In collaboration with:

UK Finance

accenture

Generative AI in Action: Opportunities & Risk Management in Financial Services

January 2025

This report has been jointly authored by UK Finance and Accenture, sponsored by the UK Finance AI Policy Committee.

Primary authors (alphabetically by surname):

- Sapan Dogra (Accenture)
- Michael Erras (Accenture)
- Callum Farrell-Morris (Accenture)
- Peter Hairs (Accenture)
- Carsten Maple (Alan Turing Institute)
- Walter McCahon (UK Finance)
- Tom Niven (Accenture)
- Bella Thornely (Accenture)
- Luma Zitani (Accenture)

Contributors and reviewers (alphabetically by surname):

- Heather Adams (Accenture)
- Claire Aldworth (Accenture)
- Molly Bloore (UK Finance)
- Liam Carlisle (Accenture)
- Ray Eitel-Porter (Accenture)
- Suhail Kapoor (Accenture)
- Anna Kharchenkova (Accenture)
- Ed Knight (Accenture)
- Chris Lane (Accenture)
- Joe Lovatt (Accenture)
- Adam Markson (Accenture)
- Phillip Mind (UK Finance)
- Jason O'Brien (Accenture)
- Lukasz Szpruch (Alan Turing Institute)
- Kuangyi Wei (Accenture)
- Mark Welsh (Accenture)
- Jen Wilson (Accenture)

Contents

01

Report overview

Foreword: UK Finance

The arrival of generative AI has sparked both excitement and nervousness among the public and policymakers. While enthusiasm is palpable at its fluency and ability to adapt to a myriad of tasks, concern remains around the risks and potential for unforeseen issues.

Yet the financial sector has a long history of responsible innovation. Its recent adoption of cloud technology stands as testament to how firms can successfully understand and manage risk when embracing new technologies.

And the experiences so far with generative AI indicates that this continues to be the case. This report shows that firms are innovating and expanding their use cases beyond chatbots, finding practical applications across different functions. They're also doing so prudently – expanding their use of the technology in step with improvements to their technical understanding and enhancements to risk management frameworks.

Rather than considering hypothetical use cases that may arise at some point in the future, this report is focused on the near-term. It is intended to illuminate how generative AI is being used in production currently or will likely be used in the near future. Similarly, it examines some of the most challenging features of generative AI and what measures are available, or emerging, to help manage the risks posed. We've looked closely at three use cases and three sets of risks, aiming to explore these well rather than taking a broad but shallow view.

Accenture has brought a wealth of expertise and knowledge to this endeavour, which would not have been possible without them. We are pleased to have worked together, alongside our members, to publish this report.

Although this paper is focused on the near-term, we're excited to see how else generative AI will be used over the medium and long terms. In particular, we look forward to seeing how it can be combined with other innovations, such as predictive AI or Smart Data, to enable new services.

Jana Mackintosh
Managing Director, Payments and Innovation, UK Finance

Foreword: Accenture

Generative AI has received a huge amount of attention across both the business and public domain. Some of the predictions for generative AI are well-founded, some are balanced, and some are relatively speculative given how quickly the technology and its use in business is evolving.

Regardless of longer-term predictions, a lot can be learnt from how the financial services industry has approached the topic over the past two years. Firms in the sector have often been at the forefront of experimenting, innovating and applying new technologies to how they run their businesses, while doing so in a considered and controlled manner. Generative AI is no exception.

We are delighted to have worked with UK Finance and its members on this paper. We have drawn on our 'on the ground' project experience of delivering AI solutions and strategies, while working closely with academic research teams, UK Finance members and industry researchers. The longer-term future of generative AI has been explored in many papers, and we're left with little doubt that it will have a significant impact on financial services. The focus in preparing this paper has been on the practical near-to mid-term uses and the pragmatic ways in which firms are navigating the potential risks.

The report highlights how, in a relatively short time, financial services firms have demonstrated the technology's potential to perform a variety of common tasks that were once the domain of highly skilled individuals and teams. We are also seeing examples of generative AI deployments driving real adoption, efficiencies and tangible benefits.

At the same time, firms are increasingly aware of the limitations, risks and uncertainties. Some of the considerations and concerns associated with generative AI appear unprecedented, in contrast to other technologies that were more familiar to existing teams and better addressed by established risk management approaches. This has left businesses both excited to explore what's possible but also cautious and measured in their approach.

We hope that this paper helps firms on their overall AI journey and brings a balanced perspective as they navigate the opportunities, overcome practical hurdles and realise the potential value of this exciting technology.

Peter Hairs
Managing Director,
UKIA Financial Services

Executive Summary

It's now just over two years since the launch of OpenAI's public chatbot ChatGPT on 30 November 2022. This event showcased the rapid progress of generative AI technologies in the public domain. Since then, firms across all industrial sectors have been adopting and experimenting with generative AI-based tools, evaluating the capabilities of the technology in the context of their own business models, risk appetites and organisational practices.

As a highly regulated sector, the financial services industry is currently working to explore the opportunities offered by generative AI, while complying with extensive existing and emerging regulations and standards.

This report aims to provide a factual overview of this emerging technology in financial services and is informed by UK Finance member experiences, expert discussions and Accenture research. It lays out seven near-term use case categories prevalent in the industry today and explains how to understand the different layers of a generative AI system.

We highlight that human expertise is at the core of effectively utilising and controlling the technology, with solutions typically relying on human oversight for training, interpretation and sensitive decision-making. Also essential is understanding the nature of the models underpinning these systems and the dependency on high-quality, structured and unstructured data to produce accurate insights.

Identifying near-term opportunities

Realising the value of generative AI starts with identifying the specific business functions or tasks that the technology is currently suited to. In financial services, generative AI can be applied across various functions including market analysis, financial crime and fraud detection, customer service automation and regulatory compliance.

Many related papers speculate about how generative AI could revolutionise the industry, automate complex activities and transform customer experience. Although much of this optimism may be warranted in the medium-term, the conditions to realise these possibilities have yet to be established as we await fully scalable technological foundations, internal operating models and greater clarity as to potential regulatory change.

In contrast, this paper focuses on uses that are either already deployed today or are in advanced stages of development. Through our assessment of these use cases, the prevalent use of generative AI is focussed on seven specific areas:

- Customer engagement and personalised marketing
- Knowledge management and information retrieval
- Software development and data management
- Intelligent workflow and email processing
- Fraud and financial crime
- Legal, contractual and compliance text analysis
- Desktop and meeting productivity

A key observation is that most live generative AI use cases at the end of 2024 are focused on exploiting relatively well understood capabilities of the technology, involve active human oversight and are focused on relatively low-risk processes or tasks. This report outlines three illustrative case studies in more detail to help characterise the current use and provide specific insights.

Return on generative AI investments

Illustrating the path from discussion and experimentation to implementation, the level of investment into generative AI from financial services is steadily growing, representing 12 per cent of technology investment in 2024, growing to 16 per cent in 2025[1]. Generative AI is being increasingly used to optimise relatively diverse processes, from improving risk management to enhancing customer service. UK-based financial institutions are investing in generative AI to automate the resource-heavy tasks listed above.

Encouragingly, survey-based industry research suggests that firms are increasingly seeing benefits from their investment. Satisfaction with realised return on investment (ROI) is high, ranging from 75 per cent of executives from large corporations to 86 per cent[2] of small and medium-sized businesses. Yet there are still many factors to consider when investing in deploying generative AI. Currently, there is limited ability to quantify the ROI and baseline costs required to introduce the technology.[3] A 2024 industry study highlighted that the most cited barriers to further adoption were implementation costs, quality and accuracy concerns, data security and privacy, as well as trust and perceived user acceptance. Despite these issues, it's encouraging to see firms are increasingly benefiting from their investments.

Managing uncertainties

Perceived regulatory uncertainty is another hurdle to large-scale adoption. Although broad regulatory approaches are becoming clearer, certain elements and details are yet to be confirmed. The landscape continues to develop rapidly, and the landmark EU AI Act has generated global attention. The UK by contrast has set out its pro-innovation approach in the 2023 AI Regulation White Paper. This outlines a principles-based framework, instead of a wide-ranging AI regulation in the EU style. The focus in the UK is on continuous updates to sector-specific regulations as required, with the Financial Conduct Authority (FCA) and Bank of England (BoE) / Prudential Regulation Authority (PRA) responses in April 2024 confirming this. An AI bill is planned in the UK, but its scope has not been determined.

It's therefore essential for banks, insurers and asset managers to continue adapting their compliance strategies to meet established regulations while also demonstrating the effectiveness of these changes. Those who have previously invested in strong compliance and risk frameworks will be particularly well positioned for this process and able to innovate safely.

1 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024 - analysis of an unpublished financial services sub-set of the data revealed that the 2024 investment is 12% and the expected 2025 investment will be 16% of overall technology spend on generative AI

2 Google - The ROI of Gen AI, A global survey of enterprise adoption and value, n.d.

3 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024 - analysis of an unpublished financial services sub-set of the data

Generative AI's risks

Generative AI requires both the management of risks that are well understood and others that are heightened in new or unique ways. To categorise these, we have utilised the National Institute of Standards and Technology (NIST) generative AI risk taxonomy that has emerged as a more mature standard specific to generative AI. We focused on three risk topics that are among the most relevant to the typical use cases being explored today and critical to financial services: (1) accuracy of outputs, (2) data privacy & security as well as (3) appropriate integration of third-party solution components.

We discuss the prevalence of these risks, in addition to mitigation approaches used in our case studies and a later in-depth section to inform the discussion regarding safe adoption of generative AI solutions.

The wide availability of generative AI also brings risks that the technology could be misused by bad actors, enhancing threats such as cyber-attacks or fraud. While this issue is important, this paper focuses instead on the risks associated with generative AI use by legitimate companies.

Illustrative generative AI case studies

To help illustrate real-life applications where firms are actively deploying generative AI, we have selected three case studies that provide a reasonable cross section of current generative AI use. These are covered in more detail in section four of this report:

Case study one: Customer complaints

Managing customer complaints is essential for maintaining consumer trust and regulatory compliance. The process involves recording, transcribing, investigating and resolving customer complaints, ensuring issues are addressed fairly and promptly. This labour-intensive and regulated process, rich in data, presents a significant cost and a strategic opportunity for deploying generative AI.

One firm has focused on deploying generative AI to this process. Following an initial pilot phase, generative AI was scaled to production to support:

1. Producing call transcripts.
2. Summarising key investigation fields from various sources.
3. Analysing documents provided by customers.
4. Drafting response letters including holding letters and final response.
5. Generating personalised feedback for agents based on complaint response.

Benefits included a productivity increase of 30-40 per cent and an improvement in both customer and employee experience.

The generative AI solution was not given decision-making powers, which rested with a member of staff who remained accountable for ensuring fair customer outcomes. Case managers' knowledge was actively utilised to refine and improve model performance. In addition, updating operating procedures and privacy documentation were considered critical measures for ensuring customers are informed that AI is being used to support case manager productivity.

Case study two: Know Your Customer

Know Your Customer (KYC) and Customer Due Diligence (CDD) processes are fundamental in financial services. They are governed by strict regulations and cover the entirety of the customer lifecycle within the firm, relying on processing a significant amount of structured and unstructured personal data. A generative AI accelerator tool was deployed to process documentation, extract mandated KYC information and populate this into an output format that was managed by existing systems of record. A very high level of accuracy was achieved.

This tool reduced processing times by 90 per cent for relevant clients.

To manage risks, the generative AI tool was also capable of running quality checks on the outputs by comparing them to source material and remediating errors or blank fields. A manual quality check was then done by an operator, who assessed the tool's output before concluding the process. To mitigate privacy and data security risks, the solution was fully hosted on a private cloud environment, with a private API (Application Programming Interface) call-out to a closed Large Language Model (LLM) in the firm's own private instance. Additionally, access rights were tightly controlled and documentation encrypted at rest and in transit. Data minimisation was applied and the generative AI environment configured for zero retention after a 30-day period.

Case study three: Software development toolkit

Financial institutions rely heavily on technology to run their businesses, drive strategic innovation, streamline operations and enhance customer experiences. The software development lifecycle (SDLC) is a promising area for generative AI-driven optimisation as a result. The SDLC is a structured process used by software developers to design, develop, test and deploy software applications. The aim was to accelerate progress through a large-scale data migration from an on-premises data centre to the cloud.

The firm adopted a generative AI toolkit for the requirements analysis and testing phases, including generating system requirements with an LLM and firm-specific inputs, as well as code conversion and testing ahead of human review. The deployment was based on a multi-agent team whereby the generative AI solution adopted different personas (e.g. designer, developer and tester) to critique and react to work produced by other agents and raise the quality of output.

This tool accelerated these SDLC phases by over 50 per cent with accuracy over 95 per cent.

This more complex arrangement, compared to case studies one and two, relied upon additional third parties contributing different solution components. To mitigate the associated third-party risks, the cloud service provider involved guaranteed the ringfence of client data and code. Specific processing capacity was also agreed to enable larger models and training performance, with Service Level Agreements (SLAs) underpinned in the vendor's contract. Previous model versions were back tested to avoid model drift. The modular approach enabled effective orchestration by a human in the loop (HITL), who also assessed the performance of each generative AI agent.

Conclusion and outlook

There is a meaningful progression being made from modest proofs of concept (PoCs) with generative AI solutions to real-world deployments that are delivering tangible value for business processes while managing and mitigating the associated risks. Firms are innovating but doing so carefully. These deployments illustrate a generally conservative risk appetite across the industry, as can be expected in a highly regulated environment.

The sector has many years of experience in safely deploying innovative technology and is home to mature governance, risk and compliance (GRC) capabilities. Financial services, along with their technology and delivery partners, have also demonstrated their ability to adapt these capabilities to emerging risks and eventually integrate their management into business as usual (BAU) processes. This is most recently demonstrated by the experience with risks across cloud and cyber security.

Firms should start considering these capabilities as strengths, equipping the sector with a competitive advantage. Given their track record of successfully integrating new technology, financial services firms should feel confident in their ability to responsibly adopt generative AI solutions and begin reaping the benefits.

This would help firms across the sector embrace wider-ranging applications than we have seen so far and enable scaled adoption across a second wave of use cases, in which greater value can be unlocked from generative AI solutions. For this to be possible, firms will need to continue to actively adopt generative AI in a way that achieves the efficiencies and savings promised by the technology, build in-house expertise and evolve their risk management and governance landscape.

Action at industry level can enable and accelerate responsible innovation with generative AI. Collaboration between regulators, the firms they oversee and those offering these solutions to the market could help clarify areas of uncertainty. Similarly, engaging customers early to understand the level of acceptance and address their concerns will be key to building trust and enabling adoption.

02

Today's generative AI landscape across financial services

Current adoption and near-term trends

By late 2024, generative AI had evolved from a niche area of data science research into a focal point for technology and process innovation for many firms. The release of ChatGPT in 2022 followed by a suite of enterprise-grade generative AI tools in 2023 marked a significant inflection point, gathering widespread attention across all industries, including financial services. Furthermore, customer expectations and the competitive landscape are likely to evolve because of the use of AI in everyday life.

Many organisations are now gaining hands-on experience with lower risk and lower complexity use cases, focusing on areas where they can deliver net business value while getting to grips with the technology. There is also an increased understanding of the specific nature and practicalities of the technology, helping firms focus their investments in AI.

In the UK, investment in generative AI has increased significantly over the past year, reflecting a global trend toward adopting AI-driven innovations across industries. The UK government, venture capitalists and tech companies have identified the potential of generative AI, supporting investments in research, startups, and infrastructure. This effort is bolstered by the country's strong academic ecosystem and its expertise in advanced AI technologies. The launch of initiatives like the National AI Strategy in 2021 highlighted the UK's strategic goal to become a global leader in AI.

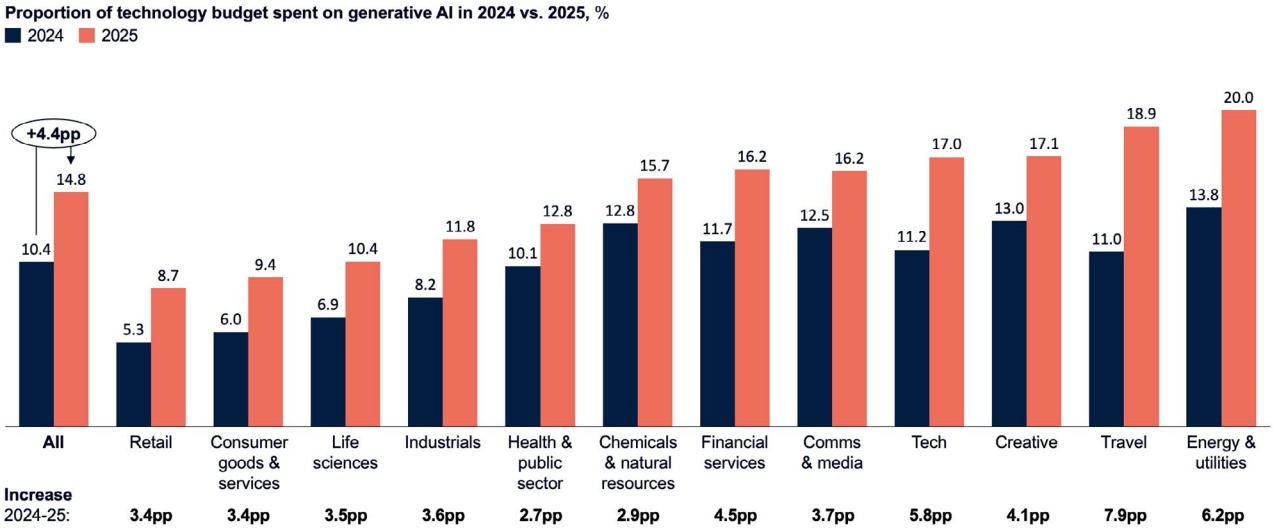
AI's potential is increasingly being realised through applications in healthcare, creative industries and finance, with notable growth in investments focused on expanding the uses of generative AI.

Many financial institutions have been exploring and adopting generative AI to some extent in 2024 and are likely to increase investments in 2025. A recent Accenture survey shows a material level of investment averaging 12 per cent of the technology budget, increasing to approximately 16 per cent of total technology spend in 2025.[4,5] While some firms focus on specific uses, 38 per cent of UK respondents to a recent survey have developed a broader AI 'roadmap' (ibid.). These incorporate multiple prioritised generative AI and broader AI-based sub-initiatives, focused on value, feasibility and risk appetite. While this constitutes a considerable proportion of firms, lacking this comprehensive level of planning still constitutes a constraint to adoption for many.

4 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024

5 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024 - Analysis of an unpublished financial services sub-set of the data revealed that the 2024 investment is 12% and the expected 2025 investment will be 16% of overall technology spend on generative AI, circa 1% higher than cross-industry averages

Figure one: Proportion of technology budget spent on generative AI in 2024 vs. 2025, %



Source: Accenture

In the UK financial services industry, generative AI is being actively applied in multiple areas of day-to-day operations. This can range from automated document summary and generation, to efficiency in fraud detection systems and processes, to software development. Generative AI-based tools are enabling institutions to analyse large unstructured datasets, quickly summarise complex bodies of information, provide natural language interfaces and significantly reduce the effort required for analysing and developing complex IT landscapes.

Most generative AI systems in financial firms are typically built on applying pre-trained LLMs to unstructured datasets to analyse or generate text. Although use of generative AI for image and video creation is finding applications in specific areas such as marketing, this has not been a focus for most financial services firms. The most common use cases are summarised in table one of this report.

Currently, a broad spectrum of solutions is being used, ranging from general-purpose generative AI tools to highly tailored and specialised systems designed for specific business needs. These include:

- General-purpose copilots: Widely available, multi-function tools serve as foundational assistants helping with common tasks such as drafting reports, summarising information and analysing datasets. Examples include Microsoft 365 Copilot, Google's Gemini for Enterprise and Anthropic's Claude. Publicly available, non-enterprise copilots may also be used for lower risk use cases.
- Vendor software with generative AI functionality: Software packages offering integrated generative AI functionalities designed to enhance specific tasks. Examples in technology management include Microsoft Security Copilot and GitHub Copilot, with many more functionalities anticipated in upcoming releases. Some of these solutions have specialised LLMs trained for a specific purpose using curated datasets and have

integrated generative AI use into familiar user interfaces.

- Low-customisation solutions: Some firms opt for low-customisation self-build solutions, allowing quicker development of generative AI tools. These offer limited flexibility to train or adapt the underlying models but do provide lower cost ways to tailor the model's behaviour to firm-specific needs (e.g. prompt injection).
- Highly customised solutions: Some institutions invest in highly customised generative AI systems with additional datasets used to re-train or refine the statistical behaviour of the underlying model. These can be trained on proprietary and/or financial services datasets, for example to offer financial service-specific insights and predictive capabilities.

In addition to investing in specific generative AI solutions, financial services firms are investing in initiatives to prepare their workforce for AI-supported work and in establishing broader AI capabilities. These include most notably:

- Skills and AI literacy: Many firms are investing in upskilling their workforce to work alongside AI tooling. AI literacy programs are becoming increasingly prevalent, aimed at educating employees on the benefits, limitations and ethical considerations of AI. Building human expertise in tandem with AI deployment helps organisations to effectively manage and optimise these tools.
- Generative AI risk management enhancements: Adopting generative AI is not without risks and financial institutions are mindful of embedding responsible AI practices within their deployments that ensure regulatory compliance, data privacy and consideration of specific operational risks. See section four for a more in-depth exploration of these topics.
- PoCs and hands on learning: Some institutions have initially opted for small-scale PoC projects to explore generative AI's potential without committing to large-scale rollouts early. These firms often opt for packaged AI solutions or partnering with third-party providers to experiment with pre-built generative AI systems. The focus here is on learning and quick wins, deploying generative AI in targeted areas before deciding whether and how to scale in-house or to continue relying on external providers.

Anticipated productivity gains from generative AI

Research has concluded that in the next 15 years, generative AI could present a significant productivity opportunity for the UK across sectors. According to Accenture modelling, the software and platforms sector is projected to experience the highest productivity boost, exceeding 30 per cent, with cost savings of £17.6 billion. Similarly, financial services sub-sectors such as capital markets, banking and insurance are anticipated to see substantial gains above 30 per cent, with potential cost savings of £9.7 billion, £12.7 billion and £3.4 billion respectively.[6]

These gains highlight generative AI's capability to streamline processes, improve automation and enhance decision-making in data-intensive sectors. In contrast, more traditional industries like energy, chemicals and automobiles show lower, albeit still significant, potential productivity gains, with estimated improvements below 15 per cent. Unsurprisingly, industries that heavily depend on complex datasets and digital platforms, especially in financial services and technology, are positioned to benefit most from generative AI advancements.[7]

6 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024
7 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024

Balancing generative AI with broader investments

Considering generative AI investments in isolation doesn't paint the full picture for many organisations. Many AI-supported processes and application solutions can have both predictive (also known as traditional) AI and generative AI components or solution options. Both types of AI present unique value propositions.

Traditional AI is often geared toward predictive customer and business performance analytics, fraud detection and risk management. In contrast, generative AI is gaining traction for its content-generation and natural language capabilities, incorporating both language understanding and content generation, for example to automate communications and improve internal efficiency.[8] Organisations are allocating their digital budgets almost evenly between the two technologies, ensuring they harness the benefits of both.[9]

The benefits of investment in these technologies can be maximised when complemented by solid digitised business and customer processes and strong data foundations. This is enabled through broader investments in scalable technology, accessible data sources and removing the general drag caused by complex and fragmented application landscapes. Continued investment in strategies that enhance previously unorganised, unlabelled and dispersed data is likely complementary to scaling the impact of generative AI. Generative AI in particular is also increasingly geared towards cloud-native platforms and a lack of cloud adoption could constrain firms in some uses of AI.

Finally, many firms are considering how generative AI can be a catalyst to incorporate a broader re-imagining of processes and services offerings. This ranges from rethinking how customers will interact with and access their services, to greenfield thinking about how functions and teams could operate alongside AI capabilities.

8 UK Finance - The impact of AI in financial services, 2023

9 Google - The ROI of GenAI. A global survey of enterprise adoption and value, n.d.

Sub-components of a generative AI solution

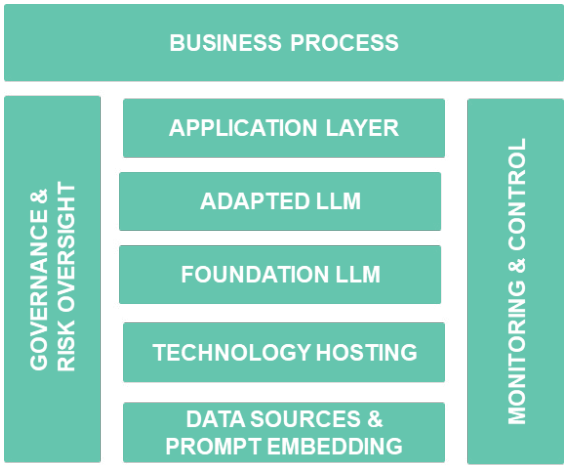
The typical architecture of generative AI in finance revolves around leveraging open or closed source LLMs and underlying data, including public data and enterprise data. Transformer-based models like GPT 4.0 pre-process vast amounts of data to train neural networks, resulting in a statistical model that can produce entirely new outputs, interpret and convert multiple languages, and infer context and intent using natural language. While these models are at the heart of generative AI tools, there are multiple components that are critical in building an enterprise solution.

Specific generative AI solutions will vary, but most implementations can be seen as comprising of a common set of building blocks. This framework helps reveal tooling considerations that in turn help deliver well-managed, robust and reusable generative AI solutions. The typical sub-components include:

- Business process layer: Generative AI solutions need to incorporate adapted process designs, training, guidelines and protocols that users need to follow to properly interact with AI systems and interpret their outputs effectively. This is essential to ensure AI tools align with their intended purpose while maintaining safety and performance through proper human supervision and verification.
- Application layer: The interface that enables interaction between the AI, the user and other functionality, including chatbots, virtual assistants and other AI-driven applications. This also includes embedded interfaces in broader end-user systems such as customer relationship management (CRM) platforms or case management software.
- Foundation LLM: Large-scale AI model trained on vast amounts of data that can be adapted or fine-tuned for various specific tasks and applications, rather than being built for a single purpose.
- Adapted LLM: Represents customised models that have been fine-tuned for specific use cases, like instruction-tuned LLMs, enhancing relevance and task performance.
- Technology hosting: Includes compute power, cloud services and hosting platforms (such as Azure, AWS, Google Cloud) that provide scalability and processing efficiency.
- Data sources and stored prompts: Critical inputs that supply the AI system with structured, reliable data to enable accurate outputs and adaptive learning.
- Monitoring and control: Integral for maintaining system health, tracking performance, and ensuring adherence to safety and ethical guidelines.
- Governance and risk oversight: A framework for risk management and compliance, safeguarding the solution against potential misuse and aligning with policies and regulatory standards.

While some firms are developing highly customised generative AI solutions, there is increased maturity in capabilities provided by third-party service providers, making it easier for businesses to integrate generative AI solutions into their technology ecosystems or simply buy software that has capabilities pre-configured. Increasingly, large technology firms not only provide specific components, but integrate these sub-component layers in their offerings. Financial institutions often adopt a 'buy for standard, build for differentiation' strategy, purchasing models for common use cases and building tailored solutions for unique or advanced applications. As uptake increases, many firms are establishing centralised AI and generative AI centres of excellence (CoEs) to help develop and govern AI deployment and promote reuse and standardisation of solutions.

Figure two: Generative AI solution components



Source: Accenture

Recent advancements in LLMs

Advancements in LLMs continue, addressing some of the limitations and aiming to constantly improve the performance and quality of the technology. While foundation models can achieve human-level performance in various tasks, such as summarising text, they still lack long-term memory, planning and reasoning. Further innovation in AI models and agentic capabilities will reduce these limitations. Recent advances include:

- Multi-modal models: These models are designed to process and generate data across multiple modalities, such as text, images, audio and video. Multi-modal models can integrate and interpret information from different types of data sources to produce more comprehensive outputs.
- Multi-agent solutions: Multi-agent solutions involve the use of multiple AI 'agents' that work together to achieve a common goal. Each agent can assume a different persona and task specialism to communicate and collaborate to solve complex problems. This interactive approach can enhance the overall performance of a generative AI solution.
- Specialised and smaller language models: These are scaled-down versions of LLMs, designed to perform specific tasks with fewer computational resources. While LLMs are trained on vast datasets, small language models are optimised for efficiency and require less computational power. They are particularly useful for applications where quick responses are needed or where the computational overhead of large models is not feasible.
- Reduced cost: As the cost of training and fine-tuning LLMs has decreased considerably, the landscape will start to shift away from almost exclusively buy strategies. The reduction in costs will lead to more in-house development, allowing firms to tailor models to their needs.

There will no doubt be further advances and innovations, both in generative AI models and ways of deploying and combining generative AI with complementary technologies, enhanced datasets and other forms of AI and analytics.

Prevalent use cases in financial services

Common use case themes

With many UK financial services firms now 18 months or more into experimenting with generative AI, a number have moved on from initial PoCs and are now establishing their generative AI-focused teams to take selected solutions into full deployment and adoption.

Across all UK industries in 2024, the most common uses of generative AI were in IT (60 per cent of survey respondents), customer service (60 per cent) and marketing (59 per cent).[10] As part of preparing this report, our team reviewed hundreds of established generative AI use cases from the public domain and their experience with financial services firms to identify common applications and approaches in the sector. The aim was to identify uses of generative AI that are either already live or expected to be fully adopted in the next 12 months.

Amidst the diverse activity among financial institutions, seven common themes have emerged, encapsulating the most prevalent use cases across the sector, summarised in table one below.

Some key observations are that most near-term uses involve single-agent deployments targeting productivity and efficiency gains and improvements to customer and colleague experience. There are relatively few examples within financial services that are aimed at increasing sales or revenue, although as some of the hyper-personalised marketing use cases mature this could change over time.

Most deployments are either internally facing, providing a capability for employees, or are closely monitored by an employee acting as a competent supervisor. Given that there are risks and uncertainties that still exist with these technologies, alongside a need for employees to familiarise themselves with using generative AI, it is appropriate that firms are starting with these types of uses.

10 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024

Table one: Common generative AI uses in financial services

Use case	Description
Customer engagement and personalised marketing	Generative AI agents that directly engage customers or support customer-facing processes such as call centre operations, complaints management and marketing.
Knowledge management and information retrieval	Generative AI-powered knowledge management solutions, providing employees with faster, more targeted access to enterprise data and documents.
Software development and data management	Generative AI solutions which assist across the SDLC, assisting with code generation and translation, code reviews, technical testing and data/metadata analysis and management.
Intelligent workflow and email processing	Generative AI agents that support high-volume email and document processing and/or workflow orchestration, such as assisting in lending operations.
Fraud and financial crime	Generative AI agents assisting in the collation, analysis and quality-checking performed as part of fraud and financial crime processes, such as KYC.
Legal, contractual and compliance text	Generative AI agents that assist in processing legal or compliance texts and associated artefacts like drafting agreements or assessing regulatory and policy text.
Desktop and meeting productivity	Generative AI desktop assistants or those integrated within core enterprise software, such as Google Gemini, Microsoft 365 Copilot and Claude for Enterprise.

Typical ROI and productivity gains

Economy-wide industry research indicates that most organisations are seeing a return on their generative AI investments: 74 per cent of enterprises realised ROI within the first year.[11]

Satisfaction with realised ROI is high, ranging from 75 per cent of executives from large corporations to 86 per cent of small and medium-sized businesses. While the benefits of individual generative AI solutions are being realised, many UK financial services organisations struggle to quantify the overall ROI of their broader investments. Only 37 per cent of business leaders in the UK say their organisation has the performance management infrastructure to measure and track the value of AI.[12]

Some of the most common use cases such as coding copilots for software development, AI-enabled chatbots, and content creation support offer significant productivity improvements, in some cases exceeding 30 per cent.[13] They also enhance the overall work experience by allowing people to spend more time on tasks they enjoy. In an experiment with the Accenture sales team, generative AI not only increased productivity but also boosted confidence by 34 per cent and the belief that the respondent was making a meaningful impact by 31 per cent. Generative AI also added to job satisfaction rather than detracting from it.[14]

11 Google - The ROI of Gen AI, A global survey of enterprise adoption and value, n.d.
12 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024
13 Google - The ROI of Gen AI, A global survey of enterprise adoption and value, n.d.
14 Accenture - Work, workforce, workers reinvented in the age of generative AI, n.d.

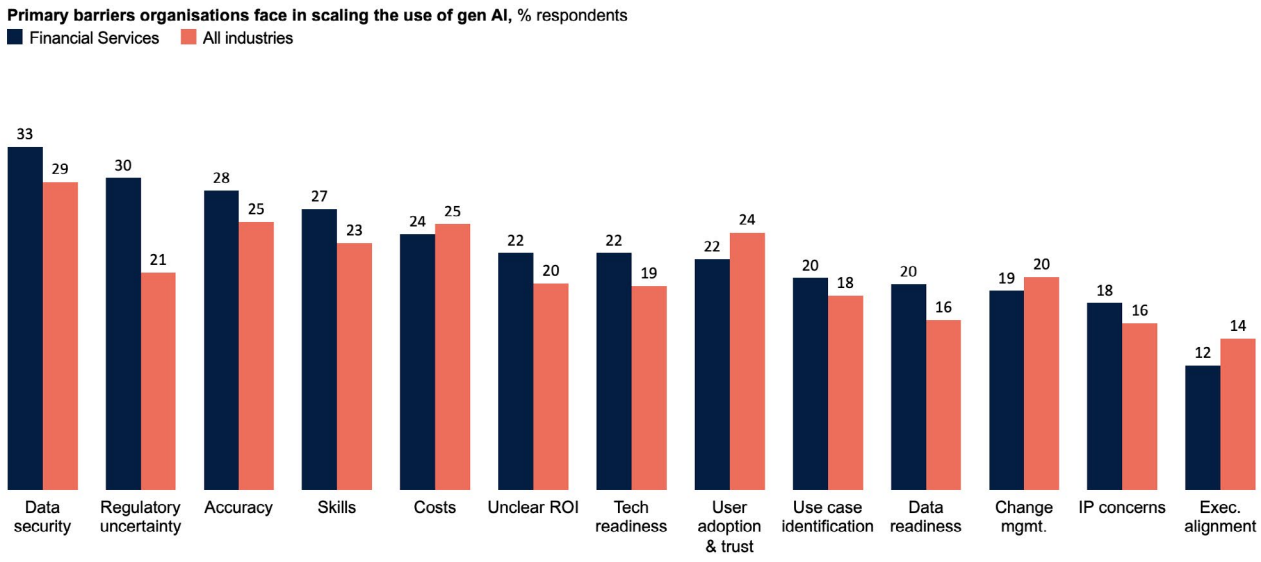
Overview of generative AI related risks

Financial services firms have developed strong capabilities to effectively manage complex financial and non-financial risks. When new technologies like cloud computing have been introduced and new risk types have emerged, such as those around conduct, businesses have adapted current risk frameworks, risk assessment methods and training for first-line risk owners and second-line risk managers.

This cycle has now been updated for generative AI risks. Some existing risks need to be reappraised in the context of this new technology but have clear mitigations, such as data security. In other cases, certain features or tendencies in generative AI affect risks in novel ways. For example, the potential generation of false outputs is a new source of conduct risk in some use cases.

Consistently, industry surveys show a set of recurring risks, concerns and issues to adopting and scaling generative AI, with themes such as security, accuracy and regulatory uncertainty commonly being cited, as well as broader concerns regarding costs and skills availability.

Figure three: Perceived barriers to scaling generative AI



Source: Accenture

Several broader AI risk taxonomies and definitions have emerged to help the understanding of generative AI-related risks and their integration into firms' risk management practices. Since this paper focuses on generative AI, we have utilised the generative AI-specific NIST risk classification to identify and discuss key risk topics associated with three detailed case studies considered.[15] To help focus the paper on the highest priority risk topics, three themes have been identified in collaboration with UK Finance and its members, informed by Accenture research and general industry perspectives.

Table two: NIST generative AI risk grouped by theme

NIST risk	Risk definition	Financial services relevance
Confabulation/ hallucinations/ fabrications	The production of confidently stated but erroneous or false content, known colloquially as hallucinations or fabrications.	High relevance
Human-AI configuration	Arrangement or interaction of humans and AI systems that can result in mistrust of AI outputs, automation bias or over-reliance on technology, misalignment between the goals or outcomes of the AI and those of its human users, deceptive or obfuscating behaviours by AI systems based on programming or anticipated human validation, anthropomorphisation.	High relevance
Intellectual property	Eased production of allegedly copyrighted, trademarked or licensed content used without authorisation and/or in an infringing manner; eased exposure to trade secrets or plagiarism/ replication.	Consideration for certain use cases
Toxicity, bias and homogenisation	Difficulty controlling public exposure to toxic or hate speech, disparaging or stereotyping content; reduced performance for certain sub-groups or languages other than English due to non-representative inputs; undesired homogeneity in data inputs and outputs resulting in degraded quality of outputs.	Consideration for certain use cases
Data privacy	Leakage and unauthorised disclosure or de-anonymisation of personal data, e.g. biometric, health, location or other sensitive data. Potential for inadvertent processing of personal data or unintended generation of inferences.	High relevance
Information security	Lowered barriers for offensive cyber capabilities, including ease of security attacks, hacking, malware, phishing and offensive cyber operations through accelerated automated discovery and exploitation of vulnerabilities; increased available attack surface for targeted cyber-attacks, which may compromise the confidentiality and integrity of model weights, code, training data and outputs.	High relevance
Value chain and component integration	Non-transparent or untraceable integration of upstream third-party components, including data that has been improperly obtained or not cleaned due to increased automation from generative AI; improper supplier vetting across the AI lifecycle; or other issues that diminish transparency or accountability for downstream users.	High relevance

15 NIST - Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, July 2024

NIST risk	Risk definition	Financial services relevance
Environmental	Impacts due to high resource utilisation in training generative AI model and related outcomes that may result in damage to ecosystems.	Consideration for certain use cases
Information integrity	Lowered barrier to entry to generate and support the exchange and consumption of content which may not be vetted, may not distinguish fact from opinion or acknowledge uncertainties, or could be leveraged for large-scale dis- and mis-information campaigns.	Lower relevance
Chemical, biological, radiological or nuclear information	Lowered barriers to entry or eased access to materially nefarious information related to chemical, biological, radiological or nuclear weapons, or other dangerous biological materials.	Lower relevance
Dangerous or violent recommendations	Eased production of and access to violent, inciting, radicalising or threatening content as well as recommendations to carry out self-harm or conduct criminal or otherwise illegal activities.	Lower relevance
Obscene, degrading, and/or abusive content	Eased production of and access to obscene, degrading, and/or abusive imagery, including synthetic child sexual abuse material and nonconsensual intimate images of adults.	Lower relevance

Using the NIST framework as a reference point, three broader risk themes emerged from discussions with UK Finance members and practitioners from banks, insurers, and asset managers. While this report will reflect on the relevant risks for financial services in the case study discussions, section four will focus in more depth on these three themes:

- Reliability of outputs: The risk that a generative AI solution provides incorrect, fabricated or inappropriate outputs for the given use. This can result from multiple aspects of the above taxonomy, including confabulation and bias.
- Data privacy and security: The risk that a generative AI solution involves inappropriate processing of personal data, leaks or generates information unintentionally or is hacked.
- Third-party considerations: The risk that third parties in the generative AI value chain are inappropriately controlled and do not conform to the expectation of the accountable solution deployer.

Strengthening governance and risk frameworks

To manage these risks, many firms are strengthening their AI governance and risk frameworks in line with their risk appetite in parallel to exploring, adopting, and scaling generative AI. Efforts have typically focused in the following areas:

- Executive ownership and sponsorship: Firms are examining how accountabilities and risk ownership relating to generative AI affects the roles of senior executives across Chief Information Officer, Chief Data Officer and Chief Operating Officer functions. In many cases, this has also led to realignment of accountability and cooperation among key risk, compliance and legal stakeholders.
- Governance and oversight forums: Many firms have established one or more governance forums focused on AI, both predictive and generative. Some firms adapt existing governance forums with additional AI supervisory responsibilities,

often aiming to concentrate complementary but rare expertise from across the firm.

- Policies and standards uplift: Many firms are uplifting current enterprise-wide policies and standards to define responsible AI usage in the organisation. Affected policies and standards typically include data privacy, cybersecurity, model risk management (MRM), third-party risk management and change management.
- AI inventories, risk assessments and processes: Most firms are creating an AI inventory to capture generative AI usage and enable risk assessment at a use case level. Risk frameworks are also being adapted to manage the effect of generative AI and AI more generally on different risks, in line with established processes such as MRM.
- Guardrails, controls and monitoring standards: Firms are designing and implementing guardrails and controls, and monitoring processes. These enhancements are to prevent misuse, ensure performance, maintain security and protect against potential harm while keeping the technology aligned with intended business purposes and ethical standards.

Related key considerations and common mitigation techniques are explored further in section three's case studies and section four's risk discussion.

Regulatory landscape

Global regulatory responses

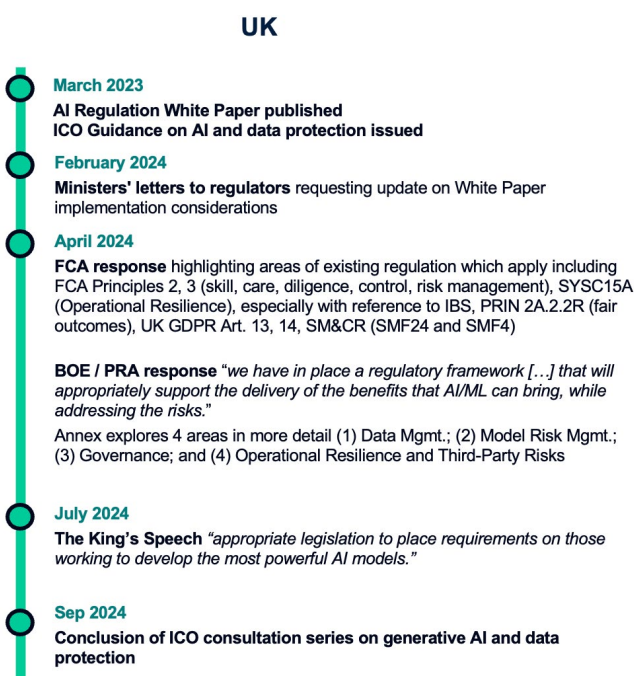
The rapid development of new solutions incorporating generative and predictive AI technologies has triggered regulatory responses worldwide, differing in terms of governance approach and regulatory instruments deployed.[16] Policymakers and regulators are considering how to balance the benefits of AI technologies against the potential harms to individuals, businesses and society. The emerging regulatory landscape for AI also interacts with new industry standards, such as the NIST generative AI risk framework.

There is an ongoing question of how regulation can keep pace with the rapid development of AI technologies, including generative AI. Different approaches are emerging globally.

The European Union has adopted the first comprehensive AI regulation with the EU AI Act entering into force in August 2024, and most of its rules becoming applicable by 2026.[17] Its extraterritorial scope makes it particularly relevant. The Act introduces a definition for an 'AI system' (aligning with the definitions from the OECD and Biden Administration Executive Order 14110), a risk-based approach for the classification of AI systems, as well as corresponding requirements and obligations on certain operators involved in the AI value chain.

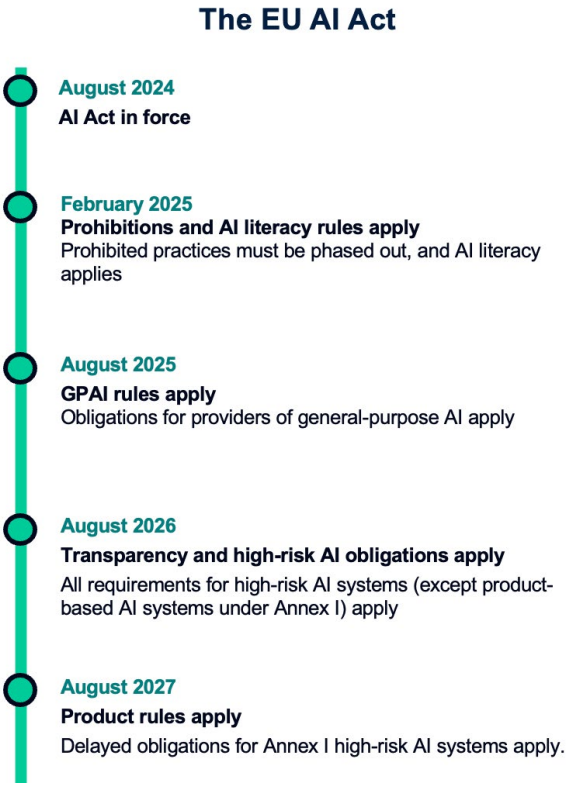
The Act also prohibits uses that are deemed an 'unacceptable risk' to the rights and freedoms of EU citizens, such as untargeted scraping of facial images or aspects of emotion recognition in the workplace. Specific mitigation strategies are required for high-risk use cases, which include AI for credit scoring or insurance pricing. The Act also makes other (limited risk) AI systems subject to transparency and disclosure obligations to ensure users are aware they're interacting with an AI system, where relevant.

Figure four: Key developments in UK and EU AI regulation



16 World Economic Forum – AI Governance Alliance: Generative AI Governance: Shaping a Collective Global Future (in collaboration with Accenture), January 2024

17 European Commission - AI Act enters into force, 2024



Source: Accenture

The UK's regulatory approach

In parallel, the UK has set out its pro-innovation approach to AI in a 2023 AI Regulation White Paper. This sets out a principles-based framework. Although this is not binding on regulators, the government provides regulators with central support to implement the cross-sectoral AI principles. This approach builds on existing industry-specific regulations and the regulators overseeing these, rather than creating a cross-cutting AI regime.

UK regulators, including the FCA[18] and BoE,[19] responded by affirming the relevance of these principles and referencing specific regulatory requirements governing AI use in the financial sector. These include FCA Principles two and three (skill, care, diligence, control, risk management),[20] SYSC15A (Operational Resilience), PRIN 2A.2.2R (fair outcomes), UK General Data Protection Regulation (GDPR) Articles 13 and 14 (transparency), and key roles under the Senior Manager & Certification Regime (SMF24 and SMF4). Additionally, the Information Commissioner's Office (ICO) has issued guidance on AI and data protection rules, covering best practices for fairness, transparency and accountability when using AI systems that process personal data. This includes explaining AI decisions, mitigating bias and assessing risks, supported by tools like the ICO's AI and Data Protection Risk Toolkit.[21] The ICO has also published draft thinking on applying certain data protection rules to generative AI and plans to update its guidance in due course.[22]

The new government elected in July 2024 indicated that it would legislate for AI. While the King's Speech acknowledged the importance of AI and the need for appropriate legislation, specifically for the most powerful AI systems, an AI bill was not introduced at this time.[23] As such, a degree of regulatory uncertainty in the UK remains, though the government's 13 January 2025 response to the AI Action Plan promises further clarification to remove perceived hurdles to safe adoption.

18 FCA - AI Update, 2024
19 Bank of England - The Bank and the PRA's response to DSIT/HMT: update on our approach to AI, 2024
20 FCA - Handbook, n.d.
21 Information Commissioner's Office - Guidance on AI and data protection , 2023
22 Information Commissioner's Office response to the consultation series on generative AI
23 The King's Speech 2024 - GOV.UK (www.gov.uk)

Adapting to regulatory expectations

While this paper does not seek to interpret regulatory expectations, we observe that banks, insurers, and asset managers are preparing to adapt their current risk management processes to comply with emerging AI regulatory approaches and demonstrate their effectiveness.

This is particularly complex in relation to customer-facing applications enabled through emerging technology and requiring an updated interpretation of regulations. For example, firms perceive the Consumer Duty to be one of the largest sectoral regulatory constraints to the use of AI[24] and will need to address AI risks as they seek to leverage AI's benefits. This is understandable, as the technology offers transformative opportunities to enable and improve good outcomes for customers, if deployed effectively. Conversely, an inappropriate deployment without comprehensive tests regarding customer outcomes could lead to unfair treatment and harm. Senior managers responsible for risk management and technology, as well as those sponsoring generative AI – or indeed predictive AI – use cases, must show they have taken 'reasonable steps' as expected under the Senior Manager & Certification Regime.

The BoE and the FCA have re-affirmed their view that a technology-agnostic approach to regulation is appropriate and, according to input received from industry, regulation is not impeding the growth and productivity benefits of AI in the UK. However, the BoE also noted the fast-emerging nature of the technology and the need for ongoing monitoring to ensure the continued viability of this approach. Areas highlighted as requiring further consideration included the complexity of working with third-party providers, senior manager responsibilities and the more limited explainability of generative AI.[25]

Another specific area of interest is the relationship between the established financial services discipline of MRM and the use of AI solutions. The BoE set expectations regarding MRM in its supervisory statement (SS) 1/23[26] in the form of five principles to manage the risk effectively across all model and risk types. These principles are applicable to all types of models that are used to inform business decisions, whether developed in-house or externally (including vendor models), regardless of technology.

In its response to the Department for Science, Innovation and Technology's (DSIT) request for an update on regulators' approach to AI, the BoE reiterated that its expectations regarding MRM cover AI models, including the management of risks from off-the-shelf products, supported by the related expectations regarding Third Party Risk Management (TPRM) from SS2/21.[27] However, the BoE also acknowledged that "the growing complexity of AI/ML models, such as LLMs, challenge the concepts of explainability and transparency."

Therefore, financial institutions are considering how they will update their well-established approach to MRM in the context of predictive AI and generative AI, including both foundational LLMs and adapted LLM components. In addition, they are aware that they need to strengthen their wider risk management frameworks, covering the wider components of their generative AI solutions outside the scope of model risk.

---

24 Bank of England - Artificial intelligence in UK financial services, 2024

25 Bank of England - Engaging with the machine: AI and financial stability − speech by Sarah Breeden, 2024

26 Bank of England - Model risk management principles for banks, 2023): Principles: (1) Model identification and model risk classification; (2) Governance; (3) Model development, implementation and use; (4) Independent model validation; (5) Model risk mitigants

27 Bank of England - HMT Letter, 2024

03

Generative AI case studies

While UK financial services are actively exploring generative AI applications, most initiatives remain in early testing phases. Currently, only 10 per cent of PoC projects advance to production.[28] Successfully deployed use cases have undergone extensive testing, demonstrating high productivity gains with effective management of risks and operational limitations.

This report examines three real-world anonymised case studies, highlighting the scope, objectives and emerging approaches to risk management.

- Generative AI assistant used within the customer complaints processes.
- Generative AI used to accelerate customer due-diligence processes.
- Generative AI used in the SDLC.

These case studies showcase common generative AI implementations across the financial sector in 2024, including retail and commercial banks, investment managers and insurers. They focus on three core business areas: customer engagement, risk management and software development.

As with most technology and process deployments, these implementations were subject to comprehensive risk assessments as part of the individual firm's risk policies, as well as additional considerations for generative AI. The use case designs, controls, guardrails and procedures were then designed to bring the residual risk to an acceptable level. A summary table outlines the most relevant generative AI risks and describes the specific treatment or mitigant used.

Each case study demonstrates proven value, successful implementation and practical risk management approaches. They provide insights into real-world generative AI applications, detailing both benefits and challenges encountered, while highlighting opportunities for future solution expansion.

28 Accenture - Generating growth how generative AI can power the UK's reinvention, 2024

Case study one: Customer complaints agent

Case study overview

Managing customer complaints is crucial for maintaining consumer trust and regulatory compliance. The UK financial services sector spends many millions of pounds annually on this process, covering compensation, administrative efforts and regulatory fines.[29]

Additionally, financial institutions received 70 per cent more customer complaints in the first quarter of 2024/25 than in the same period of the previous financial year.[30] The process involves recording, transcribing, investigating and resolving customer complaints, ensuring issues are addressed fairly and promptly. This is a manually intensive and highly regulated process, requiring skilled workers to process large amounts of information. As such, this provides a significant opportunity to improve operational productivity and reduce costs by deploying generative AI.

Generative AI can be used to enhance complaints handling with the following objectives:

- Improving customer satisfaction through a reduction in complaints handling times and an improved customer experience.
- Increasing productivity through operational efficiency gains.
- Reducing risk, through the increased support in identifying potentially vulnerable customers.
- Reducing operational costs, through a reduction in complaint escalation to the Financial Ombudsman Service (FOS).
- Improving employee satisfaction, enabling more time to be spent on higher-value activities.

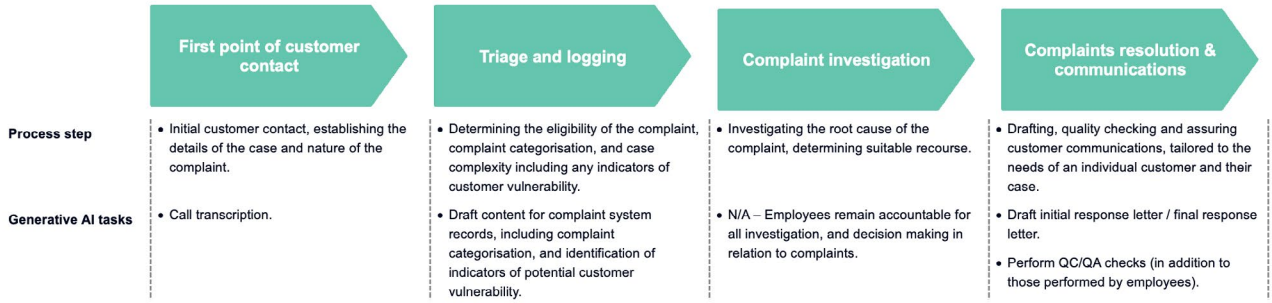
Given the sensitivity of complaints management, there are concerns that poorly drafted response letters that use inappropriate tone and content could lead to more referrals to the FOS and misunderstood complaints could disproportionately impact vulnerable customers.

This is particularly relevant considering that 47 per cent of people display a characteristic of vulnerability.[31] Nonetheless, there is potential now for process-driven generative AI deployment, aiming to reduce manual errors such as mis-categorisation and poor-quality management information (MI). Firms may look towards future potential additional applications to improve customer outcomes, such as personalised chatbots and creating higher quality, more tailored communications.

29 Institute of Customer Service - UK Customer Satisfaction Index July 2024, 2024
30 Financial Ombudsman Service - Quarterly complaints data: Q1 2024/25, 2024
31 FCA - Financial Lives 2022 survey: insights on vulnerability and financial resilience relevant to the rising cost of living, 2022

For this case study, a financial services organisation developed an initial PoC for generative AI to augment the typical processes carried out by a human complaints case manager, as shown in figure five. The tool assists a human case manager who still makes final decisions and interacts with the customer.

Figure five: Using AI to enhance complaints handling



Source: Accenture

Following the successful PoC, the solution was scaled into production, demonstrating a clear ROI through productivity benefits of 30-40 per cent, and average handling time reduction by 30-50 per cent. Having achieved these productivity advances, the firm is now investigating opportunities to utilise these for additional time engaging with customers and more permanent running cost reduction. Staff surveys also showed that customer and employee experiences were improved due to human case managers being better informed ahead of customer interactions.

Risks and practical mitigations

The proposed change to the process underwent a thorough risk assessment, identifying both generative AI-specific issues and more common risks to be considered. Practical mitigation techniques were applied. The risk profile was found to be within acceptable limits, posing a negligible change to the residual risk. A summary of the risks, and the techniques applied to mitigate these, is shown in table three.

Table three: Case study risk mitigation

The solution was deemed low-risk due to the following mitigation techniques:

Risk category	Relevance to the case study	Mitigation techniques
Unreliable outputs	High relevance due to risk of detrimental customer impact in a heavily regulated and sensitive process.	HITL call monitoring and quality control of generative AI outputs (including FOS and relevant tone of voice guidelines). Extensive testing of the end-to-end solution. Updated operating procedures applied. Generative AI adheres to a mandated summarisation framework with quality assurance embedded.
Information security	Medium relevance due to presence of existing information security controls.	Solution aligned to private cloud and on-premises information security. Additional controls for lower security environments (e.g. user acceptance testing).
Data privacy	High relevance due to processing of personal data that may be highly sensitive.	Controls aligned with relevant data privacy policies, including limited data retention and controlled employee access rights. Application of regulation-aligned vulnerability frameworks to ensure only relevant data is processed (e.g. data minimisation principle). HITL system ensures that GDPR 'automated decision-making' rules not triggered. Record of data sources enabling traceability. Developed and tested but not used: Personal data masking solution to suppress some sensitive data provided but not necessary to conduct the task.
Intellectual property	No intellectual property as inputs or outputs to the process.	N/A.
Human-AI configuration	High relevance due to generative AI being used in a supporting capacity for a heavily regulated and sensitive process.	No decision-making undertaken by generative AI. Transparency notices provided to employees and customers. Continuous monitoring, including new metrics for generative AI performance.
Value chain and component integration	Low relevance due to extensive HITL involvement in the process.	Commercial agreements and minimal control standards in place with third-party technology supplier.
Environmental impact	Low relevance due to energy consumption not significantly increased.	N/A (scale of processing relatively low, this risk may need to be reassessed as usage volumes scales).

Outcomes, insights and lessons learned

This case study highlights how risk mitigation techniques for generative AI deployments can be established and integrated into existing technology solutions. Regulatory processes and requirements remain highly relevant when introducing generative AI, yet the technology accentuates risks in new ways that must be mitigated.

Firms should consider the following broader learnings when introducing generative AI customer complaints agents:

- Synthetic data creation: Generative AI was used to create synthetic data from a sample of customer call transcripts, providing a dataset that can be used to both accelerate the machine learning process and train employee case handlers. This reduced the manual effort and time required for such tasks, highlighting an underappreciated capability of the technology. With appropriate guardrails to prevent data leakage, generation of synthetic data for these purposes is more privacy-safe than use of real datasets. But this additional application underscores the need to consider upfront efforts to enable solution testing before scaling.
- Workforce and talent: Scaling a generative AI PoC into production can reveal insights into workforce readiness for AI adoption and the capabilities of AI CoEs. This case study highlighted the need to upskill subject matter experts in AI governance and ensure AI CoEs have end-to-end capabilities for holistic and efficient change.
- Future applications: As generative AI technology matures, additional capabilities could be integrated into the process to provide higher quality, more personalised communications to customers who have complained, which may serve to avoid escalation to the FOS and potential reputational damage. These communications could include agentic chatbots or avatars, text-to-voice responses and the ability to interpret and respond to uploaded images.

Generative AI has been successfully implemented with low residual risks in this case study. The application of the technology could be deepened and, in clear cases and with appropriate testing and impact assessments, further automation and decision-making explored.

Case study two: Know Your Customer

Case study overview

KYC and CDD processes are fundamental in financial services, enabling firms to verify the identity of their clients, understand how they operate, assess risks and act as a barrier to ensure that illegal activities such as money laundering and terrorist financing are identified and curtailed. KYC processes are governed by strict regulations and cover the entire customer lifecycle.

The KYC process relies on ingesting a significant amount of structured and unstructured data identifying the customer entity received through multiple channels and mediums into the firm's core customer systems. This process can be time and resource intensive, requiring a skilled workforce to manually process large volumes of fragmented information.

Generative AI is increasingly being considered an enabler to optimise the effectiveness, efficiency and speed of KYC processes. With the ability to analyse diverse datasets and extract key information from multiple sources, generative AI can materially enhance how financial institutions perform KYC operations. However, since personal data is being used for a purpose that has a potentially high impact on customers, deployment of the technology requires careful management of data privacy and security.

In this case, a generative AI accelerator tool was deployed to ingest documentation, extract mandated KYC information and populate this into an output format which could be readily integrated into existing record systems. A performance assessment was conducted to examine the efficacy of the solution, and a very high level of accuracy was observed. The solution was fully hosted on a private cloud environment, with a private API call-out to a closed LLM in the firm's own private cloud.

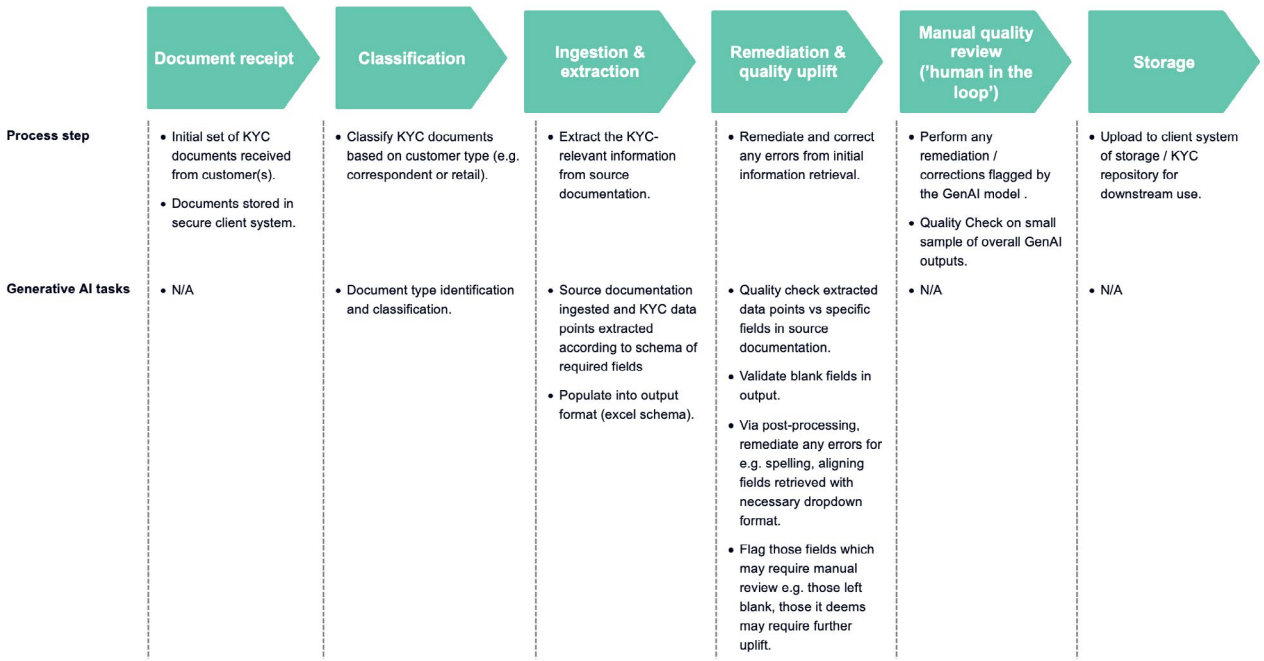
The generative AI tool was also capable of running quality checks on the outputs by comparing them to source material, remediating errors and missing fields. This was followed by a manual quality check by an operator who assessed the tool's output before the process concluded.

To mitigate privacy and data security risks, the solution was entirely hosted on a private cloud environment, utilising a secure Large Language Model Application Programming Interface (LLM API) call-out to communicate with a closed LLM within the firm's own private instance. Further, access rights were tightly controlled and documentation encrypted both at rest and in transit. Data minimisation was enabled, and the generative AI environment was configured to have zero retention after a 30-day period.

Following the initial training of the model, the generative AI solution achieved the required level of accuracy. This gave the firm the confidence to move away from conducting quality checks on all records to a sample approach. For those clients in scope, the solution reduced processing times by 90 per cent on average.

Figure six reveals where the generative AI tool is embedded into the Initial and Ongoing Due Diligence (IDD & ODD) steps within the KYC process:

Figure six: Generative AI in KYC document processing



Source: Accenture

Risks and practical mitigations

To ensure that their processes remain secure, compliant and trustworthy, firms need to adopt robust risk mitigation strategies. In this case study, the firm managed the risks to an extent it deemed acceptable, as shown within table four.

As with other sensitive use cases, firms should consider the appropriate decision-making authority of human experts. Focusing on the manual elements of the KYC process while maintaining appropriate guardrails such as HITL and sample validation of the output, allowed the firm to control this risk while also realising an acceptable level of accuracy and ROI.

Table four: Case study risk mitigation

Risk category	Relevance to the case study	Mitigation techniques
Unreliable outputs	High relevance due to outputs informing a heavily regulated process.	Ongoing sample monitoring and quality assurance by a human KYC analyst. LLM uses feedback loops to improve outputs using better data.
Information security	High relevance due to new technology interfacing with highly sensitive data.	Private cloud and closed LLM. API callouts within private cloud environment. Private cloud architecture aligned with relevant information security policies.
Data privacy	High relevance due to processing of sensitive personal data with the potential for significant impacts on customers.	Private cloud houses a secure 'storage container' aligned with relevant data privacy policies including limited data retention and employee access rights. Documentation encrypted at rest and in transit. The system assembles documentation but does not make decisions, ensuring that GDPR automated decision-making rules are not triggered. Data Protection Impact Assessment (DPIA) conducted to provide a holistic review of privacy risks. Record of data sources, enabling traceability.
Intellectual property	No intellectual property as inputs or outputs to the process.	N/A.
Human-AI configuration	Medium relevance due to automated outputs being produced by technology rather than manual intervention.	Ongoing sample monitoring and quality assurance by a human KYC analyst. LLM process configured to apply confidence ratings as an additional quality control layer.
Value chain and component integration	Medium relevance due to critical components of the process becoming reliant on third-party technology.	Commercial agreements in place with third-party technology supplier.
Environmental impact	Low relevance due to the low scale of deployment, reducing impact on overall computing capacity.	N/A (scale of processing relatively low, this risk may need to be reassessed as usage volumes scales).

Outcomes, insights and lessons learned

The measured effectiveness of this case study was encouraging with 98 per cent average field retrieval reported and a high level of accuracy (>95 per cent). In comparison, the original process, before the generative AI tool was introduced, required two human reviews: a KYC analyst and a quality checker, known as a 'four-eye check'. This is in addition to several potential iterations of file reviews to correct errors identified before the process achieved similar retrieval and accuracy results.

The use of generative AI to accelerate the KYC process accentuates certain risks, considering the need for accuracy and completion of output. The nature of data processed also requires heightened care and diligence.

Solution architecture, technology and design process choices would allow firms to employ practical mitigation techniques at each stage, while delivering accuracy and productivity improvements. This enables firms to refocus their skilled teams on more value-adding tasks. Below are some additional considerations to support the adoption process for financial institutions, highlighted by this case study:

- Managing data privacy: Firms are expanding their adoption of generative AI in a considered manner utilising existing infrastructure. This includes ensuring the generative AI solution is fully contained within the bank's existing private cloud environment and access controls to manage data privacy and security concerns.
- Unique cybersecurity considerations for LLMs: Adoption of generative AI is providing firms with an opportunity to re-evaluate their cybersecurity measures to ensure appropriate coverage of the unique aspects of LLMs, for example, the proliferation of API callouts that will need to be assessed for vulnerabilities when interacting with sensitive data and third-party technology.
- Accuracy of output: Retrieval Augmented Generation (RAG) can be built into the KYC process for the generative AI tool to review field content against original documentation, including expected values defined by the user. This can improve the accuracy of outputs and provides the ability to verify source references.
- Further adoption of generative AI in financial crime processes: Beyond generating KYC documentation outputs for downstream use, the technology could be further deployed to help throughout the KYC/CDD lifecycle. However, further review of risks and associated mitigations would be necessary for each extension since the risk profile changes depending on the specific utilisation. For example, consideration of bias would be needed if a generative AI tool is assessing the customer risk rating to ensure that the solution appropriately considers diverse client groups. Human operators remain an effective safeguard in use cases where the AI is providing recommendations for KYC or other compliance activities.

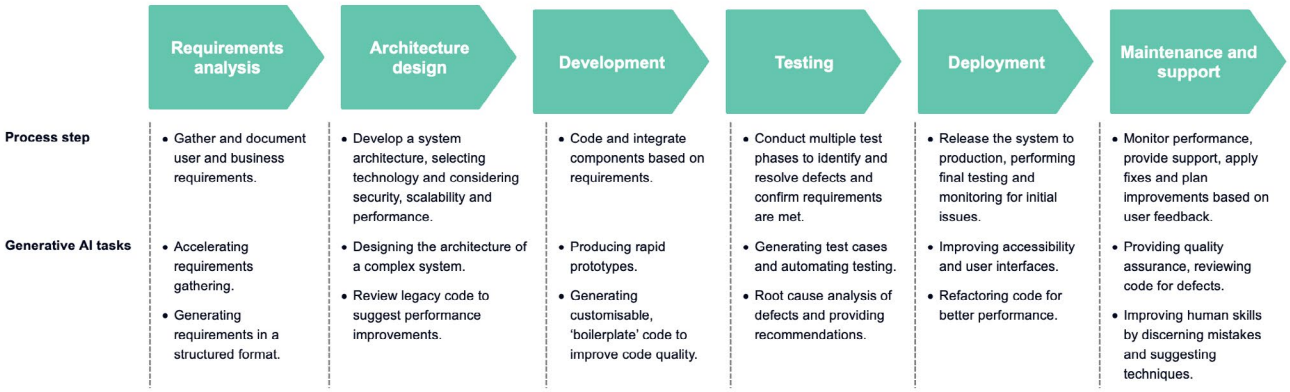
Case study three: Software development lifecycle

Case study overview

The financial services industry spends billions each year developing and maintaining complex technology and software estates. Generative AI has the potential to make a profound impact, enabling significantly more efficient and effective software development.[32,33] As financial institutions increasingly rely on technology to drive strategic innovation, streamline operations and enhance customer experiences, the SDLC is becoming a promising area for generative AI-driven optimisation.

The SDLC is a structured process used by software developers to design, develop, test and deploy software applications. In the financial services sector, the SDLC typically consists of the following broad phases across requirements gathering, design, development, testing and production deployment. Generative AI can augment each phase in several ways, as summarised in figure seven:

Figure seven: Process overview



Source: Accenture

32 MIT Technology Review Insights - Transforming software development with generative AI, 2024
33 Oracle - 7 Ways GenAI Can Help Improve Software Development, 2024

The use cases being explored in this case study relate to the requirements analysis and testing phases of the SDLC, particularly those involving complex, legacy technology estates where change initiatives are difficult to implement quickly and safely.

Like many similar organisations, the UK-based financial services firm at the centre of this case study was seeking to accelerate digital progress through a large-scale data migration from an on-premises data centre to the cloud. The firm adopted a generative AI toolkit for the requirements analysis and testing phases of the SDLC, incorporating compatible functionality from GitHub Copilot, an AI extension to a popular coding tool.

At the requirements gathering stage the toolkit was used to combine enterprise-specific documentation, architecture and data using an LLM. This produced requirements in a structured format, along with relevant context of the firm's existing and target state. These requirements were then reviewed and validated by an experienced analyst or subject expert prior to approval and sign-off.

While the outputs needed refinement from highly skilled people, the generative AI tool accelerated this phase of the SDLC by approximately 50-60 per cent based on the firm's experience of similar projects.

During the testing phase, the firm used generative AI to accelerate and improve the testing and deployment phases. In this case, a separate generative AI solution from the toolkit was leveraged to produce, assess and refine software code ahead of the migration of data into the cloud-based production environment.

To do this, the tool provided a multi-agent system where an LLM assumed different personas in the SDLC, such as software developer, engineer, designer and test analyst. This approach seeks to mimic the way a human software development team might traditionally work to enhance productivity and quality. By working on different tasks at the same time, these AI agents enabled a faster, more efficient code production process, while operating a feedback loop to improve the code, reduce defects and quality test the outputs.

As in the previous case study, the primary benefit of the tool was in accelerating the processes, with any code and test outputs being reviewed by a well-qualified HITL prior to deployment.

Risks and practical mitigations

To ensure that the software solutions being enhanced by the technology were secure, compliant and efficient, the risk profile was assessed and mitigated to acceptable limits as illustrated in table five.

Clearly, value chain and component integration was a key risk category in this example, given the potential for software development to occur in a black box, with over-reliance on the LLM reducing explainability and removing elements of control from the organisation. To mitigate this, the generative AI tool was intentionally designed in a modular structure, with the 'multi-agent' approach enabling an experienced HITL review of outputs and code security at each stage.

Table five: Case study risk mitigation

Risk category	Relevance to the case study	Mitigation techniques
Unreliable outputs	High relevance due to code being developed and deployed into production.	Code review and iteration by an experienced human developer. Generative AI uses 'feedback loops' to improve outputs using better data.
Information security	Medium relevance due to sensitive information present in the code being developed.	Private cloud and closed LLM to ringfence the environment. Experienced developers ensuring alignment with relevant information security policies.
Data privacy	Medium relevance due to personal data being processed during code development	Personal data obfuscated in accordance with relevant data privacy policies. Data entering the environment is transient and deleted after each session.
Intellectual property	Medium relevance due to processing of code and IT architecture documentation during requirements generation and code development.	Multi-tenancy cloud environment to prevent unauthorised access (on-premises instances also an option). Contractual terms to prevent code harvesting for LLM training purposes. Data entering the environment is transient and deleted after each session.
Human-AI configuration	High relevance due to reliance on new technology creating a black box of unexplainable components.	Modular approach used, segmenting stages of the SDLC into component parts. Experienced human technical architects, developers and engineers providing quality assurance.
Value chain and component integration	High relevance due to proliferation of third parties and, conversely, concentration of LLMs with key suppliers.	Commercial agreements in place with third-party technology supplier (including SLAs for LLM availability and upgrades). Reserved computing resource (e.g. GPU capacity) to assure performance. Back-testing of previous solutions when LLMs are upgraded.
Environmental impact	Medium relevance due to LLMs using significantly more computing resource than traditional AI.	Combination of small, medium and large models used to optimise energy usage. Monitoring of compute utilisation managed via FinOps team and made available to wider sustainability reporting.

Outcomes, insights and lessons learned

The benefits of using generative AI to augment the SDLC already appear to outweigh the risks when appropriate mitigants are employed. Firms are encouraged to make the following considerations when embarking on new change initiatives in this way:

- Accelerating the process while relying on human expertise: Adopting generative AI for requirements analysis yielded time savings of 50-60 per cent based on the firm's experience of similar projects, yet the outputs needed refinement from highly skilled human analysts. While in this case there was no plan to replace human effort with the generative AI tool, such adoptions in the future should consider the impact on workforce replacement and develop appropriate plans for redeployment. This case study experience reflects a wider trend among GitHub Copilot adopters; an industry survey identified coding speed increases of up to 55 per cent, while 85 per cent of developers felt more confident in the quality of their code, and 90 per cent reported higher job fulfilment.[34]
- Protecting the SDLC for major change initiatives: Firms using generative AI in the SDLC will face increased exposure to third parties, with a limited number of established providers of cloud and LLM technology currently. As indicated in table eight, enhancing procurement processes, understanding hardware limitations, such as GPU capacity and ringfencing the environment can alleviate the common concerns.
- End-to-end environmental impact: The use of generative AI in the SDLC has been limited to date, so questions remain as to the levels of resource consumption – including energy, water, and other environmental impacts – once applied at scale. Understanding the end-to-end view, including time savings once the SDLC is optimised, will be critical in determining the extent of generative AI adoption.

34 Research: Quantifying GitHub Copilot's impact in the enterprise with Accenture - The GitHub Blog

04

Key risks and mitigation approaches

This section provides a further focused examination of the three risk themes emerging from discussions with UK Finance members.

Risk topic one: Reliability of outputs

Introduction

Many of the typical systems used in financial services are built on coded rules, managed logic and controlled datasets, offering the functionality to examine the reasons for a particular system behaviour.

In contrast, generative AI models are built on deep neural networks and utilise both structured and unstructured data inputs to produce non-deterministic outputs. This means that standard testing techniques, such as parameter sensitivity and input-output mapping, are less applicable. Furthermore, the LLMs at the centre of generative AI solutions have been trained on huge datasets that are simply impractical to assess or test for inaccuracies or bias – meaning traditional approaches to assessing and testing data inputs are generally obsolete.

Further, in general there are no inbuilt checks to ensure the reliability of the output. This is fundamentally different to both the functioning of traditional software models on the one hand and to the trust we put in human operators on the other. The linguistic quality and correctness of an LLM output can appear plausible and be misinterpreted as factual correctness. Outputs may therefore exhibit bias, inaccuracies or inappropriate language, but these deficiencies may not be recognised.

Despite these challenges and fundamental limitations, in many cases these risks can be managed. Numerous approaches have been adopted in the use cases considered in this report. Depending on the specifics of a given use case, sensitivity to this risk will vary significantly, as will the specific ways in which the behaviours of a given model can produce unreliable results.

One of the most common approaches to managing this risk, while the technology is relatively new, is to have a suitably qualified human checking the outputs and correcting them if necessary. However, this technique contains some constraining factors that themselves need to be managed, including cognitive load, fatigue and variation between different team members.

There are also considerable efforts across the industry and academia to better understand and improve toolkits to help teams audit and control LLMs. Efforts are also being made to provide model developers, designers, and risk management teams with tools to understand, measure and ultimately reduce reliability risks.

The testing, fine tuning and design of generative AI solutions is an evolving space, with some promising approaches summarised in table six below. These include building in steps to allow authoritative information sources to be combined with generative AI model outputs and providing end-users with tools to fact-check the outputs. This helps development and testing teams to measure and improve the models' performance and, in some cases, add AI-supported steps within the models themselves.

Emerging approaches to reliability testing and model reviews

Given some of the characteristics of generative AI outlined above, traditional model review and testing techniques are not always possible with LLMs. Instead, testing may be based on pragmatic evaluation of repeated testing evidence. Testing of LLMs in current use cases typically involves the generation of a multitude of possible scenarios and inputs, then measuring the system's response against a set of agreed acceptable or correct outputs. The likelihood of unreliable outputs can then be inferred. Like traditional models, the hardest risks to control are in the outliers so LLMs need to be well tested for these edge cases on a continual basis.

Testing cannot be exhaustive, and scenarios need to be carefully crafted to cover as many outliers as possible and reduce the use case-specific risk level.[35] This can be combined with conscious efforts to explore the limits of the models, sometimes known as 'red-teaming'. Organisations will need to identify a risk threshold, considering the potential consequences of inaccuracy and the level of autonomy granted to the overall system, and design their testing approach accordingly.

Other mitigation techniques involve restricting the range of allowable inputs, constraining the outputs or both. While these approaches can lead to a more controlled solution, in some cases overly constraining LLMs undermines the advantages of an LLM solution and the use case may be better suited to a rules-based 'Q&A' style solution, a predictive AI model or other natural language processing (NLP) tools. These alternatives are usually cheaper and simpler to manage. Such decisions would normally need to be taken at the solution design phase and in the context of the intended use case to maximise the benefit of using an LLM.

Many practitioners would agree that financial institutions should not aim to understand the inner workings of generative AI models or seek to eliminate the risk of unreliable outputs entirely. Instead, firms can carefully design the use cases, configure the models and embed 'checks and balances', much like designing a human-based process. This includes, but is not limited to:

- Ensuring that datasets are diverse, balanced, and representative, thereby helping to reduce biased outputs. Given that firms lack control over, or full transparency into, the underlying data used to train the foundational model, the focus here should be on finetuning with additional data relevant to the intended use cases.
- Having a good understanding of which models are being used, and for what purpose, and monitoring how each performs against tests and production usage.
- Controlling and testing the upgrades to new LLM versions; while newer versions typically perform better than their predecessors, there is no guarantee of better performance for a specific use case. Firms need to work with third-party providers to ensure model upgrades are well sign-posted to allow for use case specific testing and allow informed decisions over whether to replace their existing model.[36]

By combining these design principles with the risk mitigation techniques described in table six, firms will be better prepared to manage the risk of generative AI solutions producing unreliable outputs.

35 The Alan Turing Institute guidance and research

36 Accenture research and expert reviews

Table six: Example mitigation approaches for unreliable outputs

Retrieval Augmented Generation (RAG)	Combines retrieval-based methods with generative models to enhance the quality and relevance of generated text. In this approach, a model retrieves relevant information from a large database or knowledge source before generating a response based on the provided knowledge, as opposed to standalone generative AI models that rely solely on their internal training data. There are a variety of LLM providers and third-party solutions such as Trustwise[37] and Zilliz[38]
Fact-checking and expert multi-agent systems[39]	Refers to the process of verifying the accuracy of information as the model generates responses. This involves steps to assess the claims made by the AI against reliable, up-to-date sources to ensure that the output is factual and trustworthy. There are a variety of LLM providers and third-party solutions such as FactCheckExplorer[40] and ClaimBuster[41] A growing area of interest related to this approach is to use multi-agent generative AI systems where human-only or rules-based monitoring is not always possible. One or more expert AI agents can assist with oversight and evaluation of other generative AI models' performance, behaviour, and outputs.
Automated source citations and attributions	Involves the systematic identification and referencing of sources used to generate information or content. This process supports the generation of reliable information and enables appropriate credit to the original authors or sources. This may include clear tagging of outputs to indicate sections that are based on inference rather than verifiable data.
Confidence scoring and uncertainty estimation	Confidence scoring in generative AI measures the model's certainty in its outputs through probability distributions and statistical metrics, helping identify when outputs may be unreliable or require human review. High confidence scores typically indicate the model is working with familiar patterns or well-structured data, while low scores may flag potential hallucinations, out-of-distribution inputs, or edge cases. It's crucial to note that high confidence scores don't always correlate with accuracy, which is why confidence scoring should be combined with other validation methods and regular performance monitoring.
Model fine-tuning with domain-specific data	Refers to the process of taking a pre-trained generative AI model and further training it on a smaller, specialised dataset that is specific to a particular domain or field. This approach helps adapt the model to better understand and perform tasks relevant to that domain, strengthening the statistical relationships in the model based on trustworthy information.
Ongoing performance monitoring	Ongoing monitoring and evaluation of the performance of a generative AI system against a set of criteria and comparison to 'ground truth' to track and improve model reliability. As changes, or incorrect outputs, are found this can be used to extend and refine the training dataset. This creates an improved benchmarking test suite to support model evaluation and monitoring of live systems, creating a re-enforcement feedback loop.

37 Trustwise, n.d.
38 Zilliz, n.d.
39 The Alan Turing Institute - The impact of Large Language Models in Finance: Towards Trustworthy Adoption
40 Google - Fast Check Tools, n.d.
41 ClaimBuster - Automated Live Fast-Checking, n.d.

Human reviews and human-machine task routing	One of the common techniques involves including a trained operator within the process and routing complex or higher risk cases to them, focusing the AI on more suitable tasks. This can be complementary to a feedback, monitoring and evaluation process and provide input into further system fine-tuning. This emphasises the complementary strengths of both parties, where humans provide creativity, intuition, and contextual understanding, while machines offer speed and data processing capabilities. Examples include routing of complex cases to a human for review or focusing generative AI on summarisation, drafting and other tasks it is well suited to within the use case. However, the limitations inherent in human involvement also need to be considered, including fatigue, cognitive load and performance differences between team members.
Extensive prompt engineering and testing	Well defined prompts can help to guide the LLM to answer the query. By defining how the LLM should use data, and the tooling available to it to support queries, the likelihood of model hallucinations can be reduced.

Risk topic two: Data privacy and security

Introduction

The security and safeguarding of data emerged as a major concern in a recent UK Accenture survey (see figure three). This includes both data privacy, where personal data is used in a way that's not in line with the responsibilities and interests of individuals and firms, as well as security risks, where cyber criminals exploit technological vulnerabilities. Given their close relationship, these considerations are summarised together in this section of the paper.

Key considerations for generative AI

Generative AI solutions rely on large training sets which can include data not originally provided to the deploying organisation but procured from an external party. Further fine tuning and prompting of the model provides an additional opportunity to directly or indirectly ingest sensitive or proprietary data into the system.

While these risks exist in many systems and some precautions are well established, there are features of generative AI technology that need additional consideration, such as the potential to generate inferences containing sensitive information about a person. These aspects introduce new risks to the safeguarding and legitimate processing of data through their leakage, inadvertent processing or the unintended and unlawful generation of inferences.

For example, it has been shown that it's possible to extract personal data from LLMs' training data sets[42] (known as data memorisation) and that generative AI solutions may infer the presence of special category data attributes, such as religion or medical conditions. Generative AI solutions can also contain security vulnerabilities which, if exploited successfully, could lead to breaches of confidentiality, integrity or availability of data.

Prompt injection is one potential security risk in which a malicious prompt is used to reveal data, either directly or indirectly. One example might be obtaining information in a document previously provided to the model.[43] It should also be considered that firms operating in more than one jurisdiction must contend with variations in regulations and enforcement regimes.

A frequent area of concern is whether data used to prompt and test a model will be retained by the model provider for the purpose of refining the model without prior agreement. Most commercial LLM providers state that they do not use client data for training without prior agreement. However, unless the model is hosted in a ringfenced environment, proprietary company data may leave the firm's security perimeter. Care must also be taken when updating these solutions with new features, 'plug-ins' or supporting services, to ensure that these do not introduce unintended data flows to external recipients.

Given these concerns, financial services firms must ask themselves to what extent their existing data privacy and security measures can help address generative AI related concerns and where more work is required to do so within an acceptable risk appetite.

42 Carlini et al - Extracting Training Data from Large Language Models - 2021

43 Generative AI's Biggest Security Flaw Is Not Easy to Fix | WIRED, Exercise caution when building off LLMs - NCSC.GOV.UK

Extension of existing data privacy and security guidance

Data security and privacy concerns have long been a priority for the financial services industry where this topic is relatively mature.

Over the past decade, awareness of the importance of safe personal data processing has increased and been formalised, with many jurisdictions granting special regulatory protection, such as the GDPR and the UK Data Protection Act. The UK's data protection regulator, the ICO, has issued guidance materials and tools to help firms adapt their compliance approaches for AI.

This includes an AI and data protection risk toolkit, though this is not specifically tailored for generative AI. The guidance is based on foundational principles of data protection: lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, security and accountability. Firms deploying generative AI capabilities that involve personal data must consider how these principles will continue to be upheld and evidenced.

A well-established risk framework and solid data protection foundations will help firms confidently address the main questions and work through how generative AI impacts data privacy and safety. However, the ICO is aware that more guidance is required in relation to generative AI specifically and conducted a five-part consultation series on this technology in 2024 covering topics including purpose limitation, lawful basis and allocation of controllership throughout the AI supply chain.

Regarding cybersecurity specifically, firms continue to strengthen their overall security measures for cyber and cloud, utilising organisations and frameworks such as OWASP and MITRE ATLAS. The UK's National Cyber Security Centre (NCSC) also provides useful guidance on AI system security.[44]

While mitigation techniques are being updated for generative AI and further guidance is pending, firms might also need to consider adjusting their risk tolerance. For example, the data memorisation risk cannot be fully mitigated at present, with the underlying data collection and utilisation for training the model controlled by the LLM provider. But given the potential benefits described in this report, organisations may be willing to accept a higher risk level.

There are an increasing number of techniques and approaches being used by teams developing AI solutions. In assessing the current generative AI deployments across financial services, a combination of data protection techniques, carefully considered architectural designs and thorough testing techniques have emerged:

44 NCSC - Guidelines for secure AI system development

Table seven: Example mitigations techniques for data privacy and security

Privacy-enhancing technologies (e.g. personal data filters)	Filters to detect and remove or obfuscate sensitive information from data inputs and outputs, ensuring that personal data is not inadvertently exposed or processed.
Public cloud controls and risk assessment processes	Fully understanding and leveraging the shared responsibility model, where some security responsibilities are ceded to the cloud service provider (CSP) with a clear delineation between the CSP and the use case/platform deployer. Evaluating the security and privacy risks associated with using a CSP hosting generative AI models to ensure that the provider meets the organisation's security standards, and all gaps are closed.
Ringfenced architecture solutions	Ensuring the data of organisations using the LLM is isolated and secure from other tenants who share access to the same LLM. In other words, effective data segregation for different organisations using the LLM in a multi-tenant configuration.
Penetration and adversarial testing	Testing the security of the models, data storage and APIs to ensure that they are resilient against unauthorised access, data breaches, and other security threats. Complementation of penetration testing can be combined with adversarial testing, whereby techniques used by real-world adversaries are used to test and understand how the model behaves when faced with malicious or harmful input.
Shadow AI discovery	Identifying and monitoring unauthorised or unmanaged AI applications within the organisation to ensure all AI systems (including generative AI/LLMs) are managed and comply with relevant data privacy and security policies.
Data protection impact assessments (DPIAs)	Conducting thorough DPIAs when personal data are used, reflecting on all relevant aspects while remaining cognisant of any areas of lesser maturity, given generative AI's infancy.
Data minimisation and obfuscation	Considering carefully the business purpose of the generative AI model and admitting for processing only the minimum amount of data needed for this purpose – as is already common practice for AI use cases.
User access restrictions	Managing and restricting access to AI systems and data based on user roles and responsibilities, including implementing strong authentication mechanisms, role-based access controls (e.g. ABAC, RBAC, Zero Trust), and regular audits.
DevSecOps security disciplines	Standard security disciplines remain relevant, such as regularly applying security patches and conducting vulnerability assessments to identify and remediate security weaknesses in AI systems and infrastructure. This includes code quality and security assessments.
'Red team' exercises	In addition to cyber security testing, 'red teams' in a generative AI context seek to identify unintended outcomes and impacts from deploying these systems, including testing for issues such as privacy breaches, bias and unacceptable speech.
Contractual terms/ data processing agreement/data use agreement	Establishing legal agreements with AI service providers that define the terms and conditions for data processing, including data privacy and security obligations.
Secure hosting	Deploying open-source models within an organisations cloud or on-premises network perimeter to retain full control over the data presented to and received from the model – no data exists outside the organisation's network.

Risk topic three: Third-party considerations

Introduction

Given the cost and complexity of developing and maintaining LLM-based solutions, most generative AI models and packaged solutions are provided, at least in part, by third-party vendors. Considering the role of third parties involved, data and algorithms cannot be traced to source through standard audit processes. Furthermore, the visibility of the model architecture and information on testing performed by the vendor is often limited and the vendor may update the model or software without prior consultation.

This further amplifies concerns about the level of control over generative AI solutions and the ability to manage the associated risks. The extent of these risks and the level of dependency on third party-managed controls varies based on specific hosting and architecture patterns, the nuances of the use case and data requirements, as well as the contractual agreements with third parties.

Key considerations for generative AI

Given the nature of the technology, various generative AI-specific assessments will typically be needed before these models are brought into production. Some are a natural extension of existing procurement and vendor selection practices, while others must be adapted to deal with the nuances of generative AI.

There are today four main ways in which generative AI can be accessed, customised and integrated into financial institutions' operations, each of which present different third-party considerations:

- Non-customisable third-party solution: Presents a risk that training data is not representative or is not of the quality needed for a given use case. Also involves restricted contractual protections, guardrails and updates that are at the vendor's discretion.
- Customisable third party: Requires high-quality contextual data from the user. As such, these products offer control of domain-specific data. This can reduce risk through retraining and improve fairness, output explainability and accuracy, being tailored to specific use cases. However, as with non-customisable models, there are limitations around contractual protections and downstream data usage.
- Customisable open-source: Requires high-quality contextual data from the user, offers control of domain-specific data and gives freedom to improve the model ad-hoc. This can permit improved fairness, code and model architecture transparency, output explainability and accuracy. Contractual provisions may be required to codify expectations regarding auditability, monitoring and accountability across the end-to-end pipeline.
- Generative AI embedded in applications from established enterprise platforms, such as ServiceNow, PeopleSoft, Salesforce, or specialist companies such as fintech or HR. These products pose additional risks, although these will vary for each application in terms of business and customer impact. Initial deployment and application upgrades are usually controlled by the vendor, limiting control and testing by the financial institution. Buyers are often unable to choose the terms of upgrades, with the third party sometimes adding

generative AI functionalities without consulting its clients. It is also challenging to manage across multiple vendors feeding into the same architecture.

In addition to challenges associated with managing third parties as providers of generative AI models, firms may also engage third parties that provide other services which are increasingly enabled through generative AI. Reliance on external vendors may expose financial institutions to a variety of supply chain vulnerabilities, including service interruptions or model biases that could compromise decision-making, operational resilience and the firm's reputation.

Given these challenges are an evolution of TPRM and externally provided software services, firms may take confidence from the maturity of their existing TPRM processes and controls. However, while the existing TPRM capacity and capabilities may suffice to support generative AI management currently, firms should assess and mature their TPRM operating models in advance of anticipated scaled deployment.

Even in limited use cases, generative AI-specific concerns remain, and firms cannot assume they will be able to impose specific contractual and commercial demands on vendors to manage risks in their preferred way. Despite this, generative AI model vendors, as well as other critical third-party suppliers, may become subject to stricter regulatory rules to bolster resilience in the financial services sector. The regulatory incentive to reduce systemic risk could manifest in controls which help to satisfy the third-party risk management requirements of financial institutions.

Table eight provides practical mitigation techniques that can help to manage generative AI-specific third-party risks.

Table eight: Example approaches to managing third-party risks

Test environment and metrics	Arrangement of access to test environment for users to complete independent scenario testing ahead of adoption. Select robust, reliable, representative metrics to test the model as it stands at T0. As the model evolves and updates are pushed by vendors, there is a need to understand how the model performance will evolve and when to retrain.
Audit arrangements	Operating throughout the supply chain of multiple generative AI and interacting non-generative AI vendors, based on verifiable evidence, consistent definitions and empirical testing outputs. For example, if the learning model is federated, the audit toolkit needs to ensure all the information is securely aggregated and must be able to provide a data privacy guarantee. Some assurance providers are already working towards this type of solution.
Vendor assessment and due diligence	Improved due diligence on third-party providers through vendor assessments, review of contracts and SLAs (e.g. limitation of upgrades made by the vendor without the knowledge of the buyer, ringfencing of buyer-supplied data to limit downstream use by the vendor, acceptable testing, etc.) and audit rights over vendor systems. Reviewing vendor test results against key risk areas, such as data privacy and security, bias and ethical risks, legal and regulatory compliance.
Deployer driven approaches	Performing independent testing/control of input received from the third party for relevant risks. Monitoring model performance for unexpected output and behaviours, security vulnerabilities, etc, setting up an incident response plan in case of breaches. Maintaining records of version control and any updates, any monitoring and incident responses.
Third-party gateways	Gateways/checkpoints/firewalls at logical points between different AI systems, to test and identify possible risks before the output flows between them (e.g. vendor-provided output feeding into an in-house system). When implementing new software via a modular solution, this enables a HITL review of code security at each stage.
Provider risk assessments	In addition to scenario testing of data quality, input and output control and performance optimisation, risk assessments can be undertaken within scenarios to determine robustness, resilience and security. Any use of a third party to provide critical services using generative AI would necessitate adequate incident management and business continuity planning.

05

Conclusions and outlook

Generative AI deployment in financial services has seen a meaningful progression from experimentation to real-world deployment with tangible value for business processes in the last 24 months. Firms' innovation has nonetheless been careful and responsible, with use cases being implemented only as risk mitigations develop.

As can be expected in financial services, these deployments have so far represented a relatively conservative risk appetite. Considering the speed of technological evolution, more value can likely be unlocked in the future. For example, there remains an opportunity to deploy the technology further in more complex customer-facing use cases and for higher-risk internal tasks. This will however also require appropriate safeguards, such as retaining the ability to introduce a HITL where necessary.

For this to be possible, firms need to find the right balance and navigate the moving parts of not only the emerging technology itself but also evolving best practices in risk management and governance. This would enable firms to confidently operate within the existing regulatory environment and evidence their compliance. Continued investment in data, data governance, cloud and cyber security will all bear fruit when scaling both generative and predictive AI. Only on these foundations will more sophisticated applications such as multi-agent deployments and a combination of generative AI with predictive AI be possible at scale.

Current conservative use of generative AI in financial services has meant limited environmental impact to date. But scaling the technology will require a stronger focus on how more expansive deployment will impact firms' sustainability performance.[45]

In parallel to phased adoption, firms should now focus on the education and awareness of their workforce, boards and customers. There is an opportunity to develop capability with customer input, which can help build trust and a reputation for responsible practices.

There is potentially also a strong role for industry level collaboration to facilitate responsible uptake. The principles-based regulatory model preferred by the UK should have greater flexibility to adapt to technological developments than more prescriptive or rigid regimes. But over time there may be a need for guidance on specific AI issues. Within the limitations of competition law, industry bodies need to work with firms and regulators to facilitate knowledge sharing and identify any emerging areas of regulatory uncertainty.

Clarifying regulatory expectations and industry best practice over time in relation to the sharing of responsibilities between AI providers and financial services firms implementing AI solutions will help provide certainty, consistency and efficiency. Similarly, best practices need to emerge around the information AI providers ought to make available to their clients for due diligence purposes, while accommodating IP concerns.

The UK government has promised an AI bill, focusing on the developers of the most advanced models. A principles-based, risk-driven and outcomes-focused bill may assist in resolving uncertainties. Similarly, the BoE has signalled at least a potential to regulate key AI providers directly, along with other possible regulatory clarifications.[46] These initiatives provide an opportunity to resolve any outstanding areas of uncertainty.

UK Finance will engage keenly as this policy area develops.

45 How Do We Make Generative AI Green? | Accenture

46 Bank of England - Engaging with the machine: AI and financial stability − speech by Sarah Breeden, 2024

06

References

Accenture - Generating growth how generative AI can power the UK's reinvention, 2024. [Online]
Available here
[Accessed 24 September 2024].

Accenture - Generative AI for customer growth, 2024. [Online]
Available here
[Accessed 25 September 2024].

Accenture - Reinvent the enterprise with generative AI, n.d. [Online]
Available here
[Accessed 11 September 2024].

Accenture - Work, workforce, workers Reinvented in the age of generative AI, n.d. [Online]
Available here
[Accessed 10 September 2024].

Anon. - Financial IT Innovations in Fintech, 2024. [Online]
Available here
[Accessed 23 October 2024].

Anon. - UK financial institutions double investment in AI over the past 12 months, n.d. [Online]
Available here
[Accessed 15 November 2024].

Bank of England - Artificial intelligence in UK financial services, 2024. [Online]
Available here
[Accessed 3 November 2024].

Bank of England - Engaging with the machine: AI and financial stability − speech by Sarah Breeden, 2024. Bank of England. [Online]
Available here
[Accessed 3 November 2024].

Bank of England - Model risk management principles for banks, 2023. [Online]
Available here
[Accessed 3 November 2024].

Bank of England - The Bank and the PRA's response to DSIT/HMT: update on our approach to AI, 2024. [Online]
Available here
[Accessed 27 September 2024].

Carlini, Nicholas, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, Colin Raffel - Extracting Training Data from Large Language Models, 2021. [Online]
Available here
[Accessed 19 October 2024].

ClaimBuster - Automated Live Fast-Checking, n.d. [Online]
Available here
[Accessed 9 September 2024].

Department for Science, Innovation & Technology - A pro-innovation approach, 2023. [Online]
Available here
[Accessed 3 October 2024].

Department for Science, Innovation & Technology – AI Opportunities Action Plan Government Response, 2025. [Online]
Available here
[Accessed 13 January 2025].

European Commission - AI Act enters into force, 2024. [Online]
Available here
[Accessed 4 October 2024].

FCA - AI Update, 2024. [Online]
Available here
[Accessed 9 October 2024].

FCA - Financial Lives 2022 survey: insights on vulnerability and financial resilience relevant to the rising cost of living, 2022. [Online]
Available here
[Accessed 23 September 2024].

FCA - Handbook, n.d. [Online]
Available here
[Accessed 10 September 2024].

Financial IT - UK Financial Institutions Double Investment in AI Over the Past 12 Months, 2024. [Online]
Available here
[Accessed 3 September 2024].

Financial Ombudsman Service - Quarterly complaints data: Q1 2024/25, 2024. [Online]
Available here
[Accessed 17 May 2024].

FinTech Magazine - Can technology help banks navigate a recession?, 2023. [Online]
Available here
[Accessed 4 September 2024].

Forbes - Reasons Why Generative AI Pilots Fail To Move Into Production, 2024. [Online]
Available here
[Accessed 14 October 2024].

Google - Fast Check Tools, n.d. [Online]
Available here
[Accessed 9 September 2024].

Google - The ROI of Gen AI, A global survey of enterprise adoption and value. [Online]
Available here
[Accessed 12 September 2024].

Information Commissioner's Office - AI and data protection risk toolkit, 2023. [Online]
Available here
[Accessed 12 September 2024].

Information Commissioner's Office - Consultation series on generative AI and data protection, 2024. [Online]
Available here
[Accessed 17 October 2024].

Information Commissioner's Office - Generative AI: eight questions that developers and users need to ask, 2023. [Online]
Available here
[Accessed 17 October 2024].

Information Commissioner's Office - Guidance on AI and data protection, 2023. [Online]
Available here
[Accessed 12 September 2024].

Information Commissioner's Office - Guidance on AI and data protection, 2023. [Online]
Available here
[Accessed 17 October 2024].

Information Commissioner's Office - Response to the consultation series on generative AI. [Online]
Available here
[Accessed 12 December 2024].

Institute of Customer Service - UK Customer Satisfaction Index July 2024, 2024 [Online]
Available here
[Accessed 17 October 2024].

J.P. Morgan - How AI will make payments more efficient and reduce fraud, 2023. [Online]
Available here
[Accessed 13 September 2024].

Liu, Y. et al. - Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models, 2023. [Online]
Available here
[Accessed 11 September 2024].

MIT Technology Review Insights - Transforming software development with generative AI, 2024. [Online]
Available here
[Accessed 21 October 2024].

National Cyber Security Centre - Cloud security guidance, 2023. [Online]
Available here
[Accessed 2 September 2024].

National Cyber Security Centre – Guidelines for Secure AI System Development, 2023. [Online]
Available here
[Accessed 27 November 2024].

Oracle - 7 Ways GenAI Can Help Improve Software Development, 2024. [Online]
Available here
[Accessed 2 September 2024].

Touvron, H. et al. - LLaMA: Open and Efficient Foundation Language Models. 2023. [Online]
Available here
[Accessed 11 September 2024].

Trustwise, n.d. - Trustwise. [Online]
Available here
[Accessed 12 September 2024].

UK Finance - The impact of AI in financial services. [Online]
Available here
[Accessed 5 October 2024].

World Economic Forum (in collaboration with Accenture) – AI Governance Alliance: Generative AI Governance: Shaping a Collective Global Future, January 2024
Available here
[Accessed 17 October 2024].

Zilliz, n.d. [Online]
Available here
[Accessed 11 September 2024].

This document has been jointly authored by UK Finance Limited ("UK Finance") and Accenture, and reflects the views of UK Finance and Accenture only.

Please note that this document is intended to provide general information only. It does not represent legal, financial, investment, tax, regulatory, business or other professional advice. This document does not represent or warrant that the information within the document is accurate. Nothing in this document shall operate to be binding on UK Finance and Accenture, nor does this document give rise to any enforceable obligations or duties on UK Finance and Accenture. UK Finance and Accenture (and in the case of UK Finance, only, any of its members) and/or any of their respective officers, employees or agents, shall not be responsible or liable to any person for any loss, damages or costs arising from or in connection with any use of the document or any information or views contained herein. Users of the document should ensure that it is suitable for their use and that appropriate due diligence has been conducted, including in relation to compliance with relevant applicable laws.

This document may refer to marks owned by third parties. All such third-party marks are the property of their respective owners. No sponsorship, endorsement or approval of this content by the owners of such marks is intended, expressed or implied.

UK Finance holds all copyright and other intellectual property rights in this document, and this document should not be commercialised, used or reproduced in whole or part without the express written permission of UK Finance.

© 2025, UK Finance