

f**AI**nance 5.0

GenAI and Agentic AI in the Transformation of the Financial Sector





WHO WE ARE:

Authorship & Team

The Authorship team is made up of multidisciplinary professionals with expertise in data, technology, business, and innovation. The entire Artefact team works collaboratively to deliver high-impact solutions, always aligned with our clients' needs.

Victor Pontello

Data & AI Consultant Director |
LatAm Financial Services Lead

Felipe Longo

Senior Data Consultant

Gustavo Lourenço

Senior Data Consultant

Ilton Chaves

Data Consultant

Pedro Lauand

LatAm Partner

Larissa Ferrari

Senior Data Consultant

Pedro Bertasso

Senior Data Consultant

Vincenzo Spatuzza

Senior Data Consultant

Paolo Gozdink

Marketing Specialist Brazil & LatAm

Lay d'Arc

Graphic Designer



ABOUT AUTORSHIP

Artefact

Artefact accelerates the adoption of data and Artificial Intelligence to positively impact people and organizations. We offer a wide range of services, from strategy to operations, implementing AI solutions across industries to help companies capture the competitive advantage of data and AI transformation.

in  



Summary

Introduction	1
The AI Landscape in the Financial Sector	4
The AI Landscape in the Financial Sector	
AI by the Numbers: Impact and Adoption in the Financial Sector	5
The State of the Art of Technology in the Financial Sector	8
Agentic AI & Opportunities	13
What Is Agentic AI and Why Does It Represent Such a Great Opportunity?	
GenAI & Agentic AI Applications	24
GenAI and Agentic AI: Applications and the Art of the Possible	
How AI Is Redefining the Financial Sector	36
Success Cases: How AI Is Redefining the Financial Sector	
Optimizing Customer Experience with GenAI: A New Paradigm in Customer Service	37
Challenges and Obstacles	38
Solution Development	38
Achieved Benefits	40
Highlights	40
Operational Efficiency: Optimizing Middle and Back Office Processes with GenAI	41
Challenges and Limitations	41
Developed Solution	42
Achieved Benefits	43
Reflections	44



✓ Challenges and Strategies

45

Challenges and Strategies for Implementing AI in the Financial Sector

Artefact's Convictions on the Use of GenAI and Agentic AI in the Financial Sector	46
Key Challenges in Implementing AI in the Financial Sector	48
Best Practices to Overcome Challenges and Ensure Successful Implementation	51
Machine Learning: Enhancing Fairness and Interpretability	54
Generative AI: Mitigating Hallucinations and Ensuring Data Security	54
Agentic AI: Strengthening Governance and Ethical Boundaries	55

✍ Strategy for AI Implementation

56

Strategy for AI Implementation in the Financial Sector

Diagnosis and Strategic Planning	57
Stakeholder Engagement	59
Pilots and Iterations: An Agile Approach to Reduce Risk and Maximize Value	61
Data Security and Privacy as a Strategic Priority	64
Cloud-Based Solutions	66
On-Premise Solutions	67
Critical Decision Factors	68
Trends and Recommendations	68
Special Considerations and Best Practices	69

📐 Strategic Framework & Business

72

Strategic Framework for Agentic AI Implementation by Business Area

Introduction: From Vision to Strategic Action	73
The Strategic Imperative: Why Focus on Agentic AI Now?	74
Key Dimensions of an Agentic AI Strategy	76
Essential Strategic Questions by Dimension	77
Visual Framework: The Strategic Cycle of Agentic AI Implementation	81
Adapting the Framework by Business Area	82
Conclusion: Navigating the Agentic Future with Strategy	83



 Artefact as a strategic partner	84
Why Artefact?	
Clients we have already impacted with GenAI and Agentic AI	86
 Technical Appendix	89
Framework para a Implementação de Agentic e GenAI on-premise	
Hardware Requirements	91
Tech Stack	93
Security and Compliance	96
Scalability and Performance Optimization	98
Monitoring and Maintenance	100
Overview of the Agentic AI / GenAI On-Premise Implementation Flow	102
 Glossary	103
 Links & References	106

INTRODUCTION

The financial sector is among the most dynamic and transformative areas of the global economy

In a constantly evolving landscape, banks, brokerage firms, and fintechs face the challenge of rapidly adapting to new market demands, technological advancements, and consumer expectations, all while remaining competitive in an increasingly complex environment.

In this eBook, our goal is to demonstrate how Artificial Intelligence (AI) and its various applications can become strategic allies for financial institutions. We will explore everything from basic applications, such as process automation and data analysis, to the latest trends and innovations driven by Agentic AI, including real-time monitoring of financial transactions and proactive fraud detection—redefining how companies interact with data and make decisions.

While GenAI has revolutionized the tech world by enabling systems to create content and interact with humans in the realms of speech, writing, and even art, Agentic AI adds a layer of autonomy and decision-making. It allows these systems to act proactively, learn from experience, and perform complex tasks with minimal intervention—further enhancing the impact of technology and AI on everyday life and relationships. In the financial services sector, this means assistants that not only generate insights but also apply them—optimizing operations, detecting fraud, and improving risk management in real time, among many other opportunities.

With years of experience working alongside the largest players in the industry, Artefact offers strategic solutions and the implementation of relevant case studies that comprehensively and innovatively integrate these technologies. Moreover, this prior experience, combined with the innovative and entrepreneurial spirit that runs through the company, allows us to identify a wide range of use cases that once seemed impossible but are now not only feasible thanks to technological evolution—they are set to become the reality of the financial services sector in the near future. These initiatives are designed to drive sustainable growth, foster continuous innovation, and maximize consistent and long-lasting results.



We hope this material helps you understand the **strategic value of Artificial Intelligence** and how your company can leverage technology to drive business outcomes by applying these innovations to successfully navigate **the financial market**.

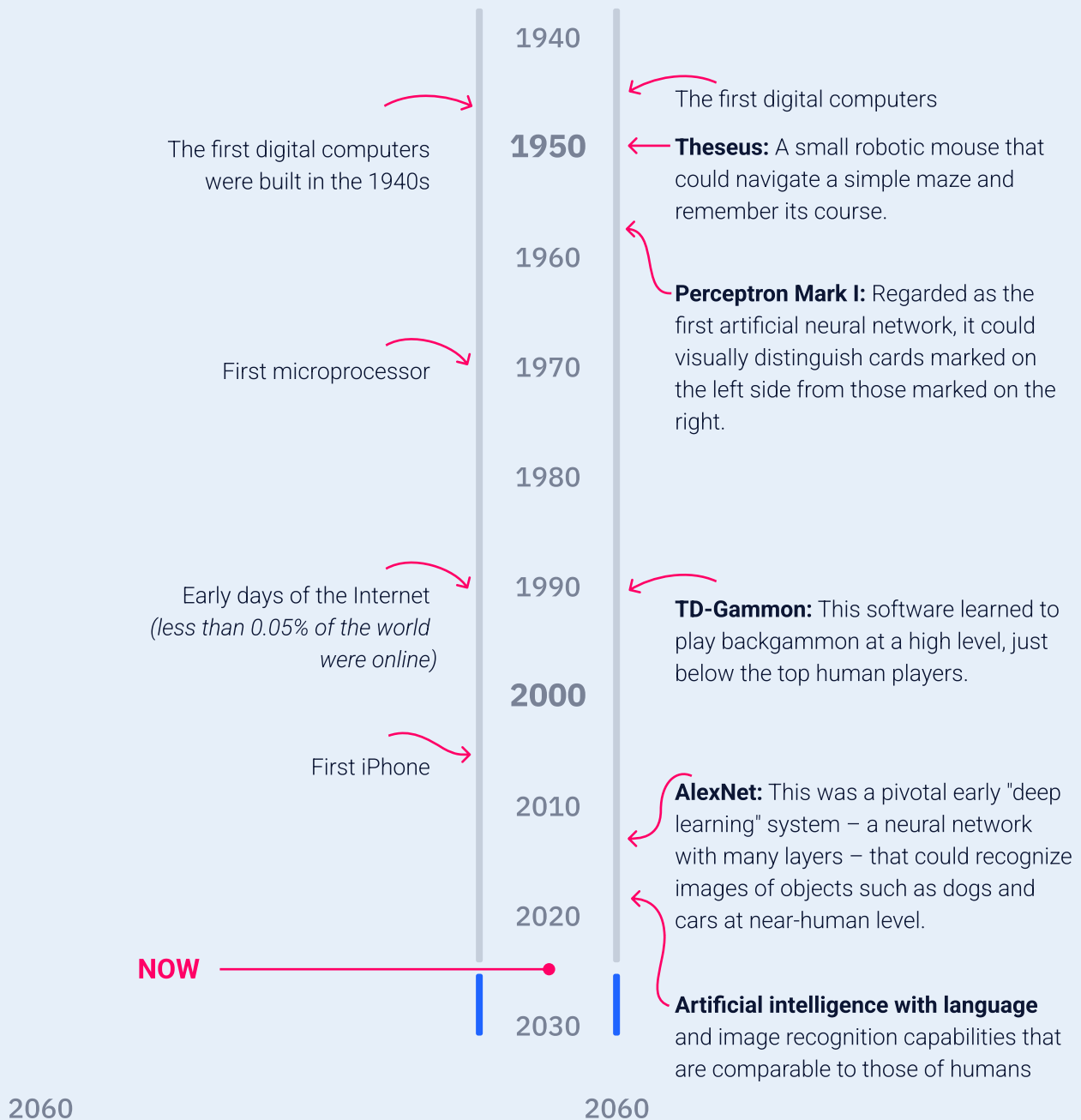


Enjoy your reading!



LINHA DO TEMPO

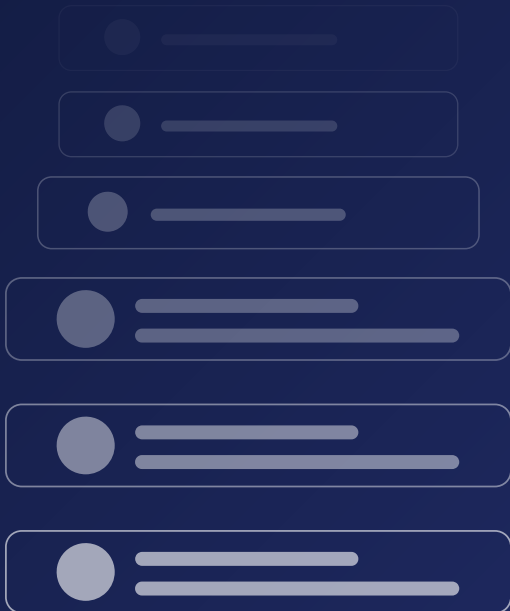
Notable Artificial Intelligence Systems ^[1]



[1] Source: The brief history of artificial intelligence: the world has changed fast – what might be next?

CHAPTER 1

The AI Landscape in the Financial Sector



50K

40K

30K

20K

10K

Jan

Feb

Mar

Apr

May

AI in Numbers: Impact and Adoption in the Financial Sector

Financial institutions have always been leaders in data usage, driving technological innovations in the sector. From mainframe computers in the 1980s to today's advanced solutions like Generative Artificial Intelligence (GenAI) and Agentic Artificial Intelligence (Agentic AI), the financial sector has been a pioneer in applying these technologies.

TECHNOLOGY EVOLUTION – TIMELINE

	1980's	Machine Language
	1998	Assembly Language
	2007	Instructions Languages
	2011	Object Oriented Language
	2022	GenAI
	2023	Agentic AI

The results achieved by financial institutions using GenAI in 2024 are remarkable and prove the potential of this technology to generate value in the sector. Concrete cases show significant returns: optimization of trading and portfolio with a 25% ROI, improvement in customer experience and engagement reaching 21%, and efficiency gains in critical operational areas, such as document processing and report generation, hitting 11%. These numbers make it clear that GenAI is already a transformative reality. Companies that know how to seize this opportunity now will be ahead of the competition. [2]



What are the top generative AI use cases by ROI?



25% Trading and portfolio optimization

21% Customer experience and engagement

11% Document processing

11% Report generation

[2] Source: NVIDIA – State of AI in Financial Services: 2025 Trends.

Besides the gains observed with GenAI, the evolution toward intelligent agent-based solutions is opening new and powerful opportunities. These agents are reshaping work in three main areas: cost reduction, by simplifying operations and eliminating unnecessary steps; lead time reduction, by removing human bottlenecks and automating processes; and service quality improvement, through the optimized use of tools and work environments. Concrete examples reinforce this impact: a 25% reduction in R&D (Research & Development) cycle time, up to 10 times savings in call center operations, up to 50 times increase in marketing content production speed, and boosts of up to 40% in IT team productivity. These results prove that the strategic adoption of agents is a decisive step toward generating sustainable and measurable value in the financial sector.

WHAT (VALUE DRIVERS) CAN BE EXPECTED?

Agents are reshaping the way we conceive work and unlocks 3 value opportunities:



Cost Reduction

Rationalization & Reduction



Lead Time Reduction

Process streamlining & Removing human bottlenecks



Service Quality

Optimized usage of different tools & environments

SOME EXAMPLES

[3]

25%

reduction in R&D cycles time

X10

Call Center cost cuts with Agents

X50

speed increase of Marketing Blog post creation

40%

IT department productivity boost

[3] Source: Artefact – Conhecimento interno aplicado a projetos no setor financeiro.

Historically, the financial sector has been built on a fundamental pillar: trust. In a highly regulated and globally interconnected environment, this trust has always been tied to the strategic management of customer data. However, the dynamics have changed. Loyalty, once guaranteed by a less consumer-centered market, now needs to be earned daily.

In this new scenario, Generative AI, combined with Agentic AI, emerges as a key tool to strengthen trust, optimize operations, and reinvent the customer experience. Whether in service personalization, security enhancement, or back-office efficiency, the strategic adoption of these technologies is redefining the future of the financial sector.

The State of the Art of Technology in the Financial Sector

Financial institutions are only beginning to explore the true potential of Generative AI (GenAI), a technology that goes beyond information processing, enabling the creation of genuinely new content from simple commands. Unlike previous technological revolutions, where humans were always the agents of transformation—whether operating machines, driving harvesters, or programming systems—technology now takes on

the role of creator. Powered by Large Language Models (LLMs), GenAI can generate insights, develop strategies, produce reports, and even write code without the need for active human intervention. This is the real game changer: automation is no longer limited to repetitive tasks but expands the boundaries of creativity and decision-making, inaugurating a new paradigm for the financial sector. [4]



[4]] Adapted from internal analysis – Artefact. Generative AI Survey (institutional document).

LLMs are advanced artificial intelligence models trained on large volumes of textual data, capable of understanding and generating natural language with high accuracy. After specific adaptations, these models can be used for interaction, content creation, and automation of textual processes. In the GenAI ecosystem, there are also models focused on other data formats, such as images, videos, and audio, as well as Multimodal Models that integrate different types of inputs simultaneously, further expanding application possibilities.

Just as the microcomputer, the Internet, and the smartphone transformed our relationship with technology, GenAI is redefining the interface between humans and machines, making interaction more accessible, intuitive, and productive. Institutions that strategically adopt this technology are already reaping benefits such as productivity gains, cost reductions, and greater innovation agility. Players advancing quickly in GenAI implementation are gaining a significant competitive advantage, while others who have yet to grasp the urgency of this transformation are falling behind. [3]



But the sophistication of this technology goes beyond simple content generation. The next step for AI in the financial sector involves more autonomous and intelligent systems capable of acting in a structured and strategic way.

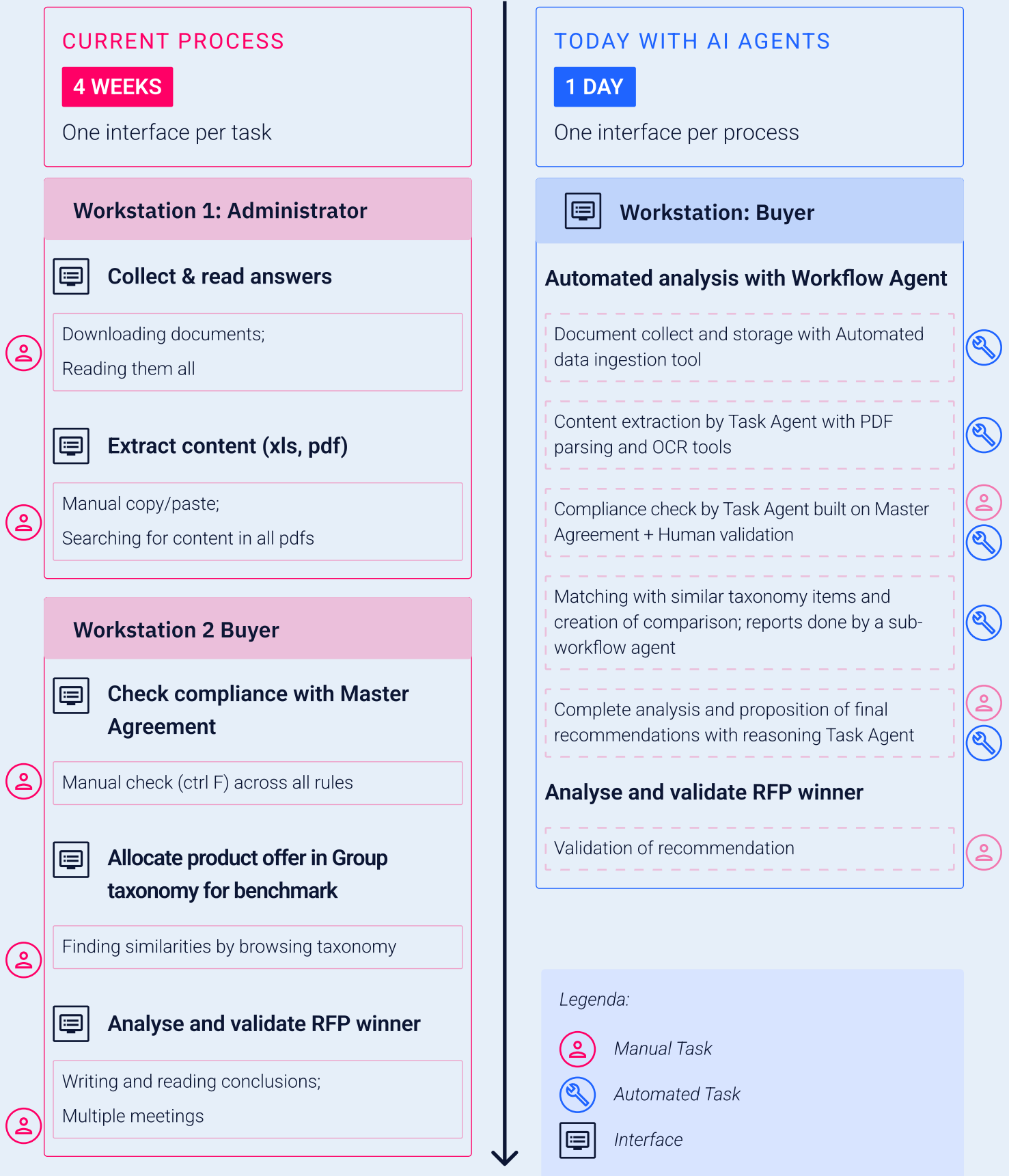
This advancement is reflected in the concept of Agentic AI, which expands GenAI's capabilities by allowing models not only to generate responses but also to make informed decisions, perform tasks continuously, and dynamically adapt to the context. Whether interacting with the real world, accessing up-to-date information, or using APIs and tools beyond training data, Agentic AI ushers in a new era of autonomy and efficiency.

The image below illustrates the significant transformation that the use of Agentic AI can bring to the process of selecting a winner in an RFP (Request for Proposal), comparing the traditional model with the agent-optimized model. [3]

The analysis and validation of RFPs (Requests for Proposal) is still, in many banks, an operationally intensive, decentralized process heavily dependent on manual tasks. Traditionally, this process can take up to four weeks, involving multiple interfaces, several professionals, and a fragmented workflow. From downloading and reading documents in various formats to checking contract compliance, grouping by taxonomy, and finally deciding on the winning supplier, each step requires considerable human effort—especially when performed at scale. This traditional model is characterized by low efficiency, high operational burden, and little integration between stages.



Difference between analyzing and validating an RFP **with** or **without** agents



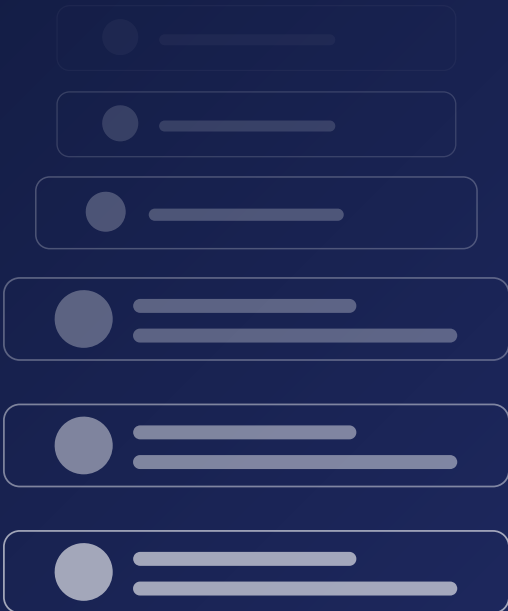
With the introduction of autonomous AI agents, this scenario is radically transformed. The use of Agentic AI allows the process to be consolidated into a single workstation, centralized with the buyer, and reduces the total execution time to just one day. Specialized agents automate the ingestion and extraction of data from PDF or Excel documents, perform compliance analysis against master contracts, compare responses to pre-established benchmarks, and propose final recommendations based on logical reasoning.

Instead of multiple professionals performing repetitive and error-prone tasks, human involvement is focused solely on the final validation step—centered on critical analysis and decision-making. The operational efficiency gain is evident, but the strategic impact goes beyond that: by freeing up time and cognitive capacity of teams, banks gain greater agility to respond to market opportunities, reduce risks, and make more informed decisions.

This transformation clearly illustrates the value of Agentic AI: more than automating tasks, it repositions the human role in financial operations, elevating the quality of decision-making processes and accelerating the pace of innovation.

CHAPTER 2

Agentic AI & Opportunities



50K

40K

30K

20K

10K

Jan

Feb

Mar

Apr

May



What is **Agentic AI** and why does It represent such a great **opportunity**?

The transformation we observed in the RFP process is just one among hundreds of possible applications of a new generation of artificial intelligence: Agentic AI. If traditional AI automated specific tasks, and GenAI introduced creativity and on-demand content generation, Agentic AI takes this revolution a step further—with agents capable of making decisions, acting autonomously, and adapting to their environment in real time.

It is no longer just about answering questions or generating text based on commands: it is about executing complex end-to-end missions, connecting systems, accessing dynamic data, operating tools, and learning from every interaction.

This ability to operate with purpose, context, and coordination is what makes Agentic AI such a massive strategic opportunity for the financial sector. And while the concept may seem new, its evolution is already underway—with technical milestones advancing year after year and rapidly bringing us closer to a scenario where intelligent agents can act as true digital collaborators. The timeline below illustrates how this transformation is happening at an accelerated pace.

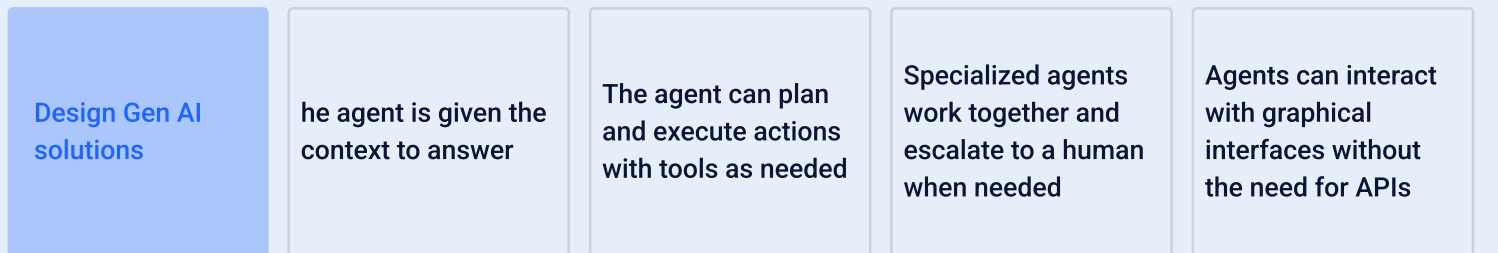
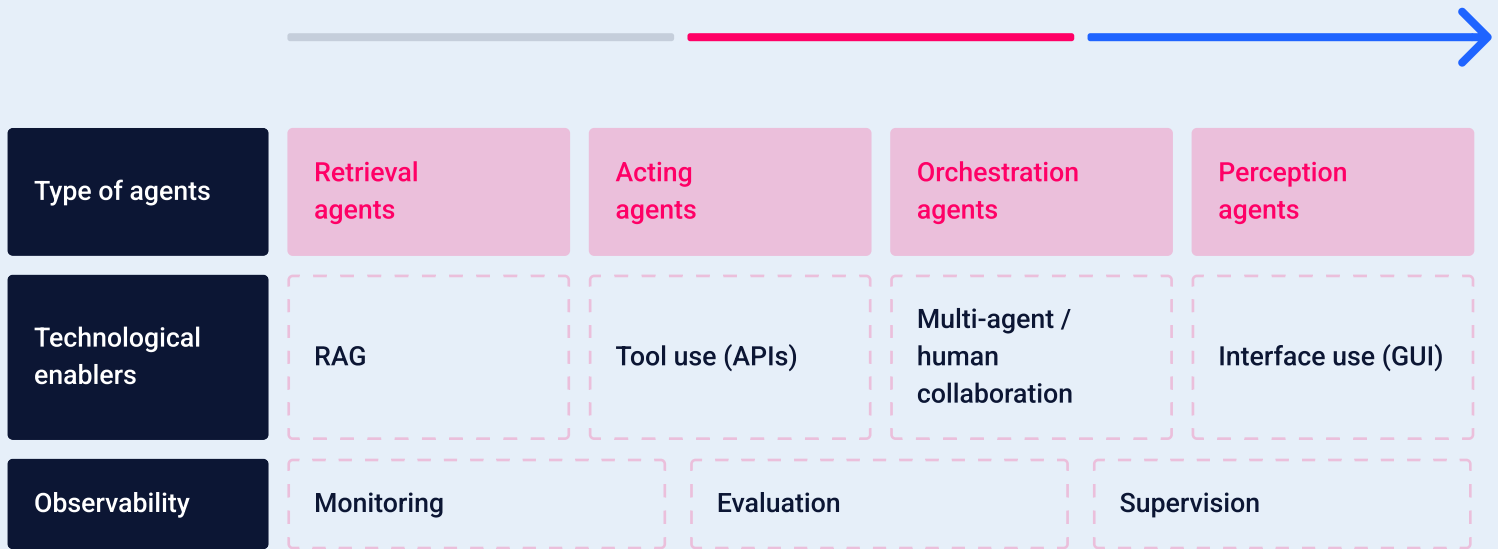


“Agents” have rapidly evolved, their new abilities pave the way for game changing process automation opportunities.

2023: Assist

2024: Act

2025: Automate



This evolution journey can be understood through three major milestones



Assist



Act



Automate



Each phase represents a concrete expansion of AI agents’ capabilities and the opportunities they unlock for the financial sector:



2023: Assist

In this initial stage, retrieval agents emerge—specialized in retrieving relevant information based on provided contexts. Supported by technologies such as RAG (retrieval-augmented generation) and tools like LangChain, these agents are capable of fetching data, consolidating information, and responding accurately. At this point, observability is still limited to basic performance monitoring of the models.



2024: Act

With advances in the use of APIs, acting agents emerge—capable of executing concrete actions within external tools. This evolution marks an operational turning point: agents not only respond but also act based on what they interpret, integrating with systems like CRMs, ERPs, and analytics platforms. In the same year, orchestration agents appear, coordinating multiple agents or collaborating with humans.



2025: Automate

In 2025, agents evolve into more complex interactions with systems, giving rise to perception agents. The key difference lies in their ability to navigate graphical user interfaces (GUIs), without relying solely on APIs to perform actions. This enables the automation of workflows involving multiple applications or legacy systems—something especially relevant for financial institutions with more heterogeneous IT architectures. Technologies such as Runner H and Claude make this advancement possible, while active supervision mechanisms ensure operational control and security.

This technology roadmap clearly demonstrates that Agentic AI is not a distant trend. It is already shaping new ways of operating, making decisions, and creating value. For leaders in the financial sector, understanding these phases is essential to structuring robust strategies, identifying priority use cases, and ensuring the scalable and secure adoption of this transformative technology.

To truly understand the potential of Agentic AI, it is crucial to grasp how these agents operate in practice. Unlike solutions based solely on large language models (LLMs), agents are autonomous entities composed of three main components:

☒ **Perception;**

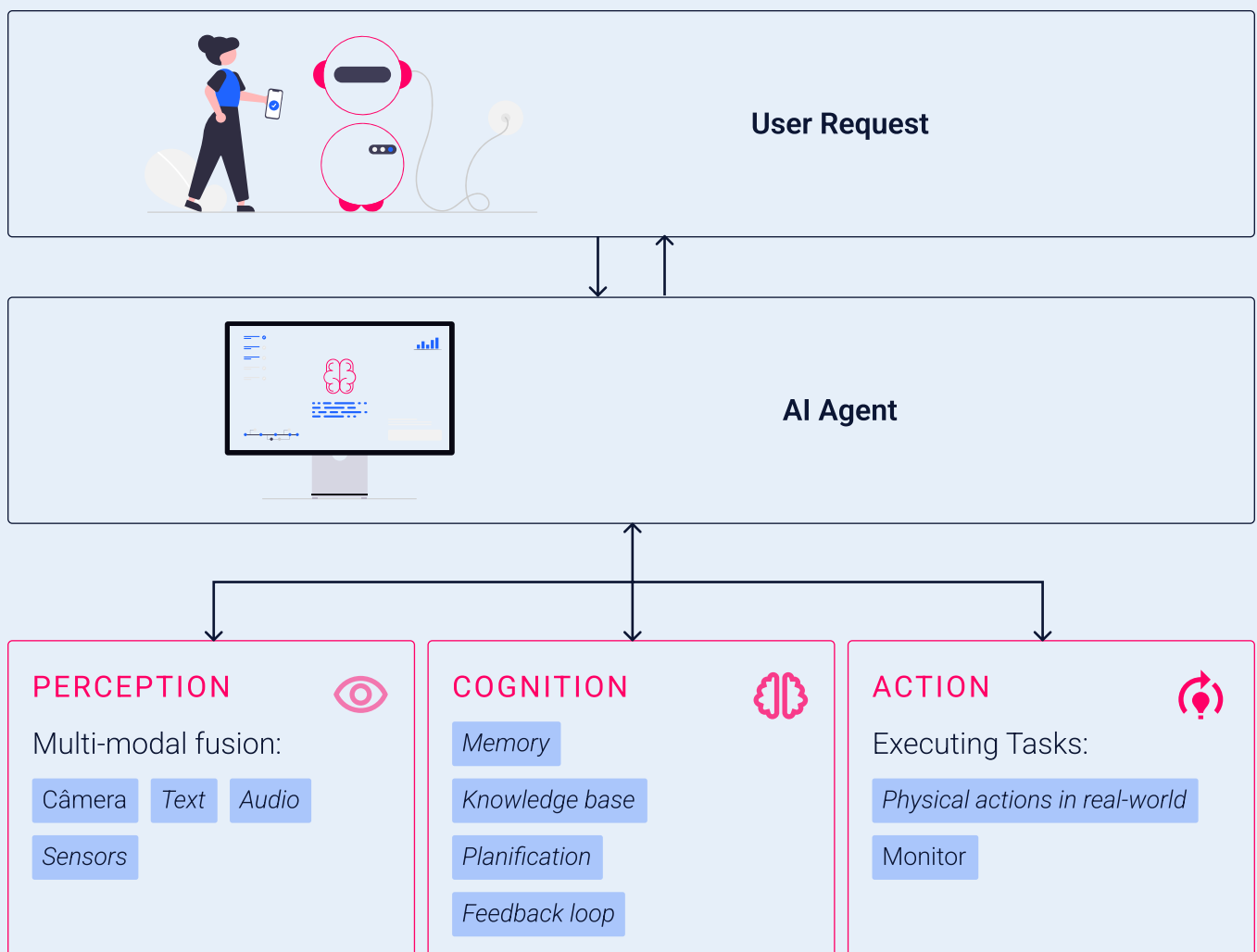
 ☒ **Cognition;**

 ☒ **Action.**

Each of these elements plays a vital role in how the agent interacts with its environment, makes decisions, and performs tasks in a continuous and adaptable manner.

AGENT, NEW DEFINITION

An autonomous entity that perceives, reasons, acts and adapts to changes.



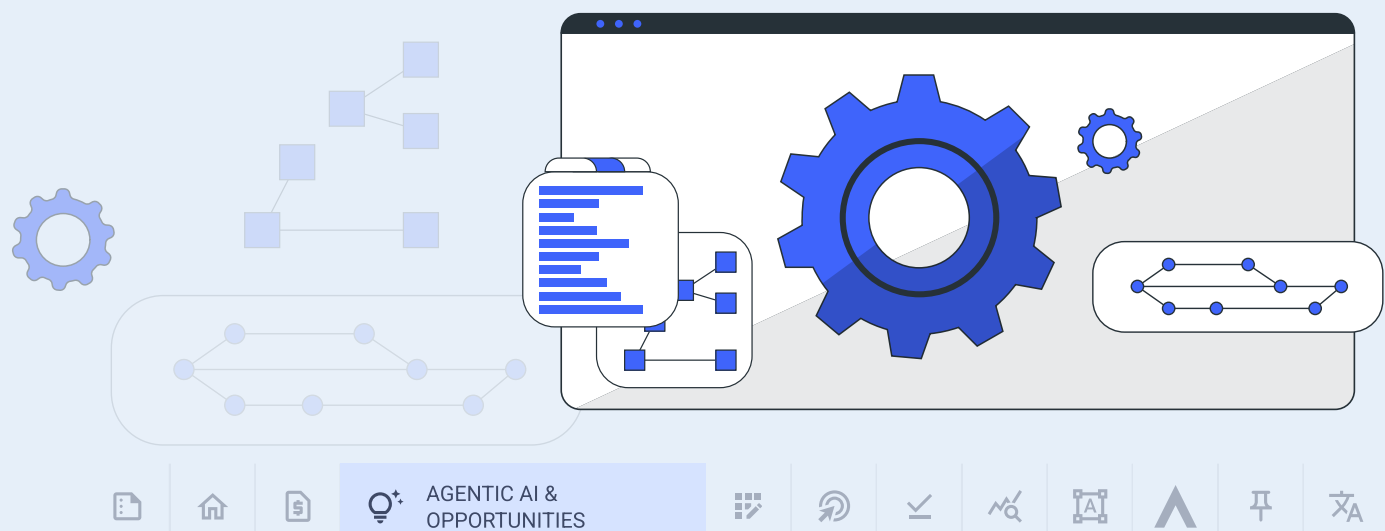
Perception enables the agent to gather information from various sources through what is known as multimodal fusion. This includes text, audio, images (via cameras), and a range of sensor data. In business contexts, for instance, this means an agent can cross-reference structured system data (such as ERPs and CRMs) with unstructured information (such as emails, customer service recordings, and documents), creating a rich and contextualized view of the situation.

Based on the perceived data, the cognition component comes into play. It combines memory, knowledge bases, and planning capabilities to interpret the incoming signals, make decisions, and adapt. A key aspect here is the feedback loop, which allows the agent to learn from its own actions and refine its behavior over time—an essential ability for dealing with complex environments such as the financial market.

Finally, the agent is able to carry out tasks based on its developed reasoning. This may include interactions with systems via API, navigating graphical user interfaces, or even performing physical actions in connected environments. Additionally, the agent retains the ability to monitor the results of its actions, ensuring safer, more efficient, and auditable operations.

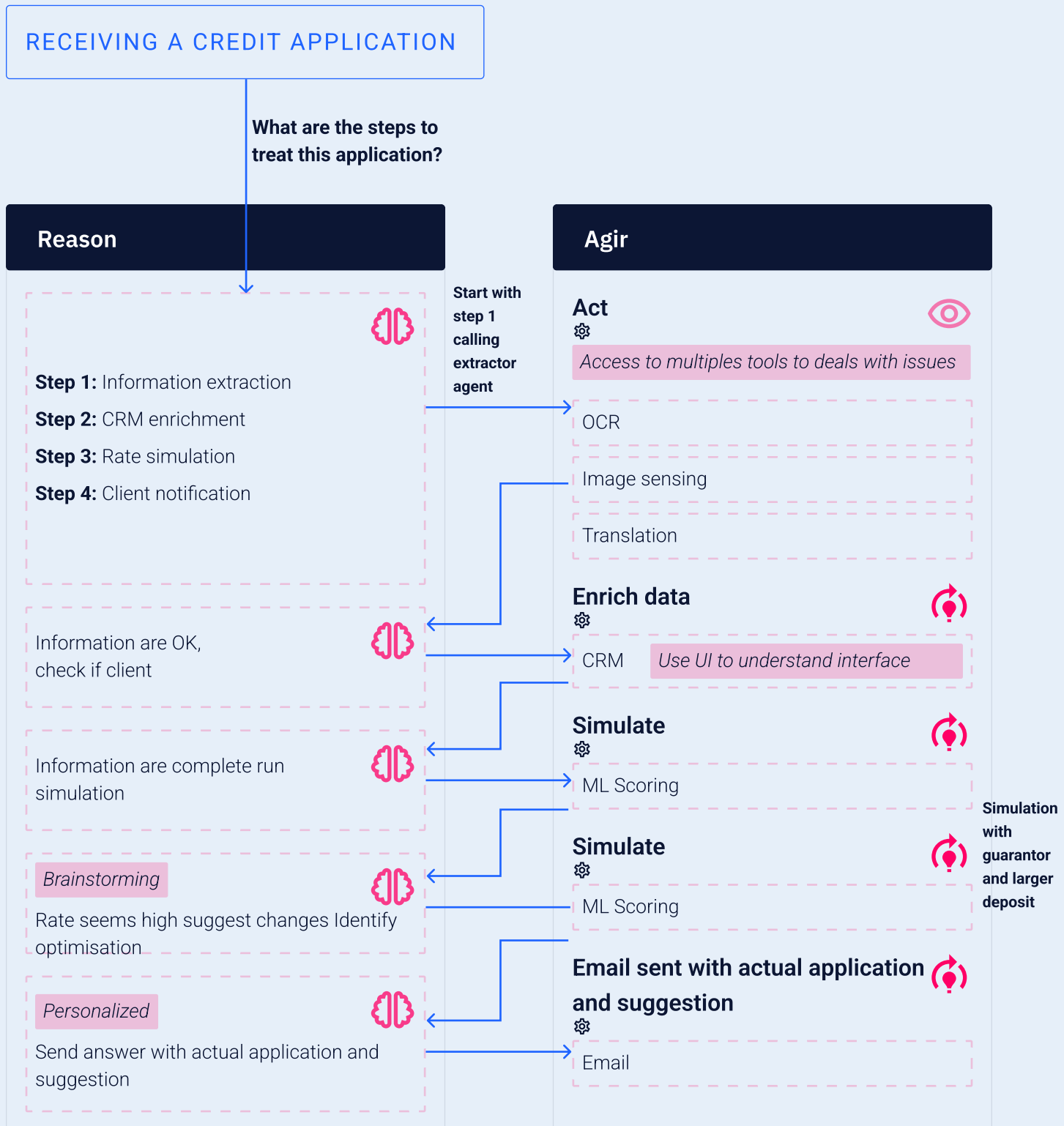
This architecture enables agents to act proactively, respond to changes in real time, and evolve based on experience—which makes them one of the most promising innovations in modern AI.

Agentic AI allows for the end-to-end automation of complex workflows through coordination between multiple agents. A clear example, highly applicable to the financial sector, is the automated processing of credit applications.



In this workflow, agents operate in an orchestrated manner, combining reasoning and execution in a logical sequence of steps, as illustrated in the following diagram:

By creating a chain of agents, we can automate an entire workflow end-to-end with React (Reason & Act) - *Automated credit application processing*



Upon receiving a credit application, a chain of agents is activated, each responsible for a specific part of the process:



1. Information Extraction

An agent automatically extracts relevant data from the application using OCR, translation, and interface interpretation. It transforms raw inputs into structured information.



2. Data Enrichment (CRM)

Next, another agent validates and enriches this data with information from the CRM, adding contextual depth to the customer profile.



3. Rate and Terms Simulation

With the data prepared, a third agent runs a credit simulation based on machine learning models to predict the ideal rate.



4. Offer Optimization and Personalization

If the rate is not optimal, the system identifies alternatives (such as requiring a guarantor or a higher down payment), re-runs the simulation, and optimizes the offer in a personalized manner.



5. Customer Response

Finally, an agent automatically drafts and sends a response to the customer, including the simulation results, analysis, and recommended offer.

This example illustrates how chains of agents can replicate cognitive processes—but with superior speed, accuracy, and scalability. Each agent understands its role, queries systems, makes decisions based on rules or models, learns from outcomes, and collaborates with others to fulfill the final objective.

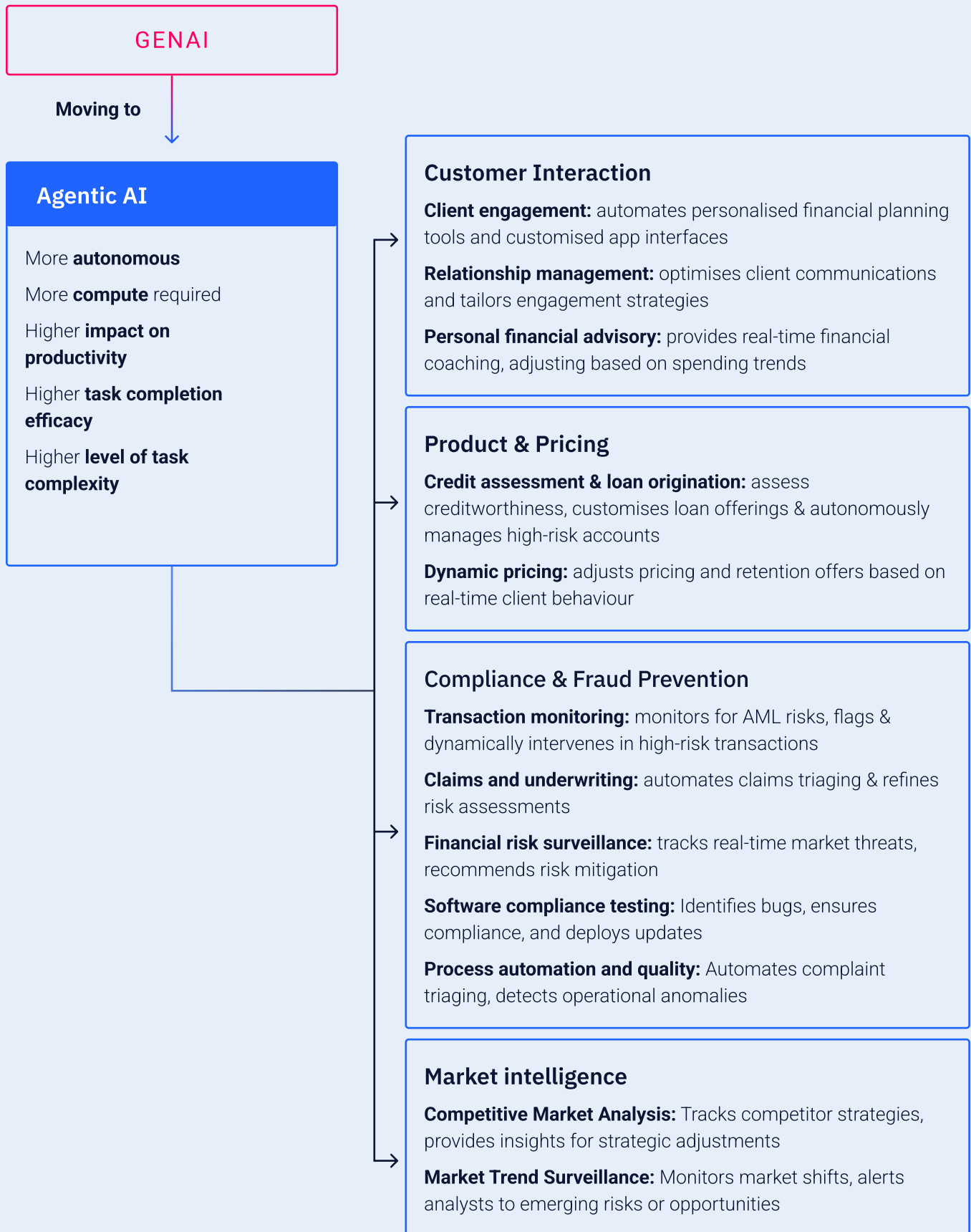
However, for this level of sophistication to be implemented safely and effectively, several critical considerations must be addressed. It is essential to establish robust governance frameworks to ensure that agents operate within ethical and regulatory boundaries, preventing errors and data breaches. This includes building “guardrails” that define what agents are allowed to do and prevent hallucinations. Additionally, integrating agents increases the surface area for cyberattacks, requiring solutions such as contextual authentication and continuous monitoring. Despite their autonomy, agents also require human oversight in high-impact decisions, ensuring that their actions remain aligned with organizational goals.

By balancing autonomy, security, and governance, the use of agents in the financial sector promises to usher in a new era of efficiency and innovation, profoundly transforming the way financial services are delivered and consumed.

According to Citi GPS, Agentic AI represents a shift toward a “do it for me” economy, where technology takes on both decision-making and task execution [6]. The greater autonomy of Agentic AI, compared to GenAI, allows it to handle repetitive, data-intensive processes—optimizing workflows, enhancing compliance, and improving decision-making. The image below provides a clearer view of the transition from GenAI to Agentic AI and its potential applications in the financial services sector. [5]

[5] Extraído de: Citi GPS. *Agentic AI: Finance & the ‘Do It For Me’ Economy*.

[6] Adaptado de artigo: *InsiderFinance Wire – Sahaj Godhani, Medium*.



In the financial sector, Agentic AI means, for example, intelligent assistants that identify risks, adjust portfolios, automate complex processes, and even manage customer interactions without the need for human intervention.

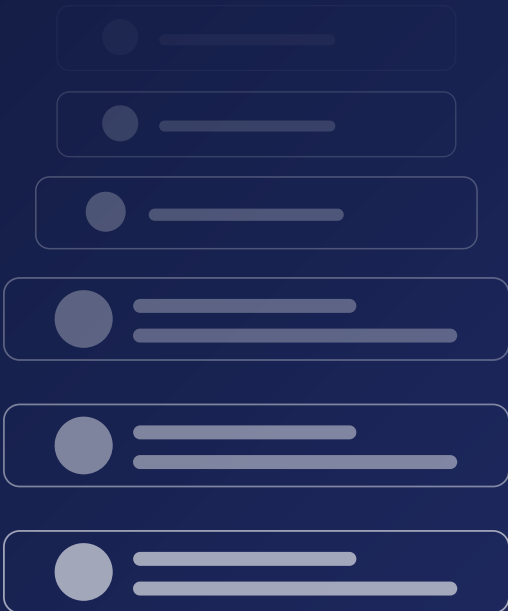
With this level of technological sophistication, institutions can scale operations, deliver mass personalization, and significantly increase efficiency—while maintaining customer trust in an increasingly dynamic market.

The future of the sector will not be defined merely by who adopts AI, but by who knows how to leverage its highest sophistication and strategic intelligence.



CHAPTER 3

GenAI & Agentic AI and their Applications



50K

40K

30K

20K

10K

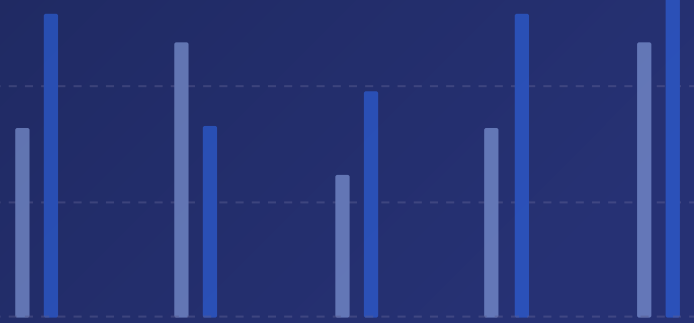
Jan

Feb

Mar

Apr

May



GenAI and Agentic AI:

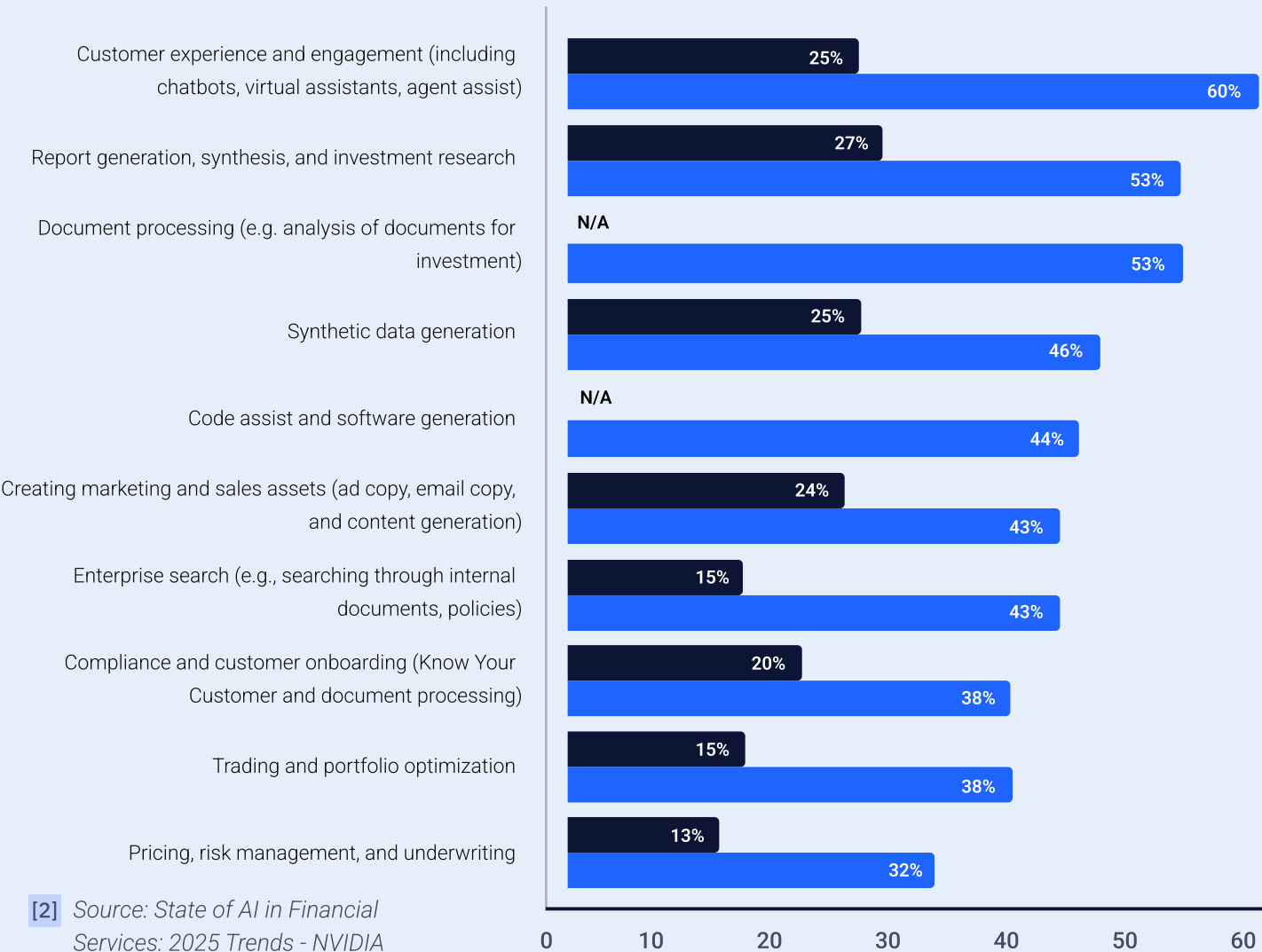
Applications and the Art of the Possible

In the financial sector, the adoption of generative AI-based solutions has gained accelerated traction, with companies seeking to explore new use cases that drive efficiency, personalization, and innovation. As shown in the image [2], there was a significant leap from 2023 to 2024 across various areas. Notably, Customer Experience grew from 25% to 60%, and Report Generation and Research jumped from 27% to 53%. Other areas, such as Document Processing, Code Assistance, and Enterprise Search, also advanced rapidly, reflecting the financial market's growing bet on the transformative potential of GenAI.

Top Generative AI Use Cases

2023

2024



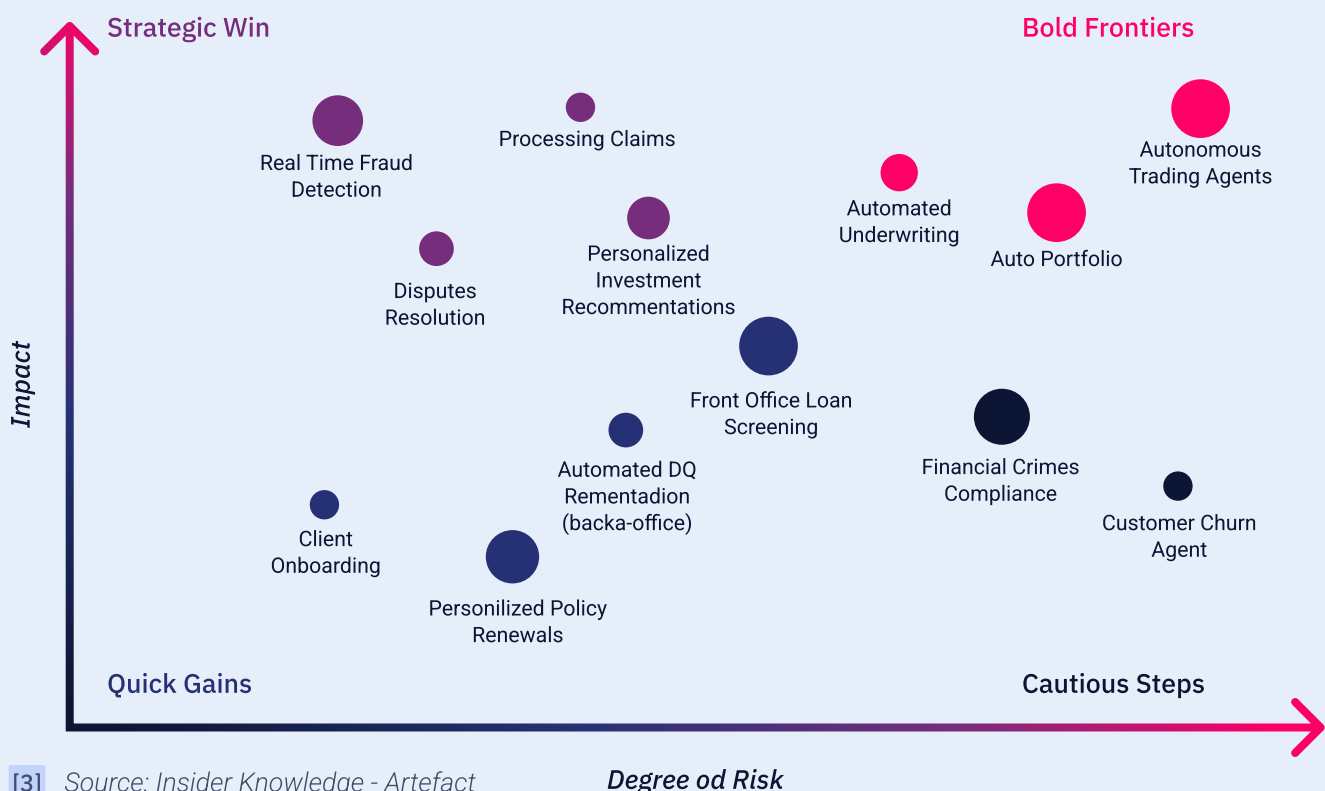
[2] Source: State of AI in Financial Services: 2025 Trends - NVIDIA

GenAI applications in the financial sector are already a reality, as the study shows. Moreover, this technology lays the foundation for a range of even more transformational use cases that can be developed through Agentic AI.

To guide the strategic adoption of Agentic AI solutions in this sector, it is essential to consider three central dimensions: risk, impact, and complexity. The image below illustrates this logic, where the vertical axis represents the potential impact of the solution, the horizontal axis indicates the degree of risk involved, and the size of the bubble reflects the effort required to build the agent.

Your Agentic AI roadmap will be refined by a combination of risk, impact and complexity

Tamanho da bolha = grau de esforço



[3] Source: Insider Knowledge - Artefact

Degree of Risk





In this scenario, two cases stand out as quick wins. Client Onboarding allows automating the customer's initial journey with agents that collect documents, validate information, integrate systems (such as CRMs and fraud prevention), and notify the internal team or the customer themselves. Personalized Policy Renewals have agents review policy histories to suggest personalized renewals based on profiles and send proactive communications to the customer.

In the Strategic Wins layer, applications that require moderate effort but deliver high business impact stand out. Cases like Real-Time Fraud Detection, Disputes Resolution, and Claims Processing optimize critical processes with significant gains in speed and accuracy, without the level of risk of the more advanced frontiers.

The Bold Frontiers include use cases that combine high impact with greater risk and complexity —such as Autonomous Trading Agents, Auto Portfolio Rebalancing, and Automated Underwriting. These applications represent the future of autonomous intelligence in the sector, requiring robust architectures and sophisticated governance.

To facilitate the prioritization and structuring of initiatives with Agentic AI, it is useful to organize use cases into broad value areas.

Below, we present a structured view grouping the main cases into four key categories:

-  **Risk Management**
-  **Process Automation**
-  **Customer Service and Personalization**
-  **Decision Making and Insights**

For each case, it is indicated where it generates the most value, with marks of full or partial alignment. This allows quick visualization of how each initiative connects to strategic and operational objectives, helping prioritize and implement Agentic AI-based solutions.

Risk Management agents focus on mitigating financial, operational, and regulatory risks, with the ability to operate continuously and learn constantly. Process Automation includes agents that optimize routine or operational tasks, reducing errors and increasing efficiency with more responsive and fluid workflows. Customer Service and Personalization involves agents that interact with channels and people to offer personalized and integrated real-time experiences. Finally, Decision Making and Insights covers agents focused on advanced analysis and decision support, allowing artificial intelligence to actively participate in strategy building and business operations.



Use Cases	Risk Management	Process Automation	Customer Service and Personalization	Decision Making and Insights
Real Time Fraud Detection	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Financial Crimes Compliance	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Front Office Loan Screening	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Claims Processing	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> [with feedback loops]
Automated Underwriting	<input checked="" type="checkbox"/> [partial]	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Automated DQ Remediation (back-office)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Disputes Resolution	<input checked="" type="checkbox"/> [partial]	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Client Onboarding	<input type="checkbox"/>	<input checked="" type="checkbox"/> [partial]	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Personalized Policy Renewals	<input type="checkbox"/>	<input checked="" type="checkbox"/> [partial]	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Personalized Investment Recommendations	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Customer Churn Agent	<input checked="" type="checkbox"/> [partial]	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Auto Portfolio Rebalancing	<input checked="" type="checkbox"/> [partial]	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Autonomous Trading Agents	<input checked="" type="checkbox"/> [high complexity]	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> [autonomous decisions]

The above categorization allows for a clear visualization of where each Agentic AI use case can generate the most value within financial operations. To deepen the analysis, below we detail each of the four main areas, presenting concrete examples of agents, their functions, and the expected impact. This view helps to make tangible how agents operate in practice, highlighting the efficiency gains, risk mitigation, and personalization they can bring to different areas of the sector.

Risk Management

Key cases:



Real Time Fraud Detection

Agent that monitors financial transactions in real time, identifying suspicious fraud patterns (such as behavior outside the customer's profile) and immediately triggering automated blocks or investigations.



Financial Crimes Compliance

Agent specialized in compliance that continuously checks transaction compliance with regulations such as AML (anti-money laundering) and KYC (know your customer), cross-referencing data with sanction lists and risk profiles.



Front Office Loan Screening

Agent that performs preliminary credit or loan analysis in the front office, combining structured and unstructured data to calculate scores, validate documentation, and suggest initial decisions.

These cases have high impact and involve considerable risk, being classic areas for using **artificial intelligence to mitigate financial and operational risks.**

With Agentic AI, it is possible to:



Create agents that monitor transactions in real time, detecting anomalous patterns and triggering automatic or semi-automatic investigation workflows.



Enable continuous compliance with agents that check adherence to policies and regulations.



Automate the pre-screening of clients or loans based on risk scores, history, and behavior.

Process Automation

Key cases



Claims Processing

Agent responsible for automatically receiving, classifying, validating, and forwarding claims, using document recognition, business rules, and historical data.



Automated Underwriting

Agent that automatically analyzes the applicant's profile, queries external databases, applies actuarial rules, and provides an initial decision to accept or reject the risk.



Automated DQ Remediation *(back-office)*

Agent focused on identifying and correcting data quality (DQ) issues, such as inconsistent or missing data, maintaining database integrity and properly feeding analytical systems.



Disputes Resolution

Agent that acts in mediating and resolving disputes (for example, in payments or banking transactions), organizing evidence, assessing the situation, and suggesting solutions or next steps.

These cases directly benefit from intelligent automation of repetitive steps, optimizing time and costs. The proposal with Agentic AI is to:



Develop specialized agents that perform tasks such as document validation, data cross-checking, and application of business rules.



Incorporate adaptive logic, allowing agents to learn exceptions and continuously improve.



Integrate these agents with BPM (Business Process Management) or ERP systems for end-to-end process orchestration.

Customer Service and Personalization

Key cases



Customer Churn Agent

Agent that monitors customer behaviors (such as decreased product usage, complaints, or pattern changes) to predict churn risk and suggest personalized retention actions.



Client Onboarding

Agent that manages the new customer onboarding process adaptively, requesting documents based on the profile, explaining the steps, and automatically integrating data into internal systems.



Personalized Policy Renewals

Agent that evaluates the policyholder's history, profile changes, and current context to offer personalized and proactive renewals, adjusting coverage and pricing as needed.



Personalized Investment Recommendations

Agent that suggests personalized investments based on risk profile, goals, preferences, and market events, and can even interact with the user in natural language.

These cases represent initiatives for **personalization and proactive customer engagement**.
Using Agentic AI:



It is possible to create agents that identify early signs of churn and initiate mitigation actions, such as personalized offers and contact.



Onboarding agents make the process smoother, adapting the flow to the customer's needs based on profiles and behavior.



Recommendation agents build personalized journeys based on historical data, preferences, and the user's current context.

Decision Making and Insights

Key cases



Auto Portfolio Rebalancing

Agent that evaluates investment portfolios or products and, based on predefined criteria (such as risk profile or external events), proposes and executes automatic reallocations.



Personalized Investment Recommendations (Also present in Personalization)

Here the focus is on explainability and real-time insight generation to assist the investor in autonomous decision-making.



Claims Processing (with feedback loops)

In this context, the claims agent collects continuous feedback from decisions and outcomes, learning from mistakes and successes to refine its future decisions.

These cases focus more on improving the ability to make well-founded, accurate, and efficient decisions by building agents that organize themselves into a workflow to cover all necessary points. Thus, with Agentic AI:



Financial agents can automatically reassess portfolios based on market events and risk profiles.



In synergy with analytical tools, agents provide recommendations with explanations (XAI), increasing user confidence.



Agents can act as decision copilots, preparing insights, simulations, and even suggesting actions within a decision-making flow.

As Agentic AI use cases multiply, a strategic question gains strength: how to transform these initiatives into scalable solutions integrated into the organization's daily routine? The answer lies in building intelligent digital products where agents are designed to operate within real workflows, interacting with systems, data, and people. This approach ensures that AI stops being just a "recommendation layer" and becomes an active part of the operation and customer experience.

INTEGRATION WITH BUSINESS WORKFLOWS VIA AGENTIC AI

The key to extracting value from these cases is to treat them as digital products composed of intelligent agents, each integrated into the company's core processes. This involves:

- Defining specific roles for agents in each journey (e.g., risk assessor, claims analyst, investment recommender).
- Designing interactions between agents, humans, and systems, with governance and supervision.
- Monitoring performance KPIs and continuous learning of agents as part of an agile product evolution cycle.



Thus, in the following table, we have examples of digital products that combine the previously presented use cases with a focus on the most specific challenges of four segments within financial services: Retail Banking, Investment Banking, Insurance, and Payments. This way, we can see more clearly how Agentic AI would be applied in building digital products that address the specific challenges of each business.

Key Categories for Value Generation

Categories	1. Process Automation	2. Risk Management Customer	3. Engagement and Personalization	AI-Based Decision Making and Insights
Retail	Virtual credit desk analyst capable of orchestrating and automating the entire end-to-end credit granting process.	Virtual operational risk manager that monitors transactions, investigates profiles, and responds to real-time events.	Digital financial concierge that understands segments and profiles, executes, and recommends products based on individual context.	Autonomous decision-support agent that collects data, generates analyses, explains trends, simulates scenarios, and suggests actions.
Investment Banking	Deal Book Automation Agent that automates data collection, analysis, and generation of materials to support deals.	Risk Intelligence Agent, capable of monitoring the market in real time, identifying critical exposures, and recommending actions.	Virtual Relationship Manager that knows the client, anticipates their needs, responds consultatively, and executes tasks.	Virtual Deal Hunter that analyzes trends, identifies opportunities, and suggests origination ideas with structured explanations.
Insurance	Claim advisor capable of receiving, analyzing, classifying, deciding, and communicating the progress of a claims process.	Risk Sentinel, capable of monitoring real-time risk signals, identifying anomalous patterns, generating alerts, simulating impacts, and suggesting mitigating actions.	24/7 digital concierge, capable of anticipating needs, mapping profiles, understanding lifestyles, and personalizing insurance recommendations.	Autonomous agent that uses historical data, market trends, and customer behavior to generate insights and optimize insurance and product portfolios.
Payments	Autonomous financial reconciliation agent that orchestrates, automates, analyzes, and acts on reconciliation and payment settlement processes.	Risk Guardian, acting as a digital sentinel, constantly monitoring transactions, detecting risk patterns, and acting in real time.	PayBot Concierge combining technical support, financial guidance, and customized offers based on transaction patterns.	Payments Strategist capable of analyzing transaction volumes, giving recommendations, simulating scenarios, and integrating with customers using natural language.

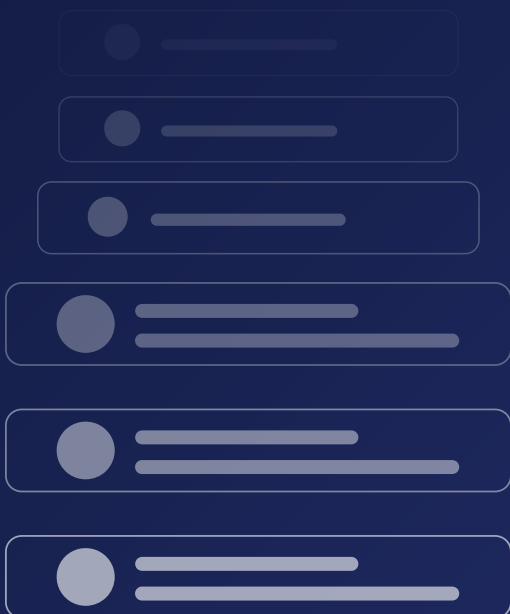


Agentic AI elevates the construction of digital products in the financial sector by enabling autonomous agents to perform complex tasks—such as monitoring risks, personalizing offers, or automating processes—always aligned with business objectives. Unlike traditional AI, which delivers analyses or predictions, Agentic AI transforms these models into engines of continuous action, expanding their impact and accelerating value generation in areas like credit, insurance, and investments.

In the coming years, intelligent agents are expected to work in increasingly collaborative and integrated ways, redefining how financial institutions orchestrate data, decisions, and experiences for customers and operations.

CHAPTER 4

Success Use Cases in the Financial Sector



50K

40K

30K

20K

10K

Jan

Feb

Mar

Apr

May



Success Use Cases: How AI is Redefining the Financial Services Industry

Among all the applications made possible by the evolution of GenAI and AgenticAI tools, Artefact has already developed practical implementations for several clients in the financial services industry at a global level. These use cases are not just projections or future trends - they are real applications, already implemented by major players in the sector. Leading institutions are using GenAI to transform their operations and customer interactions, achieving concrete results in efficiency, personalization and innovation.

OPTIMIZING CUSTOMER EXPERIENCE: A NEW PARADIGM IN CUSTOMER SERVICE

The digital transformation has imposed increasing challenges on finance companies when it comes to managing their interactions with customers. In a highly competitive market, the ability to respond to customer queries with speed, precision and in a language that suits each customer context has become not only a strategic differentiator, but a determining factor for customer loyalty. In this context, Artefact identified a significant opportunity: to build a chatbot that would replace the need for constant human interaction, guaranteeing accurate, secure and timely responses at a major European investment bank.



Challenges and obstacles

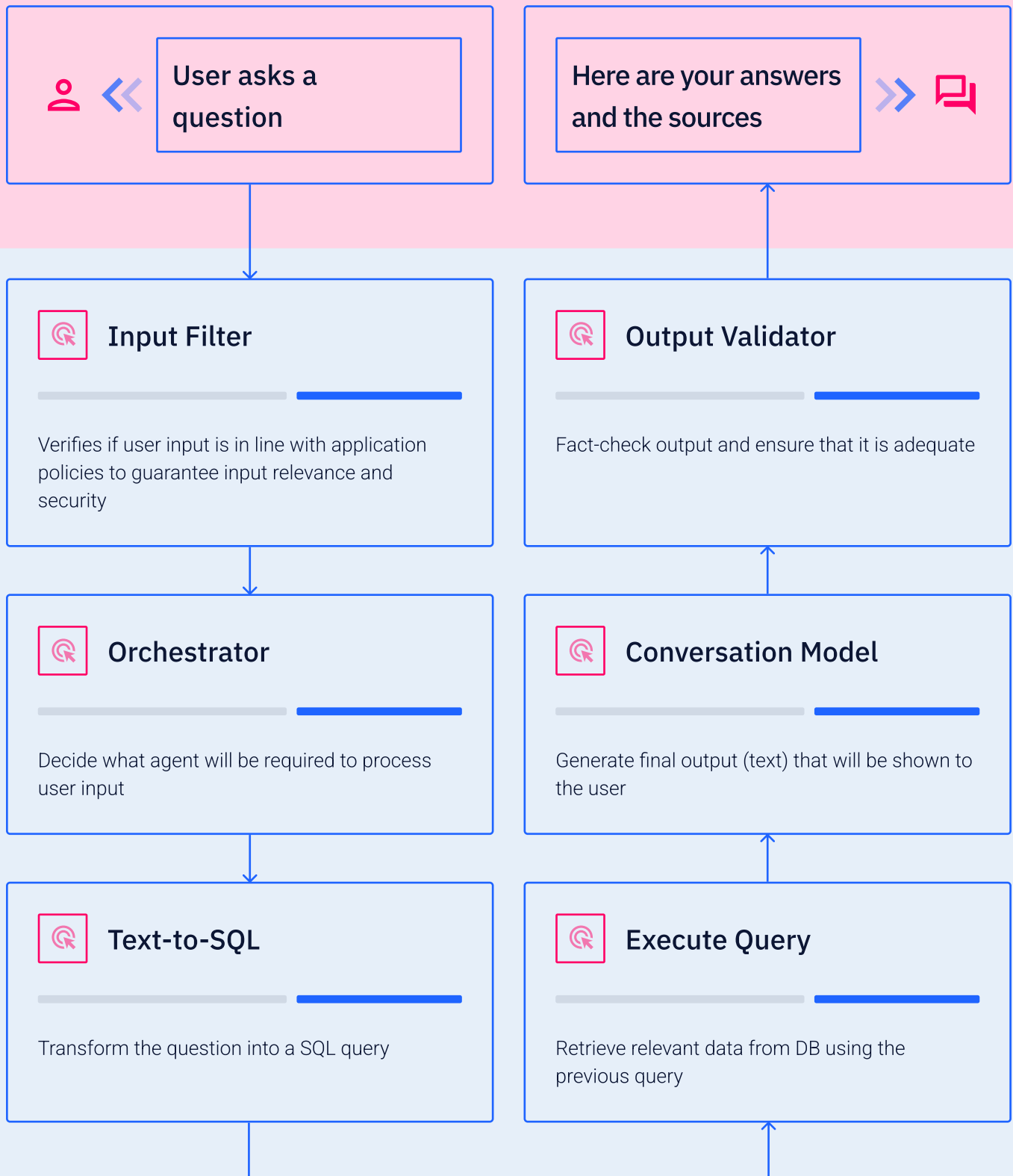
Although promising, the use case faces well-defined barriers in its implementation, regardless of the company's structure. The first of these is meeting users' growing expectation for accessible, fast and quality responses, a critical element for customer satisfaction. At the same time, the solution needs to comply with strict security standards, which include not only the protection of sensitive data, but also adherence to regulations specific to the financial sector.

Developing the solution

The proposed solution is based on a GenAI model designed to operate with high efficiency and security. The model is anchored in a knowledge base of Questions & Answers (Q&A), structured from the organization's relevant data. The solution also uses advanced filtering algorithms to interpret the user's question, extract the most pertinent information from the knowledge base and generate contextualized and accurate answers. In addition to the structures required for the model, there is the ability to adapt the proposed solution to any cloud (Azure, AWS and GCP) and on-premise environments, preserving data integrity and meeting security requirements.



USER INTERFACE



Adapted from the Artefact One Pager – Augmented Business Intelligence Agent applied to the pharmaceutical sector

Benefits achieved

As a result of implementing the conversation bot in different scenarios, our clients have achieved a 90% reduction in the average response time to users, promoting a more agile and satisfactory service. In financial terms, the project managed to generate savings of more than 5 million euros in the customer service team's operating costs, reaffirming GenAI's strategic value in optimizing processes and allocating resources more efficiently.

Highlights

The implementation of this solution highlights how GenAI technologies can redefine paradigms in the financial sector, combining operational efficiency and excellence in customer service. By overcoming technical and operational challenges, our clients not only modernized their interaction with consumers, but also set a standard for innovation that can serve as a benchmark for the entire industry.

What's more, by integrating the chatbot with internal system APIs - such as balance inquiries, process status or registration data - the solution already incorporates an operational agent within an Agentic AI ecosystem, capable of interacting autonomously with different systems to perform tasks and deliver even more personalized and resolute responses.

This case illustrates how GenAI acts as a technical enabler in solutions; enabling an interface that goes beyond simply generating answers and positions itself as a strategic driver of intelligent solutions. In a scenario where data, systems and interactions need to be orchestrated with intelligence, agility and security, initiatives like this demonstrate the potential of artificial intelligence to transform operations, improve the customer experience and generate tangible gains in the efficiency and performance of applications for the financial sector.

Compiled and adapted from functional demonstrations and MVP studies applied in digital and traditional banks, focused on chatbots, conversational APIs, and cognitive assistants.

Operational efficiency:

Optimizing middle and back office processes with GenAI

The financial sector, characterized by its operational complexity and high regulation, faces significant challenges in managing unstructured data. Bank transaction files, compliance documents and due diligence reports represent a massive amount of information that requires an enormous amount of time to analyze manually, and when not managed properly, can limit organizational efficiency in automated processes. To overcome this barrier, a new solution based on generative artificial intelligence (GenAI) was developed, focused on transforming the way data is processed and used in the middle and back office.

CHALLENGES AND LIMITATIONS

The project faced specific challenges when dealing with unstructured data, which can be divided into two main factors:

1

The propensity of generative AI models to hallucinate - answers that deviate from the reality of the data - requiring the development of rigorous mechanisms to guarantee the quality of the information extracted.

2

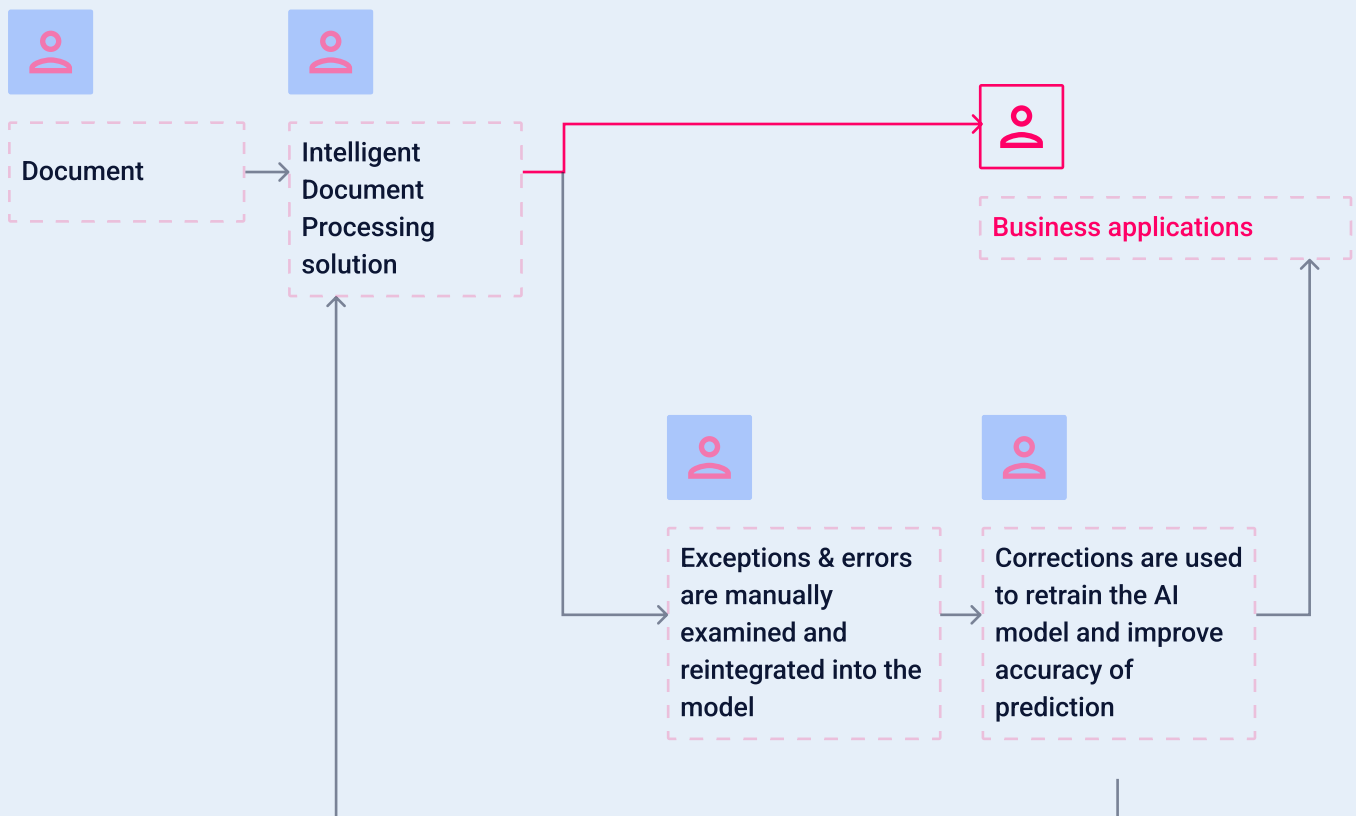
Alignment of the solution with regulations and compliance policies specific to the financial sector, especially in this field where extremely sensitive data is processed, guaranteeing security and compliance at all stages of processing.



Solution developed

The approach combined GenAI's ability to interpret unstructured data with advanced information processing and extraction techniques. The solution centered on a model trained to transform unstructured data into structured formats and direct it to treatments taking into account the specific objectives of each area.

The practical impact of this application was significant. Companies were able to obtain risk matrices based on compliance documents, consolidated summaries of bank transactions and detailed due diligence analyses. In addition, the solution accurately categorized customer complaints, providing automated recommendations for resolving problems in chats and calls, thus reducing the need for human interaction.



Content adapted from technical materials and executive summaries on the use of AI applied to intelligent document processing in the financial sector.

Benefits achieved

The results of the project include:

- Substantial increase in **productivity** due to ease of analysis and synthesis of large volumes of documents;
- **1/3 reduction in calls to the call center**, with automated handling of complaints and categorization of problems;
- Structuring of highly detailed risk matrices, improving **compliance management**;
- **90% accuracy in analyzing banking transactions**, raising the level of trust and operational efficiency.



These solutions have not only solved immediate challenges, but have also brought efficiency gains to diverse sectors such as open banking, private equity (PE) and insurance companies. The adaptability of the model has allowed it to be applied in different business contexts, setting a new standard for the use of GenAI in critical data management.

Reflections

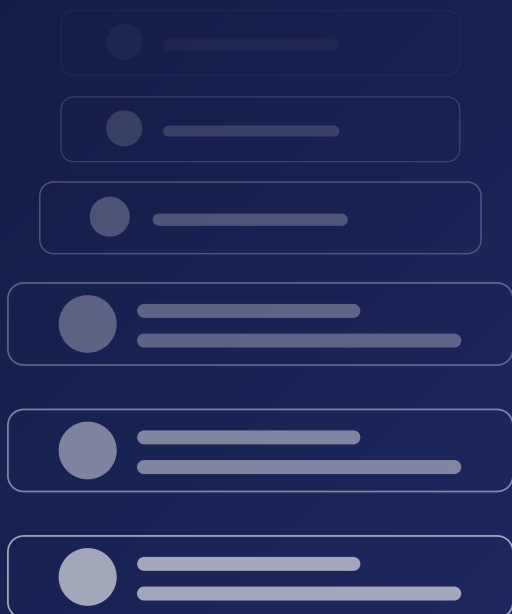
This project demonstrates how generative AI technologies can transform internal processes in the financial sector, offering faster, more accurate analysis in line with regulatory requirements. By intelligently automating previously manual and complex tasks, companies have been able to optimize their operations and redirect efforts to strategic areas, reaffirming GenAI's role as a driver of innovation and efficiency.

Considering the architecture already implemented, if the solution were to not only interpret incoming documents, but also autonomously search for new data in internal systems or external databases via APIs - such as automatic queries to update compliance reports or validate bank transaction information - it would already incorporate the characteristics and potential related to an agent. Thus, the model would evolve from a passive analysis solution to a dynamic agent, which not only processes data, but also acts continuously to enrich and validate its analysis in real time.

Adaptado a partir de iniciativas práticas e estudos de caso sobre aplicação de GenAI e NLP em automação de documentos, análise de risco e atendimento ao cliente no setor financeiro.

CHAPTER 5

Principais Desafios no Setor Financeiro e como superá-los



50K

40K

30K

20K

10K

Jan

Feb

Mar

Ari

May

Main challenges in the financial sector and how to overcome them

Before showing the main challenges in implementing GenAI and Agentic AI in the financial sector, it is important to highlight some convictions obtained from the implementation of real use cases and which are decisive in overcoming the main challenges and ensuring the success of the application of GenAI and Agentic AI in practice:

Convictions of Artefact on the use of GenAI and Agentic AI in the Financial Sector ^[3]

Successful adoption of GenAI and Agentic AI requires careful planning and overcoming significant challenges, such as integrating legacy systems and effectively managing data in a regulated environment. To guide you through the GenAI adoption process, Artefact has seven core beliefs for industries in the financial services sector:

Rapid proof-of-concepts with long-term strategic vision

Data quality as the foundation for success

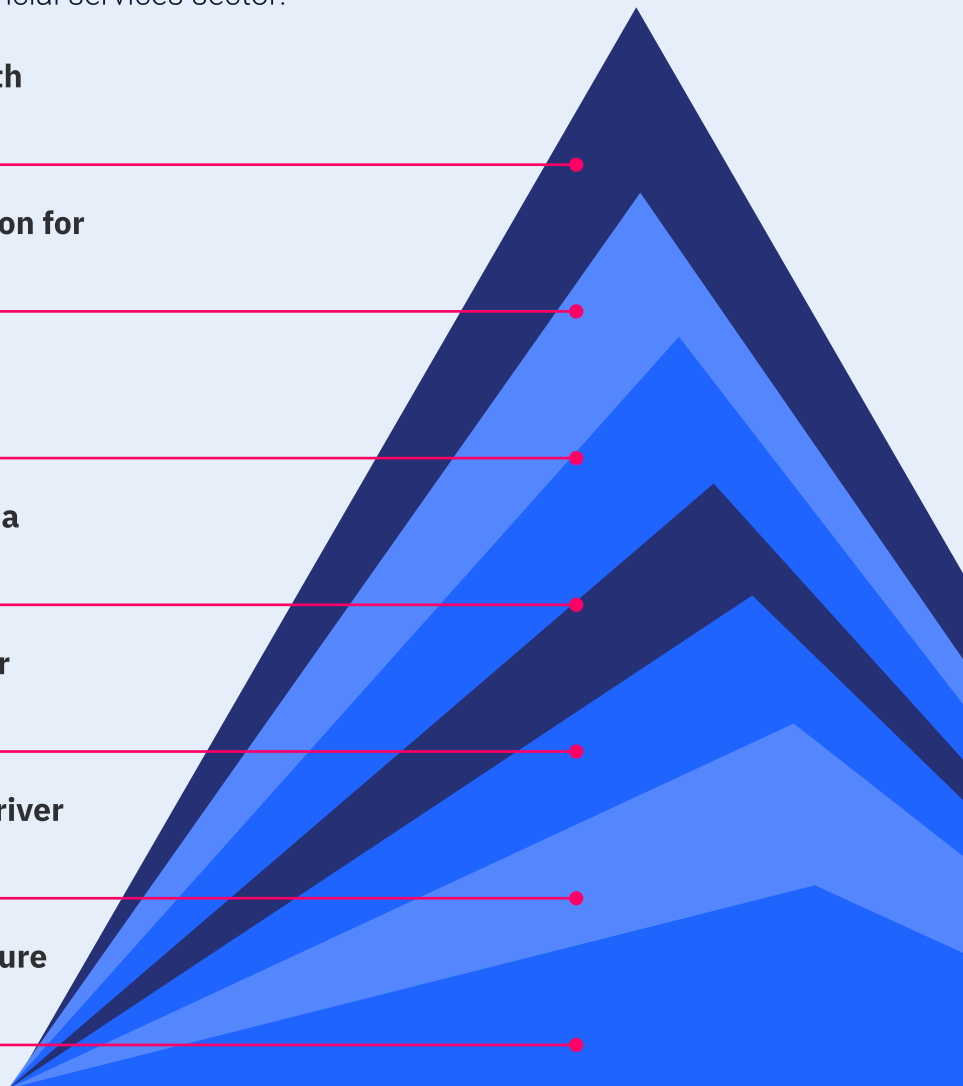
Intellectual property as a competitive differentiator

Compliance and security as a fundamental pillar

Orchestration framework for smooth integration

Continuous feedback as a driver of evolution

Change management to ensure organizational buy-in



Rapid proof-of-concepts with long-term strategic vision

The creation of rapid proof-of-concepts is essential to gain initial traction, but must be integrated with a long-term strategic vision. This alignment ensures that immediate execution is not limited to isolated initiatives and allows companies to capture sustainable value from GenAI over time.

Data quality as the foundation for success

The true value of GenAI is unlocked when the technology is fed with high-quality data, preferably first-hand. This data provides a solid foundation for generating reliable insights and accurate results, which are key to meeting the demands of a demanding market.

Intellectual property as a competitive differentiator

The creation of proprietary intellectual property is crucial to ensure future scalability. This includes the development of advanced prompt engineering and the systematic recording of outputs to build a knowledge base that can be used in the continuous refinement and improvement of the technology.

Compliance and security as a fundamental pillar

The financial services industry operates in a highly regulated environment, and the adoption of GenAI must ensure compliance with local and international regulations. In addition to managing risks related to data security and intellectual property, it is essential that GenAI solutions are designed to meet regulatory requirements while preserving stakeholder trust.

Orchestration framework for smooth integration

In addition to technical readiness, an orchestration framework is indispensable for integrating GenAI cohesively into existing workflows and systems. This framework ensures that GenAI solutions are applied effectively without disrupting established processes.

Continuous feedback as a driver of evolution

Continuous feedback mechanisms, with human participation in the process, are essential to guarantee constant improvements, alignment with business objectives and the creation of intellectual property. This cycle of improvement is a determining factor for long-term success.

Change management to ensure organizational buy-in

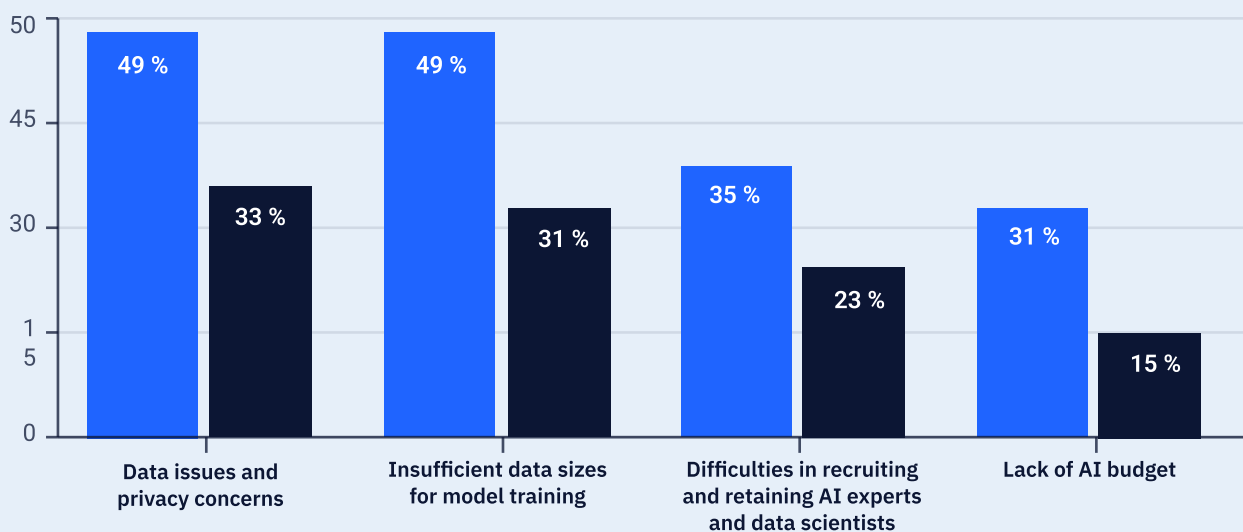
The implementation of GenAI requires more than technology; it demands a cultural and organizational transformation. Change management is crucial for involving stakeholders, training teams and ensuring that the adoption of the technology is effectively integrated into processes. Without a structured change management plan, the risk of resistance and low adherence increases significantly.

With these convictions, Artefact strengthens its ability to guide companies in the financial sector in a strategic and efficient adoption of GenAI, maximizing the positive impact of the technology while overcoming the challenges associated with its implementation.

Main Challenges for Implementing AI in the Financial Sector

The traditional challenges to implementing AI in the financial sector, such as data concerns, budget constraints and difficulty recruiting talent, have been on a downward trend.

According to NVIDIA's State of AI in Financial Services: 2025 Trends report, issues such as data privacy and the insufficient size of training bases, which were once significant barriers, saw a significant reduction from 2023 to 2024. This movement indicates a growing maturity in the market: organizations are moving from the stage of experimentation to the consolidation of large-scale projects, reflecting an environment that is more prepared to capture the strategic value of artificial intelligence. [2]



[2] Source: State of AI in Financial Services: 2025 Trends - NVIDIA

2023

2024

Despite the reduction of initial obstacles, in the financial sector attention to risk remains fundamental. The highly regulated and sensitive nature of the industry requires AI initiatives to be designed with additional rigor to ensure trust, compliance and market stability.

The following image reinforces this need, showing that the implementation of AI and GenAI goes far beyond technology - it involves the active management of strategic risks. Successfully navigating these challenges requires robust governance structures, transparent AI systems and a skilled workforce capable of effectively using and overseeing these technologies.

Safeguarding AI/GenAI value while addressing the risk imperative

In financial services, AI risk isn't just about technology – it's about trust, compliance, and market stability



AI Use Cases	Inherent Risk Type	Risk Description	Existing Control	Residual Risk Score
Fraud Detection	Bias and Fairness	Risk of producing biased fraud predictions due to imbalance training data	Implemented diverse and representative training sets with regular audits	LOW
Loan Approval	Transparency and Explainability	Lack of clarity in loan approval decisions could lead to customer dissatisfaction or regulatory scrutiny	Integrated explainability tools (ex. SHAP values) for decisions; human oversight	MEDIUM
GenAI Financial Report Summary Generator	Hallucination	Risk of generating inaccurate or misleading financial reports, summaries, and/or insights	Fact-checking workflows integrated into output review; fine-tuned model on verified financial data; human review for critical reports	HIGH

The framework presented organizes the main risks into four main categories: Regulatory and Compliance, Market and Credit, Reputation and Operational. Within these categories, specific risks such as model governance, algorithmic decision risk, data quality and systems integration are mapped. Each type of risk is exemplified with use cases: in fraud detection, the main challenge is to mitigate data bias; in credit approval, to guarantee transparency and explainability; and in financial reporting with GenAI, to control hallucinations - a risk that, even with mitigation measures, still has a high residual criticality.

This panorama reinforces that, even in a technologically mature environment, the success of AI in the financial sector depends directly on the ability of organizations to protect the value generated by managing risks in a structured and continuous manner.

In addition to the risks inherent in the models, the need to deal with sensitive information and meet the requirements of regulatory bodies makes the implementation of AI even more complex. Sectors such as fraud prevention and credit analysis require high levels of accuracy, transparency and compliance, since errors can have a significant financial and reputational impact.

Despite technological advances and the growing maturity of the market, the implementation of Generative AI (GenAI) in the financial sector still faces considerable challenges, especially in relation to scalability. Only around 10% of financial institutions manage to scale their AI initiatives successfully. Although tools such as ChatGPT or Copilot make it easier to create proofs of concept (POCs), the major obstacle lies in transforming these solutions into products that generate real impact. For GenAI to have a significant effect, it needs to be able to answer strategic business questions in a robust and scalable way, going beyond simple integration with existing processes. Without this structuring, measuring return on investment, reducing costs and increasing customer satisfaction becomes a complex challenge.



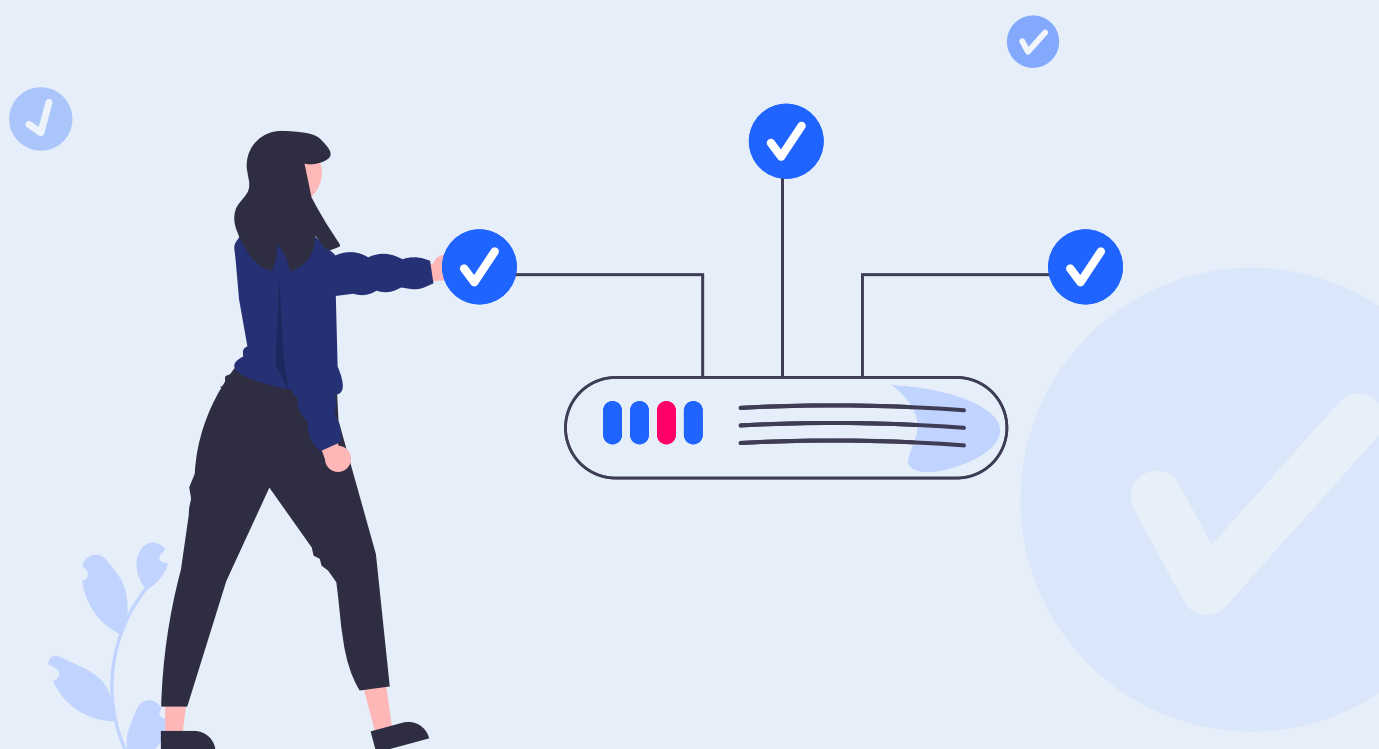
[7] Source: AI for Finance Event 2024 – Lecture by Joffrey Martinez (Artefact).

Best practices to overcome challenges and ensure successful implementation

To ensure a successful implementation of AI in the financial sector, institutions must adopt a strategic approach that tackles the main challenges in an integrated manner.

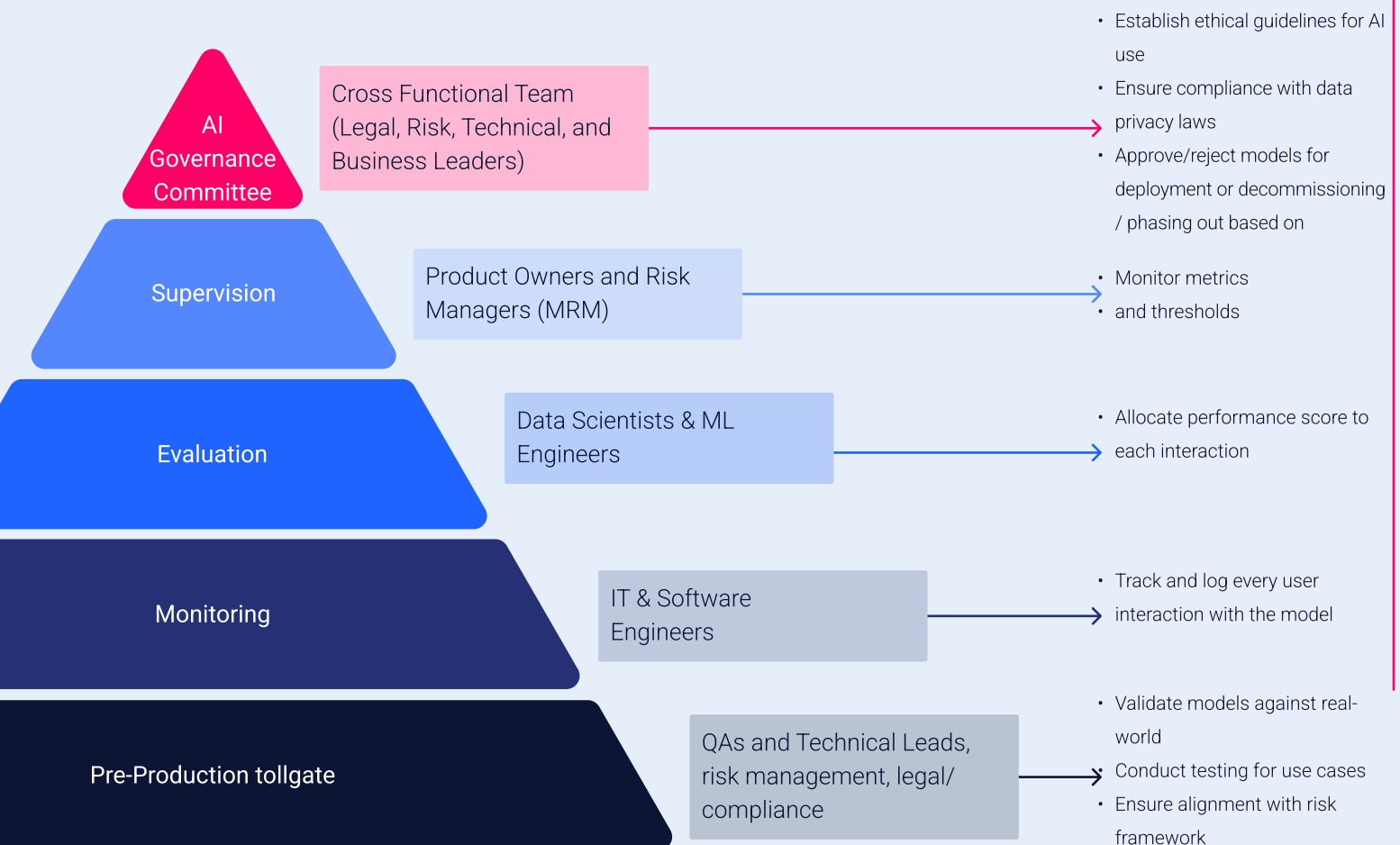
Building solid governance, with clear policies and adequate oversight mechanisms, is indispensable for managing risks and ensuring regulatory compliance.

In addition, investing in the development of talent specialized in AI can mitigate the shortage of qualified professionals, facilitating the effective adoption of this technology. Improving data quality is also crucial: robust data governance policies, improved collection practices and protection of privacy and security are determining factors for success. Finally, transparency in AI systems increases trust, ensures effective oversight and, together with well-defined ethical practices, reduces bias and promotes fair decisions.

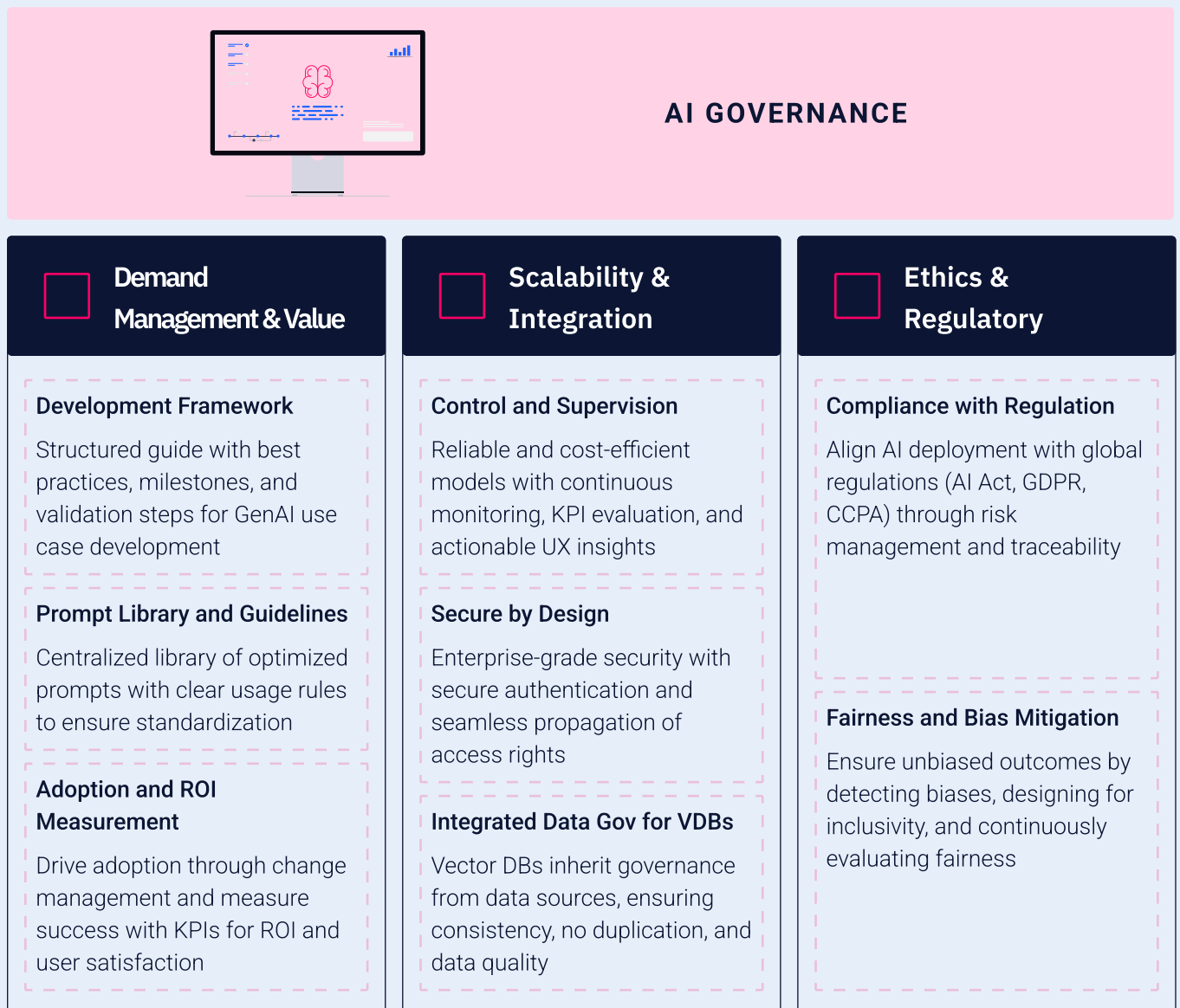


Structured risk management and the clear definition of roles and responsibilities remain essential pillars for the sustainable success of AI initiatives. The AI governance framework described in the chapter offers a practical, hierarchical view, showing how different levels of responsibility - from governance committees to technical teams - collaborate to ensure the safe and effective implementation of these technologies. This structure highlights the importance of aligning multidisciplinary skills, such as legal, risk and engineering, with well-defined processes, ranging from pre-production validations to ongoing supervision in production. This approach not only mitigates risks, but also promotes transparency and compliance, ensuring that AI initiatives are aligned with the financial sector's strategic and regulatory objectives.

Govern to grow: defining roles and responsibilities for sustainable AI success



Effective AI governance must be built on multifunctional pillars that integrate business value, technical scalability and ethical compliance. As illustrated in the image, three fundamental fronts underpin this approach: demand management and value generation, which includes clear frameworks for use case development, centralized libraries of optimized prompts and metrics to measure adoption and ROI; scalability and integration, with a focus on security by design, continuous oversight and integrated governance of vector data (VDBs) to ensure consistency and quality; and ethics and regulation, which ensure alignment with global legislation, such as GDPR and the AI Act, and promote bias mitigation practices for fairer and more inclusive decisions. This integrated framework is essential for scaling the use of GenAI with efficiency, security and adherence to corporate governance and regulatory standards.



Machine Learning: Improving Impartiality and Interpretability

To overcome the challenges of algorithmic bias and model interpretability in machine learning, financial institutions should prioritize algorithms with impartiality recognition and explainable AI (XAI) techniques [8] [9] . Implementing rigorous testing and validation procedures can help identify and mitigate biases in training data and model outputs [8] . Using techniques such as SHAP and LIME values to quantify the weight of features in model outputs can provide insights into the decision-making processes of machine learning models, improving interpretability and transparency [8] . In addition, engaging with stakeholders and regulators to establish clear guidelines for AI ethics and transparency is essential for the responsible deployment of AI. [8]

Generative AI: Mitigating Hallucinations and Ensuring Data Security

To meet the challenges of GenAI, it is necessary to focus on data security, privacy and accuracy. Implementing Augmented Retrieval Generation (RAG) can improve response accuracy and context by extracting information from internal company sources. Banks must implement robust data security measures and obtain explicit customer consent for the use of AI in order to adhere to strict data privacy regulations [8] . In addition, active engagement between banks and regulators is necessary to navigate the evolving regulatory landscape and mitigate potential inaccuracies in AI predictions by establishing transparent and effective frameworks [8] .

[8] Data from: The Alan Turing Institute – The AI Revolution, 2024.

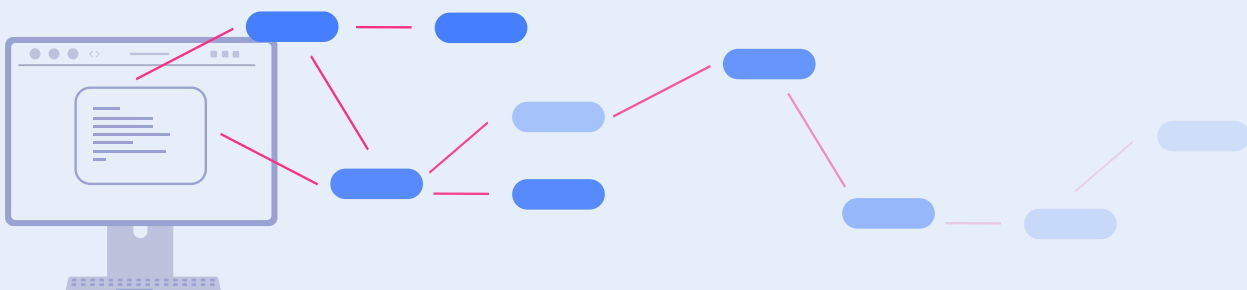
[9] Source: WhatNext.Law – Legal challenges of AI in finance.

Agentic AI: Strengthening Governance and Ethical Boundaries

To overcome the challenges of Agentic AI, governance, ethical boundaries and security infrastructure must be strengthened. Financial institutions must establish clear ethical guidelines for AI decision-making, incorporating human oversight and accountability mechanisms [8] [9]. Implementing robust security measures and so-called "guard rails" is crucial to prevent misuse, such as money laundering or insider trading [8]. Algorithmic bias in Agentic AI can be addressed through careful monitoring, mitigation strategies and transparent decision-making processes [8].

Finally, the scalability of AI as a whole involves ethical, environmental and social issues, as well as relying heavily on effective change management, especially given the organizational complexity of these institutions.

Promoting a culture of responsibility and ownership over data is essential, but often comes up against a lack of alignment between different areas and resistance to change. Thus, the successful implementation of AI requires a holistic approach that balances technological innovation, social and environmental responsibility and the fostering and enhancement of a data-driven organizational culture.



CHAPTER 6

Strategy in implementing AI in the Financial Sector



50K

40K

30K

20K

10K

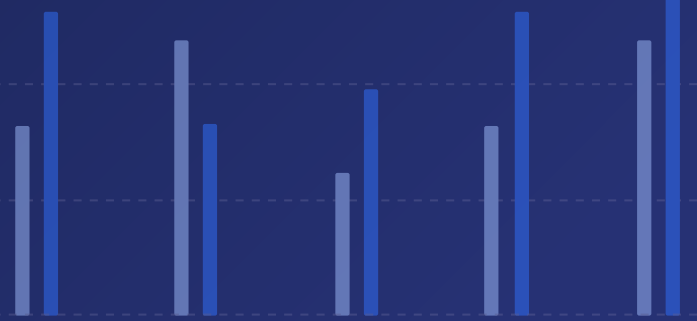
Jan

Feb

Mar

Apr

May

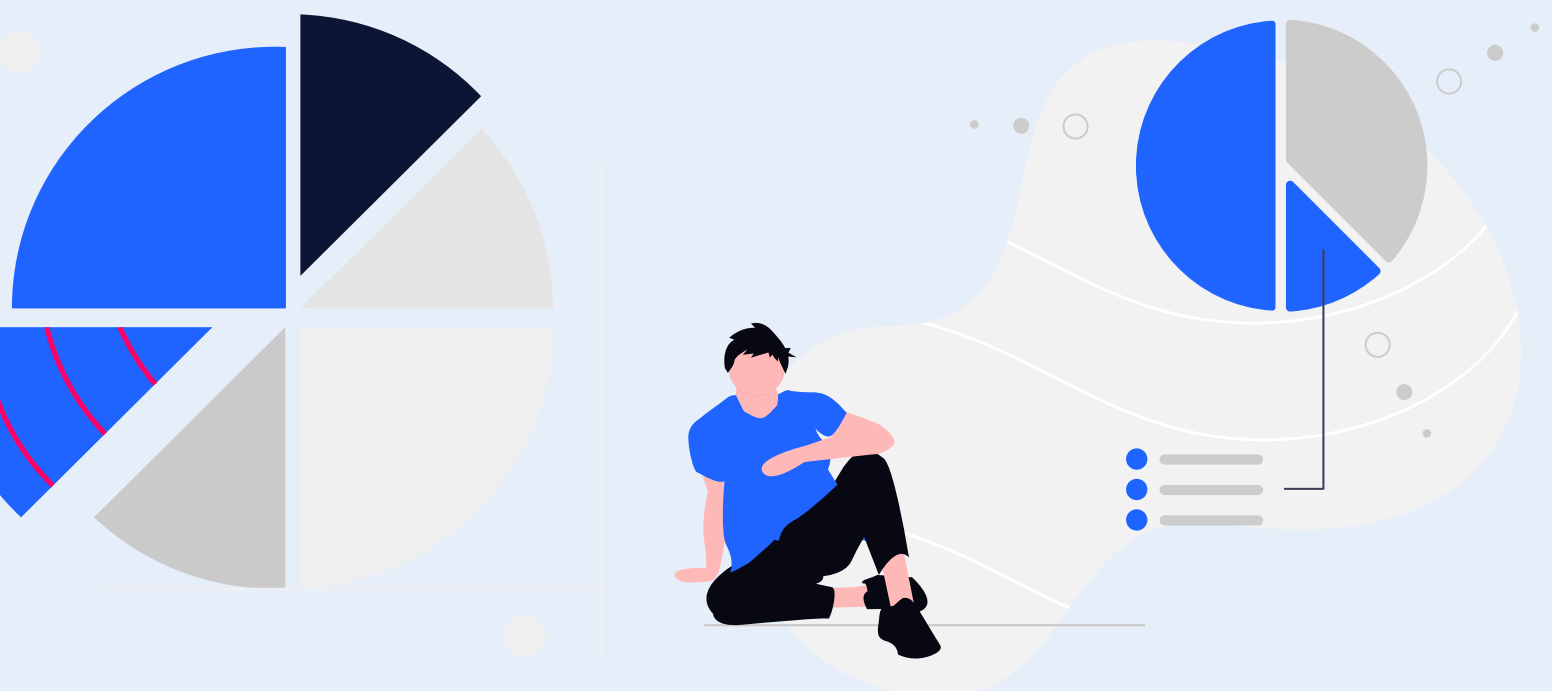


Strategy in implementing AI in the Financial Sector

Successful implementation of AI use cases begins with an accurate assessment of organizational maturity and the existing operating model. This diagnosis is key to identifying gaps and potential risks that could jeopardize the project. Based on this analysis, financial institutions can adopt the following strategies:

DIAGNOSIS AND STRATEGIC PLANNING

To ensure the successful adoption of any technology, it is essential to understand where the company is today in its digital maturity journey. Only with an accurate diagnosis is it possible to structure a solid plan that directs the organization to the desired level. This understanding allows the implementation of generative artificial intelligence to be done strategically, maximizing value and minimizing risks.



At Artefact, we assess companies' digital maturity through three fundamental dimensions:



Strategy

Data is valued as an essential asset to drive business strategy, ensuring that decision-making is guided by robust insights.



Delivery

The organization must be able to transform data into concrete results, applying analytical intelligence to optimize processes and generate positive impact.



Data Management and Governance

Proper data management ensures quality, compliance and security, indispensable factors in a highly regulated environment such as the financial sector.

These dimensions are assessed within 3 aspects:



People

Evaluates the leadership's commitment to a well-structured and funded data strategy, employee training for the strategic use of data and executive sponsorship for effective governance and management, ensuring organizational alignment and value generation for the business.



Processes

Assesses the systematic use of data assets to generate value, improve the customer proposition and optimize operations. In addition, it analyzes the integration of analytics and reliable insights into business processes, as well as the existence of structured governance, with clear definition, documentation, execution and monitoring of data management processes, guaranteeing their reliability.



Technology

Assesses the existence of a clear technical strategy and a roadmap that defines the capabilities needed to achieve the organization's data objectives. It also considers whether the technology architecture supports the delivery of priority use cases and is future-proof. It also assesses compliance, security, stability and data access control.

In two-dimensional form, we can exemplify the dimensions and aspects of maturity analyzed according to the image below:

	Strategy	Delivery	Data management and governance
	Data is valued as a strategic asset	Data drives delivery	Data mgmt & governance ensures compliance & drives data quality
People	Leadership are actively owning an ambitious and actionable data strategy, which has the right level of funding and organisational support.	Everyone in the organisation, business users and data specialists, has the right skills, knowledge to use data to deliver business outcomes.	Data management and governance have Executive sponsorship.
Process	Data assets are used systematically to create new value, enhance the customer proposition and drive operational improvement.	Business processes are powered by analytics and reliable insights.	Data management and governance processes are owned, defined, documented, delivered and tracked. Data is trustworthy and can be relied upon.
Technology	There is a clear technical strategy and roadmap that defines the capabilities required to deliver the organisation's data.	Technology stack enables delivery of priority use cases and is appropriately future-proof.	Data is compliant, secure, stable and access-controlled.

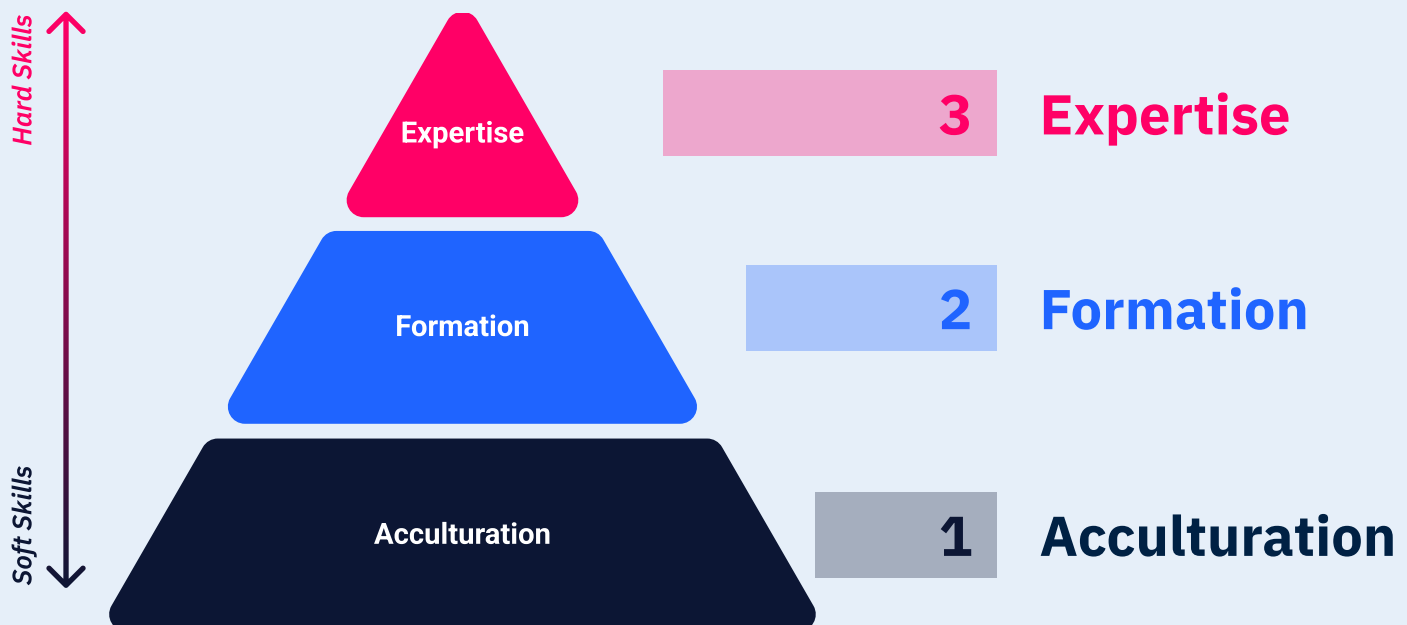
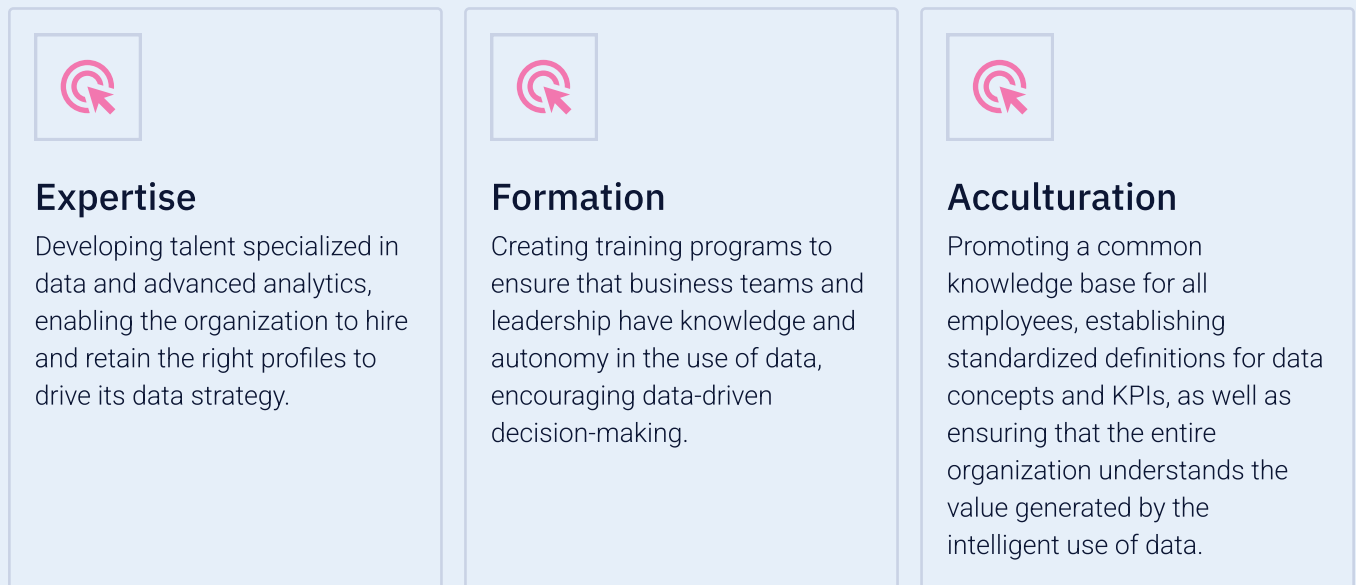
With this structured approach, financial institutions can not only implement AI efficiently, but also ensure that technology is aligned with business strategy, strengthening competitiveness and organizational resilience.

Stakeholder engagement

Stakeholder engagement is an essential element for the successful adoption of AI in the financial sector. Involving leaders from the business, technology, compliance and security areas from the start of the process makes it possible to align expectations and ensure ongoing support throughout the project. In addition, a collaborative approach that promotes engagement and shared responsibility between teams is key to integrating different perspectives and building a unified vision of strategic objectives.

Digital transformation and the adoption of AI go beyond technology: they require structured change management within the organization. This involves both the development of soft skills, such as cultural adaptation and a data-driven mindset, and hard skills, such as the technical training needed to operate the new artificial intelligence tools.

At Artefact, we accelerate this process of organizational change for our clients through our framework, which has three essential levers:



Pilots and Iterations: an agile approach to reduce risk and maximize value

When we talk about AI solutions, it is essential to recognize that there are inherent risks, such as hallucinations, model bias and interpretability challenges. To mitigate these challenges, the change management mentioned earlier must create an environment conducive to experimentation, allowing teams to test and validate new solutions in a controlled manner.

BUSINESS DISCOVERY

The first step in this journey is the Business Discovery phase. In this initial stage, it is essential to map possible use cases of the technology within the organization, trying to be as exhaustive as possible. This stage must be directly aligned with the company's strategic vision, ensuring that the adoption of AI has a clear purpose and generates real impact.



VALUE ESTIMATE

After mapping, it is necessary to estimate the value generated by each Use Case. This allows for prioritization that ensures that efforts are focused on the initiatives that will bring the greatest return in the shortest possible time, always seeking to add value in an agile and efficient manner. The right prioritization allows the first projects tested to serve as proof of concept for future initiatives, creating a sustainable cycle of innovation.

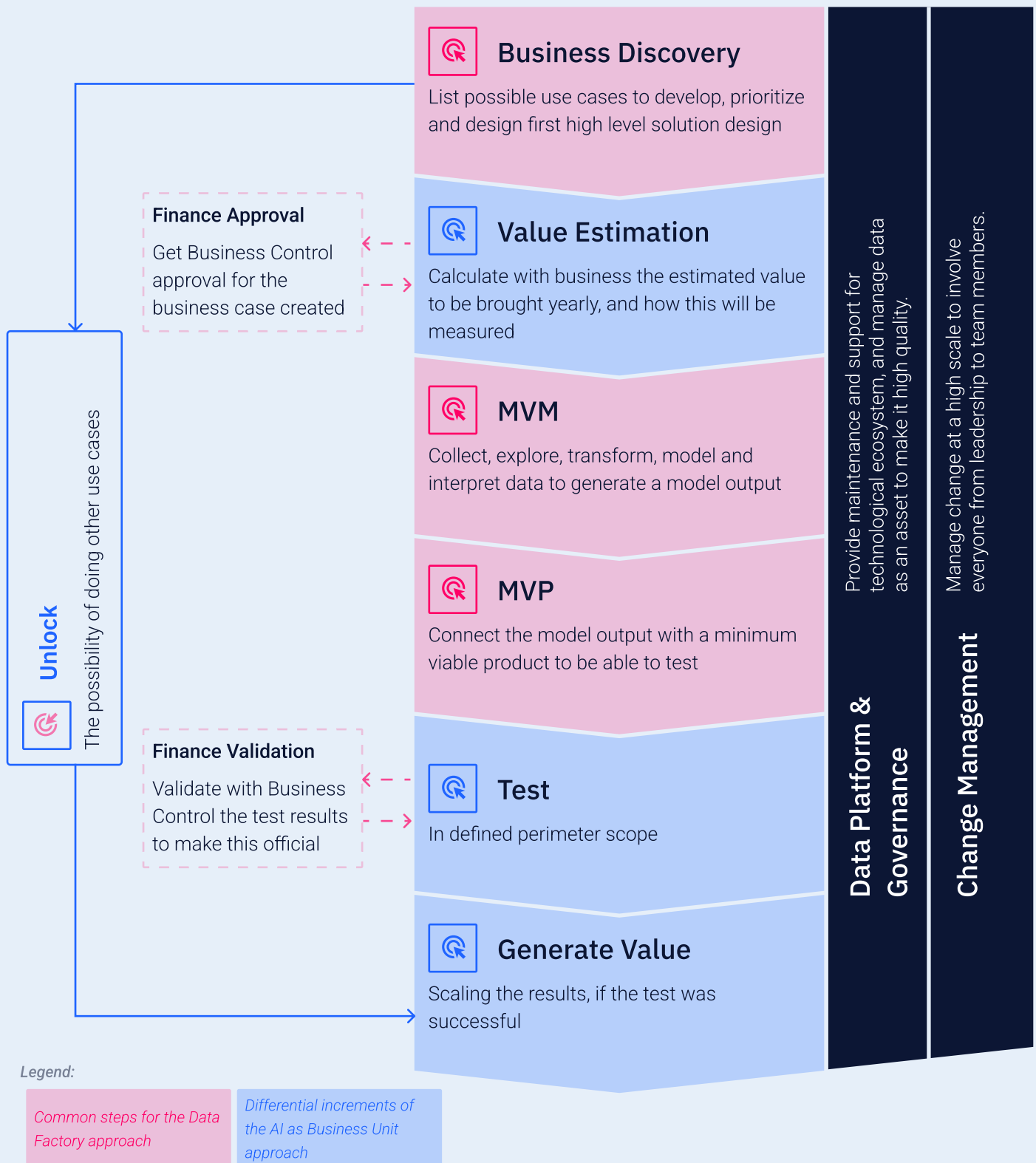


With the cases prioritized, we move on to the build and test phase, following the logic of MVM (Minimum Viable Model), MVP (Minimum Viable Product) and iterative experimentation. This is when the principle of "start small and make mistakes fast" comes into play. This methodology reduces costs and minimizes risks, as it allows for continuous adjustments before a more robust investment in the final solution. Through multiple iterations, the models are refined to ensure a well-defined scope, with less propensity for errors and greater reliability.



Once the test generates positive and conclusive results, it's time to scale the solution. At this stage, the technology is applied more widely within the organization, maximizing its impact and generating tangible value for the business. This learning and validation cycle creates a multiplier effect: the gain in efficiency and financial return obtained from the first projects allows new initiatives to be funded, further accelerating the company's digital transformation.

By combining controlled experimentation, continuous iteration and strategic scalability, financial institutions are able to exploit the potential of Gen AI in a safe, efficient and sustainable way, ensuring that innovation brings concrete, long-term results.



By following these steps, financial institutions can lay a solid foundation for the adoption and creation of use cases, creating an environment where innovation and security go hand in hand. In this way, it is possible to maximize the value generated by technology, promoting efficiency, compliance and tangible results for the business.

Data Security and Data Privacy as a Strategic Priority

In the financial sector, data security and privacy are non-negotiable factors. Banks and other institutions deal with highly sensitive information, such as financial histories, personal data and transactions, which requires stringent protection measures. With the advent of Artificial Intelligence (AI), including its variants such as Machine Learning and Generative AI, this need has become even more critical.

The choice of technological infrastructure to support AI solutions is a crucial strategic factor that directly impacts on the ability to implement and maintain robust security measures. In this context, financial institutions are faced with a fundamental choice: adopt cloud-based solutions or maintain an on-premise infrastructure.



When implementing AI solutions in the financial sector, institutions face a crucial decision between adopting cloud-based infrastructures or maintaining on-premise systems. Each approach offers distinct characteristics that can significantly impact the success and effectiveness of AI initiatives.

The following table presents an objective comparison between cloud-based and on-premise solutions, highlighting the main aspects, advantages and disadvantages of each approach. This analysis aims to help financial institutions choose the most suitable infrastructure to support their artificial intelligence initiatives, considering factors such as cost, scalability, security, regulatory compliance and technological dependence.



ASPECT	CLOUD SOLUTIONS	ON-PREMISE SOLUTIONS
Infrastructure	Managed by third-party providers.	Fully managed by the institution.
Scalability	High, with agile expansion on demand.	Depends on existing infrastructure.
Initial costs	Low, pay-per-use model.	High, requires significant investment in hardware and software.
Maintenance	Automatic updates by the provider.	Responsibility of the institution.
Security and privacy	Compliance risks and external storage.	Greater control over sensitive data.
Regulatory compliance	Can be complex in terms of data localization.	Facilitates compliance with specific regulations.
Technology	Access to cutting-edge frameworks and tools offered by leading providers.	Infrastructure adapted to the specific needs of the institution.
Dependency	High, linked to the reliability of external providers.	Low, operation independent of third parties.



In the following chapters, the details of each approach will be explored in greater depth. Specific characteristics, the most suitable use cases and best practices for implementing cloud and on-premise solutions will be discussed, providing a comprehensive overview to support strategic decision-making in the context of financial institutions.

Cloud Solutions

This refers to the use of computing services provided by third-party providers over the internet. In this model, the infrastructure, platforms and software are managed by the cloud service provider, allowing financial institutions to access AI resources on demand, without the need to maintain physical hardware on their premises.



✓ PROS:

Scalability and Flexibility: Agile expansion of processing capacity, adjusting to business demands.

Access to cutting-edge technology: Providers such as AWS, Google Cloud and Azure offer platforms optimized for Gen AI, with market-leading frameworks and tools.

Pay-per-use model: Reduces initial costs, allowing institutions to invest proportionally to usage.

Automatic Updates: Guarantees access to the latest versions of security and technology, without the need for internal intervention.

⊗ CONS:

Security and Privacy Concerns: Sensitive data stored outside the internal infrastructure can generate compliance challenges and security risks.

Dependence on External Providers: The operation is tied to the reliability and continuity of the services offered by third parties.

Regulatory Compliance: Some laws require data to remain in the country or be treated with specific controls, which can be complex to meet on certain cloud platforms.

On-Premise Solutions

This involves implementing and maintaining the entire AI infrastructure within the financial institution's own physical premises. In this scenario, the organization is responsible for all aspects of the infrastructure, including hardware, software, security and maintenance, offering total control over the systems and data.



✓ PROS:

Autonomy and Absolute Control: Operations remain entirely within the institution's infrastructure, ensuring greater protection of sensitive data.

Ease of Compliance: The location of data within the internal environment simplifies compliance with specific regulations.

Customization: Infrastructure adapted to the specific needs of the institution.

⊗ CONS:

High Initial Costs: Significant investments in hardware, software and specialized teams.

Limited Scalability: Expansions can be slow and costly, depending on the existing infrastructure.

Maintenance and Support: All updates and security measures depend on the organization's internal capacity.

Critical Decision Factors

We can summarize informed decision-making in a series of seven critical factors:

01

Regulatory Compliance: Evaluate legal restrictions on data location and processing.

02

Scalability: Consider future growth and the ability to adapt quickly.

03

Security: Analyze the controls needed to protect sensitive data.

04

Costs: Compare initial and long-term operating costs.

05

In-house expertise: Assess the IT team's ability to manage AI infrastructure.

06

Integration: Consider compatibility with existing systems.

07

Performance: Evaluate latency and performance requirements for critical applications.

Trends and Recommendations

Considering current trends in the financial sector and the benefits offered by cloud computing, many institutions are leaning towards cloud-based solutions or adopting a hybrid approach. The cloud offers significant advantages in terms of agility, scalability and access to cutting-edge technologies, which are crucial for staying competitive in the rapidly evolving AI landscape.

However, for institutions with exceptional security requirements, very specific regulations or a need for total control over the infrastructure, an on-premise or hybrid solution may be more appropriate. The hybrid approach, in particular, is gaining popularity as an intermediate step, allowing institutions to take advantage of the benefits of the cloud while keeping certain critical systems on-premise.

Special Cautions and Best Practices

The table below shows special precautions and good practices that should be observed when implementing artificial intelligence solutions in both cloud environments and on-premise infrastructures. These aspects are fundamental to guaranteeing security, compliance and operational efficiency, taking into account the particularities of each technological approach.

DIMENSION	CLOUD SOLUTIONS	ON-PREMISE SOLUTIONS
Data Protection	End-to-end encryption (data in transit and at rest)	Physical security of servers and infrastructure
Access management	Multi-factor authentication and principle of least privilege	Physical and logical access control policies
Auditing and Monitoring	Regular audits and real-time monitoring tools	Intrusion detection and internal monitoring systems
Resilience and Backup	Robust redundancy and disaster recovery strategies	Regularly tested business continuity plans
Updates and vulnerabilities	Dependence on SLAs and updates from the cloud provider	Frequent updates and vulnerability scans carried out in-house
Segmentation and Isolation	Logical segmentation via network and identity policies	Physical and logical network segmentation to protect critical systems
Regulatory Compliance	Contractual guarantee with providers to meet financial regulations	Compliance directly managed and audited by the internal team
Team training	Ongoing training in secure practices for cloud environments	Training for secure infrastructure administration and local incident response



For Cloud Solutions:

Data Encryption: Use end-to-end encryption for data in transit and at rest.

Access Management: Implement strict access controls, including multi-factor authentication and the principle of least privilege.

Regular Auditing: Carry out frequent security and compliance audits.

Continuous Monitoring: Use real-time monitoring tools to detect suspicious activity.

Redundancy and Backup: Implement robust backup and disaster recovery strategies.

Contract Compliance: Ensure that contracts with cloud providers meet the regulatory requirements of the financial sector.

Staff Training: Train staff in security practices specific to cloud environments.

For On-Premise Solutions:

Physical Security: Implement rigorous physical security measures to protect servers and infrastructure.

Regular Updates: Keep systems and software up to date with the latest security patches.

Network Segmentation: Use network segmentation to isolate critical systems.

Internal Monitoring: Implement intrusion detection and internal activity monitoring systems.

Continuity Plan: Develop and regularly test business continuity plans.

Vulnerability Management: Conduct regular vulnerability scans and promptly correct any problems identified.

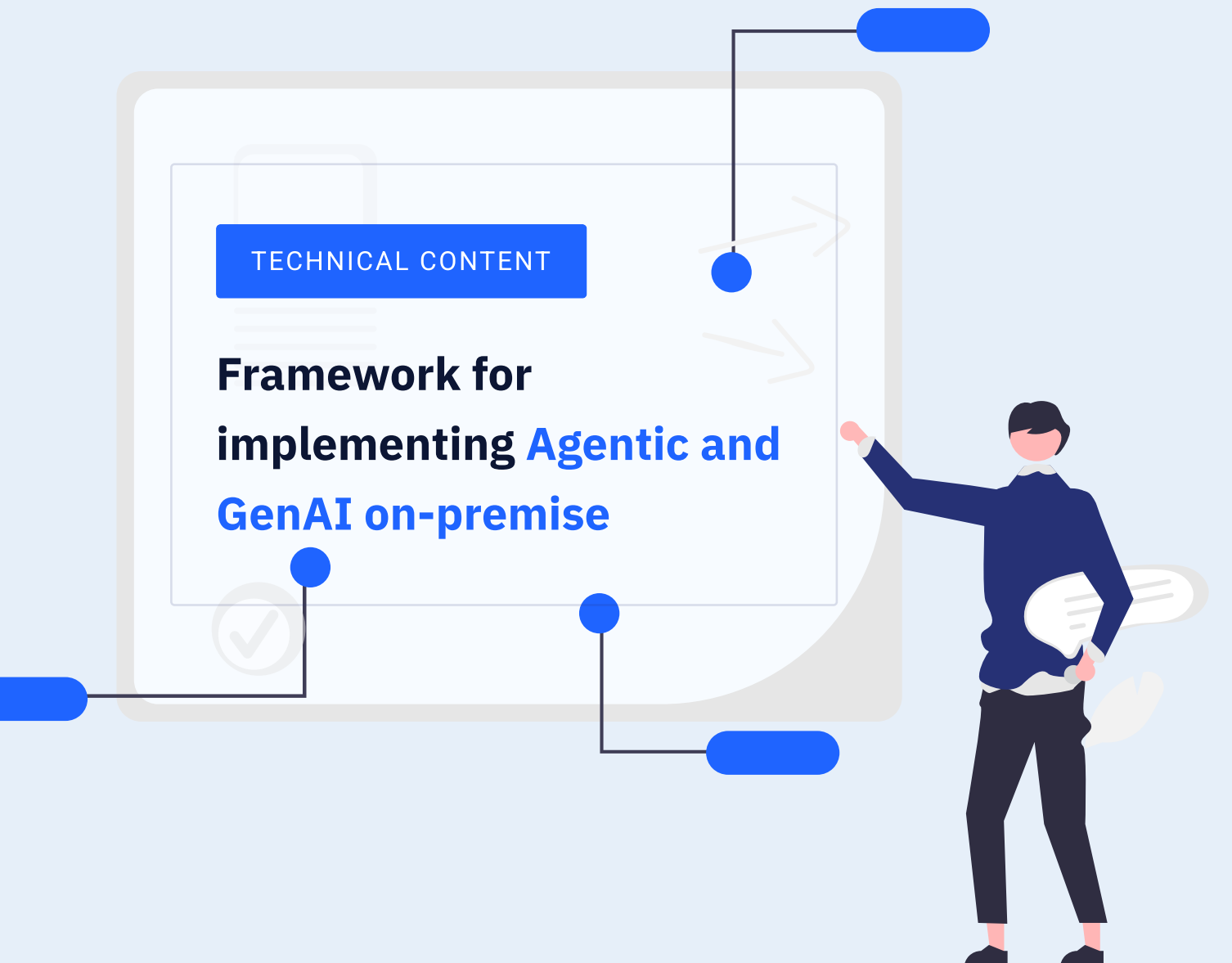
Physical and Logical Access Control: Implement strict access control policies for both physical and logical access to systems.

[SEE APPENDIX](#)

Technical Implementation
Framework (p. 89)

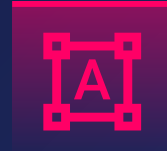
From Strategy to Architecture: Your AI Journey Starts Here!

Deepen your vision and prepare your organization for the practical and secure adoption of Agentic AI and GenAI in on-premise environments. We have created a complete technical framework, with detailed recommendations to transform intention into action with responsibility, scalability, and governance.



CHAPTER 7

Strategic Framework for Implementing Agentic AI by Business Area



50K

40K

30K

20K

10K

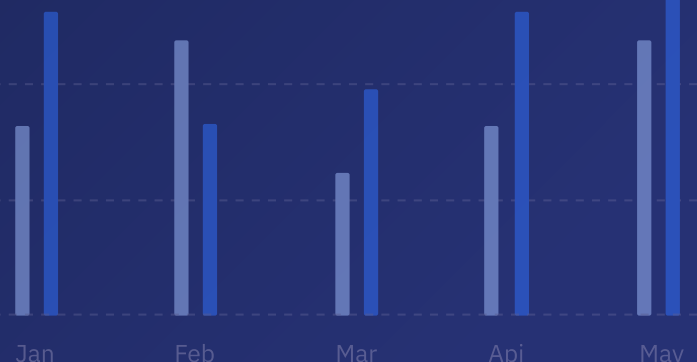
Jan

Feb

Mar

Apr

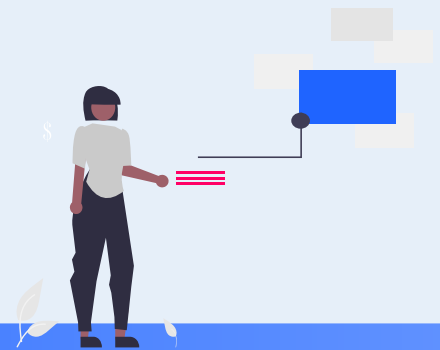
May



Introduction: From Vision to Strategic Action

The previous chapters demonstrated the transformative potential of Artificial Intelligence (AI), culminating in the emergence of Agentic AI - systems capable not only of analyzing and generating insights, but of acting autonomously to perform complex tasks and achieve defined objectives. While Generative AI (GenAI) revolutionized content creation and interaction, Agentic AI represents the next leap: cognitive automation and proactive execution.

However, unleashing the true potential of Agentic AI in the diverse financial landscape requires more than technological adoption; it demands a targeted and specific strategy for each business area. Whether in Retail Banking, Investment Banking, Insurance or Payments (as detailed in the Table on page 34), the challenges, opportunities and regulatory imperatives vary significantly.



This chapter provides a strategic framework to guide financial institutions in building an Agentic AI implementation roadmap tailored to their specific business units. The aim is to enable leaders to ask the right questions, consider the critical dimensions and structure an approach that maximizes the value and mitigates the risks inherent in this powerful technology, drawing on the concrete examples already possible.



The Strategic Imperative: Unlocking the Potential of Agentic AI Across Four Key Fronts

Agentic AI represents a fundamental shift, empowering systems not just to analyze, but to act autonomously and in coordination to achieve complex objectives. The strategic imperative for the financial sector lies in exploring how this ability to act radically transforms the four key value fronts, as categorized (and initially exemplified) in the Table on page 34. The potential goes far beyond the current examples:

01 Process Automation - Towards Intelligent and Adaptive Orchestration:

Agentic AI elevates automation from simple task execution to dynamic end-to-end workflow orchestration. The future horizon involves agents that not only follow processes (such as credit granting in Retail or claims management in Insurance), but that optimize them in real time, learn from exceptions and negotiate autonomously within safe parameters. Imagine agents managing complex interactions between institutions in Payments or automating the resolution of exceptions in trade cycles in Investment Banking, adapting the process dynamically to emerging conditions and information.

02 Risk Management - From Response to Prevention and Autonomous Adaptation:

The real strategic leap in risk management with Agentic AI is to move from real-time detection and response to predictive prevention and autonomous adaptation of defenses. In the future, agents in Retail will be able to anticipate default risks and initiate mitigation plans even before the problem manifests itself. In Insurance, agents will be able to model emerging risks (such as weather-related events) and trigger preventive responses. In Investment Banking, agents will be able to autonomously adjust limits and collateral based on complex predictive signals. In Payments, the ability to continuously learn and adapt defenses will enable the autonomous dismantling of increasingly sophisticated fraud networks.

03 Service, Engagement and Personalization - The Agent as Proactive Financial Navigator:

Proactive:

Agentic AI enables a level of truly proactive and holistic engagement, far beyond today's personalized recommendations. The future sees agents in Retail managing the customer's complete financial lifecycle, anticipating needs and optimizing products in an integrated way. In Insurance, agents will be able to use data (with consent) to proactively suggest coverage adjustments or preventive actions. In Investment Banking (Wealth Management), agents will be able to autonomously execute portfolio optimizations in line with the client's long-term objectives. In Payments, agents will be able to act as autonomous financial managers, proactively optimizing spending and savings.

04 AI-based Decision Making and Insights - From Analysis to Autonomous Strategic Initiative:

Autonomous:

The most advanced frontier of Agentic AI lies in autonomous decision-making in complex strategic domains. This includes agents in Investment Banking capable of developing, testing and executing new trading strategies. In Retail and Insurance, agents will be able to carry out dynamic price and underwriting optimization, responding autonomously to market conditions. In Payments, agents will be able to analyze vast data sets to identify and even start exploring new business models or strategic partnerships, turning insights directly into strategic action.

Therefore, when defining the Agentic AI strategy for your business area, it is crucial to look at current examples (such as those in the Table on page 34) for inspiration and validation, but to aim for the vast untapped potential. The question is not just "How can we automate what we do today?", but rather "What new capabilities for action, optimization and value creation can the autonomy of agents enable us to achieve tomorrow?". The institutions that answer this question in a visionary and strategic way will lead the transformation of the financial sector.

Key Dimensions of the Agentic AI Strategy

To build a robust roadmap, the Agentic AI implementation strategy for a specific business area must address the following interconnected dimensions:

01 Strategic Alignment and Value Proposition: Connect Agentic AI to the area's fundamental business objectives and clearly define the expected value.

02 Identification and Prioritization of Use Cases: Map processes and opportunities where agents' autonomy and ability to act will have the greatest impact, inspired by existing examples and aiming for future potential.

03 Technology and Data Readiness: Evaluate and plan the infrastructure, data, APIs and tools needed to support agent operations.

04 Governance, Risk and Compliance (GRC): Establish the ethical, regulatory and security boundaries for the autonomous operation of agents.

05 Human Capacity Building and Change Management: Prepare teams to collaborate, supervise and trust the new agent systems.

06 Performance Measurement and Continuous Improvement: Define metrics to track impact and establish feedback loops for the evolution of agents and strategy.



Essential Strategic Questions by Dimension

For each dimension, business leaders should seek clear and specific answers:

01 Strategic Alignment and Value Proposition:

What are the top 2–3 strategic objectives of our business area (e.g. reduce operating costs by X%, increase customer retention by Y%, launch Z new products) for the next 1-3 years?

How can Agentic AI's autonomy and ability to act (going beyond GenAI analytics) directly accelerate the achievement of these objectives, considering the four key fronts (Automation, Risk, Service, Decision)?

What is Agentic AI's unique value proposition for our customers and internal operations in this specific area? What problem does it solve better than current solutions, especially with "Beyond the Horizon" potential in mind?

How does Agentic AI fit into our broader digital and data strategy?

02 Identifying and Prioritizing Use Cases:

Which end-to-end processes in our area (involving multiple steps, systems and decisions) are best suited to be orchestrated by autonomous agents? (See Table on page 34 for starting examples and categories such as Process Automation, Risk Management, Service/Personalization and Decision Making/Insights).

Within our specific business area (Retail, Investment, Insurance or Payments), which of the "Beyond the Horizon" examples described in the section "The Strategic Imperative: Unlocking the Potential of Agentic AI on Four Key Fronts" represent the greatest transformational opportunities or solve the deepest pains?

Where would the ability of an agent to interact with interfaces (GUI), APIs, various databases, and perform actions based on these interactions and continuous learning bring the greatest efficiency gains, reduction of errors or creation of new value?

Which complex cognitive tasks, requiring judgment and adaptation, could be augmented or possibly delegated to specialized agents, freeing up human expertise for strategic oversight and innovation?

How to prioritize the identified use cases (current and future) based on potential impact (transformational value), implementation complexity, associated risks (including autonomy risks) and strategic alignment? (Use Risk x Impact x Complexity matrix - Chap 3, p. 26).

03 Technological and data readiness:

Do our core systems (CRM, ERP, trading platforms, etc.) have robust, accessible and secure APIs so that agents can interact and perform actions? Are there any gaps?

Do we have access to high-quality, up-to-date, relevant and sufficiently diverse data (structured and unstructured) to train agents capable of generalizing and operating effectively in complex scenarios? (Ref. Conviction 2, Chap 5)

What is our preferred architecture (Cloud, On-Premise, Hybrid - Chapter 6 - Data Security and Privacy as a Strategic Priority) for hosting and operating these agents, considering security, latency, data sovereignty and compliance?

What tools for agent orchestration (e.g. LangChain, CrewAI - Chap 2), monitoring, observability and control (for human intervention) do we need to implement?

04 Governance, Risk and Compliance (GRC):

How will we define the ethical and operational "guard rails" for autonomous agents in this area? Which actions are permitted, which require "human-no-loop" supervision, and which are strictly prohibited? (Ref. Chap 5 and Governance Framework p.53)

What specific regulatory requirements (e.g. KYC, AML, Basel, Solvency II, LGPD/GDPR, suitability rules) apply to the autonomous actions that agents will perform? How will we ensure continuous and auditable compliance?

How will we guarantee the auditability, explainability (XAI where applicable) and traceability of decisions and actions taken by autonomous agents?

What is the process for dealing with errors, biases, unexpected behavior or model drift in the agents? Who is responsible for intervention, retraining and decommissioning, if necessary?

05 Human Empowerment and Change Management:

What new skills will be required for our teams to effectively design, train, supervise, validate and collaborate with autonomous AI agents? (Ref. Change Management Framework p.60)

How will we communicate the introduction of Agentic AI and manage the impact on existing roles and responsibilities, emphasizing human-machine collaboration?

What training and acculturation programs (Ref. "Stakeholder Engagement" in Chapter 6) are needed to build trust, demystify the technology and ensure effective adoption by all stakeholders?

How will we foster a culture of controlled experimentation, continuous learning and shared responsibility in the operation of Agentic AI?

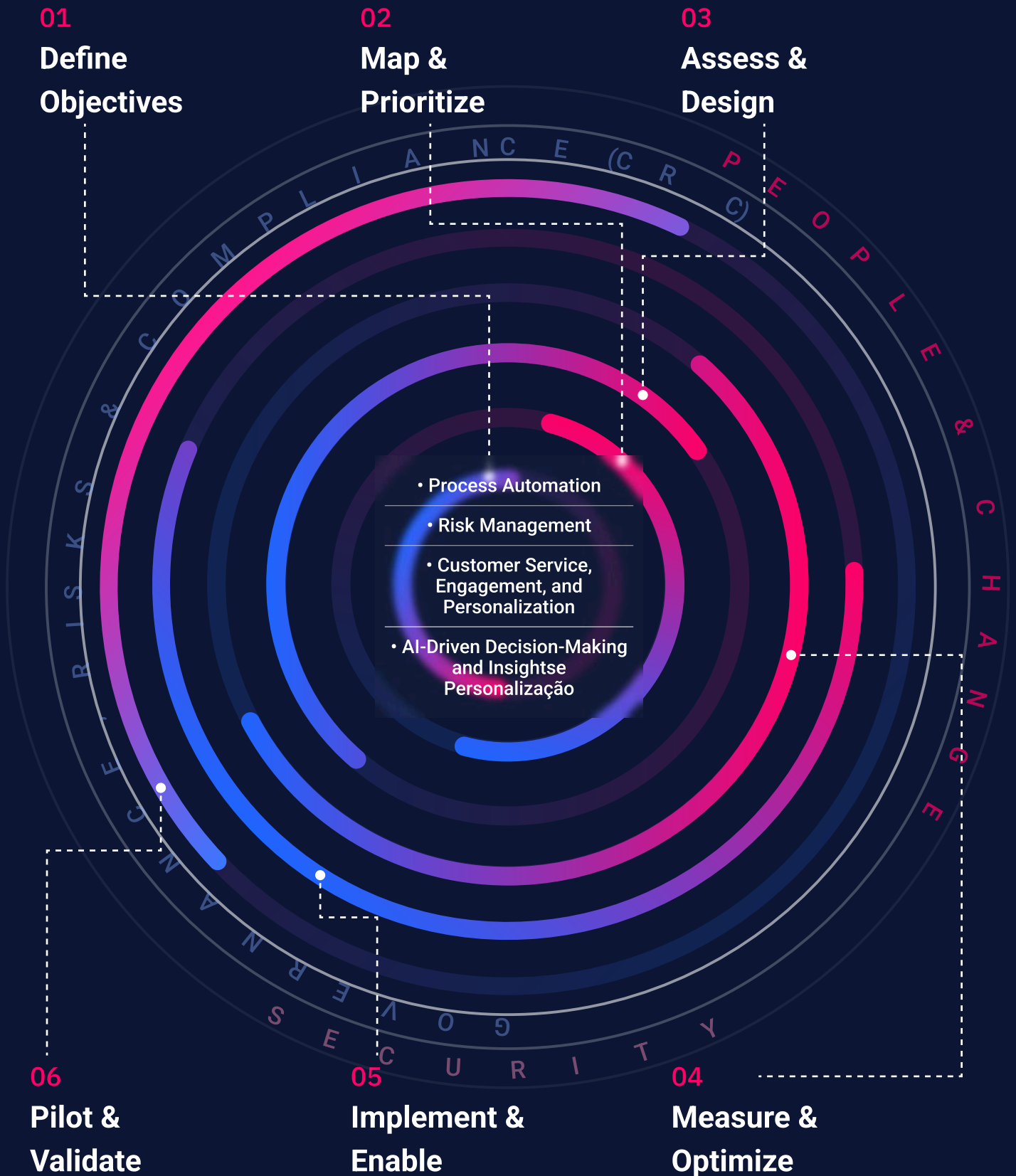
06 Measure performance and improve continuously:

What are the specific KPIs that will measure the success of the Agentic AI implementation against the defined business objectives? (Examples: end-to-end cycle time, cost per automated process, NPS, error rate in autonomous decisions, ROI, new revenue generated). (Ref. Conviction 6, Chap. 5).

How will we implement robust feedback mechanisms (human and automated) to continuously refine agent performance, behavior and safety? (Ref. Feedback Loops, Chap. 2 and 3).

How often will we review agent performance, the effectiveness of the overall strategy and alignment with the regulatory and market environment, adjusting the action plan as necessary? (Ref. Agile Approach, page 61).

Visual Framework: The Strategic Cycle for Agentic AI Implementation



Adapting the Framework by **Business Area** (with examples from Table p. 34)

Although the framework is general, the emphasis on each stage and the answers to the questions will vary, as exemplified by the basic ideas (Table on p. 34) and future potential ("The Strategic Imperative: Unlocking the Potential of Agentic AI on Four Key Fronts"):



Retail Banking

Intense focus on automating the customer and credit journey, personalizing interactions and proactive recommendations, and managing operational and default risks predictively. Critical CRM in consumer data (LGPD/GDPR) and explainability for credit.



Insurance

Key use cases in end-to-end automation and claims optimization (including negotiation), dynamic and predictive underwriting, proactive policy personalization and continuous data-driven risk/prevention management. GRC focused on policyholder data, underwriting fairness and fraud prevention.



Investment Banking

Priority on automation of complex trade processes, real-time market risk intelligence with autonomous action, consultative and active relationship management (including discretionary execution) and proactive opportunity identification/execution. Emphasis on rigorous GRC, cybersecurity and explainability (XAI) for high-value decisions.



Payments

Extreme focus on automation of complex reconciliation/settlement, autonomous detection/dismantling of sophisticated fraud, and proactive financial support/management for clients. Agent speed, accuracy, security and resilience are paramount, with GRC focused on transaction security and regulatory compliance (PSD2, etc.).

Conclusion: Navigating the Agentic Future with Strategy

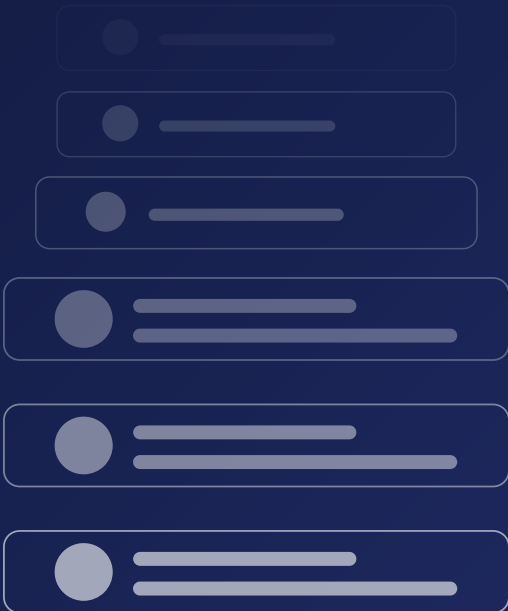
The successful implementation of Agentic AI will not happen by chance. It requires a deliberate, strategic approach tailored to the nuances of each business area within the financial institution, looking beyond current applications to the transformative potential of intelligent autonomy. By using this framework to ask the right questions, assess readiness, prioritize visionary opportunities and, crucially, establish robust governance and a culture of learning and accountability, organizations can navigate this new frontier with confidence.

Agentic AI offers the promise of radically transforming efficiency, customer experience, risk management and the very nature of financial decision-making. A well-defined strategy, anchored in future possibilities and adapted to your specific context, is the compass needed to turn this promise into a tangible and sustainable reality, positioning your business area and your institution at the forefront of financial innovation.



CHAPTER 8

Artefact as a strategic partner



50K

40K

30K

20K

10K

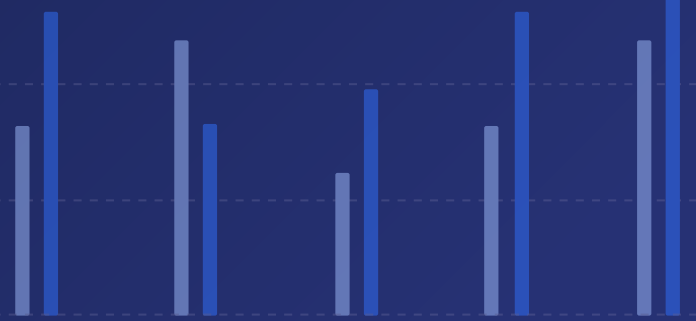
Jan

Feb

Mar

Apr

May



Why trust **Artefact**?



Proven Industry Experience

We have vast experience in the Financial sector, delivering practical solutions and measurable results for major leading companies. We operate at all stages of the value chain, with an emphasis on innovation, efficiency and sustainable growth.



Structured and Personalized Methodology

Our approach combines detailed analysis with tailored solutions, balancing strategic vision and practical execution. We continuously monitor results to ensure our clients' success.



Vision of AI as a Business Unit

We incorporate the vision of Data and Artificial Intelligence as value-generating assets, so that technology is at the service of the business. With a focus on solving business challenges, the applications must provide tangible and measurable results, whether through more assertive decision making, process optimization and automation or the generation of actionable insights in real time, positioning your company ahead of the competition.



Commitment to Concrete and Sustainable Results

Our goal is to deliver solid, far-reaching results, promoting continuous innovation and strategic efficiency to position your company as a benchmark in the market.



ARTEFACT AS A
STRATEGIC PARTNER



Businesses we have already impacted with GenAI and Agentic AI



Demand and Supply Forecasting

With Artefact's forecasting solution, we increased accuracy and reduced the time required from teams. In just three months, we achieved significant financial gains.



Stock-Out Prediction

With Artefact's stock-out prediction, we prevented risks, increased efficiency, and safeguarded revenue across the entire supply chain in just five days.

Some of our clients

AI/Data Acceleration for BCEF

Data strategy

Definition of the AI/DATA Factory operating model
Identification of +20 use cases (focus on cost saving)

AI/Data factory: Multiple use cases delivered / in progress

Optimization of corporate credit scoring
Fraud fishing optimization (+X M€ additional fraud blocked)
Self driving CRM (+X M€ PNB)
GenAI conversational agent (inheritance tax)

Data foundations

Building MLOps best practices
Providing expertise on AI/Data platform structuring

AI/Data Acceleration for RBF

Data strategy

Definition of data monetization strategy and identification of priority UCs (BAAS program)

AI/Data factory: Development of multiple use cases

Dashboards
Customer clustering
Upsell / cross sell

Data democratization

Large-scale training of professionals in data analytics on 2 concrete examples: production of dashboards and use of ESG data

AI/Data Acceleration

Acculturation: Training for the Group's 140,000 employees

- Implementation of digital content and microlearning covering 7 major data management themes
- Development of a handbook on data management
- Defining structure for the training content by adapting it to AXA's context
- Training launched at group level in 3 months

Data marketing program

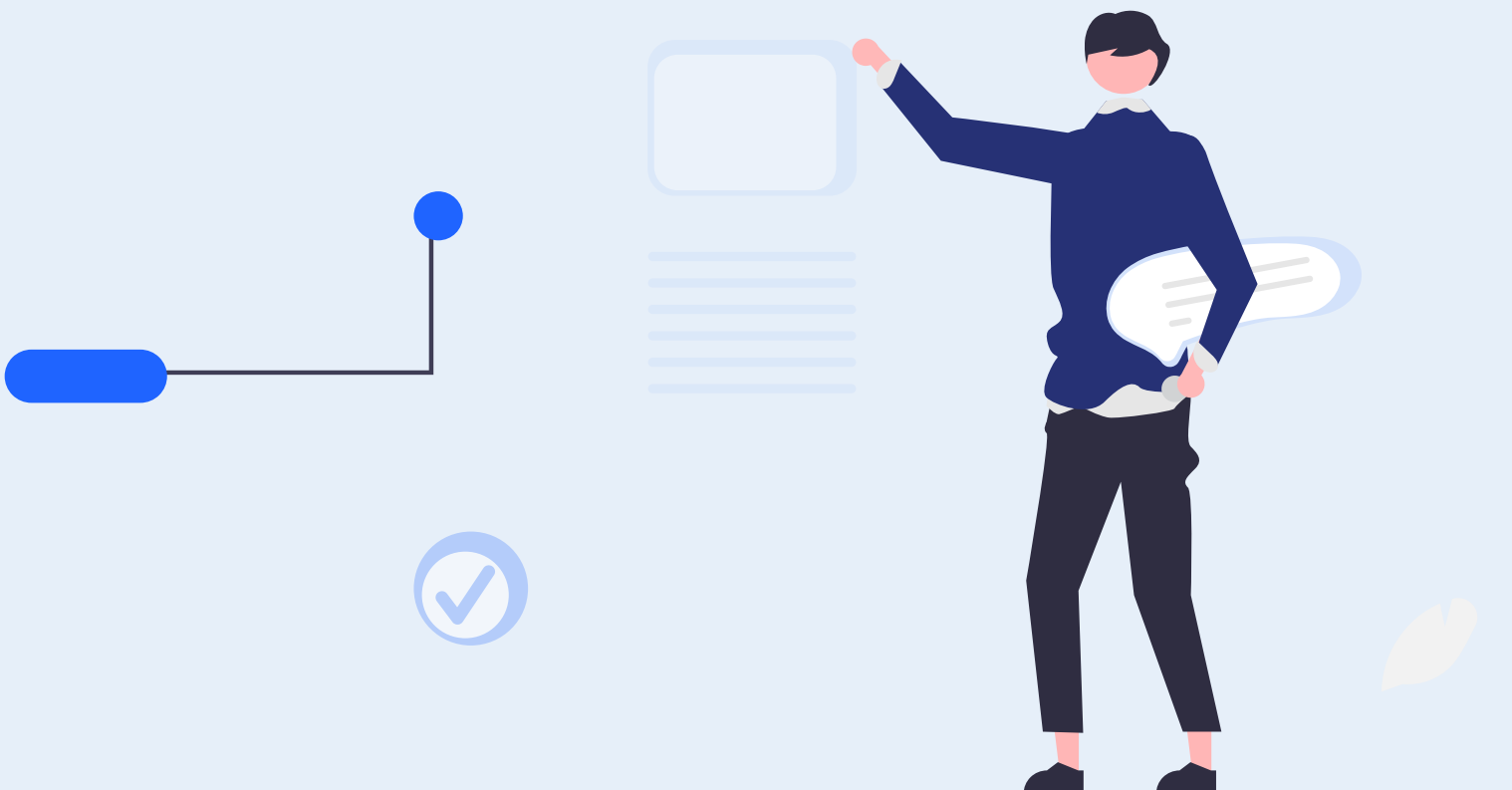
- Data-marketing acceleration to improve sales on the French network - tools/orga/measurement (in progress)

Appendix

TECHNICAL CONTENT

Framework for implementing **Agentic and GenAI** on-premise

In this exclusive appendix, we present the essential technical pillars to guide a safe and efficient journey toward adopting GenAI and Agentic AI in the financial sector. Recommended architectures, security practices, compliance measures, and scalability strategies — all tailored to the context of on-premise solutions.

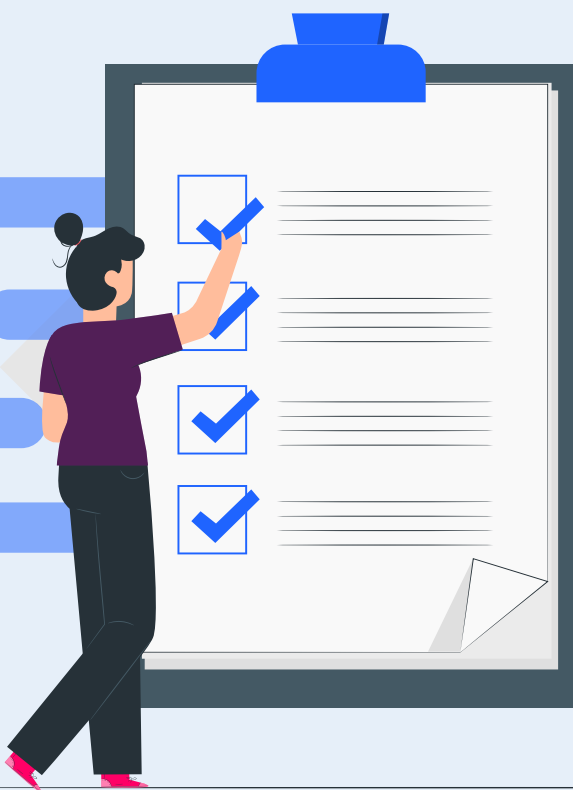


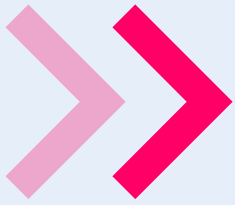
Framework for implementing **Agentic and GenAI** on-premise

A adoção de soluções de GenAI e Agentic AI em ambientes on-premise oferece vantagens estratégicas importantes, especialmente no que diz respeito à segurança da informação, conformidade regulatória e controle sobre os recursos computacionais. No entanto, essa abordagem também impõe desafios técnicos e operacionais que exigem um planejamento detalhado.

Desde a definição da infraestrutura de hardware até as camadas de segurança, rede, armazenamento e manutenção, cada componente da arquitetura deve ser cuidadosamente alinhado à realidade do negócio, à sensibilidade dos dados processados e à frequência de uso dos modelos. Isso é ainda mais crítico quando se trata de automações autônomas e fluxos contínuos de decisão, como os viabilizados por agentes inteligentes.

Com base em experiências práticas e conhecimento técnico especializado, este capítulo apresenta os principais aspectos a serem considerados para a implementação bem-sucedida de soluções GenAI e Agentic AI em ambientes on-premise.





Hardware requirements

In terms of computing power, smaller projects can be run effectively on servers with CPUs that have a high core count, achieving a good balance between cost and performance. However, as demand increases - either due to the volume of data or the complexity of the models used - it becomes advisable to use specialized GPUs (such as the NVIDIA A100 or H100) to drastically reduce training time. Two typical scenarios are therefore outlined:



Low demand: Use of high throughput CPUs, with the potential for gradual scaling as the need grows.



High demand: Infrastructure based on state-of-the-art GPUs, involving multi-server clusters and dynamic orchestration (e.g. Kubernetes).

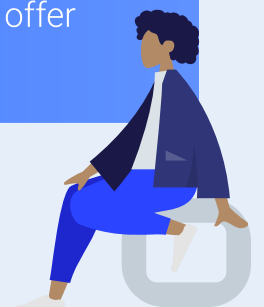
With regard to storage, initial projects are usually well served by less complex solutions, such as SSDs, local HDDs or even a small NAS (Network Attached Storage) for data sharing. However, when the volume of information and criticality increase, higher demands for speed and fault tolerance arise. In this scenario, distributed storage systems (e.g. Ceph or GlusterFS) and NVMe SSD drives come into play, providing high I/O performance and resilience. Again, two levels of demand can be highlighted:



Low demand: Local SSDs or a simple NAS, associated with basic backup policies.



High demand: Distributed storage with NVMe SSD, designed to offer redundancy, availability and scalability.



As far as the network is concerned, the disparity between lower demands and more complex operations is quite clear. For smaller environments, 1GbE or 10GbE connections usually do the trick, and for basic isolation, simple VLANs or VPNs may be enough. However, when you want high performance and low latency (for example, in HPC - High-Performance Computing - scenarios), it becomes essential to use connections of 25GbE or even Infiniband. In these situations, it is also necessary to invest in advanced network segmentation and robust firewalls to preserve security on a large scale:



Low demand: 1GbE or 10GbE networks and simple segmentation via VLAN/VPN.



High demand: $\geq 25\text{GbE}$ or Infiniband connections, with advanced segmentation and continuous traffic monitoring.

By structuring the on-premise AI architecture from this modular perspective - adequately differentiating smaller workloads from large ones - resource allocation becomes more precise. This avoids initial oversizing and also guarantees the possibility of expansion without the need to interrupt services. This lays a solid foundation for Generative AI and "Agent" AI projects that prioritize safety, performance and scalability, as well as creating a path for future adaptations.





Tech Stack

Defining a suitable Tech Stack is a fundamental step towards achieving maximum performance and the desired scalability in on-premise AI projects. The main recommended components are listed below:

Operating System

Linux distributions such as Ubuntu or CentOS are suitable for AI workloads, as they combine **stability, broad community support and ease of optimizing advanced hardware** (GPUs and multi-core CPUs).

AI frameworks

When choosing an AI framework, align your decision with the needs of the project, team expertise and scalability. For deep learning, TensorFlow offers scalability, PyTorch flexibility and Keras ease of use. For traditional machine learning, Scikit-Learn is intuitive, while XGBoost and LightGBM are ideal for structured data. For computer vision, Caffe stands out. Frameworks such as LangChain, OpenAI and Hugging Face are recommended for LLMs and GenAI. Also consider community support, integration with your stack and long-term maintenance.

Containerization, Orchestration and Packaging

The use of containers such as Docker facilitates the creation of isolated environments for each model or service, simplifying both development and deployment. In parallel, Kubernetes is the standard choice for orchestration, due to its ability to manage horizontal scalability and provide fault tolerance mechanisms - critical factors in scenarios with multiple AI services running simultaneously.

To maximize the use of hardware resources and the selected software stack, certain configurations are especially important:



Kubernetes Cluster with GPU Support: Adopting NVIDIA's Kubernetes device plugin enables containerized applications to make use of GPU resources on demand, optimizing training and reducing inference time.



Resource monitoring: Tools such as Prometheus and Grafana allow real-time monitoring of CPU, GPU, memory and network metrics, making it possible to identify performance bottlenecks and guide continuous improvements.

In addition, containerization ensures portability and isolation of the environment, simplifying both the development and maintenance of different model versions. Tools such as Docker and inference platforms (e.g. FastAPI or Flask) help make up the implementation ecosystem. They also make it possible to package the model, since when the Docker image is created, all the necessary dependencies are included, from AI libraries (PyTorch, TensorFlow, Transformers, etc.) to the model in its optimized state.

Dockerfile example: Below is an illustrative example of a Dockerfile that installs the basic dependencies, copies the model to the container and exposes the inference service:

Unset

- `# Dockerfile Example`
- `FROM nvidia/cuda:11.8-base`
- `RUN pip install torch transformers fastapi uvicorn`
- `COPY model/ /app/model/`
- `COPY app/ /app`
- `WORKDIR /app`
- `CMD ["uvicorn", "main:app", "--host", "0.0.0.0", "--port", "8000"]`

This example uses the NVIDIA image with CUDA 11.8 as a base, installs the necessary libraries, copies the model and application files into the container and starts an Uvicorn server to expose the inference API.

Model selection: To maintain full control over the deployment and ensure adaptability, it is recommended to use open source models such as Llama 3 or Mixtral. These can be fine-tuned with proprietary data, should it be necessary to adapt behavior or performance to specific scenarios.



Open Source Models: Avoid over-dependence on suppliers and increase transparency regarding internal workings.



Fine-tuning: When necessary, apply fine-tuning techniques to corporate data, preserving the model's competitiveness and relevance to the business domain in question.

Model optimization: Once the model has been selected, it can be optimized to reduce inference latency and the consumption of computing resources. Among the usual strategies are quantization and parallelism:



Quantization: Converting models to lower precision formats, such as FP16 or INT8, reduces memory usage and improves inference speed, especially on the latest generation GPUs.



Parallelism: In very large models, the use of model or tensor parallelism can distribute processing across several nodes or GPUs, reducing the total execution time.

Implementation: After creating the container, the next step is to integrate the model into the Kubernetes environment. This can be done via YAML files or Helm Charts. In both cases, essential aspects must be observed:

Scalability Management: The Horizontal Pod Autoscaler (HPA) automatically adjusts the number of pod replicas as the workload varies. It monitors metrics such as CPU or GPU usage and creates or removes pods to maintain performance.

Monitoring and Logs: Link deployment to monitoring solutions (such as Prometheus and Grafana) to track CPU, GPU, memory and latency usage metrics in real time. This helps identify bottlenecks and make scalability and optimization decisions.

Security and Compliance

Data security and regulatory compliance are cornerstones in on-premise AI projects, especially when handling sensitive information or under strict regulations. Below are the main practices for protecting data integrity, adapting the implementation to the relevant legislation and strengthening network security.

Data Protection

To keep corporate data in a secure and controlled environment, a multi-layered approach is recommended:



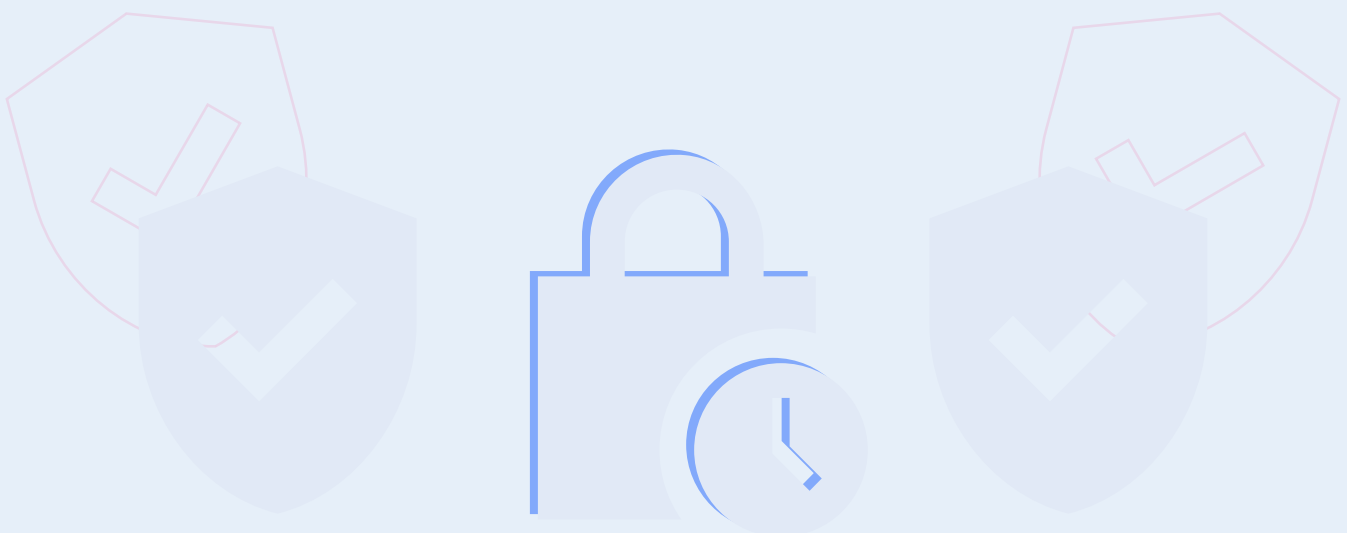
Encrypted Storage: It is crucial to implement encryption at rest and in transit, keeping data unreadable even in situations of improper access.



Private Networks: Restrict cluster communication to internal private networks, preventing the exposure of critical ports and services to the external environment.



Role-Based Access Control (RBAC): Configuring Kubernetes with RBAC ensures that each role within the organization receives specific permissions, preventing unauthorized users from gaining access to sensitive resources.



Compliance

Depending on the sector, legal and regulatory requirements can vary considerably. Each AI project must align itself with the corresponding standards, such as:



GDPR: Aimed at protecting the personal data of European Union citizens, it requires transparency in data processing.



HIPAA: Applicable to healthcare in the US, it sets standards for protecting and maintaining the privacy of medical information.



ISO 27001: International certification aimed at information security management, including control, monitoring and continuous improvement policies.



LGPD: Brazilian law that regulates the processing of personal data, guaranteeing the rights of data subjects and imposing responsibilities on organizations regarding the collection, use and storage of data.

These benchmarks guide the creation of robust internal policies, guaranteeing data confidentiality, integrity and availability.

Network security

The network infrastructure in on-premise environments needs to be configured to contain external threats and minimize internal risks:

Virtual Private Cloud

(VPC): Use private IP addresses and segment critical subnets so that only trusted services access the AI cluster.

Firewalls and Intrusion Detection Systems (IDS):

Define firewall rules to filter sensitive ports and protocols and apply IDS in order to block intrusion attempts or detect abnormal traffic.

Continuous Monitoring:

Integrate tools such as Prometheus and Grafana to observe security events in real time and generate alerts in the event of suspicious activity.

Scalability and Performance Optimization

Agentic AI or Gen AI projects can expand rapidly in terms of data volume, number of users and model complexity. To keep up with this growth without compromising quality, it is essential to apply scalability and performance optimization practices. Here are three key recommendations:

Load Balancing

Load balancing distributes requests evenly among the available resources, avoiding overload. Tools such as NGINX or Traefik can be configured to route traffic efficiently in a Kubernetes cluster, increasing reliability and response speed.



NGINX or Traefik: Adjust the routing rules so that pods with more processing capacity receive more requests proportionally.



Fault tolerance: If a node goes down, traffic is rerouted without affecting the service.

Caching

Caching mechanisms are extremely useful when certain inferences or model results are accessed frequently. Tools such as Redis make it possible to store the most frequently requested answers in memory, considerably reducing response times.



Reduced Latency: By serving answers directly from memory, there is no need to reprocess the model with each request.



Reduced Load: Caching reduces the pressure on AI servers, freeing up resources for other inferences or more complex tasks.

Parallelism

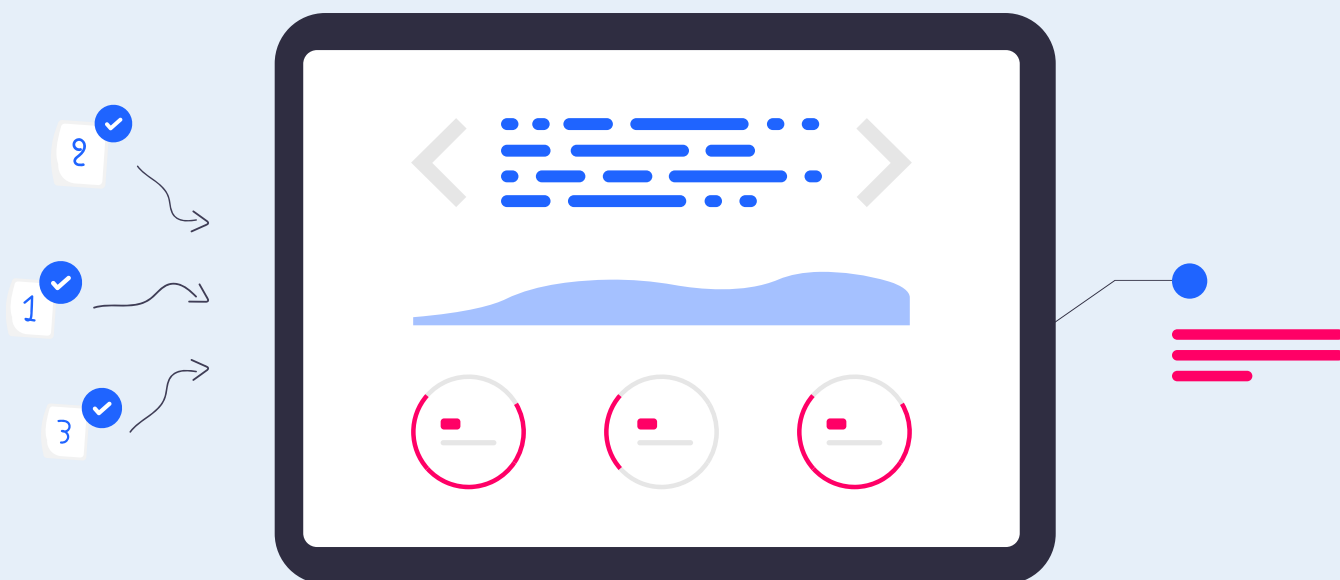
To deal with large volumes of data or highly complex models, parallelism makes it possible to divide up the workload, speeding up execution. Frameworks such as Ray or Dask simplify the distribution of training or inference tasks across several nodes, increasing efficiency and reducing processing time.



Distributed Computing Frameworks: Ray and Dask offer accessible APIs for parallelizing operations without the need to deal directly with communication between nodes.



Horizontal Scaling: As the number of nodes increases, performance tends to grow linearly, making it possible to run progressively larger workloads.



Monitoring and Maintenance

To maintain the efficiency and availability of Generative AI or "Agent" AI solutions in an on-premise environment, it is essential to establish solid monitoring and maintenance procedures. These practices allow performance bottlenecks to be detected, logs to be monitored and models to be updated, ensuring that the solutions continue to meet business needs.

Monitoring Tools

A comprehensive monitoring architecture facilitates the identification and agile resolution of problems. Prometheus and Grafana are often used together to collect metrics (CPU, memory, GPU usage, latency, etc.) and display dashboards that portray the overall health of the system.

In parallel, a log stack, such as ELK (Elasticsearch, Logstash, Kibana), unifies log management at a single point. This simplifies the diagnosis of incidents, making it possible to quickly identify abnormal behavior or failures, as well as keeping a detailed history for auditing or future investigations.



Prometheus + Grafana: Real-time monitoring and visualization of the cluster's most critical metrics.



ELK Stack: Centralization and analysis of logs, making it possible to correct problems and check history more efficiently.



Model retraining and updates

As input data changes and new insights emerge, keeping the model up to date is crucial to preserving the accuracy and relevance of the AI solution. Integrating CI/CD (Continuous Integration and Continuous Delivery) pipelines into the retraining process makes it possible to automate steps such as loading data, verifying the performance of the new model and, eventually, deploying it in production.



Tools such as ArgoCD or Jenkins help control model versions, releasing incremental updates and performing rollbacks when inconsistencies or significant performance losses occur. This cycle of continuous improvement speeds up development, minimizes risks and ensures that the model in production remains appropriate to the demands of the business.



CI/CD pipelines: Allow frequent retraining with updated data and automated validation tests.



Deployment automation: Solutions such as ArgoCD or Jenkins facilitate the delivery of revised models, ensuring traceability and version control.



Overview of the Agentic AI / Gen AI On Premise Implementation Flow

STAGE	TOPIC	SUBTOPIC	POSSIBLE TOOLS/ FRAMEWORKS USED
Hardware	Computing Power	Low Demand	High throughput CPUs
		High demand	State-of-the-art GPUs (Example: Kubernetes)
	Storage	Low Demand	Local SSDs or a simple NAS
		High demand	NVMe SSD
	Networking	Low Demand	1GbE or 10GbE networks (VLAN/ VPN segmentation)
		High demand	Connections ≥25GbE or Infiniband
Tech Stack	Operating System	---	Linux (Ubuntu or CentOS)
	Frameworks	AI	PyTorch, TensorFlow or Keras
		Gen AI	LangChain, OpenAI and Hugging Face
		ML	Scikit-Learn, XGBoost or LightGBM
	Containerization, Orchestration and Packaging	---	Docker, Kubernetes
	Model selection	---	Llama 3 or Mixtral
	Model optimization	---	Quantization and/or Parallelism
	Implementation	---	YAML files or Helm Charts
Security and Compliance	Data Protection	---	RBAC
	Compliance	---	---
	Network Security	---	VPC
Scalability and Performance Optimization	Load Balancing	---	NGINX or Traefik
	Cache	---	Redis
	Parallelism	---	Ray or Dask
Monitoring and Maintenance	Monitoring Tools	---	Prometheus + Grafana or ELK Stack
	Retraining and Model Updates	---	ArgoCD or Jenkins

Glossary

A - E

Advanced Personalization

The use of AI to tailor products, services, and experiences based on individual customer needs and preferences.

Compliance

A set of rules and regulations that companies must follow to ensure legal and regulatory adherence, especially in the financial sector.

Agentic AI

A type of artificial intelligence that adds autonomy and decision-making capabilities to systems, allowing them to act proactively, learn from experience, and perform complex tasks with minimal human intervention.

Customer Onboarding

The process of integrating new customers into a financial institution, ensuring regulatory compliance and access to personalized products and services.

API (Application Programming Interface)

A set of rules and tools that enables applications to integrate with AI systems, facilitating communication and data exchange.

Data Security

Practices and technologies used to protect sensitive information from unauthorized access, leaks, or cyberattacks.

Artificial Intelligence (AI)

A field of computer science focused on developing systems capable of performing tasks that typically require human intelligence, such as learning, decision-making, and pattern recognition.

Digital Transformation

The process of integrating digital technologies across all areas of a company to improve operations, efficiency, and customer experience.

Chatbot

An AI-based program that interacts with users in natural language to answer questions, provide information, or perform automated actions.

Encryption

A method of securing data through encoding, ensuring privacy and protection in digital transactions.

Fraud Detection

The use of artificial intelligence and data analysis to identify suspicious activities and prevent financial crimes such as money laundering and banking fraud.

GenAI (Generative Artificial Intelligence)

A subfield of AI focused on creating new content—such as text, images, code, and audio—based on models trained on large volumes of data.

GUI (Graphical User Interface)

A visual interface that facilitates interaction with agentic AI through graphical elements such as buttons, menus, and panels for intuitive navigation.

Data Governance

A set of processes and policies that ensure the quality, security, and regulatory compliance of data use within an organization.

LLMs (Large Language Models)

Advanced neural network-based language models trained on massive datasets to understand and generate natural language, enabling applications like chatbots and virtual assistants.

Language Models

AI algorithms trained to understand, process, and generate text based on a vast corpus of natural language data.

Machine Learning

An AI technique that allows systems to learn from data and improve their performance without explicit programming.

Predictive Analytics

A data analysis technique that uses statistics and algorithms to forecast future trends and patterns based on historical data.

Process Automation

The use of technology to perform repetitive and operational tasks efficiently, reducing the need for human intervention.

RAG (Retrieval-Augmented Generation)

A technique used in generative AI systems where the model combines text generation with the retrieval of relevant information from a database or external documents to increase the accuracy and relevance of responses.

RPA (Robotic Process Automation)

A technology that uses software to automate repetitive business processes, increasing operational efficiency.

XAI (Explainable AI)

Explainable Artificial Intelligence is a branch of AI focused on bringing transparency to the outcomes of intelligent systems, enabling justification and auditability of processes with embedded AI.

Unstructured Data

Information that does not follow a predefined format, such as text, images, audio, and video, making it more complex to process.

Links & References

[1] Article

The brief history of artificial intelligence: the world has changed fast – what might be next? Our World in Data. Available at: <https://ourworldindata.org/brief-history-of-ai>

[2] Research

NVIDIA. *State of AI in Financial Services: 2025 Trends*.

[3] Institutional Material

Artefact. Internal knowledge.

[4] Research

Artefact. Generative AI Survey – The Technology, the Rewards & the Risks. Internal document ([Artefact-GenAI-Survey.pdf](#)).

[5] Article

Citi GPS. Agentic AI: Finance & the 'Do It For Me' Economy. Jan. 2025.

[6] Article

GODHANI, Sahaj. Agentic AI will transform financial services with autonomy, efficiency, and inclusion. Medium – InsiderFinance Wire. Available at: <https://medium.com/insiderfinance-wire/agentic-ai-will-transform-financial-services>

[7] Video | Even

MARTINEZ, Joffrey. AI in Financial Services: Key Market Trends and Insights for 2024. AI For Finance Event 2024. Powered by Artefact.

[8] Report

The Alan Turing Institute. The AI Revolution – Opportunities and Challenges for the Finance Sector. Nov. 2024. Available at: https://www.turing.ac.uk/sites/default/files/2024-11/the_ai_revolution_-_opportunities_and_challenges_for_the_finance_sector_-_report_1.pdf

[9] Article

AI implementation in the financial sector: legal challenges. WhatNext.Law. 8 nov. 2024. Available at: <https://whatnext.law/2024/11/08/ai-implementation-in-the-financial-sector-legal-challenges/>

Technical materials and internal studies

Compiled from MVPs and internal demonstrations of generative AI applied to the financial sector, focusing on document automation, conversational APIs, chatbots, and copilots.

Artefact – One-Pagers and functional prototypes applied:

DraftAI – MVP / Cognitive ChatBot Service / Conversational bot with banking integration / Augmented Agent Business Intelligence / Intelligent Document Processing.

Studies on GenAI and NLP applied in:

Mapping of emerging risks / Call understanding with NLP / Deep dive into Private Equity documentation / Acceleration of the Due Diligence process / Credit automation via Intelligent Document Processing.