EUROMONEY

The AI in Banking Best Practices Playbook

How banks are deploying AI today and tomorrow

Contents

Foreword
Beyond the hype

Dominic O'Neill

Head of Banking US & Europe, Euromoney
dominiconeill@euromoney.com

The rapidly increasing power of generative artificial intelligence (gen AI) models is more than a passing fad, even if the full extent of its impact in financial services is yet unclear.

At its best, gen AI has the potential to take pain out of banking beyond just saving money. AI co-pilots powered by large language models (LLMs) are already easing painful tasks for bank staff, including software developers, call centres and even investment bankers. The next stage is to turn AI's increasingly human-like skills towards customers, albeit cautiously.

Some banks are already experimenting with multiple AI agents, plugged into internal and external applications, to assess customers' problems and even to solve them.

AI's ever-changing nature demands flexibility in its adoption. It also brings up difficult challenges in data security and privacy, hallucinations and other problems around predictability, the danger of reinforcing human bias, and, increasingly, geopolitical risk.

All that needs careful attention by banks. Particularly when it comes to customer-facing use cases, banks will not be able to implement gen AI tools as boldly as some other industries.

In this report, we offer an unrivalled deep look into how individual banking leaders are thinking about AI and how they are acting on it. How are they leveraging the technology? And which are the most successful and forward-looking initiatives in gen and agentic AI?

We held more than 30 in-depth conversations with those in charge of implementing gen and agentic AI at top global banks, and in many tech-leading national banks. We also spoke to banking-focused AI professionals at LLM vendors, and smaller AI-focused fintech firms.

While all banks spoken to are using gen AI for basic, internal tasks, there is a clear awareness that AI is far from realising its potential. Inevitably, there is a gap between claims of radical innovation and reality. This is particularly true in agentic AI, given its still hazy definition.

The bombast around AI illustrates how deeply banks recognise that they need to get ahead in the field before it is too late. At JPMorgan, commitment to prioritising AI at the CEO level is crucial and makes AI harder for rival banks to deprioritise. UBS might admit that it is behind JPMorgan given the latter's head start, but it is determined to catch up. The rollout of AI now appears second only to the integration of Credit Suisse in UBS chief executive Sergio Ermotti's concerns.

Success in AI requires much more than a few bullish CEO pronouncements, of course. The key challenge is how to gear the entire organisation. That means deep-seated preparation, as well as a lucid strategy to leverage the best of the competing array of models and vendors.

Get gen AI right, and your role in finance could be more important than ever. Get it wrong – as banks inch towards agentic systems – and your business risks being consigned to history.

Key learnings from the top

THE NINE-STEP JOURNEY TO IMPLEMENT AI IN BANKS

01 | Adopt a decisive but flexible approach to AI development

02 | Strengthen a central AI function plugged into businesses

03 | Upskill across the organisation, not just in AI research

04 | Develop processes for selecting third-party AI models

05 | Build platforms to boost model flexibility and security

06 | Consider smaller models which meet your needs

07 | Start with low risk but impactful use cases

08 | Ease into customer-facing opportunities

09 | Build trust to succeed in agentic banking

CHAPTER I

Preparing the organisation

Be flexible

01

Success Factors

- Adopt an LLM-agnostic approach
- Balance agility with decisiveness
- Use traditional AI where appropriate
- Focus on end goals

One of the most exciting and challenging characteristics of AI is how rapidly it is developing, as big tech firms – and geopolitical blocks - race to get ahead. Technological advancement has rarely happened this fast. The regulation around the use of AI models is also subject to change, partly due to geopolitical tensions: with potentially drastic implications for how banks should roll out AI and which vendors they should use.

How can banks keep up with the dizzying pace?
The answer is flexibility. Effectively deploying generative AI starts with being ready to switch quickly between LLMs and, for most banks, between the companies that supply them. That does not just mean adapting the type of model and vendor to the specific use case (open or closed-source, proprietary or third party).

Beyond this, banks need to be prepared to switch between models in individual use cases, including when they are under development, whether at a pilot stage or after. Having the right organisational approach to AI is vital.

Banks must properly balance the freedom for individual business lines to come up with use cases that are adapted to their specific demands, with a certain centralisation of AI strategy, governance and costs. Without that element of centralised control, organisations could once again be left with a field littered with legacy technology.

Keeping up with the latest **ChatGPT** models is hard enough. But change in AI goes much deeper than developments at OpenAI. Before the launch of ChatGPT in late 2022, banks and other companies were building their own AI models. That changed abruptly three years ago. The release of China's **DeepSeek-R1** model in January 2025 made it even more obvious that banks could develop more customised advanced language models in an economically viable manner.

The latest development of models with greater capacity to reason - and the buzz around so-called agentic AI, with the potential for greater automation – further illustrate how implementing AI is a constantly shifting endeavour.

Rather than keeping up with tech changes for the tech's sake, it's important to focus on the end goals. If a steady state in technology and regulation is a hard ask, the answer may be in technological solutions built on more traditional forms of AI.

GEN AI IN PRACTICE — NatWest

How NatWest's AI projects moved with the technology

The fast-moving nature of AI can disrupt efforts to use the technology, requiring fresh and nimble thinking around projects. That means combining an overall vision of what to achieve with an open mind about how to get there.

These are insights that systemic UK bank NatWest knows well. Before the launch of ChatGPT, it was developing a call-summarisation system for its Coutts wealth management advisors, tailored to the heightened security demands of these customers, and initially using its own natural language processing technology. The release of ChatGPT meant NatWest decided to discontinue the project and move on, as third-party LLMs were suddenly able to do the same much more effectively.

The speed of AI development has made other NatWest's projects even more relevant, although the turbo technological change still makes it necessary for technologists and business colleagues alike to divorce themselves emotionally from specific plans and methods.

In 2024, the bank started to develop a complaints-automation procedure based on OpenAI's GPT-3.5-Turbo model, first released in late 2023. Nine months later, it was using the o1 model, having been through four OpenAI model generations. NatWest could not predict what developments at OpenAI would come out during the design phase, but the model's ability to help NatWest automate its complaints handling was far greater towards the end of the project than at the beginning.

"Everything you build in AI should allow you to be agnostic as to which LLM you use, and as to the vendor. It should allow you to quickly adapt to new technological developments."

Christoph Rabenseifner, Chief Strategy Officer for Technology, Data and Innovation, Deutsche Bank

Words of advice

“You need to become very agile in the way you deal with AI as a bank, but you also need to be decisive. In this race, you could be in a situation where you never get anything done because there's always something new around the block. Then there's never any value generation. There is a point at which you need to say, ‘This is good enough, let’s bring this into production’. At some point, you will replace that solution, but if you constantly try to jump onto the next thing, you never get any roll-out. If you're too fast-moving, you constantly chase yourself. But your platform will very quickly get outdated if you roll something out, and you insist on not changing it for the next three years, as you might have done in the past. This is a very important and difficult balance between being at the bleeding edge, versus being production-ready: and being safe enough, having done enough of your own engineering, to make sure the solution really fits into your world.”

Dirk Marzluf

Group Head of Technology & Operations
Banco Santander

Centralise strategy, devolve execution

02

Success Factors

- Build a plugged-in central AI function
- Formulate internal AI standards
- Keep a lid on costs
- Invest more in broader use cases

As AI grows in importance for the future of banks' competitiveness and profitability, banking leaders need to build up their central AI functions. These often brand new central AI capabilities are key not just to oversee strategy, but also to control rising costs and the risks of AI. They may deploy central AI expertise or develop it jointly with business leaders and technical teams. They also oversee the vital task of training the wider organisation.

Centralised AI leadership should help develop AI technology platforms across the organisation, while divisional teams should advise about risks that are specific to them. Banks need to draw up internal standards for how to deploy AI responsible, in negotiation with regulators and other industry bodies, as well as high-level internal councils of diverse leaders who can assess whether individual use cases meet that standard. Shortcuts are counterproductive, as any bank's ability to deploy gen AI should reside firstly in its confidence that it can do so safely. With such powerful capabilities in these models, putting appropriate guardrails in place is often the hardest part.

New and reinforced central teams in charge of firm-wide AI strategy might think of themselves in a shepherding role: guiding, rather than doing. **HSBC**, for example, has set up a central AI team split into three sub-teams overseeing commercial benefit and safety; a platform approach to using models that has the right balance between flexibility and technical consistency across the bank; and, finally, a team of machine-learning engineers charged with more complex projects such

as building custom-built models for mitigating fraud. This central team co-exists with data and analytics experts which sit in the individual divisions and businesses.

Having a structure which devolves much of the responsibility for executing the AI strategy to the divisional level may be more likely to result in use cases that solve problems for that division, rather than just for technologists in the central organisation. Using more traditional machine learning, data scientists in UBS' investment bank recently built a tool to come up with M&A ideas for its bankers to pitch to clients, sifting through hundreds of thousands of companies in a database, compared to the few hundred at best that a banker might think about. **UBS** estimates the tool was used in over 1000 pitches in 2024. A senior M&A banker's early involvement in the idea's formulation was crucial in making sure it would have practical application.

Devolving use-case development to the business level makes even more sense as AI will make it easier for people without advanced coding skills to build AI systems. However, the advent of generative AI also brings external models that have much wider-ranging application than the smaller systems that banks and others previously developed in-house.

That makes it more important to focus on use cases that can generate more value by being tested in one part of the organisation before being scaled across divisions. **Morgan Stanley**, for example, used an AI assistant for answering wealth management questions as the basis for similar assistants in other areas of the firm. Much of its central AI function grew out of a team working on AI in wealth management.

Identifying which use cases are most likely to get to that position of scale, and helping them achieve that, is a major reason to have a robust central AI team.

> "I'm a big believer that the CEOs of the businesses are best placed to understand where AI can help them achieve their strategic objectives. People talk about AI strategies. That's a misnomer. There are business strategies, and there are places where AI can enable the strategy."
>
> Teresa Heitsenrether, Chief Data & Analytics officer, JPMorgan Chase

GEN AI IN PRACTICE — BNP PARIBAS

Letting a thousand AI flowers bloom at BNP Paribas

Despite the vast sea of opportunities, banks should pick their battles in AI, limiting their efforts to a small number of use cases which a central team helps to select.

Previously, banks were more inclined to let a thousand flowers bloom in AI. BNP Paribas set a target to roughly double AI use cases at the bank to 1000 by 2025 in its last group three-year strategic plan, unveiled in early 2022. AI teams proliferated across the organisation. Thanks to a reliance on open-source models, the bank was able to build its own models, with its own data, relatively cheaply.

BNP Paribas may have achieved its target, but times have changed. Powerful LLMs supplied by external vendors involve much larger amounts of capital and computing power, and they are attracting greater regulatory attention, including in Europe. An overly decentralised approach to deploying LLMs could result in unacceptable levels of cost and risk, calling for a more focused approach to AI initiatives.

In BNP Paribas' next three-year plan, because of technological advance in AI, setting and achieving a financial target for value-creation in AI will be even more important. But a high number of use cases may be less of an aim.

"Implementing AI in banking effectively is not just about writing a bit of code. You need appropriate processes and structures in place. You need technological platforms in place. You need governance in place. And most importantly, you need skills and people in place."

Nimish Panchmatia, Chief Data & Transformation Officer, DBS bank

GEN AI IN PRACTICE — Deutsche Bank

Deutsche Bank's seeds a gen AI assistant in research

Around the time that gen AI first burst onto the global stage, Deutsche Bank set up a structure involving a central team in charge of AI strategy, risk and control, but in which technology developers work jointly with business divisions based on demand from the latter. The objective was for AI at Deutsche not to be pushed by the central team, but instead to respond to problems at the coal face.

One example is the gen AI assistant that it first developed for its research department, using an LLM from **Google**, to help analysts delve into bulky documents and analyse them, and to help create more digestible client materials. After work to finetune the model and build a front end for the research department, the bank has deployed a similarly customised model for origination and advisory bankers who might be preparing for client briefings or putting pitch books together.

It may be harder to scale similar use cases across the organisation than to bring a narrower use case into production. Yet that is precisely the goal. While Deutsche's central AI team counts around 300 ideas for implementing generative AI in different parts of the bank, the bank is prioritising certain lead use cases that can be rolled out in a broader cluster of businesses.

"Many of the best ideas have not been found yet, because those ideas will come from frontline staff when they truly understand what this technology can do. The ideas come from the people who do the job. That's why you need education."

Chris Purves, Co-head of Emerging Technology, UBS

1000+ Number of UBS investment-banking pitches using an M&A machine-learning tool in 2024

Words of advice

“Gen AI technology, while it seems incredibly powerful, basically only does five things. It searches, it summarises, it analyses, it generates text or images, and it translates. We're finding that once you build a solution around one of those capabilities for one business, there is a lot of reusability across financial services, whether it be a retail business or an institutional business. The work is not so much building, it's about taking content and then going to an evaluation framework, making sure that this thing is producing output consistent with what you want. As others have said, the technology is deceptively easy to build, and it's quite challenging in many cases to do the work to validate the efficacy of these solutions. That’s a big part of what the firm-wide AI team does.”

Jeff McMillan

Head of Firm-wide AI, Morgan Stanley

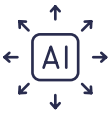
Using AI needs people first

03

Success Factors

Upskill across the firm

Have different levels of AI training



Launch AI mentors and evangelists

Attract scientific talent

The biggest job in boosting AI skills inside banks lies with the vast bulk of the workforce who are not technologists. After banks initially banned employees from using **ChatGPT** for their work, many are now working to get appropriately protected gen AI gateways onto the desktops of as many people as possible: demystifying the technology, and teaching people how to use it. This should improve use cases' adaptability to the business, as AI tools help to democratise the design and implementation of tech solutions, perhaps transferring that process towards people with more of the creativity needed to exploit the full potential of LLMs.

Some banks may take the attitude that, even if widespread rollout of AI assistants does not transform the bank's operations just yet, the effort is worthwhile so that the organisation is better prepared for a more impactful later stage. To that extent, rollouts of products such as **Microsoft Copilot** can be seen almost as part of a training budget. Making better use of those products, however, requires staff to be appropriately trained in how to prompt effectively, via training courses, programmes of AI mentors or technology evangelists, permeating every level of the bank.

The capacity of individual LLM providers to attract the best scientists, and to nurture internal talent, is one of the core determinants of whether they can develop superior models. In banking, especially since the shift in AI towards using external models such as **OpenAI**, the war for research and development talent is less intense. However, deep

understanding of AI models is vital to deploy them safely in large financial institutions.

Lofty salaries on offer at big tech firms, alongside other perks, make it hard for banks to attract top tech talent. Banks have a reputation for being bogged down by outdated technology. Neobanks may be an exception, but they must also fight to win the best people.

The newness of gen AI should also reinforce the need and opportunity to upskill staff, including training courses for data specialists. Beyond this, how can banks make themselves more attractive as career destinations for data specialists?

One answer is to move beyond traditional banking attitudes towards trade secrets. In-house data scientists will have scientific ambitions which means they will want to publish their research. Before joining a bank, they will also want to know which like-minded people works there, and how well qualified they are.

> "The big thing about generative AI is how you allow for everybody in the organisation to be able to use it, and to use it in a meaningful way, which will help with efficiency, toil removal, employee experience, and then customer experience and productivity."
>
> Nimish Panchmatia, Chief Data & Transformation officer, DBS bank

> "The barriers for people who don't have technical skills are going down. We're all going to be developers – not coders, but developing solutions that meet our business needs, in our own language. This means getting business experts engaged in a process that's historically been left to IT specialists."
>
> Jeff McMillan, head of firm-wide AI, Morgan Stanley

GEN AI IN PRACTICE — UBS

UBS goes big on its copilot rollout

UBS is conducting one of the biggest Microsoft Copilot rollouts in financial services globally, and in late February was on track to meet a target to give access to the assistant to 50,000 people, almost half of its staff base, by the end of March 2025. Yet, if not properly used, it could end up as an amusing, and costly, distraction. Usage, in turn, would depend on staff's comfort with the tool.

In tandem with the rollout of the licenses, UBS set up AI training courses touching all corners of the firm, tailored to people's existing skills and place in the organisation. At one extreme, it has launched a year-long data science training programme, attracting around 1700 enrolments across the firm. Another element is a reverse-mentoring programme. This sees senior members of the group, up to the executive board, sit down even just for an hour every few weeks to learn from someone who might be more junior to them but more in tune with the latest gen AI developments. All employees in the firm have completed a training course on the responsible use of AI. It also has prompting competitions, to help populate its growing prompt libraries.

One indicator of the seriousness of its drive to get as many bankers as possible familiar with gen AI is the way it counts AI prompts expanding across the organisation. It's a crude measure, as the bank is not tracking what its staff are asking. Yet the number of prompts is growing exponentially as its rollout of gen AI tools gathers pace. It recorded a million prompts in January alone, compared to 1.75 million for the whole of 2024.

GEN AI IN PRACTICE — DBS

Building an in-house community of data professionals at DBS

People are the most important factor in deploying AI successfully, so how could DBS, southeast Asia's biggest bank, be a more attractive destination for AI talent?

In 2023, the bank created what it calls a Data Chapter, bringing together 700 data analysts, translators and scientists from across DBS, and advertising to current and aspiring data professionals the benefits of a career at the firm. It is almost equivalent to a medieval trade guild: involving training, networking, and career advice, as well as structuring compensation for these employees.

Part of the aim was to give data professionals exposure to the variety of use cases in an institution as large and multi-faceted as DBS. This allows staff to work on different things, foster new skills and progress their career - discouraging those with itchy feet from moving away.

The bank has also sought to demystify data and analytics by rolling out basic training courses, supplemented by self-learning modules for those who wanted to deepen their knowledge of the field. Fundamentally, it has sought to change the bank's culture around data so that staff would ask 'what's the data saying', rather than acting on hunches.

"Even when it's hard to measure the impact, allowing our people to use gen AI technology has an important effect in taking out the fear and magic. We need to make sure people learn about AI, get comfortable with it, and understand what it can do. It will prepare us for the next step of adoption which depends on technology and regulation."

Dirk Marzluf, Group Head of Technology and Operations, Banco Santander

1 million

Number of staff prompts to UBS' AI tools in January 2025 vs 1.75m in all 2024

Words of advice

"It takes a while for people to really understand gen AI and to make it part of the natural cadence of their day. Just putting the tool in people's hands is very helpful in terms of training and adoption, and having people become able to more tangibly conceptualise where and how it can be used. Every day, we get an email with a new idea from someone in the organisation, or they'll send us a note and say, 'This is incredible, it saved me time on something'. People are excited about having something that makes them more efficient. It's taking the no-joy work out of your day."

Teresa Heitsenrether

Chief Data & Analytics Officer, JPMorgan Chase

CHAPTER II

Using vendors strategically

How to choose between models

04

Success Factors

- Develop a framework for model selection
- Keep your specific needs in mind
- Assess performance, security, pricing
- Ignore advantages likely to be temporary

As the number and variety of LLMs proliferates, banks need to get better at choosing between models and the companies that provide them. Even the biggest financial institutions have moved away from building their own language models over the past three years. But gen AI has come a long way from a time when the answer was just **OpenAI** or **Microsoft**.

Companies such as **Mistral** and **Cohere** are also building LLMs which may be smaller, but more targeted to the specific needs of business users, notably banks. Competitive pressure, coupled with new and more efficient methods of building LLMs, is also bringing prices down for business customers.

Selecting models and their providers needs to consider pricing, but also capability and potential. While flexibility to switch remains vital, banks should be wary of preferring a model that is too niche. For example, a model developed in a minority national language might lose to other models that will work better in a variety of languages.

Data security is another crucial element. How well the model can run in the bank's on-premises data infrastructure and private cloud, and how much the bank can delve into the model and finetune it are key questions. This means that banks will need to have a mix between closed-source models like those built by OpenAI, **Google's Gemini**, and Cohere, and open-source models from the likes of Meta and Mistral. The model selection process should include divisional staff as

well as central technologists to meet data security concerns and the goals of enhancing the experience of clients and employees.

LLM providers' pricing structures will also vary between banks, depending on their size, business model, and internal AI capabilities. Microsoft Copilot, for example, is easy to roll out, and embedded in Microsoft products that bank staff use all the time. But it is based on per-employee subscription prices, rather than usage. That cost will mount up in large retail banks with hundreds of thousands of staff. That is partly why some large banks like **JPMorgan Chase** and others in the US have built their own assistants, even if they are also contracting with Microsoft's cloud platform and using an OpenAI LLM.

Banks' on-premises and private cloud data infrastructure will also vary between banks, reflecting the differences between wider cloud strategies. Large banks have moved to a hybrid cloud strategy, with a mix of public and private cloud use, and on-premises data infrastructure.

Yet, banks subject to greater regulatory concerns on data sovereignty, such as in Europe – and banks with bigger institutional businesses – may have greater concerns about their use of models on public clouds. That must also be balanced with the desire to access the most advanced LLM capabilities quickly.

> "Our approach is to be model agnostic and to be able to switch models based on cost, privacy, and data classification. For a given use case, we could use a small language model hosted internally, or we may want reasoning and more advanced capabilities, which we can route to an alternative model."
>
> Hari Gopalkrishnan, Head of Consumer, Business & Wealth Management Technology, Bank of America

> "You need to have some internal capabilities to assess which model to adopt in which case. You need to develop an architecture that is flexible and recognise that you will have to use a different set of models depending on the different problem and use cases you want to develop."
>
> Mariona Vicens Cuyás, Head of Digital Transformation and Advanced Analytics, CaixaBank

Words of advice

“Not all models are born equal. It depends on what you're trying to do with them, and how much you're trying to optimise for performance, capability or cost. You always want to be able to answer what do you know as a company, what are you doing with AI, who's using it, and for what. Open source and semi-open-source models are cost-effective, and as the models advance, that competitive pressure is important, so that as you work with your providers, you can point to an anchor from a costing perspective. Today’s view might be different in six months, and different again in 12 months. You need to set a high bar for quality from the outset. When the models pass that test, costs are a big concern. We also prefer to have a more strategic relationship with the providers. Keeping the whole architecture as simple as possible is a priority, and a core decision-making criterion. If we had to deploy new infrastructure to support a particular model, that would not be favourable.”

David Griffiths
Chief Technology Officer
Citi

In-house platforms can help you partner

05

- Favour strategic relationships
- Maintain pricing pressure
- Stay agile and switch when needed
- Invest in a flexible LLM platform layer

Being able to build strategic relationships with LLM providers and the cloud platforms that they run on is one of the most important adoption concerns. OpenAI, Microsoft, and Google have been the most common partners for banks to date. Some banks enjoying especially close relationships with companies like **OpenAI** have been more aware of what is coming technologically, allowing them more lead time in thinking through their AI strategies and preparing the organisation to adopt LLMs.

UBS has built an especially strong relationship with **Microsoft**: embarking on an industry-leading roll-out of Microsoft Copilot and constructing an extensive employee training programme to go with it. BBVA has also based much of its work in AI on a strategic agreement with OpenAI, including gen AI systems to help with credit assessments, legal queries, and client sentiment analysis. The latter is providing BBVA with 3000 licenses to use the business versions of its GPT models, which come with greater protections around data, having piloted the partnership with its technologists.

Yet banks know they need to switch between models and providers in an agile way, to access the most advanced and most cost-effective AI technology. In early 2025, Chinese AI firm **Deepseek** released its R1 model, highlighting how banks may be able to access cheaper and more customisable open-source LLMs that perform similarly to expensive closed-source models from companies like OpenAI and **Google**. The Stargate AI infrastructure project in the US has also

triggered changes in OpenAI's partnership with Microsoft which previously meant OpenAI models were only available on Azure.

Developments like this may validate the approach of building firm-wide AI platforms, sometimes characterised as in-house LLM wrappers, which act as proprietary abstraction layers from the model providers, making it easier for banks to switch between models while ensuring security protocols – even if the construction of the wrappers slows down the rollout.

One notable platform in this mould is LLM Suite from **JPMorgan Chase**. The bank had given access to gen AI tools to around 200,000 employees, via LLM Suite, by early 2024 – about two thirds of its staff globally. This included a tool integrating gen AI into the workflow of customer-facing employees, giving them timely information that helps with their calls. It also offers a desktop question and answer function. Like Microsoft Copilot, it integrates OpenAI's models. However, the fee to Microsoft is based on usage of Azure, not Copilot subscriptions. Crucially, it also allows the bank to swap models without changing the interface.

Other banks, particularly in the US, have built similar platforms to LLM Suite, often integrating OpenAI's models initially. **Morgan Stanley** was relatively early in rolling out an OpenAI-powered employee assistant with a question-and-answer function, plugged into permissioned company databases, in 2023. It has also rolled out a product called Debrief for summarising and suggesting follow-ups for client calls. Roughly 50% of Morgan Stanley's staff are using an AI solution. Bank of America also has a tool to help relationship bankers prepare for client meetings, bringing together client data.

It's not just US banks taking this approach. **Banco Santander** has a platform that also aims to commoditise third-party LLMs and plug them into a proprietary system, called Luiza. **DBS** has deployed its own version of Chat GPT, called DBS-GPT, available to almost all its 40,000 employees, and referencing the bank's knowledge bases including unstructured data, such as PowerPoint presentations or email. Using Google's Gemini and Anthropic's Claude, DBS-GPT does not involve the bank's own LLM, but nor is it limited to OpenAI or a single LLM provider.

> "Our strategy is to make sure that we are not sticky to a specific model or to a specific company. You need to make sure that whatever investment you do is relatively independent of the model that you have underneath it."
>
> Dirk Marzluf, Group Head of Technology and Operations, Banco Santander

200,000

Number of JPMorgan Chase staff with access to LLM Suite

GEN AI IN PRACTICE — Goldman Sachs

How Goldman Sachs' handles radioactive AI

Goldman Sachs has deployed a firm-wide developer platform to channel and fine tune various AI models, connecting proprietary data in a secure way. The idea is that while the LLMs are like uranium, the developer platform should act like an AI nuclear powerplant – with all the needed safeguards. It's not a niche tool, as Goldman counts about a quarter of its employees as developers.

A new AI assistant for Goldman employees has emerged out of the platform, offering a common chat interface to staff using secured LLMs. The application was already available to 10,000 employees in early 2025. The bank plans not only to expand the number of people with access to the tool, but also to connect it to more data sources. It aims to incorporate AI models with greater reasoning capabilities, so that it can evolve to perform tasks in an agentic framework.

The firm is piloting a variety of other applications based on the developer platform, customising models to improve capabilities around data modelling and visualisation in its markets and in its asset and wealth management businesses – and building a new co-pilot designed to save time for investment bankers pitching for their next deal.

GEN AI IN PRACTICE — citi

Citi's platform-based gen AI rollout

In early 2025, Citi rolled out three new gen AI tools to a broad range of employees. Citi Stylus allows staff to summarise and ask questions about documents. Citi Assist is a chatbot for information about Citi's policies and procedures. Finally, Citi Squad is an add-on to the gen AI coding tool Github, that helps developers review and generate documentation about the code.

Behind the rollout of these tools is a desire to demystify the technology in a practical, low-risk way. The internal platform also allows Citi to develop more sophisticated gen AI use cases, serving as an entry point for multiple AI models. Citi has a close partnership with Google covering cloud technology and AI models, while still mindful of not becoming bound to specific providers. As AI keeps evolving, this allows the bank to stay nimble and adapt to the risk of regulatory fragmentation.

50% Morgan Stanley's staff are using an AI solution

"By having this layer in the middle, it makes it quite seamless to swap models in and out in the background for our users. Latency, cost, speed, accuracy, all these things matter, and different models are going to be best for different things. We wanted the optionality of making sure that we were able to leverage the best. We never want to have a sole dependency on any critical provider."

Teresa Heitsenrether, Chief Data & Analytics officer, JPMorgan Chase

Bigger is not always better

06

- Look beyond the big LLM players
- Customisation is key
- Be thoughtful of your advantages
- Adapt to your jurisdiction

Part of the wow factor of LLMs is the range of their capabilities, and the subsequent extent to which banks can adapt the technology to their needs by finetuning them. One example is the programme **ING** designed for doing customer due diligence in wholesale banking. In this, a **Google** LLM helps staff check the reliability of financial statements, such as whether they are audited by big accounting firms. The work in deploying LLMs like this lies in managing risks around reliability and data security, rather than the construction of the model itself.

LLM providers like **OpenAI** and **Anthropic** offer agreements to bolster data security and to avoid client companies' data being used to train their models. The latter is a particular concern for banks' competitive advantages, given their insights into tens of millions of customers. Data-retrieval techniques, meanwhile, can improve the accuracy of outputs.

Some banks, especially in customer-facing use cases, may prefer to use small language models, to avoid gen AI risks such as hallucination. Others may want a greater degree of customisation in the use of LLMs than what closed-source models such as Google and OpenAI can offer. This is becoming easier due to the emergence of cheaper and more transparent LLMs, including models built for institutional use, outside the US and China.

Banks will have varying levels of comfort using LLMs running on public clouds. Their choices may reflect their business models and the sort of

customer data they typically process, the regulatory intricacies of their geographic reach, and any environmental sustainability considerations, given how much energy LLMs use.

Other model capabilities may be sufficient if they are slightly behind the latest models running on public clouds. In some cases, if it allows the bank to feed sensitive data into the LLM, smaller and more customisable but less cutting-edge LLMs may help the bank be a pioneer in the use of said technology in finance. Finally, physical proximity to the LLM provider may help the bank develop a stronger relationship with the vendor.

There are also LLMs being built specifically for financial services, raising the prospect of AI's ability to answer more complex financial questions, such as providing investment advice. One example is **Aveni**, a company based in Edinburgh which is attempting to build a LLM for finance by training it on what it regards as reliably accurate financial data, with backing from UK banks **Lloyds Banking Group** and **Nationwide**.

Growing out of earlier work on reinforcement learning in electronic equities trading, **Royal Bank of Canada** has also developed a large transaction model, capable of different tasks, and trained on RBC's transactional history. RBC has already used the model to help predict clients' interest in certain products and understand credit risk based on complex data inputs. It now plans to integrate that project into LLMs built by fellow Canadian company **Cohere**.

To tap AI for credit decisions, banks need much greater understanding of how the model works than is possible in closed-source models. Even with custom-built models, they must remain wary of the risk of replicating bias via AI. Banks also need to decide whether investing time and money in custom-building is worth it given the number of non-bank players, often with much greater financial, political, and scientific resources.

> "Credit is the highest risk part on the risk slope in using AI. If the transformation is enduring, what is the rush to get it done? Let's get it done right. Nowhere is that more important than in credit, because it fundamentally affects people's everyday lives, in a deep way. We're still using traditional AI and machine learning analytics for credit, with humans in the loop. Using gen AI in credit is more of a multiyear journey."
>
> Prem Natarajan, Chief Science Officer, Capital One

GEN AI IN PRACTICE

RBC and BNP Paribas invest in homegrown LLM partnerships

BNP Paribas' partnership with French AI company **Mistral**, its favoured LLM provider, reflects a long-standing strategy at the French bank to focus on using a private cloud, built in its case with **IBM**. Although the bank is using other models, including Microsoft Copilot, it wanted to use gen AI models involving data such as contracts and transactions. which cannot run on a public cloud.

At Mistral, based in Paris and founded by former engineers from Google and Meta, the offering is more targeted to customers with such demands. Other customers include French insurer Axa (whose asset manager arm BNP Paribas is in the process of buying) and the French defence ministry. Aside the models' performance, Mistral suited BNP because it allows the bank to use a smaller model at times, benefiting from lower costs and environmental impact.

Similar concerns loom large in Royal Bank of Canada's partnership with **Cohere**, a Canadian AI focusing on regulated industries such as banking and healthcare. Even with assurances as to how its data would flow back and forth, RBC did not feel comfortable using cloud-based models for specific customer data. It also felt that a more targeted and customised model, which it could run on its own data infrastructure, would reduce the risk of hallucinations thanks to better control over inputs and outputs beyond any amount of mainstream LLM finetuning.

"We've really been holding back on what we could do with gen AI in the bank. We feel that by having these technologies run on our servers, with secure access to RBC data, we can unlock a ton of opportunity for how we use them. You can do more interesting things with them when they can see our data sets."

Foteini Agrafioti, Chief Science Officer, RBC

GEN AI IN PRACTICE

Nubank's acquisitive strategy to leverage gen AI for credit and fraud

Nubank processes vast amounts of data from its 60 million client base every day, free from the legacy IT infrastructure that plagues older banks of the same size. The Brazilian neobank illustrates the potential for gen AI for those sitting on a modern tech stack and piles of customer data.

It's still a work in progress, but there is a degree of excitement about the possibilities. The neobank's acquisition of Silicon Valley data intelligence company **Hyperplane** last year reflects the bank's firm belief that it can use gen AI principles to build its own foundational models to predict issues such as credit risk and fraud. It picks up on signals hard to spot by humans, and finds patterns within the data, including working from less structured data which traditional machine-learning tools were less able to use.

Nubank's strategic thinking was already heading in this direction when it came across the Hyperplane team, which was building such models. While Nubank has been integrating Hyperplane's functionalities into its own in the latter part of 2024, it's also been carrying out testing to bring associated applications into production. Early signs were promising.

"We are currently reliant, to a certain point, on third-party large language models. But we might get to the stage where we need to own the models ourselves. As some applications require very controlled models, and there are also potential regulatory implications, this is an open question."

Jon Ander Beracoechea, Chief Scientist, BBVA

Words of advice

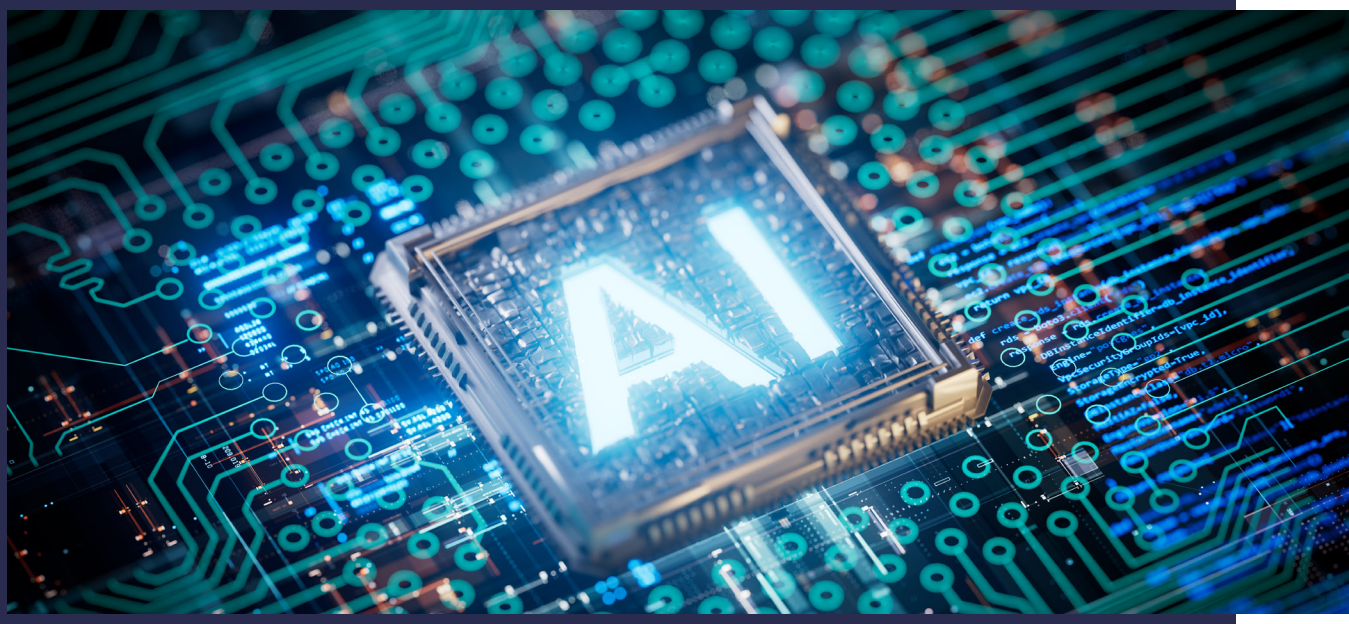
“Responsible development of AI comes with standards and governance around the cost: how do you neutralise, as much as possible the investment? How do you build assets at the group level to serve various entities that require the same kind of service? How do you ensure that the quality of the data is sufficient to run gen AI models? How do we control bias in the data and in the modelling phases of the project? It comes with transparency and explainability of the model, to be able to really understand the outputs. And it comes with accuracy. You can measure accuracy well when you work on specific data, but when you work on an answer to a question, it’s a bit more complex. Finally, it comes with environmental sustainability, making sure that we don't consume more resources than necessary.”

Hugues Even

Chief Data Officer, BNP Paribas

CHAPTER III

Embracing the future



Low-risk use cases can have high-value returns

07

- Don’t overlook simple use cases
- Start with internal deployment
- Develop skills to improve AI output
- Allow users to check at the source

Given the experimental nature of gen AI, banks must start using the technology for low-risk internal use cases, restricted to less sensitive data sources. Low risk use cases can be impactful, however, and not just because they help the bank to learn how to implement AI.

Among the early wins are LLM chat functions restricted to answering questions from less critical parts of banks’ internal knowledge bases, such as product descriptions, or policies and procedures. Given the complexity of large incumbent financial institutions, staff in branches and call centres can find it hard to locate answers to customer queries. Gen AI tools have helped reduce customer handling times.

Discover Financial Services, for example, found a tool based on **Google Cloud** reduced policy-and-procedure search time by 70% in its call centres. Gen AI tools are also able to run in the background of the call, measuring sentiment and offering help in real time, such as posting links to products that appear relevant to the conversation, as well as summarising it and suggesting follow-up actions.

Other relatively low-risk use cases include helping marketing departments with basic content, including for social media, as well as designing investment-banking pitchbooks, for example. Even in these cases, banks must develop processes and skills such as prompt engineering to validate and improve the accuracy of gen AI output. That might start with identifying risks, creating a golden source of truth, establishing validated examples, and then checking how the system

performs, until a certain level of consistency and accuracy is achieved. Gen AI answers should also allow users to see where the information comes from.

Given large banks employ tens of thousands of IT developers, software is one of the areas in gen AI realising productivity gains early on, through tools such as **Microsoft's GitHub Copilot**. **Goldman Sachs** has rolled out gen AI tools including GitHub to its entire developer population – around 12,000 people, or one in four employees – and seen efficiency gains of up to 20% as a result.

Other banks have similarly encouraging results. **HSBC** has rolled out GitHub Copilot to 10,000 developers, recording efficiency gains of between 15% and 30%. Banks are also deploying tools for a wider range of developer tasks. **Citi** has released an add-on for GitHub Copilot, called Citi Squad, which creates code reviews, suggests improvements, and helps its developers build instructions for how to run software, including via diagrams. **Morgan Stanley** has benefited from lower cost and higher quality in converting legacy code languages by using gen AI tools.

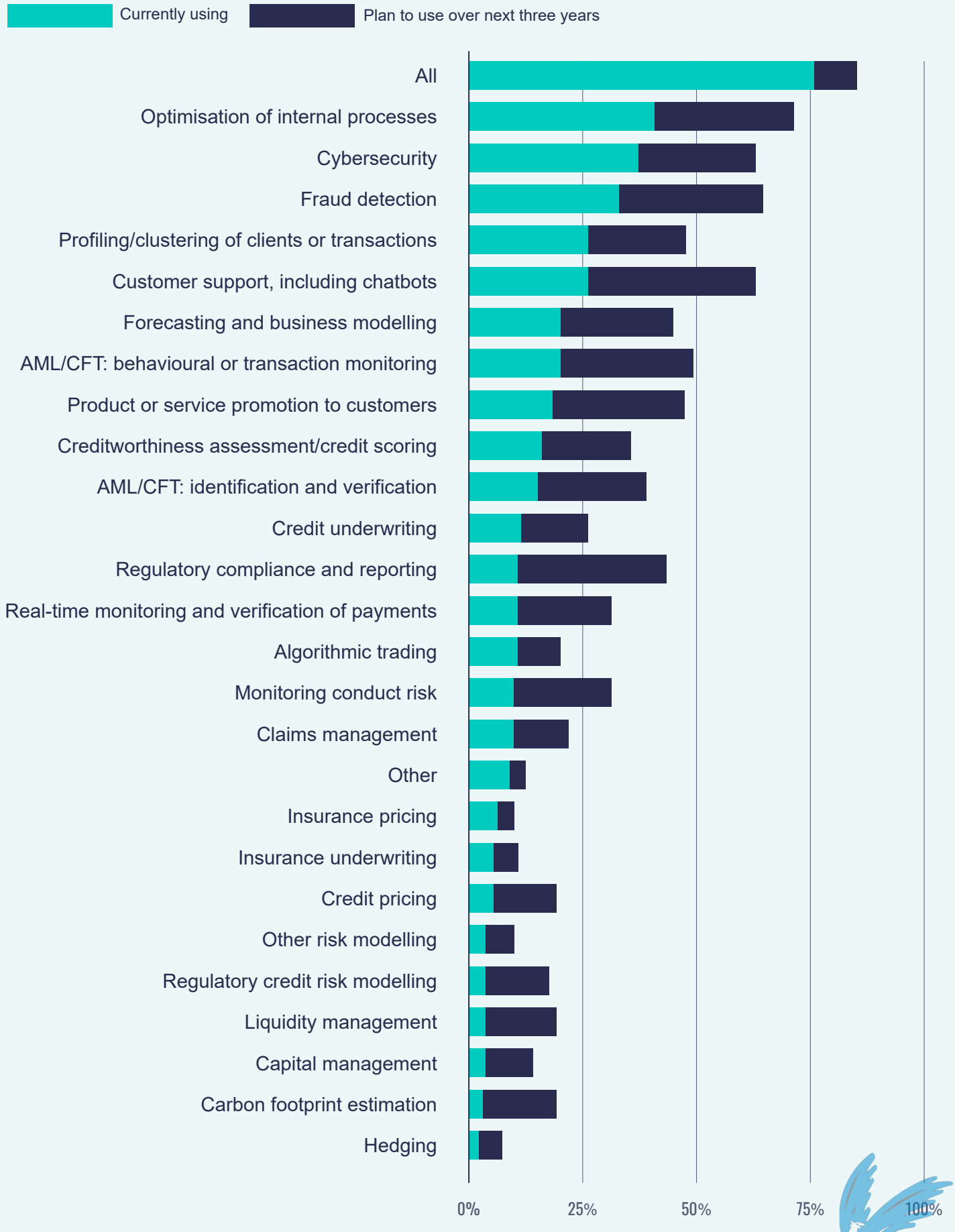
Looking ahead, agentic systems are advancing especially fast in software development, with more potential for banks to realise efficiency gains around testing code, for example.

15%-30% Efficiency gains recorded by HSBC after rolling out GitHub Copilot to 10,000 developers

"We had some banks where 30% of the contact centre was churning out every couple of months because staff couldn't find the answers they need. With gen AI, they're able to get those answers faster. Now we're seeing more financial advisors and relationship managers using gen AI to help them get ready to meet clients, so it's moving from helping banks with revenue generation, not just servicing customers."

Daragh Morrissey, AI direct, Microsoft Worldwide Financial Services

Percentage of firms currently using or planning to use AI



Source: Bank of England's Artificial intelligence in UK financial services 2024

GEN AI IN PRACTICE — BBVA

BBVA uses gen AI to update mortgage collateral data

Gen AI might not be bringing in billions of dollars in extra revenues yet, but the benefits in productivity are steadily emerging – even if, in some cases, they go almost unnoticed.

BBVA's pool of data on mortgage collateral is a case in point. The Madrid-based bank's AI strategy has included rolling out 3000 **ChatGPT** enterprise licenses across the firm. Recently its mortgage team sought to leverage access to AI tools when they saw that a database on collateral values needed rebuilding, accounting for wider market changes in home values. In a matter of days, the team – with relatively little specialist knowledge in programming or in the bank's internal IT systems - was able to build a system that would verify and update the collateral values on BBVA's systems, based on public data on the housing market. Previously, the task would have required a relatively large project to be carried out by a technical team.

70% Drop in policy-and-procedure search time at Discover's call centres, thanks to gen AI

"One of the most frustrating parts about writing code is when you might miss a parenthesis and spend five hours trying to debug your code, when you just had a small syntax error. GenAI has eliminated a lot of the frustrating parts of being an engineer, so you can dedicate your brainpower to solving complex problems. It's made engineering more fun, safer and more efficient."

Ian Glasner, Group Head of Emerging Technology, Innovation and Ventures, HSBC

Broad capabilities mean broad risks

08

Success Factors

- Expect the unexpected
- Test and learn, repeat
- Customer-facing deployment takes its time
- Humans stay behind the scenes for now

Banks' chatbots are universally frustrating. Little more than pathways to lists of frequently asked questions, they all too often fail at matching questions to answers, with human agents often none the wiser either.

Gen AI now offers the prospect of virtual agents which can interpret customer needs more easily, retrieving relevant information, or routing to an agent better equipped to help. However, banks must approach customer-facing use cases for gen AI with greater care than for copilot-type tools for staff.

LLMs offer an impression of being easier to deploy than traditional AI, because they work in natural language and have broad knowledge stores. That breadth of capabilities means LLM-powered chatbots have the potential to engage with an almost infinite range of questions and answers. But their breadth of training can also make LLMs less predictable. Some banks have consequently decided that gen AI today is simply not advanced enough for customer-facing use yet, even if they expect that those problems will be solved in time.

Such caution is warranted, particularly as the most advanced gen AI models operate on closed-source to protect their intellectual property. In more traditional AI systems built in-house, banks could have a more precise understanding of what could go wrong.

Gen AI opens a range of risks that are harder to know. This goes beyond hallucinations, where models make up answers. Data scientists at **BBVA**, for example, were recently working on an application which

would allow the bank to use voice interactions in gen AI-powered systems. In one of the tests with an audience outside of the bank, it answered in Italian out of the blue, something it had not done before.

Banks are, nevertheless, testing applications in which their customers interact directly with gen AI models, rather than just benefitting from a human agent who might be asking the model questions. They are finding ways around the associated risks by customising models based on internally curated data, while putting in place safeguards and going through multiple testing stages, each one involving new learning and improvement. In some cases, banks have got comfortable enough to put customer-facing use cases into production.

This could herald a deep shift in the way customers interact with banks. Ultimately, gen AI offers the prospect of the bank serving its customers using AI agents in much the same way as it used to do with humans.

These will be able to navigate the bank's systems according to the customer's needs, using natural language: moving beyond the self-service model of the mobile-banking era, where the customer needs to navigate through the app. In the future, the customer may also be able to converse with the bank by speech rather than text. Even if a human is involved at the end of the interaction, the hope is that solving problems will take much shorter.

> "Even if we froze all development on gen AI, with the models and tools that exist today, we're in the early days for its adoption in banking. Just knowing how to apply these tools – how to plug them into your processes, and use them effectively in production, at scale - is a learning curve, requiring a lot of work. It's like mobile banking in 2008 or 2009, when you only saw the full potential five or six years later. I think there will be a similar curve in this case."
>
> Vitor Guarino Olivier, Chief Technology Officer, Nubank

150%

Increase in customer satisfaction among customers using gen AI in NatWest's Cora+ chatbot

GEN AI IN PRACTICE — NatWest

A gen AI-powered chatbot at NatWest

Internal chat functions for employees to tap the banks' knowledge repositories via LLMs are a widespread entry-level strategy. Now banks are beginning to extend that premise to customers: limiting LLMs to searching the bank's products and policies and allowing customers to converse with that database in natural language via the chatbot.

One example in the UK is **NatWest's Cora**, which has a gen AI add-on called Cora+. NatWest's first Cora chatbot went into production in 2017. With the advent of gen AI, as part of its experimentation with various foundational LLMs, the bank saw a chance to solve queries the bot is unable to answer within the chat under the older AI model.

In its traditional form, the bot directs the customer to a relevant part of NatWest's website, where the customer can search for the answer themselves. Using gen AI, it asks the customer's permission to retrieve the information itself and present it to the customer in its own words. Not all customers elect to take this gen AI option, which still involves human checks. When customers do elect it, however, NatWest has recorded a 150% increase in customer satisfaction metrics and 50% fewer handoffs to human agents.

GEN AI IN PRACTICE — BANK OF AMERICA

Bank of America's third way to customer-facing gen AI use

Like NatWest, **Lloyds Banking Group** is working to improve its in-app search capability by feeding data on its policies and products to a gen AI model. On the continent, **ING** has also launched an LLM-powered chatbot based on a Google LLM for searching the bank's internal knowledge repository on products and policies.

Other banks are taking a different approach, however, including in the US. **Bank of America** has invested heavily in its customer chatbot, Erica. Its data science team has made more than 55,000 changes to Erica to improve performance since launch in 2018 – including tweaks, expansions, and finetuning natural language understanding capabilities to make sure the insights are timely and relevant. Bank of America recently adjusted the model to recognise when a customer would be looking for hardship assistance linked to the Los Angeles wildfires, for example.

Although Erica does not deploy LLMs to generate what it says, removing any risk of hallucination, it is now using the transformative technology to help understand and categorise customers' questions. Thanks to all its updates, Erica's accuracy rate has increased from around 65% at launch to 95% today.

"We already have lots of back-office use cases with gen AI, which have a clear ROI, because they save time and effort by automating manual processes. Customer-facing gen AI use cases are new and have a high potential, and at the same time a higher risk, so more of a 'return on future' than always an immediate return on investment."

Rohit Dhawan, Director of AI, Lloyds Banking Group

65% → 95%

Accuracy improvement in Bank of America's Erica chatbot between 2018 and 2025

Words of advice

"We always make sure that gen AI tools are suggestions for our people. If the tools do drafts, our people are ultimately responsible for the final material. I don't think we are at the stage now where you can take output from AI and consider that a finished product. It's the same thing for developers. The code doesn't automatically go into production. It's a suggestion to the developer, who needs to carry out code review. It's very important that we keep that element of human supervision. Gen AI is incredibly promising. It's already realising some of that promise. But it can hallucinate and produce results that are not as accurate as we want. As we work on improving that accuracy, we also need to make sure that our process always includes a human in the loop."

Marco Argenti

Chief Information Officer, Goldman Sachs

Experiment with AI agents, but prioritise trust

09

Success Factors

- Recognise the threat of agentic AI
- Treat agentic AI as a growth opportunity
- Experiment with less material agent actions
- Build a framework around checking requests

As LLM technology advances and brings greater capacity for human-like reasoning, agentic has become the latest AI buzzword. Beyond the hype, agentic methods bring the world closer to a future in which AI systems can mastermind more complex tasks, using multiple AI sub-agents, plugged into different computer applications. This has deep implications for banks.

The definition of an agentic system is not set in stone yet, with some agents appearing to be little more than tools for pre-populating forms or routing emails. Commonly cited examples of more complex agentic systems include tools that can search for flights and hotels, and book them, based on a natural language interaction with a traveller.

Even this example could be useful for banks, whose staff often travel frequently, and similar tools could be used in procurement. Eventually, agentic AI could be able to not just search and summarise internal bank information, but also to do things for customers, such as opening an account or applying for a loan.

Banks cannot push the boundaries as much as less regulated industries, where failures are less costly and less reputationally damaging. As with customer-facing uses of LLMs in chatbots, automating customer requests using natural language does not itself pose much trouble. The difficulty lies in making sure the system has understood the need and responds correctly. If an e-commerce site sends a customer the wrong colour coat, it will be an easier fix than if a

bank sends money to the wrong person, because a voice system has misheard the customer.

Customer-facing AI agents will need to include several layers of checks to ensure they understood customer needs, before carrying out any action. For now, those actions will be low risk, such as booking appointments. In pilot phases, there may still be some involvement by a human agent in the back end, checking the AI's language and actions before it reaches the customer.

Aside the risks around hallucination, agentic AI could implicate banks in political controversies around job losses. While Swedish payments company **Klarna** has been vocal about the number of jobs it is shedding due to AI, banks will often need to approach this question with greater caution. On the other hand, legacy banks will struggle still further to carry large cost bases if agentic AI systems make it much easier for people to search and apply for the best-value financial products, potentially doing so almost automatically when the product becomes uncompetitive. While this could pose an existential threat to banks with outdated platforms, it could also be another opportunity for respected financial brands with up-to-date customer experiences to develop their apps into supermarkets for finance, and associated products.

This is why experimentation and investment in AI-powered customer service is not a choice but an imperative. At the same time, established banks' biggest advantage over big tech and fintech will be the reputation they have built as financial providers, often over hundreds of years.

> "We bring security and trust behind the interface. We need to make the app as interactive and user-friendly as the customer needs. But the opportunity for us is to create the agent that can move money, and find the best product for you - and why not book a flight ticket, why not reserve an Uber for you? We think we have quite an opportunity to fill that gap, in a trusted way."
>
> Gavin Munroe, Chief Information Officer, Commonwealth Bank of Australia

GEN AI IN PRACTICE — Commonwealth Bank

How CommBank agents judge and resolve card disputes

Banks are already experimenting with how gen AI agents interact with customers in a dynamic manner – making a judgement from the conversation and taking actions on their behalf. One example is in the disputes process at **Commonwealth Bank of Australia** (CBA, also known as CommBank). In Australia, banks have an unusually large role in managing disputes of various kinds, including e-commerce items that didn't turn up or arrived broken. The bank categorises the dispute and interacts with card schemes to determine whether it was fraud, misrepresentation, or some other problem, before seeking a resolution.

Previously, customers could log a dispute on the bank's app via a series of drop-down menus, where the app would also ask for things like an image of the item before processing the dispute, and additional documentation. With gen AI, the bank has rebuilt the process so customers can describe the problem in an intuitive way, with an AI agent assessing the issue in real time, including images sent by customers, and where necessary asking for more information during the same interaction.

Although more complicated disputes such as fraud will still take longer and involve more human orchestration, the bank hopes that the new system will allow it to process simple claims within minutes, instead of a day or two - allowing the customer's account to be credited almost immediately. But it is moving ahead cautiously. After initially testing with tight guardrails with employees last year, it's now rolling out the tool to customers for simple disputes, initially just focusing on evaluating customer intents, before moving onto AI agents resolving the problem.

GEN AI IN PRACTICE — Capital One

Capital One's agentic base in auto finance

Visitors to some car dealership websites in the US will now be presented with a new gen AI tool, built by Capital One. Called Chat Concierge, the tool allows users on participating dealer websites to compare cars and look at financing options. It will also give you an estimated price for a trade-in vehicle and even schedule a test drive by plugging into the dealers' customer relationship systems.

The tool is based on Llama, **Meta**'s family of open source LLMs, which the bank customised using its data. The tool seeks to understand the customer's needs based on the chat – desired price range, colour or manufacturer. It then formulates a plan for those needs, before checking it adheres to certain policies, and testing it with the customer.

As this is an early instance of gen AI going in front of customers, there is still a human behind the scenes, checking key outputs – such as whether the AI is suggesting an appointment time, not on Sunday at midnight, and that the tone and language is appropriate.

Why did Capital One prioritise this use case? One reason is the relatively low level of risk involved compared to a credit decision, for example. But another is the framework's scope to be transferred to other areas of the business, as the basic agentic architecture for Chat Concierge can be applied to other workflows for both customers and staff outside auto finance.

Looking ahead

What it takes to win in AI

The fluidity of AI capabilities, coupled with geopolitical and regulatory uncertainty, requires a well-considered approach to the transformational potential of this technology. Strengthened central functions that have a firm handle on the tech, cost and governance implications of cutting-edge AI must go together with a readiness to quickly abandon avenues of implementation in favour of others, even while maintaining the same end goal in mind.

As part of that flexibility, banks with sufficient resources must decide whether they should invest in proprietary platforms and interfaces that are better able to securely swap between the latest and best third-party LLMs, according to use, location, pricing and technical development.

Alternatively, banks with systemic importance to large economies could decide to invest in deeply customised LLMs and even proprietary foundational models, involving tight partnerships with open-source model providers that are geographically closer, and more focused on business users among regulated industries. Banks must also decide on the value of experimenting with LLMs for customer-facing goals, including agentic frameworks, with minimum human involvement by the bank.

Those decisions should reflect banks' aims in AI, their internal technological capabilities and investment budgets, and their risk appetite. They will also echo their place in the world, including their business models and the regulatory environment.

Don’t get swept up in the hype

AI is expensive. If banks are putting a lot of capital into AI, before long they will need to explain to their investors the value that it has generated. That is not an excuse, however, to ignore AI and underinvest. Return on investment will be clear in use cases with a low-risk profile and a measurable impact on productivity, such as improving customer satisfaction and call times.

Incumbent banks could have an advantage in AI if that allows them to build or customise models that leverage large customer bases. But it is not just the quantum but also the quality of data that matters. Banks need to have invested in data estates that are AI-ready. This is not as simple as a wholesale migration to a public cloud, which for many banks will be neither possible nor desirable. Greater use of public cloud may make the estate more flexible, with benefits in terms of costs and access to the latest third-party models, but centralisation and standardisation of data is also important to confidently use AI.

Above all, winning in AI will come down to how much banks can get their organisations to embrace the technology: not just so they are better able to tell an AI story to investors, but most importantly so they are better able to future-proof their businesses.

Concerns around security, hallucinations, and bias mean it is vital to have good governance over the bank’s use of AI, as well as constant dialogue with regulators. That is not just to avoid accidents, but also because banks’ status as regulated entities will be vital in maintaining their pricing power.

EUROMONEY

AI in Banking
Best Practices Playbook

Euromoney sets the **benchmark for excellence** across financial services, through exclusive awards and proprietary benchmarking, bringing data-led market intelligence to strategic decision-making.

Find our latest insights on **euromoney.com/reports**

For information on our market insights and best practices offering, please contact **Sofia.Cerqueira@euromoney.com**

More information on Euromoney's awards can be found on **euromoney.com/awards**

Gain competitive insights and read about the latest market trends on banking: **euromoney.com/banking**

Contact **subscriptions@euromoney.com** to gain access today.