 Global Outlook*, the current spike in rates is unlikely to herald a return to steadier and healthier profit, given the vastly different macroeconomic conditions and weak efficiency ratios of most banks. The cost-to-income ratio (CIR)—a measure of operational efficiency—remains elevated in major advanced and EU economies (see Figure 3). In 2023, the median is 12% higher compared to the rest of the world (excluding China). When China is included, the gap grows to 16%.⁴

The result can be observed in capital markets, which were quick to discount the economic weaknesses of traditional banks' business models. Since the onset of the global financial crisis in 2008, investment analysts haven't championed banking stocks, pushing down banks' price-to-book ratios (PBRs)—a key metric of bank franchise value (see Figure 4). PBRs trend below 1 for many institutions, suggesting investor apprehensions about shareholder value, leading to higher capital costs for banks that opt to issue additional equity.⁵

Among major advanced and EU economies, only Canadian and US banks boast a PBR higher than 1.⁶ These banks benefited from faster restructuring and higher interest rates after the global financial crisis erupted.

PBRs trend below 1 for many institutions, suggesting apprehensions about shareholder value, leading to higher capital costs for banks that opt to issue additional equity.⁵

FIGURE 3

Persistently high CIRs

CIRs remain elevated in major advanced and EU economies.

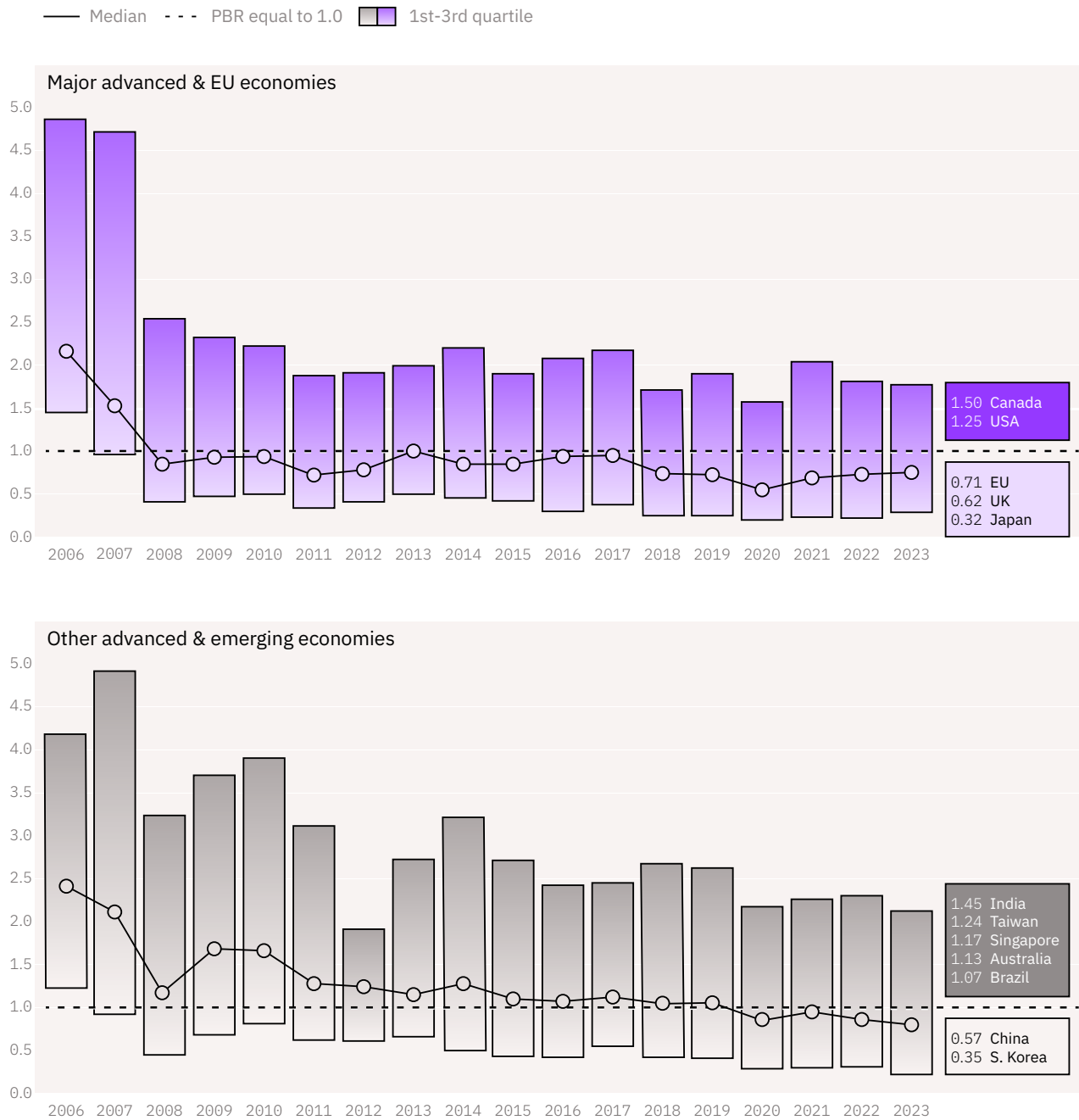


Note: Represents banks in each region with total assets of more than USD 50 billion. Source: IBM Institute for Business Value analysis of S&P Global data.

FIGURE 4

PBRs trend lower

PBRs have fallen below 1 for many institutions.



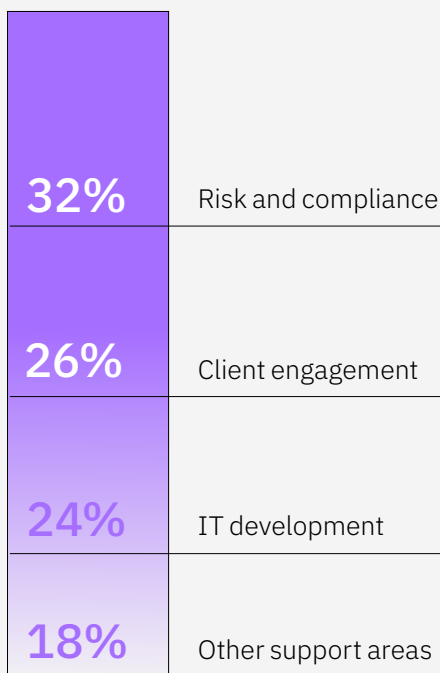
Note: Countries depicted have banking systems with more than USD 1 trillion total assets in aggregate. Source: IBM Institute for Business Value analysis of S&P Global data.

To regain investor favor, financial institutions need to demonstrate that above average financial performance is not extemporaneous. It must correspond to structural changes in business models and related investment in technology. However, many banks were hesitant to pursue deeper transformation in the last decade. Instead, they've attempted to digitalize existing methods and operations without a sufficient business and technological redesign. The 2019 adage of Mario Draghi, former Chair of the European Central Bank, still rings true: "The necessity to adjust the business model to the digitalization, to the changes in technology, is something much more compelling than being angry about negative rates."⁷

While not a novel concept, it's worthy of emphasis: banks risk losing even more ground as the landscape starts to shift more quickly than expected. The acceleration of AI, particularly following widespread access to LLMs, has propelled generative AI to the forefront of boardrooms and executive-level conversations.

This year's *Global Outlook* presents an opportunity to reflect on the structural weaknesses of global banking and how recent advances in AI can unlock new possibilities for improved performance and efficiency. And, we explore how to scale AI enterprise-wide while addressing risk and compliance concerns.

Perspective



Q: Rank the banking functions above in which generative AI unlocks the greatest value for your institution.

Risk, compliance reporting, and client engagement lead the way in AI potential

We surveyed 600 banking executives with decision-making responsibility for AI from strategy to operations. They reflected on where to find the greatest business value for 19 generative AI use cases (see "Domains and use cases" on page 44) across four business domains. The top use cases (32%) correspond to applying the technology to risk and compliance challenges and opportunities, followed by improving client engagement (26%). Our survey reveals how banks are acting on each use case by domain.

- How many banks are live or about to go live? The result may surprise you.
- How many are running a pilot or plan to start one this year? Find out what's happening.
- How many assessed the feasibility and decided not to proceed? Not everyone is ready.
- And how many have not yet formulated a generative AI approach? Generative AI is not a one size fits all.

See Perspectives on pages 16, 24, and 33 to learn more.

Foreword

The game-changer:
How generative AI can
transform the banking
and financial sectors

A critical moment

Communication:
The driving force behind
the customer experience

A reimagined workforce
experience can boost
productivity and
re-balance costs

Risk and compliance:
The double-edged sword

Manage a trusted AI
platform for value,
innovation, and risk

The 10 guiding actions

Communication: The driving force behind the customer experience

Enhancing the customer experience ranked as a top priority for 53% of the financial services CEOs surveyed by the IBM Institute for Business Value (IBM IBV) in 2023, second only to cybersecurity (58%).⁸ In an era of increasing digital disintermediation, a bank's capability to sustain a competitive level of customer experience is fundamental.

Clients, long accustomed to managing a large part of their personal and business lives on mobile apps, are reshaping the landscape. We asked 12,000 consumers where they prefer to deposit their salary and keep their savings.⁹ Already, 16% of consumers globally are comfortable with a branchless, fully digital proposition—with that figure significantly higher in other advanced and emerging economies. For example, in Brazil, 29% of respondents hold a primary account with a neobank, defined as a firm offering applications, software, and other technologies to streamline mobile and online banking.¹⁰

When asked how they prefer to apply for loans, execute investment decisions, or buy/renew car insurance, the majority of consumers cite mobile apps and websites as their most favored option: 63% for loan applications, 69% for investment execution, 56% for car insurance. When selecting a car insurer, a positive claim experience was 44% more likely to sway decisions toward a specific insurer compared to other drivers, such as brand reputation or comparing coverage.¹¹

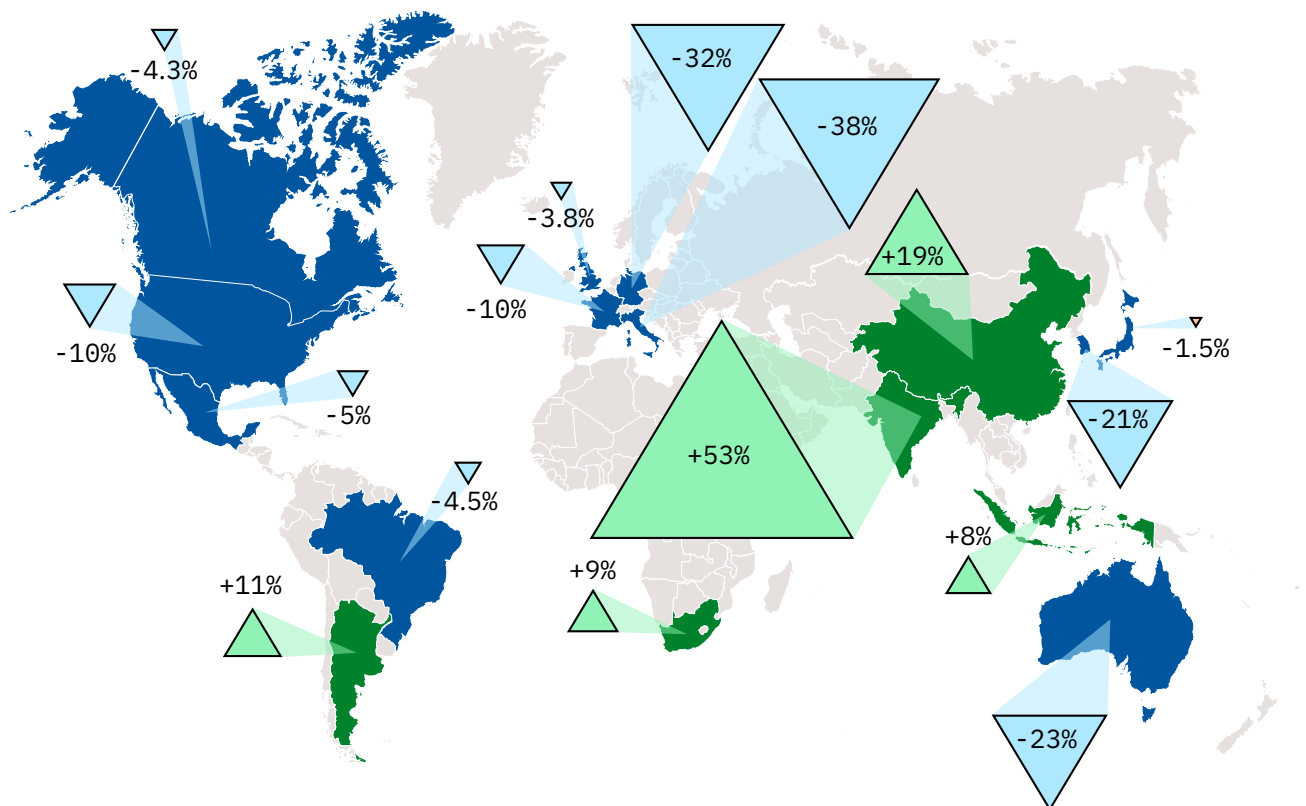
Clients seek seamless integration between digital and analog life, demanding a consistent omnichannel experience characterized by real-time, comprehensive personal relationships. When contemplating switching their main bank accounts, the quality of customer service is a significant factor. In fact, there was a 20% greater chance this feature could influence their decision when compared with other elements, such as better access to mobile apps or branch proximity.¹²

Although branches still play a crucial role in banking intermediation, their number has been shrinking since 2012, particularly in major advanced and EU economies (see Figure 5).¹³ This is primarily due to intense M&A activity, but it's also hastened by declining demand and revenue pressures, notably resulting from the low interest rates over the last 10 years. Instead, branch networks are growing in other advanced and emerging economies, along with economic wealth, banks' earning capacity, and the provision of services to previously unbanked citizens and regions.¹⁴

FIGURE 5

The rate of commercial bank branch openings—and closures

Branches are declining since 2012, particularly in major advanced and EU economies.



Note: Data for Germany is 2020, not 2021.
Source: IBM Institute for Business Value analysis of S&P Global data.

Although higher interest rates are boosting global banking profits, a widespread return to branch network expansion is unlikely, given client appreciation of the convenience and efficiency of digital banking. How have banks responded to clients' increasing digital adaptation, and the need to shift more, better, and reimagined experiences from branches to mobile contact points? As exponential technologies mature, banks have addressed three strategies supporting better customer service. Each phase corresponds to a different investment focus in their tech portfolio (see Figure 6).

Digitalization as access

Banks initially introduced online journeys to supplement branch access, coexisting without differentiating key propositions. With the advent of smartphones, digital banking matured as a primary engagement platform. We asked 12,000 consumers about their preferred method to perform basic banking transactions, such as accessing their bank account and verifying balances and transactions. 62% already said they're using a mobile app and 12% report using a bank website.¹⁵

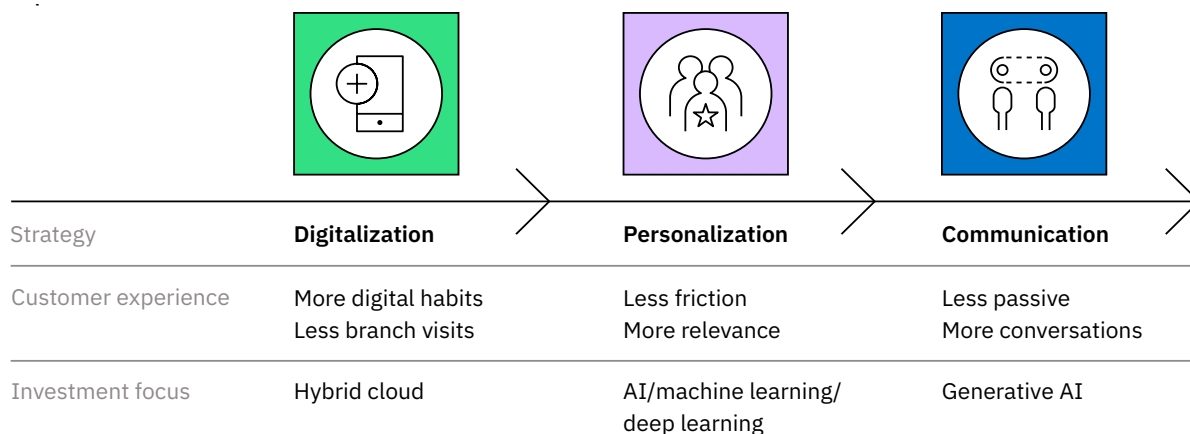
However, interfaces for basic banking services were built and optimized without resolving core banking complexities, limiting the capability to better satisfy digital requirements of the clientele. This hindered banks' effectiveness in engaging clients based on core banking transactions data, even when enriched by other sources of information.

As digital banking grew in usage and relevance, cloud technology became essential to elevating the omnichannel experience, providing advanced analytics for timely insights into client behaviors. Notably, the more open architecture of hybrid cloud underscores the vital role of innovation and the need for easier integration of the fintech ecosystem to shape a more responsive, customer-centric digital banking landscape.

FIGURE 6

Evolving with exponential tech

Three technology pillars to build better customer service



Source: IBM Institute for Business Value.

FIGURE 7

“High touch” service equates to high NPS

This is reflected in a bank’s Net Promoter Score (NPS)—a measure of how likely customers are to recommend a bank’s products and services. A high NPS indicates stronger satisfaction in customer relationships, potentially resulting in more business per client and growth via referrals.

According to Survey Sensum in 2023, specialized retailers and smartphone producers boast the highest NPS (see Figure 7). Specialty stores typically engage in “high touch” personalized customer interactions. And superior customer services usually complement the client-centric focus of high-tech gadgets. On the low end of the spectrum are utility services, such as internet and cable providers. This might indicate a gap in addressing key friction in client consumption of highly commoditized services.¹⁶

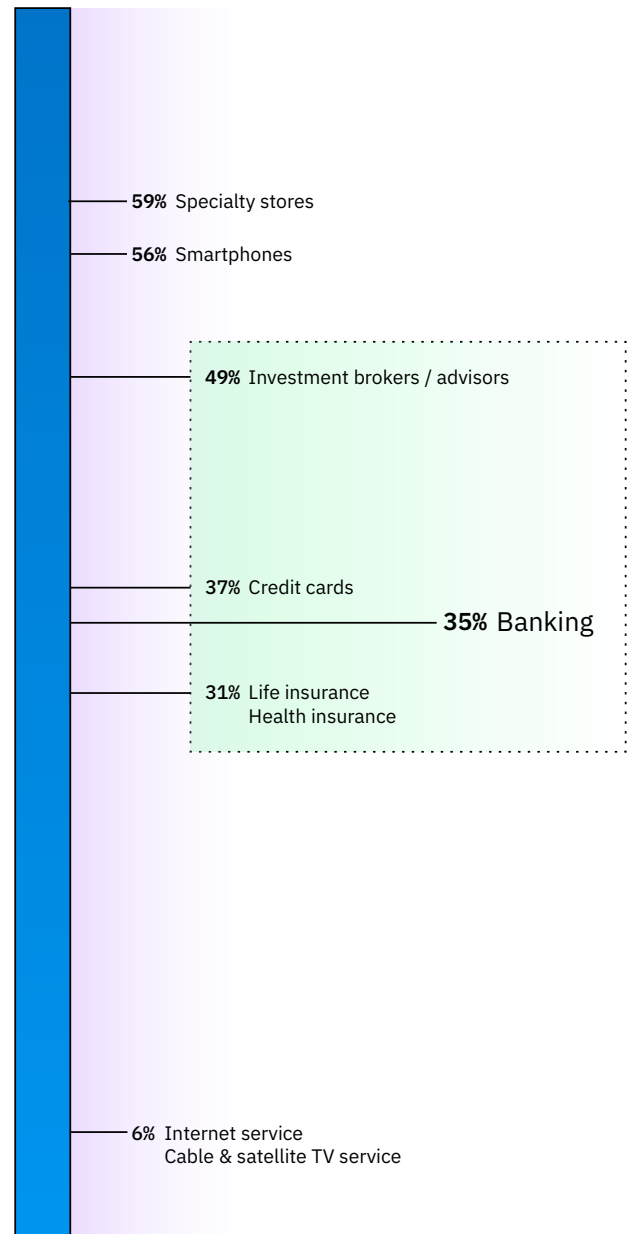
And what about financial services? Banking, together with life and health insurance, has room to improve, while investment brokers/advisors and credit cards rank higher. This reflects greater alignment with customer needs.¹⁷

Personalization as relevance

Providing digital access doesn’t automatically translate to personalized, frictionless digital services. The emergence of AI—machine learning, deep learning, and now generative AI—has granted banks new capabilities to reimagine the customer experience.

By leveraging AI processes in search engine optimization (SEO), banks can identify key frictions across the digital landscape, learning more from each client interaction and redirecting actions to improve NPS. Mobile apps can’t capture the same soft information about clients as human branch managers. But machine learning offers new avenues to extract client preferences and needs from their transactions.

Banking has room to improve.



Source: *The Role of NPS in Banking and Other Financial Institutions*. Survey Sensum. 2023.

Notwithstanding the advantages, better insights and new information alone failed to drive a comprehensive shift in client intermediation toward digital solutions. Mobile, being demand-driven, is effective for users who can self-direct, but most banking revenues originate from offers. Bank officers typically push products to clients who rely on them to make informed financial decisions (for example, investing, insuring, or funding for business ambitions).

Without sufficient conversational touchpoints, many clients might pivot to digital banks before resolving important considerations to financial decisions. For instance, while 69% of consumers prefer to execute investments online, there is a 29% greater chance that financial advisors trigger investment decisions compared to other factors, such as access to planning tools, mobile support, or timely news.¹⁸

First-generation chatbots, powered by natural language processing (NLP), empowered banks with rudimentary conversational capabilities to assist clients—but rarely provided an adequate conversational experience. This limitation restricted their capability to handle client relationships in a confined, rule-based set of domains.

As generative AI matures, this barrier is starting to dissolve. Not only can generative AI add value in processing information for better categorization, boosting the quality of customer service (see case study, “Large global payments company turns complaints into actionable insights with generative AI”), but LLMs can empower banks to handle digital conversations that better facilitate value-added relationships.

While chatbots answer questions, generative AI engages clients and employees of financial institutions—it works for them by assisting them in managing their tasks.

Case study

Large global payments company turns complaints into actionable insights with generative AI¹⁹

A global payment company sought to improve its customer service processes leveraging generative AI. Previously, the company manually categorized millions of customer complaints across thousands of products into broad groups. This complaint analysis process was time-consuming, error-prone, and unresponsive to emerging issues. Producing actionable insights required a full three weeks—inadequate for trend spotting. This reduced customer satisfaction and prompted the need for compliance reviews.

By adopting a generative AI model with more than 100 million parameters, the company was able to overhaul this inefficient process. The model was trained on public banking data sets and integrated securely into the company’s infrastructure. Reinforcement learning with human feedback (RLHF) was employed to align the model to proprietary context and products.

This new process cut the time to complete complaint analysis from three weeks to under 15 minutes, at the same time improving the granularity of categorization with 91% accuracy. Furthermore, generative AI enabled the extraction of keywords, summarization of call notes, and identification of intents—leading to higher-quality customer service.

Communication as a service

Can generative AI make digital platforms the core channel to steward client decision-making?
Can mobile technology shift from a demand-led method to an offer-oriented engagement in financial services?

The revolution of generative AI has the potential to infuse human language proficiency into machine-driven interactions, elevating the quality and relevance of engagement. This stems from its capability to create content based on analyzing large amounts of data, including text, image, video, or code. Content can be summarized, and questions answered in chats, all consumable via an AI-generated persona.

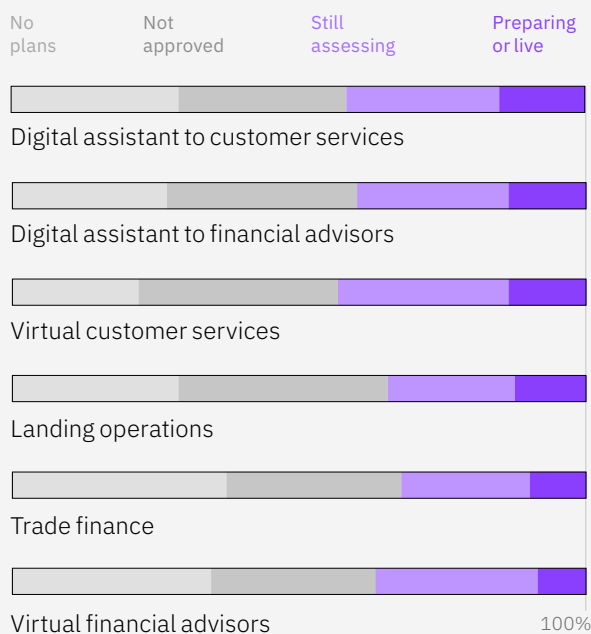
It's no surprise that generative AI can find timely application to customer services. For instance, NatWest has invested in elevating the quality of Cora—the bank's virtual assistant—with LLMs (see case study, "NatWest's Cora integrates generative AI to enhance customer experience" on page 17). Based on next-best actions, responsive technology reduces the gap between forging client intentions with executing financial decisions.

Alternatively, private bankers can also augment their capability to reach out to clients with high-end content delivered in unconventional ways. Generative AI models have been designed to assist a leading European private bank in transforming investment research and financial information into podcasts that high-net-worth individuals can consume on their mobile devices (see case study, "Generative AI transforms a leading European private bank's research reports into podcasts" on page 18). Added-value content can be personalized to the preferences, interests, and sentiments of each client.

Banks are not only looking at client touch points to improve business experiences. Many are equally invested in implementing generative AI to reimagine the internal operational experience of employees. For example, conversational capabilities can assist bankers in executing the full lifecycle of software development. As well, generative AI can boost efficiencies and processes in fields such as human resources and administrative requests. For instance, IBM deploys an AI HR assistant to free up a significant amount of time in the HR function.²⁰

Perspective

Digital assistants to client-facing workforce are the most common starting point.



Q: For the above client engagement use cases, what is your institution's approach to implementing generative AI?

By focusing on customer services and financial advice, banks can consider two co-existing approaches. They can leverage generative AI to fully automate client conversations powering chats with virtual agents. And they can augment the workforce of human agents to deliver faster and more relevant aid to clients.

Among banks that are in production or preparing to go live with these use cases, digital assistants to the workforce are preferred, compared to fully virtual services. With reference to virtual financial advice, this might reflect higher complexity and greater risk of hallucination as revealed by our group of experts (see page 37).

There is no single pattern in banks' approaches to use cases. Statistical evidence indicates that bankers see use cases as fairly independent from one another. A bank with "no plans" for one of the listed use cases might be "preparing or live" for another one of the same domain.

Case study

NatWest's Cora integrates generative AI to enhance customer experience²¹

NatWest will deploy generative AI via its virtual assistant Cora to provide customers with access to a wider range of information through conversational interactions. This innovative capability has been designed to provide a more accessible, human-like interaction for customers looking to compare products and services across the product suite, or who research information across the NatWest Group websites.

According to Wendy Redshaw, Chief Digital Information Officer of the NatWest Group's Retail Bank: "We are a relationship bank in a digital world, building trusted, long-term relationships with our customers through meaningful and personalized engagement. Building on Cora's success over the last five years, we're working [...] to leverage the latest generative AI innovations that will help make Cora feel even more 'human' and, most importantly, a trusted, safe, and reliable digital partner for our customers."

Generative AI will be used to access information from multiple secure sources previously inaccessible through chat alone, such as products, services, information about the bank, and career opportunities. As a result, customers can ask questions and receive responses in a more natural, conversational style and are also provided with links to requested information.

Generative AI will be used to access information from multiple secure sources previously inaccessible through chat alone.

Case study

Generative AI transforms a leading European private bank's research reports into podcasts.²²

In today's digital age, information is power. Clients express new preferences to stay up to date with current events and opportunities impacting their financial well-being. Relevant information must be communicated at the time of client interest, wherever and whenever needed.

As larger chunks of daily routines are conducted on mobile devices, clients increasingly prefer listening to podcasts over reading traditional reports. Sifting through these reports and curating relevant information is time-consuming for the client.

A leading European private bank identified this shift as a strategic opportunity to accelerate the development of personalized content for its clients. The company is currently testing the use of generative AI to transform research reports into captivating podcasts in a client-centric way.

Private bankers will be able to upload research reports and generate podcast summaries catering to specific client interests. Generative AI proves to be an advantageous technology for promoting inclusion and accessibility to clients with visual disabilities. The system uses a two-step approach—providing control and verifiability—to arrive at final communication.

A leading European private bank is currently testing the use of generative AI to transform research reports into captivating podcasts in a client-centric way.

Foreword

The game-changer:
How generative AI can
transform the banking
and financial sectors

A critical moment

Communication:
The driving force behind
the customer experience

A reimagined workforce
experience can boost
productivity and
re-balance costs

Risk and compliance:
The double-edged sword

Manage a trusted AI
platform for value,
innovation, and risk

The 10 guiding actions

A reimagined workforce experience can boost productivity and re-balance costs

Banks' investments in technology are not confined to user interfaces but also address the need to redesign architectures end to end. These investments account for 14% of a total worldwide expenditure in technology that exceeds \$4.5 trillion, according to 2023 analysis by Gartner.²³ In 2022, technology expenses in the banking industry averaged 7% of total operating expenses (see Figure 8), and this percentage varied across different regions (see Figure 9).²⁴ Besides technology, total *operating* expenses include workforce, real estate, and other expenses.

Yet, correlating the impact of spending in banking technology to financial performance is not easy. More so than in other industries, traditional business models strongly depend on macroeconomic conditions, and regulation more clearly defines the spaces of application. Most importantly, the total amount spent does not singularly determine positive or negligible outcomes. Rather, the *utilization* of technology expenses holds greater significance. For example, effective technology spend can increase the value of services that motivate client willingness to pay for prime experiences, protect economic value from cyberattacks, and embed ecosystem interactions into nonbanking platforms.

Contrary to assumptions, research commissioned by the ECON Committee of the European Parliament reveals some banks with lower IT spend outperformed their higher-spending counterparts, emphasizing the importance of efficient IT utilization.²⁵ Additionally, CIR dynamics (see Figure 3 on page 8) indicate many banks struggled to translate efficiency-oriented initiatives into structural gains. Similarly, investments in innovation did not achieve expected productivity outcomes.

While total operating expenses have been growing in the last 15 years, the percentage directed to technology changed less than other expense domains. For instance, workforce costs increased by almost 5% of the total amount (see Figure 8).²⁶

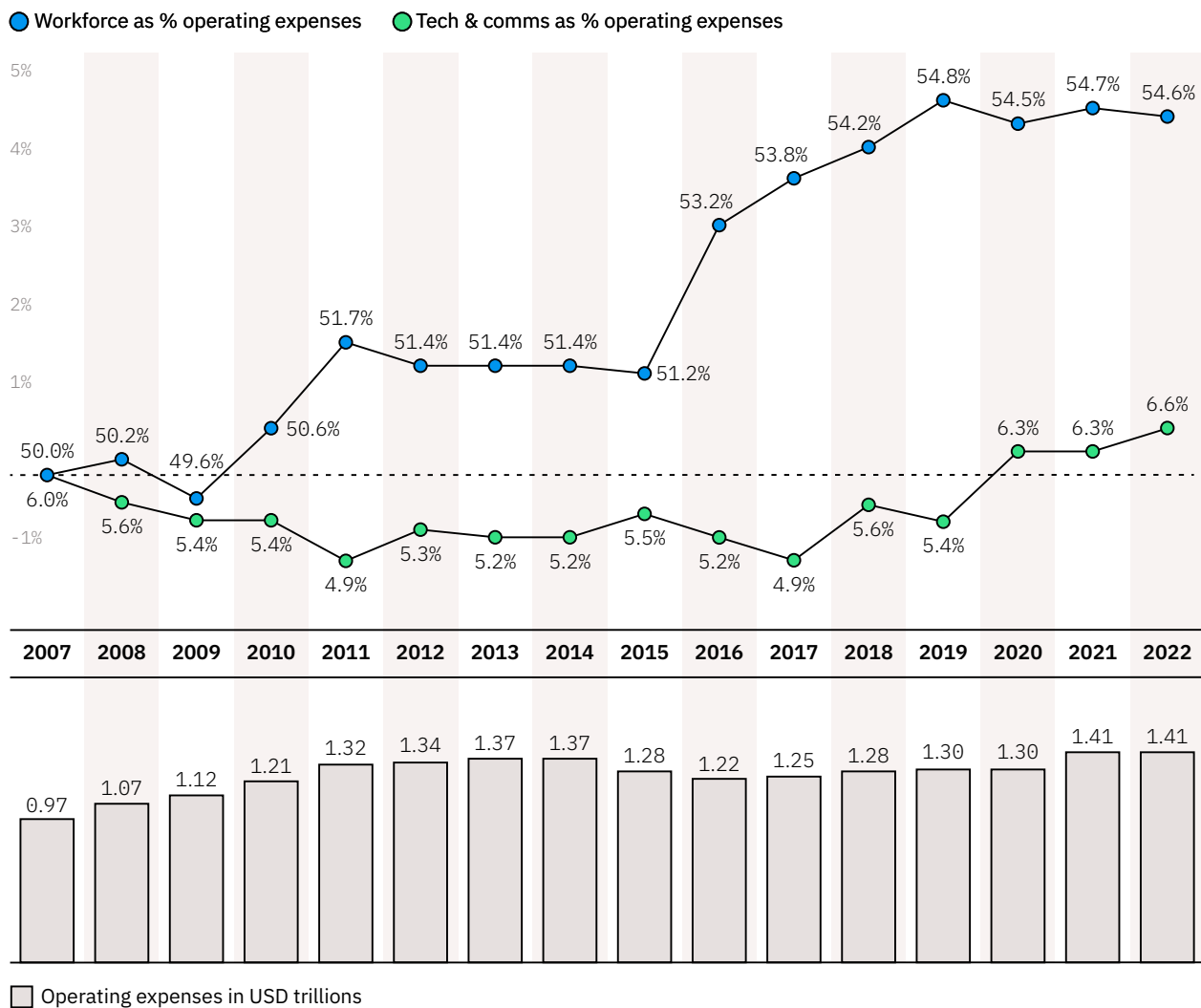
To significantly impact financial performance, technology investment must concurrently address automation and augmentation to re-balance bankers' contributions to the bottom line and generate more business value per workforce unit.

The recent advancements in AI significantly expanded the breadth of its capabilities. While the precise impact on productivity is yet to be realized, early signs indicate the potential for this to be a game changer. Yet fully exploiting the benefits is likely to require a very human response: a collaborative effort from industries and regulators, and a wholesale reimagining of business models, and workflows.²⁷

FIGURE 8

Evolution of workforce and tech expenses since 2007

Total operating expenses have increased, but the percentage of operating expenses directed to technology changed less than other expense domains.



Source: IBM Institute for Business Value analysis of S&P Global data.

FIGURE 9

Tech & comms expense ratio by buckets across regions

Institutions in other advanced and emerging economies tend to invest a smaller portion than banks in major advanced and EU economies.

Tech & comms expenses over total operating expenses	Percentage of banks with this ratio	
	Major advanced and EU economies	Other advanced and emerging economies
20 - 25%	6%	3%
15 - 20%	17%	4%
10 - 15%	32%	20%
5 - 10%	17%	20%
0 - 5%	28%	53%

Source: IBM Institute for Business Value analysis of S&P Global data.

The potential of generative AI, automation, and increased productivity

Against this backdrop, the ascent of generative AI offers new potential for transforming complex, content-driven applications. Its primary objective: augmenting the workforce and delivering better customer services per unit price, more informed advice per conversation, and accelerated time-to-market of application development.

In a recent IBM IBV survey on embedded finance, we identified a constraint in banks' capabilities to achieve return on investment (ROI). Bankers experience a gap in how they design services to meet client needs in nonbanking ecosystems and the needs of their digital partners. Nonbanking executives highlight inadequate operational support and slowness to develop APIs.²⁸

Yet, a majority of banking executives don't consider these elements critical factors to succeed in ecosystem-oriented partnerships. When asked what hinders banks' capabilities to open architectures that embed their products and services in external digital platforms, 72% of executives were still more likely to identify internal roadblocks such as lack of modularity in core banking systems and suboptimal API standards.²⁹

Complex monolithic processes persistently impede banks' capabilities to redesign value streams for a more dynamic cloud consumption, resulting in a convergence of process debt and technical debt. This is precisely where generative AI emerges as pivotal to achieving full transformation of legacy IT automating systems. It starts from understanding and documenting code that already exists, buried under complex processes that lost causality.

Expediting software development with generative AI

Generative AI tools, when integrated into a developer's process, can document pre-existing code, reverse-engineer requirements, and write new ones. Based on user input, generative AI tools can produce high-quality code recommendations.

Auto-generated code suggestions can then increase developer productivity by providing straightforward answers and handling routine coding tasks, reducing the need to "context switch" and conserving mental energy. These code development tools can also help identify coding errors and potential security vulnerabilities.³⁰

Using AI code generation software is generally straightforward and available for many programming languages and frameworks, and it's accessible to both experts and junior developers. These tools offer three broad benefits. Typically, they can:

- Enable developers to generate code faster, reducing the manual work of writing lines of code and freeing developers to focus on higher-value work
- Test and debug computer code quickly and efficiently
- Make advanced code development accessible to junior developers.³¹

With modernization proving lengthy and more costly than banks estimated, generative AI can also translate legacy code base (expensive to maintain) into newer software languages while driving new ways of working on application modernization. German core banking provider Finanz Informatik has demonstrated this well (see case study, "Finanz Informatik is on a modernization journey of its core banking platform" on page 25).

The impact of generative AI on the skills gap

In a banking world eager to embrace technology yet facing skilled labor shortages, the application of generative AI can reduce the skills gap—and enhance competitiveness.

In the absence of a comprehensive plan for redefining workforce engagement, improved efficiency doesn't translate immediately into operational savings. Realizing the full potential of AI-driven efficiencies necessitates not only reskilling but also adopting a new talent mix.

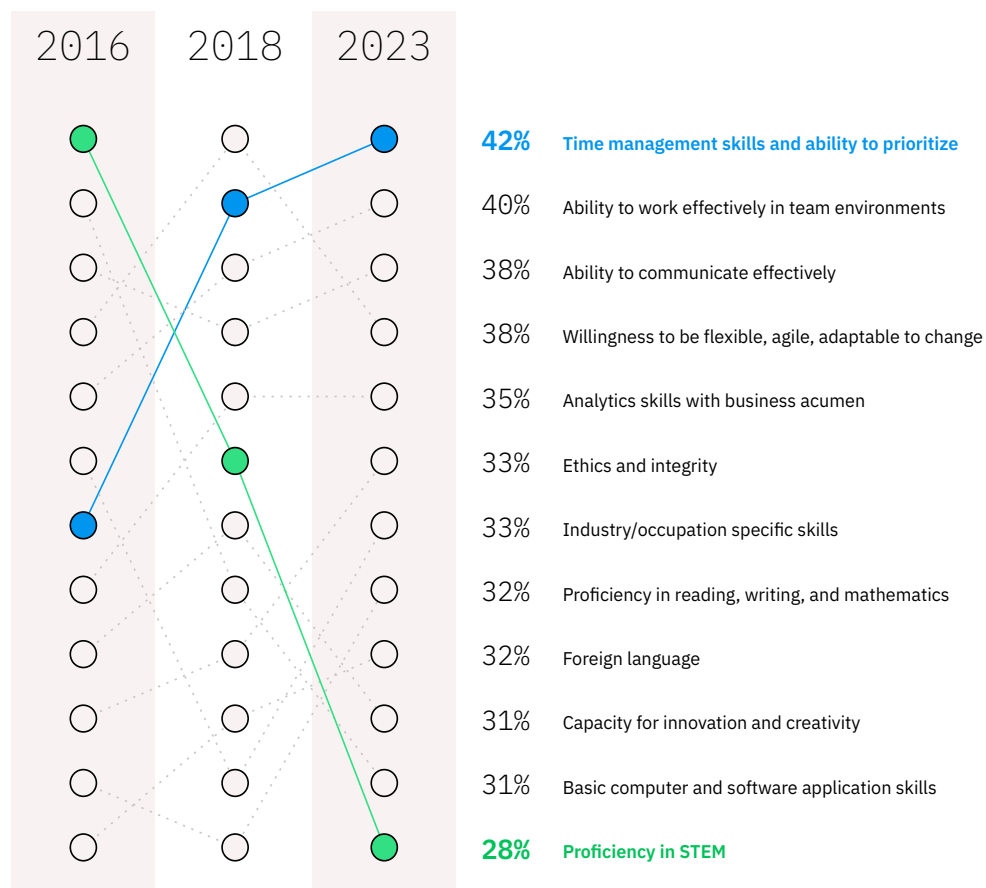
Crucially, unlocking the value of new skills and talent requires a change in the way people work, co-create, collaborate, and execute. Before fully unlocking the potential of STEM skills, institutions must first establish priorities and untangle organizational complexities—endeavors that require a strategic approach. That could be one reason why STEM and IT skills have decreased in importance since 2016 (see Figure 10).³²

AI applications are not immune from risks and compliance hurdles. In the age of generative AI, banks' approaches to data and privacy must be rethought, with considerations such as EU GDPR and IP protection assumed as givens.

FIGURE 10

Evolving priorities in skills and talent

Establishing priorities and untangling complexities require strategic skills.



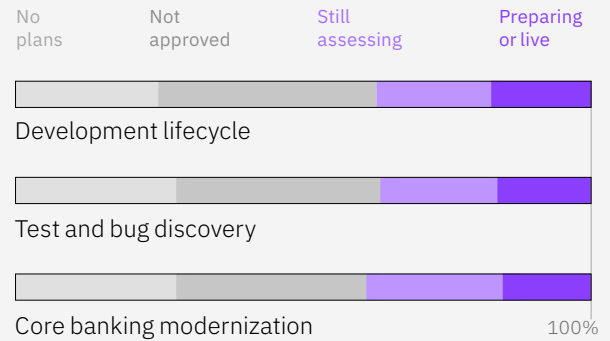
Source: *Augmented work for an automated, AI-driven world: Boost performance with human-machine partnerships*. IBM Institute for Business Value. August 2023.

Perspectives

Generative AI is gaining momentum in IT development.

IT development lifecycle is a growing area for immediate application of generative AI. This is being addressed by 17% of institutions globally.

There is no single pattern in banks' approaches to use cases. Statistical evidence indicates that bankers see use cases as fairly independent from one another. A bank with "no plans" for one of the listed use cases might be "preparing or live" for another one of the same domain.

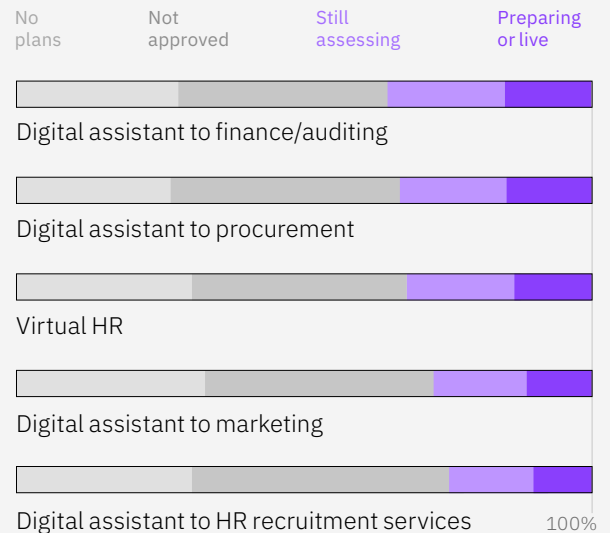


Q: For the above use cases, what is your institution's approach to implementing generative AI?

Finance and auditing applications of generative AI attract the greatest attention from banks.

Over the last year, 52% of financial institutions executed feasibility studies in support areas such as human resources, finance, and procurement. This equals the percentage of studies conducted in IT development (52%), and it's followed by 48% in risk and compliance, and 48% for customer engagement.

There is no single pattern in banks' approaches to use cases. Statistical evidence indicates that bankers see use cases as fairly independent from one another. A bank with "no plans" for one of the listed use cases might be "preparing or live" for another one of the same domain.



Q: For the above human resources, marketing, and other operations use cases, what is your institution's approach to implementing generative AI?

Case study

Finanz Informatik is on a modernization journey of its core banking platform³³

Finanz Informatik embarked on a journey to modernize its central core banking platform OSPlus, a pivotal system supporting over 350 financial institutions. Transitioning from a COBOL-coded structure to a Java-based architecture could prove challenging, yet it was deemed essential to interoperate seamlessly on both IBM z/OS® mainframe and Linux® platforms. This initially involved the mapping and prioritization of business processes, helping ensure clarity in the transition from the legacy system to a new platform.

The complexity extended beyond technology, entailing a concurrent skill transformation across the organization, leading to a better developer experience and acquiring young talent in the market.

The first investment prepares Finanz Informatik for the potential future introduction of generative AI to assist and further accelerate development design and execution, while reducing risk for the end-to-end lifecycle. An LLM calibrated to banking-specific coding domains is in discussion. This LLM could support development, from application discovery and analysis to automated code refactoring, and assistance in the COBOL to Java conversion. The first investment prepares Finanz Informatik for the introduction of generative AI to assist and further accelerate development design and execution.

The first investment prepared Finanz Informatik for the introduction of generative AI to assist and further accelerate development design and execution.



Foreword

The game-changer:
How generative AI can
transform the banking
and financial sectors

A critical moment

Communication:
The driving force behind
the customer experience

A reimagined workforce
experience can boost
productivity and
re-balance costs

Risk and compliance:
The double-edged sword

Manage a trusted AI
platform for value,
innovation, and risk

The 10 guiding actions

Risk and compliance: The double-edged sword

Sound financial institutions navigate financial and economic uncertainties for and with their clients—and effective risk management is a fundamental tenant. Banking profits rely on adept risk management. This inherently makes every banking employee a risk manager, responsible for generating sustainable risk-adjusted value—whether in approving loans, offering investment advice, or managing IT operations.

A key challenge is to risk-manage the complexity of operations. Since the introduction of interest rate derivatives in the 1980s, when interest rates skyrocketed, financial innovation took hold at an accelerated pace, opening a gap between risk management capabilities and IT responsiveness. Investments in technology were often a catch-up game, with business requirements lacking an enterprise-wide perspective and consistent architectural design.

The result: banks not only layered new technology onto weak architectural foundations; they ended up developing more tactically than strategically and building siloes. The more banks innovated, the more the complexity of their back- and middle-office operations increased, limiting their capability to reign in CIR.

The pace of “financial innovation” faltered in the aftermath of the financial crisis, following regulatory requirements for the simplification and central clearing of most derivative contracts. And banks didn’t fully grab the opportunity to swiftly simplify spaghetti-like back- and middle-office architectures. With the advent of “fintech innovation”—which focuses less on contractual payoffs and more on transformed client engagement with technology—banks realized a profound architectural gap in what makes their operations fit for an open digital world.

As many institutions were only midstream in modernizing their core environments, banks' capabilities to transform at speed have encountered another roadblock. More than 50% of banking executives reported insufficient API standards and insufficient modularity in core banking systems as the main challenges to compete with ecosystem-oriented integration strategies such as embedded finance.³⁴

The stratification of operational risks and constraints on both the business and the technology sides makes software development more expensive—and much slower—for banks compared to their Silicon Valley competitors and partners.

With its ability to create, test, and deploy code, generative AI can revitalize software development, which is what banks need at scale. However, an accelerated development experience, grounded on weak architectural foundations, risks a rapid increase in system complexity.

There's potential but also peril. Imagine a scenario in which generative AI accelerates the software development lifecycle—but without corresponding architectural clarity to master business domains, avoid compliance loopholes, and avoid enterprise-wide dependencies. The result could be heightened security and resiliency risks.

For example, consider the exponential curve of complexity created from compounding relationships. That is why architectural design and platform governance are more business-critical than ever in gauging operational risks while transformation accelerates, avoiding what could be considered an “event horizon”—a black hole that rapidly consumes any possibility of added value.

The rapid advancements in generative AI also introduce a mix of challenging operational risks, especially when LLMs are consumed through “as-a-service” models. Limited visibility into the interactions of LLMs within an extensive data supply chain, often comprised of substantial unfiltered data from public domains, necessitates dedicated supervision and stringent controls of outputs (see Perspective, “The generative AI supply chain”).

As well, this expanded attack surface heightens third-party risk and cybersecurity considerations. In fact, when asked which concerns slow the adoption of generative AI in their enterprise, 88% of banking CEOs identified difficulties in controlling data lineage as the top compliance challenge. This is 27 percentage points more than CEOs in other industries.³⁵

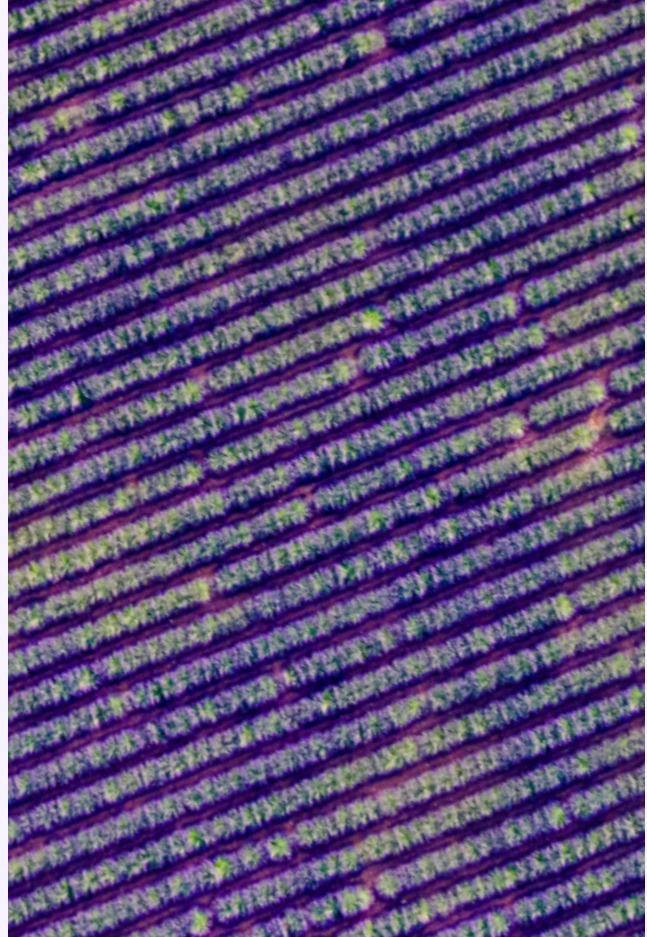
Perspective

The generative AI supply chain

The generative AI supply chain refers to the process and flow of components, data, and activities involved in the development, deployment, and maintenance of generative AI models. It encompasses various stages, including data collection and preprocessing, model training and fine-tuning, model validation and testing, deployment and integration into systems, and ongoing monitoring and updates.

The supply chain involves sourcing and curating training data, selecting and configuring AI models, implementing infrastructure and computing resources, and facilitating compliance with ethical and legal considerations. It also includes collaboration with data providers, AI developers, researchers, and stakeholders throughout the lifecycle of generative AI models.

The generative AI supply chain refers to the process and flow of components, data, and activities involved in the development, deployment, and maintenance of generative AI models.



Risk management of AI—and with AI

Banking is and always has been a data-intensive industry. The new news: generative AI models employ different data and more varied methods to source data, including online repositories, public data sets, web scraping, APIs, surveys, images, and partner data feeding.

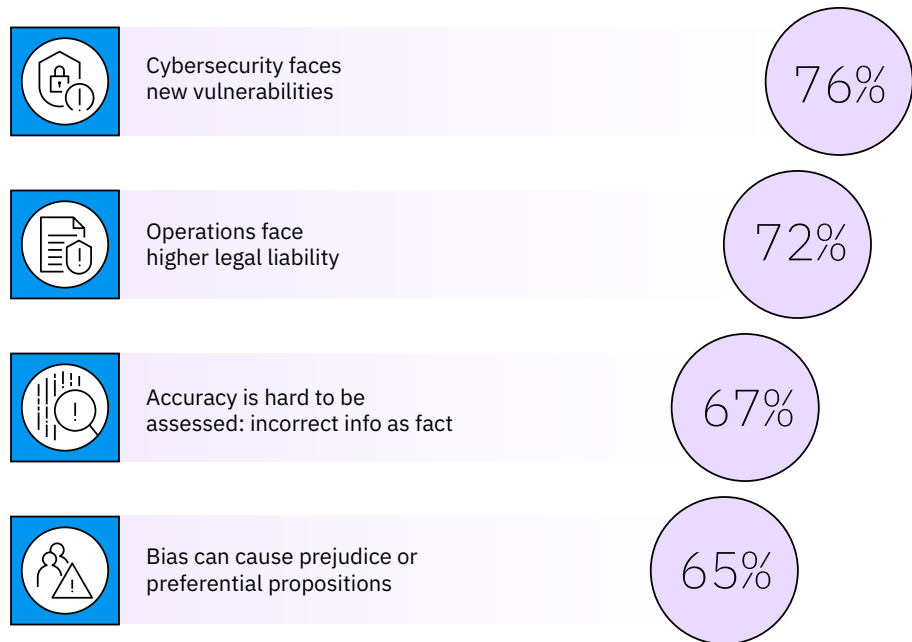
An AI model is not coded like other quantitative techniques but learns directly from data, recognizing patterns or executing insights-driven decisions without human intervention. Therefore, the data model is intrinsically linked to the AI model. This changes the way banks deal with data management and sheds a spotlight on previously peripheral risks dealing with confidentiality, integrity, and availability.

In 2023, we asked banking executives worldwide which AI-related risks pose severe threat to the economic viability, should their enterprise lack proper forms of governance.³⁶ More than 60% indicated new vulnerabilities for cybersecurity (76%), legal uncertainty related to operations (72%), difficulties in controlling outcome accuracy (67%), and prejudice from model bias (65%) (see Figure 11).

FIGURE 11

CEOs weigh in on AI risks

Governance of AI becomes strategic.



Source: Generative AI—Impact on hybrid cloud pulse survey.
IBM Institute for Business Value. 2023. IBM internal information.

For instance, legal rights for intellectual property (IP) must be carefully considered before implementing any generative AI application, which means due diligence on compliance in distant corners of the AI supply chain.

After a decade of digital innovation, AI algorithms are not new to leading banks. Solutions range from automation of mortgage approval, analysis of customer sentiment, or monitoring transaction data to detect fraud. However, development has typically occurred inside siloed banking organizations around localized use cases.

As AI takes hold enterprise-wide, this segmentation constrains the governance of a system that blurs the line between business and technology accountability. Lack of clarity about how banks must practically govern and risk-manage AI from a centralized perspective hinders the ability to move forward quickly.

Uncurtailed AI systems can exhibit unexpected behaviors or make incorrect predictions in certain scenarios, leading to adverse consequences or impacts. Human oversight is the key mitigation action for the most concerning errors, biases, or unintended consequences that may go otherwise unnoticed or unaddressed.

But generative AI can also boost the efficiency and quality of human oversight for compliance. By enhancing the understanding of complex regulation, and the application of demanding requirements and controls, generative AI can assist in smoothing the impact of mandatory changes.

Another potential impact: generative AI's hunger for data and training comes with substantial energy consumption, creating a potentially negative impact on an institution's sustainability metrics. However, improvements in extreme ultraviolet lithography (EUV) are expected to reduce environmental, social, and governance (ESG) impact by packing more transistors onto a single chip, granting AI more processing brainpower and performance while using less energy.³⁷

Still, generative AI can be useful to augment the human oversight of risks too complex to manage traditionally. Some banking domains have already embraced AI to support risk mitigation and improve compliance effectiveness—for example, automation for the identification, investigation, and remediation of data breaches.

Generative AI can be useful to augment the human oversight of risks too complex to manage traditionally.

Case study

Leading Nordic bank dramatically improves fraud detection with AI in real time³⁸

A leading Nordic bank has always prioritized fraud containment with detection systems. And to support that objective, they are investing in AI models.

Typically, card and payment transactions run on core banking platforms, while AI models scan transactions through off-platform detection systems. High latency between on-premises core systems and off-premises cloud technology can generate “time-outs,” which could halt the execution of transactions when customers execute payments. To avoid time-outs, real-time checks tend to be executed on a limited number of transactions (estimated to be approximately 10%).

The question: Can fraud-detection AI models be run in real-time on 100% of transactions?

The bank established their ability to develop and train their machine learning models on cloud and deploy them on-premises where the transactions reside. By embedding AI models in core banking platforms, vast amounts of data could be deciphered with business-critical response time, improving from more than 50 milliseconds to 1 millisecond. This performance enables fraud detection scan for 100% of transactions.

Using AI inferencing on core platforms demonstrated that AI can serve two objectives: improving fraud detection while maintaining a smooth customer experience.

A key consideration: for enterprise oversights and application of AI to be effective—as well as suitable governance deployed—organizations need a new collaboration between humans and machines.

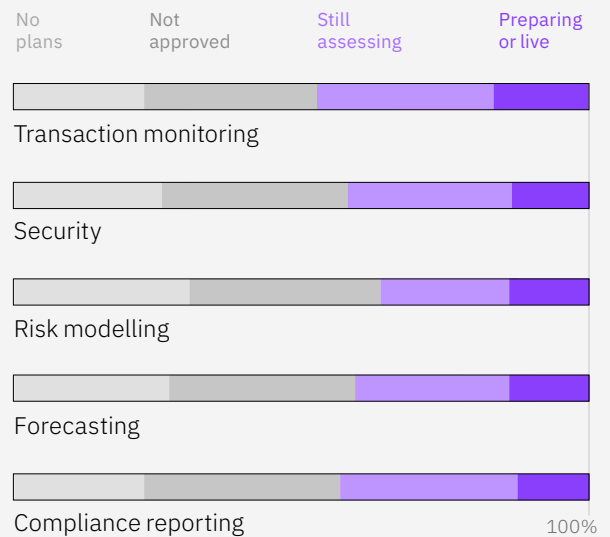
Perspective

Risk and compliance can generate value—yet more than 50% of banks hesitate

Across business domains, risk and compliance is expected to generate the greatest value for the banks. It's no surprise: this is where more banks are production-oriented.

But generative AI is a tale of two cities. On average, 15% of organizations are preparing to go live or already live with risk and compliance use cases. Yet, 56% will not proceed at this time.

There is no single pattern in banks' approaches to use cases. Statistical evidence indicates that bankers see use cases as fairly independent from one another. A bank with "no plans" for one of the listed use cases might be "preparing or live" for another one of the same domain.



Q: For the above risk, compliance, and security processes use cases, what is your institution's approach to implementing generative AI?



Foreword

The game-changer:
How generative AI can
transform the banking
and financial sectors

A critical moment

Communication:
The driving force behind
the customer experience

A reimagined workforce
experience can boost
productivity and
re-balance costs

Risk and compliance:
The double-edged sword

[Manage a trusted AI
platform for value,
innovation, and risk](#)

The 10 guiding actions

Manage a trusted AI platform for value, innovation, and risk

AI governance enables trust—and that’s not an easy task. Financial institutions already possess risk management frameworks for processes that may incorporate AI, including third-party risk, model risk management, and change management processes. Standards and best practices are still taking shape for governance and management of AI risks. Organizations are forging broader AI regulation and sorting out banking and financial markets idiosyncrasies.

Financial institutions don’t have to reinvent the wheel when it comes to their risk management framework in the era of AI. Rather, they can add the new “spoke” of AI and refine as needed to properly account for transparency, robustness, explicability, fairness, and privacy. For instance, agreements with customers and third parties might already be in place, defined before AI took center stage. This demands an understanding of how data can be anonymized—and agreements updated responsibly—to foster compliant deployment of AI.

Technology choice also significantly impacts governance effectiveness. Early adopters of AI often manage models across various tools, applications, and platforms both internal and external. Standardizing these on a common platform is essential for consistent governance. Open platform transformations in particular offer advantages by facilitating continuous and consistent integration of proprietary and third-party models—recognizing no single AI model can address all use cases.

Ultimately, AI governance is a collaborative process involving all functions of a banking organization. It requires new ways of working, amplified by changes in data and unstable models still subject to exponential innovation.

Practically, good governance is essential in the discovery phase of use cases to help manage compliance against the expected goals and risk profile of the bank. It is also essential to manage implementation and scaling enterprise-wide.

Four guiding principles

Risk management and governance, end to end

We identified four guiding principles to manage the end-to-end process.

1

Manage for value

Clearly identify, articulate, and quantify the business and economic value of AI initiatives across the enterprise or the specific domain of application. Align initiatives on a prioritized roadmap, focusing on potential value at scale. Given the multitude of use cases and the high costs that challenge the ROI of AI innovation, investment discipline is paramount.

2

Manage for the complexity of innovation

Recognize that use cases may vary in terms of innovation and corresponding feasibility. The most innovative use cases might pose greater challenges in terms of data and control. Support federated innovation across business units and functions, allowing different parts of the business to explore ideas within a controlled, cohesive environment supported by a Center of Excellence.

3

Manage for risk

Define and communicate a risk profile as the core element of governance. Enforce a consistent level of acceptable risk through adequate guiding principles, processes, and the IT system configuration. As new types of risk emerge, the risk and compliance processes require continuous calibration to include hallucination, bias, legal, and regulatory compliance.

4

Manage for scale

Establish an effective governance framework relying on a common enterprise platform for data accessibility, model, and use case approval. As AI capabilities might require costly, tailored tools (for example, building, tuning, and/or maintaining large foundation models), they are best developed and risk-managed at scale across an enterprise-wide common platform definition.

Governance is not a one-size-fits-all approach, and every bank must cater to its own perspective. Well-conducted AI governance is the foundation of a well-conducted business—and a compelling competitive advantage—as it directs business action with clarity and effectiveness. In this regard, we reflected with 110 IBM experts about a high-level mapping to value, complexity, and risk for the same 19 use cases included in the international survey of 600 banking executives. The exercise is only illustrative but serves as a worthy starting point for custom reviews inside individual organizations (see Figure 12).

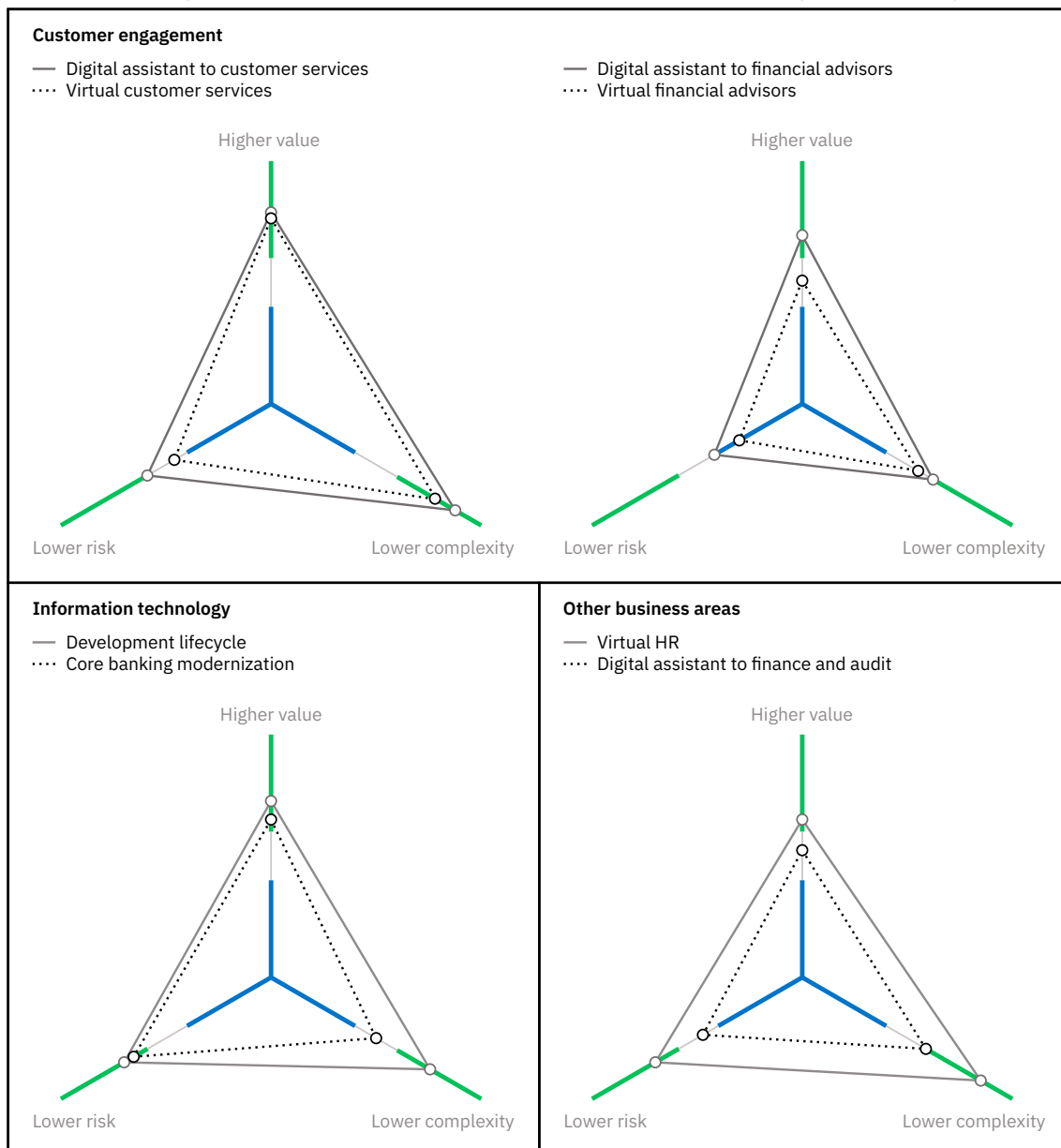
A variety of elements like data access, choice of AI platform, talent availability, and local regulation characterize your bank’s trade-offs for each use case.

FIGURE 12

Mapping use cases to value, complexity, and risk—representative samples

Examples to use as starting points in each analysis of use cases—with full data set below.

— More favorable (higher value, lower complexity, lower risk) — Less favorable (lower value, higher complexity, higher risk)



Source: Based on an internal survey of 110 IBM banking experts to assess a high-level mapping of risk use cases.

Each use case has been evaluated in terms of potential trade-offs between expected value (new revenue, improved user experience, cost reduction, and better risk management of operations), complexity (readiness of the organization, talent gap, technology maturity), and risk (hallucination, bias, security, and privacy). Each evaluation ranges from 0 to 100 (see Figure 13). Use cases with value assessed between 0 to 40 are deemed less favorable (in blue) and from 60 to 100 more favorable (in green). Use cases with complexity or risk are estimated between 0 to 40 are deemed more favorable (in green) and from 60 to 100 less favorable (in blue).

FIGURE 13

Mapping use cases to value, complexity, and risk—full data set

■ More favorable (higher value, lower complexity, lower risk) ■ Less favorable (lower value, higher complexity, higher risk)

Generative AI use cases		Value	Complexity	Risk
Customer engagement	Digital assistant to customer services	79	13	41
	Digital assistant to financial advisors	69	37	58
	Virtual customer services	76	22	54
	Virtual financial advisors	51	45	65
	Lending operations	64	42	64
	Trade finance	64	46	53
Risk and compliance	Transaction monitoring	74	44	63
	Risk modeling	63	58	66
	Forecasting	58	44	49
	Compliance reporting	74	43	53
	Security	72	37	59
Information technology	Development lifecycle	73	24	30
	Core banking modernization	65	45	34
	Test and bug discovery	71	18	23
Other business areas	Virtual HR	65	15	30
	Digital assistant to HR recruitment services	57	16	50
	Digital assistant to marketing	62	16	37
	Digital assistant to procurement	57	20	36
	Digital assistant to finance and audit	53	41	53

Source: Based on an internal survey of 110 IBM banking experts to assess a high-level mapping of risk use cases.

Foreword

The game-changer:
How generative AI can transform the banking and financial sectors

A critical moment

Communication:
The driving force behind the customer experience

A reimagined workforce experience can boost productivity and re-balance costs

Risk and compliance:
The double-edged sword

Manage a trusted AI platform for value, innovation, and risk

The 10 guiding actions

The 10 guiding actions

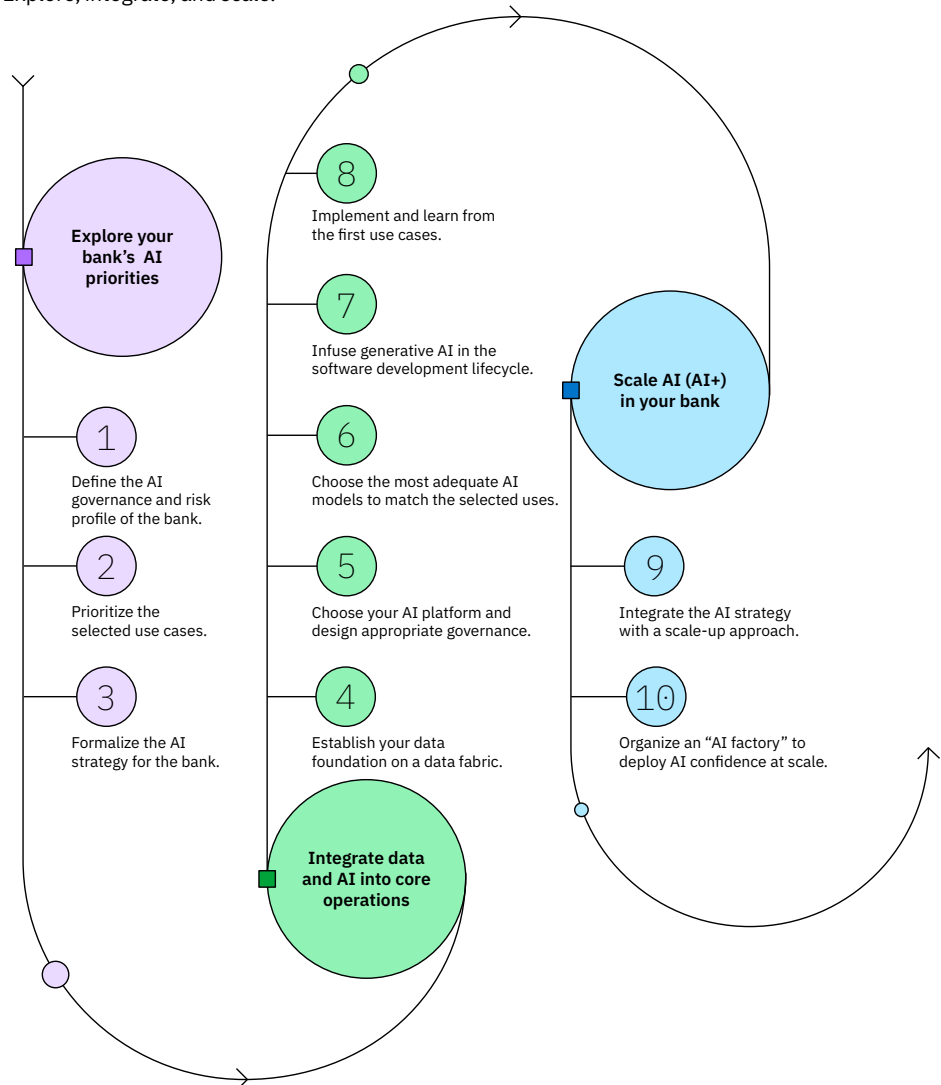
Banks are facing shorter-than-ever technology cycles. To manage, they need an expedited, structured process to not only move forward, but to move forward quickly.

Based on IBM's experience with leading financial institutions working with AI, we share 10 actions that can guide decisions about building generative AI foundations. These actions align with the broader processes of scaling AI across the firm, including machine learning, deep learning, and NLP (see Figure 14).

FIGURE 14

Action plan

Explore, integrate, and scale.



01 Define the AI governance and risk profile of the bank. In defining your AI governance, include key criteria for managing your bank's risk appetite and assess the potential value of AI implementations adjusted for risk and complexity. Enterprise-wide AI redefines your bank's operational risk profile in terms of new risk factors and their relative relevance.

As well, banks need to outline expectations of regulators to facilitate risk-controlled adoption across the enterprise. Demonstrating centralized AI governance of the complete AI model lifecycle is essential for regulatory compliance, transparency, explainability of results, and data security. Key factors here can include data security, ethical considerations, regulatory compliance, and the integration of AI into operations. Successfully identifying and addressing these risks is critical to maintaining customer trust, regulatory adherence, and helping ensure responsible use of AI.

02 Prioritize the selected use cases. Prioritize identified use cases based on your bank's risk appetite, weighing internally perceived value, feasibility within an acceptable timeframe, and impact on risk profile (see example on page 37). At the same time, encourage creativity through pilot projects that address business themes (for example, enhancement of customer services) and IT topics (for example, assisting code creation).

03 Formalize the AI strategy for the bank, grounded in use case prioritization and planning for skills and resources. Define a master plan for enterprise AI that covers all selected use cases, changes in the organization, funding needs, and plans for resource management. Put each use case in the broader context of banking transformation to mutualize costs for AI platform transformation, and gain clarity about marginal contribution to enterprise value.

At this stage, it's crucial to create a framework and doctrine that guide teams in AI-related work. Key considerations include establishing goals. Does the bank aim to be a "simple AI user" through APIs accessing "black box" models? Or is the goal to be an "AI creator" and gain autonomy in building competitive advantages? These decisions identify potential skill gaps.

04 Establish your data foundation on a data fabric. A strong foundation for AI success requires more than just a methodology or set of principles; organizations also need to modernize their information architecture technology. Banks need an architecture designed for AI—one that will help them optimize and automate data access and availability, deliver high-quality governed data, and manage privacy and compliance.

Simplifying, unifying, and connecting a data fabric across complex, dispersed environments is critical to building a foundation for successful, timely AI initiatives in a transforming banking business.

05 Choose a trusted AI platform and design appropriate governance. Scaling AI enterprise-wide means applying multiple AI models, since no single model is a best fit in the execution of all use cases. Banks cannot advance without a clear decision on which AI platform aligns with their role expectations in the AI value chain. The chosen mix between open-source models and models built in-house is a pivotal factor in platform selection. And independence from any specific cloud provider must be considered to preserve autonomy and flexibility to deploy future models.

06 Choose the most adequate AI models to match the selected uses in a multimodal world. Once the platform is chosen, define which generative AI model to onboard for each use case. Consider criteria that includes accuracy, costs, performance, carbon footprint, and the modalities by which generative AI can be accessed.

This is because generative AI is not only about multiple models but also about multiple modalities. The platform must allow for a broad range of formats for both inputs and outputs. Formats can include text, images, video, voice, audio, and combinations thereof. Other important considerations include the model publisher's commitment to quality, the enforcement of privacy protection rules and IP clearance for training data, and the ability to verify the absence of ethical bias in the models used.

07 Infuse generative AI in the software development lifecycle. To leverage generative AI in software development successfully, the technology must be seen as more than just a basic tool for code generation. Generative AI tools, when integrated into a developer's process, can document pre-existing code, reverse-engineer requirements, and write new ones. Based on user input, generative AI tools can produce high-quality code recommendations. By offering clear solutions and managing routine programming tasks, generative AI can streamline and drive new ways of working. Additionally, it can detect coding errors and expose possible security risks—both significant advantages.

08 Implement initial use cases and learn from a precise impact analysis to prepare for scale-up (processes, organization, skills, and change management). The deployment of initial use cases provides an opportunity to assess impacts on the operating model, preparing for more structural updates to organizations and skills. This is when a bank's AI practice is forged by concrete actions. Lessons learned are not failures but risk-controlled opportunities to assess and refine your bank's AI culture. Immediately promote feedback loops addressing change management and skill transformation. Generative AI itself can be used to scan large volumes of client inquiries, complaints, and communications to build action-oriented summaries for management decisions.

- 09 **Integrate the AI strategy with a scale-up approach to infuse AI into business processes.** The initial use cases that adhere to the bank’s risk profile have been validated, and the AI platform transformation has demonstrated its efficacy, laying the foundations for progress. Now, the focus is on momentum and formalizing a scaling strategy for the entire organization, with some adopting an “AI-first” approach for all new initiatives.
- 10 **Organize an “AI factory” to deploy AI confidence at scale, promoting short cycles and cost control.** Recommendations include organizing around an “AI factory” for iterative testing and deployment. Now your bank can change pace from the initial +AI approach, which means *adding AI* to your operations, to an AI+ operating model in which the AI *platform is central* to all development and business methods.

After reading our extensive research, you’ve undoubtedly realized that banks must keep ahead of this rapidly evolving landscape. As our guiding actions outline, this requires continuous thought and a consultative approach to harnessing generative AI—and it’s imperative to get started now.

About the authors



Shanker Ramamurthy

Global Managing Partner,
Banking and Financial Markets
IBM Consulting
sramamur@us.ibm.com
linkedin.com/in/shankerramamurthy

Shanker is the Global Managing Partner for Banking in IBM Consulting, with a particular focus on bank transformation, core banking, and payments. He is also the President of the IBM Industry Academy and a member of the IBM Acceleration Team. He is a well-known thought leader with five patents relating to addressing complexity, has authored several white papers, and been ranked as one of the 50 most influential financial services consultants worldwide by *Euromoney* magazine.

John J. Duigenan

General Manager, Financial Services,
Banking, Financial Markets, and Insurance
IBM Technology, Global Industries
john.duigenan@us.ibm.com
linkedin.com/in/Duigenan

John is the General Manager for the Financial Services sector within IBM Global Industries. His mission is to align the value creation from IBM Technology with current industry imperatives that span clients across the sector. John and his team partner closely with clients to show how industry-specific use cases map to IBM technology capabilities, solutions, and outcomes. Use cases are a vehicle to demonstrate relevance, scale re-use, and create client references. As an IBM Distinguished Engineer, John also leads by example by being a hands-on technologist. Since joining IBM in 1998, he held technical and business roles spanning the entire product lifecycle, with extensive development, service, sales, and technology leadership. He led multibillion-dollar transformation programs at IBM's customers. John is based in New York City.

Paolo Sironi

Global Research Leader,
Banking and Financial Markets
IBM Institute for Business Value
paolo.sironi@de.ibm.com
linkedin.com/in/theptionsironi

Paolo is a Global Research Leader in Banking and Financial Markets at the IBM Institute for Business Value. He is senior advisor for selected global accounts, assisting service teams in Board-level and C-level conversations about business model adaptation in platform economies. He is one of the most respected Fintech voices worldwide and co-hosts the European edition of the *Breaking Banks* podcast. He is a celebrated book author on digital transformation and quantitative finance and economics, and a keynote speaker at major international events.

Contributors

Many thanks to the following for their contributions and insights: Sara Aboulhosn, Swati Bhide, Diane Connelly, Michael Conway, Angela Finley, Leah Generao, Vivek Kapur, Connor Loessl, Nicolas Meyerhoffer, Lucy Sieger, and Saket Sinha.

Domains and use cases

Please see below for domain categories and specific use cases. They are not exhaustive of all potential opportunities to leverage generative AI in banking.

Client engagement

Virtual customer services. Generative AI enhances the client experience with virtual agents that reduce customer service call volumes.

Digital assistant to customer services. Generative AI enhances classification of client complaints and automates content summarization to augment agent productivity and quality of live conversations.

Virtual financial advisors (retail or corporate banking). Generative AI powers virtual financial advisors to digitize the advisory interaction and reduce human touch-points.

Digital assistant to financial advisors (retail or corporate banking). Generative AI enhances search and summarization of financial information to improve the quality of advisory conversations with clients.

Lending operations. Generative AI enhances approval processes to reduce time to serve customers.

Trade finance. Generative AI enhances classification and summarization of counterparty agreements to speed processing time of trade finance operations.

Risk and security

Transaction monitoring. Generative AI empowers KYC/AML processes to reveal intricate webs of transactions and criminal networks.

Risk modeling. Generative AI enriches the discovery of behavioral patterns to improve client and counterparty segmentation for risk management.

Forecasting. Generative AI integrates financial models with analysis of alternative data to enhance forecasting.

Compliance reporting. Generative AI can ingest existing regulation and upcoming regulatory requirements to automate monitoring and compliance alerts.

Security. Generative AI can scan logs, data, and software to find vulnerabilities to address cybersecurity proactively.

IT development

Development lifecycle. Generative AI enhances software development to reduce the time and cost of building applications.

Test and bug discovery. Generative AI creates synthetic data and tests scripts for efficient deployment of new code.

Core banking modernization. Generative AI enhances developers' capability to convert software code, such as COBOL to Java, to reduce costs in modernizing monolithic core banking.

Other support areas

Virtual HR. Generative AI improves the self-service options of employees with better knowledge search and accuracy of classification, freeing up HR capacity.

Digital assistant to HR recruitment services. Generative AI improves summarization and content creation to perform recruitment campaigns.

Digital assistant to Marketing. Generative AI creates content, images, and personas to enhance marketing campaigns.

Digital assistant to Procurement. Generative AI improves the summarization of vendor qualification and the analysis of legal documentation to streamline procurement.

Digital assistant to Finance and Auditing. Generative AI improves the identification and reconciliation of gaps in financial data to streamline audit processes.

Research and methodology

In collaboration with Oxford Economics, the IBM Institute for Business Value surveyed 600 bankers with executive responsibilities in the application of data and AI across their institutions. Their organizations each had more than \$10 billion in total assets in the latest fiscal year. As well, they operate in 17 countries: Australia, Brazil, Canada, China, France, Germany, Hong Kong, India, Italy, Japan, Mexico, Singapore, Spain, Sweden, UAE, UK, and US.

This research is complemented by the financial performance analysis of worldwide banks with total assets larger than \$50 billion. Annual reports are sourced from S&P Global Market Intelligence. Regional comparisons are provided based on the country classification in the International Monetary Fund (IMF) World Economic Outlook database.³⁹ We distinguish between “major advanced and EU economies” (corresponding to G7 countries, with all EU member states) and “other advanced and emerging economies.”

In addition, we conducted an internal survey of 110 IBM banking experts to assess a high-level mapping of risk use cases, keeping in mind the potential AI risk profile of a common financial institution, operating in a highly regulated environment, bound to adequate but finite investment resources.

IBM Institute for Business Value

For two decades, the IBM Institute for Business Value has served as the thought leadership think tank for IBM. What inspires us is producing research-backed, technology-informed strategic insights that help leaders make smarter business decisions.

From our unique position at the intersection of business, technology, and society, we survey, interview, and engage with thousands of executives, consumers, and experts each year, synthesizing their perspectives into credible, inspiring, and actionable insights.

To stay connected and informed, sign up to receive IBV’s email newsletter at ibm.com/ibv. You can also find us on LinkedIn at <https://ibm.co/ibv-linkedin>.

The right partner for a changing world

At IBM, we collaborate with our clients, bringing together business insight, advanced research, and technology to give them a distinct advantage in today’s rapidly changing environment.

About Expert Insights

Expert Insights represent the opinions of thought leaders on newsworthy business and related technology topics. They are based on conversations with leading subject-matter experts from around the globe. For more information, contact the IBM Institute for Business Value at iibv@us.ibm.com.

Related reports

Embedded finance: Creating the everywhere, every day bank

IBM Institute for Business Value in collaboration with BIAN and Red Hat. September 2023.
<https://ibm.co/embedded-finance>

Embedded finance: The voice of the makers

IBM Institute for Business Value in collaboration with BIAN and Red Hat. October 2023.
<https://ibm.co/embedded-finance-makers>

Foundations of banking excellence: Practices and priorities to accelerate digital transformation

IBM Institute for Business Value in collaboration with BIAN. October 2022.
<https://ibm.co/foundations-banking-excellence>

Notes and sources

- 1 IBM Institute for Business Value analysis of S&P Global data. Unpublished information.
- 2 “October 2023 euro area bank lending survey.” Press release. European Central Bank. October 24, 2023. <https://www.ecb.europa.eu/press/pr/date/2023/html/ecb.pr231024~c42cea39db.en.html>
- 3 “Household debt and credit report.” Center for Microeconomic Data. Federal Reserve Bank of New York. <https://www.newyorkfed.org/microeconomics/hhdc.html>; “Credit Balances on the Rise as Consumers Manage Higher Costs.” TransUnion. November 9, 2023. <https://newsroom.transunion.com/q3-2023-ciir/>
- 4 IBM Institute for Business Value analysis of S&P Global data. Unpublished information.
- 5 Ibid.
- 6 Ibid.
- 7 “Mario Draghi, President of the ECB, and Luis de Guindos, Vice-President of the ECB. Frankfurt am Main.” Introductory statement. Press conference. European Central Bank. September 12, 2019. <https://www.ecb.europa.eu/press/pressconf/2019/html/ecb.is190912~658eb51d68.en.html>
- 8 *CEO decision-making in the age of AI: Act with intention*. Global C-suite Series: 28th Edition. IBM Institute for Business Value. June 2023. Unpublished information based on banking and financial markets data. <https://www.ibm.com/thought-leadership/institute-business-value/en-us/c-suite-study/ceo>
- 9 Ramamurthy, Shanker, John J. Duigenan, Hans Tesselaar, Héctor Arias, and Paolo Sironi. *Embedded finance: Creating the everywhere, everyday bank*. IBM Institute for Business Value in partnership with BIAN and Red Hat. September 2023. <https://ibm.co/embedded-finance>
- 10 Walden, Stephanie and Mitch Strohm. “What Is A Neobank?” *Forbes*. June 24, 2021. <https://www.forbes.com/advisor/banking/what-is-a-neobank/>
- 11 Ramamurthy, Shanker, John J. Duigenan, Hans Tesselaar, Héctor Arias, and Paolo Sironi. *Embedded finance: Creating the everywhere, everyday bank*. IBM Institute for Business Value in partnership with BIAN and Red Hat. September 2023. <https://ibm.co/embedded-finance>
- 12 Ibid.
- 13 IBM Institute for Business Value analysis of S&P Global data. Unpublished information.
- 14 Ibid.
- 15 Ramamurthy, Shanker, John J. Duigenan, Hans Tesselaar, Héctor Arias, and Paolo Sironi. *Embedded finance: Creating the everywhere, everyday bank*. IBM Institute for Business Value in partnership with BIAN and Red Hat. September 2023. <https://ibm.co/embedded-finance>
- 16 Khandelwal, Manisha. *The Role of NPS in Banking and Other Financial Institutions*. Survey Sensum. May 19, 2023. <https://www.surveysensum.com/blog/nps-in-banking>
- 17 Ibid.
- 18 Ramamurthy, Shanker, John J. Duigenan, Hans Tesselaar, Héctor Arias, and Paolo Sironi. *Embedded finance: Creating the everywhere, everyday bank*. IBM Institute for Business Value in partnership with BIAN and Red Hat. September 2023. <https://ibm.co/embedded-finance>
- 19 Based on internal IBM client information.
- 20 Krishna, Arvind. “IBM CEO: ‘Today’s workforce should prepare to work hand in hand with A.I.’” *Fortune*. April 20, 2023. <https://fortune.com/2023/04/20/ibm-ceo-work-ai-artificial-intelligence-careers-tech-arvind-krishna/>
- 21 “NatWest and IBM Collaborate on Generative AI Initiative to Enhance Customer Experience.” IBM Newsroom. November 6, 2023. <https://newsroom.ibm.com/2023-11-06-NatWest-and-IBM-Collaborate-on-Generative-AI-Initiative-to-Enhance-Customer-Experience>

- 22 Based on IBM client experience.
- 23 “Gartner Forecasts Worldwide Banking and Investment Services IT Spending to Reach \$652 Billion in 2023.” Gartner Newsroom. June 21, 2023. <https://www.gartner.com/en/newsroom/press-releases/2023-06-21-gartner-forecasts-worldwide-banking-and-investment-services-it-spending-to-reach-652-billion-in-2023>; “Gartner Forecasts Worldwide IT Spending to Grow 4.3% in 2023.” Gartner Newsroom. July 19, 2023. <https://www.gartner.com/en/newsroom/press-releases/2023-07-19-gartner-forecasts-worldwide-it-spending-to-grow-4-percent-in-2023>
- 24 IBM Institute for Business Value analysis of S&P Global data. Unpublished information.
- 25 De Haan, Jakob. “Low IT spending by banks: Reason for concern?” ECON Committee. European Parliament. June 2021. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/689439/IPOL_IDA\(2021\)689439_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/689439/IPOL_IDA(2021)689439_EN.pdf)
- 26 IBM Institute for Business Value analysis of S&P Global data. Unpublished information.
- 27 Keller, Christian, Mark Cus Babic, Akash Utsav, Ana Paula de Jesus Assis, and Brian Goehring. *AI revolution: productivity boom and beyond*. Barclays Research in partnership the IBM Institute for Business Value. January 11, 2024. <https://www.cib.barclays/our-insights/AI-productivity-boom.html>
- 28 Sironi, Paolo, Diane Connelly, and Connor Loessl. *Embedded Finance: The Voice of the Makers*. IBM Institute for Business Value in partnership with BIAN and Red Hat. October 2023. <https://ibm.co/embedded-finance-makers>
- 29 Ramamurthy, Shanker, John J. Duigenan, Hans Tesselaar, Héctor Arias, and Paolo Sironi. *Embedded finance: Creating the everywhere, everyday bank*. IBM Institute for Business Value in partnership with BIAN and Red Hat. September 2023. <https://ibm.co/embedded-finance>
- 30 “AI code-generation software: What it is and how it works.” IBM. September 19, 2023. <https://www.ibm.com/blog/ai-code-generation/>
- 31 “AI code-generation software: What it is and how it works.” IBM. September 19, 2023. <https://www.ibm.com/blog/ai-code-generation/>
- 32 Goldstein, Jill, Bill Lobig, Cathy Fillare, and Christopher Nowak. *Augmented work for an automated, AI-driven world: Boost performance with human-machine partnerships*. IBM Institute for Business Value. August 2023. <https://ibm.co/augmented-workforce>
- 33 Ibid.
- 34 Ramamurthy, Shanker, John J. Duigenan, Hans Tesselaar, Héctor Arias, and Paolo Sironi. *Embedded finance: Creating the everywhere, everyday bank*. IBM Institute for Business Value in partnership with BIAN and Red Hat. September 2023. <https://ibm.co/embedded-finance>
- 35 Generative AI: Impact on hybrid cloud pulse survey. 2023. IBM internal information.
- 36 Ibid.
- 37 Moore, Samuel K. “The Semiconductor Industry’s Most Important Tool Goes Green.” IEEE Spectrum. June 12, 2023. <https://spectrum.ieee.org/euv-lithography>
- 38 Based on internal IBM information.
- 39 “Groups and Aggregates Information.” World Economic Outlook Database. International Monetary Fund. April 2023. <https://www.imf.org/en/%20Publications/WEO/weo-database/2023/April/%20groups-and-aggregates>

© Copyright IBM Corporation 2024

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America | January 2024

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

IBM, the IBM logo, ibm.com, Watson, and IBM z/OS® mainframe are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at: ibm.com/legal/copytrade.shtml.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

This report is intended for general guidance only. It is not intended to be a substitute for detailed research or the exercise of professional judgment. IBM shall not be responsible for any loss whatsoever sustained by any organization or person who relies on this publication.

The data used in this report may be derived from third-party sources and IBM does not independently verify, validate or audit such data. The results from the use of such data are provided on an “as is” basis and IBM makes no representations or warranties, express or implied.

