Generative AI Adoption in Financial Services (GAIA - FS)

PRINCIPAL ACADEMIC TUTOR

Andrea Panelli, DIGEP, @Polito

ACADEMIC TUTOR

Laura Grassi, DIG @Polimi

EXTERNAL INSTITUTIONS

Reply IRISCUBE
Reply E*FINANCE CONSULTING

EXTERNAL TUTOR

Mattia Guadagni | Manager, e*finance consulting Reply
Davide Centonze | Partner, e*finance consulting Reply

TEAM MEMBERS

Simone Brandimarte
Management Engineering,
Politecnico di Torino

Stefano Fumero
Computer Engineering,
Politecnico di Torino

Luigi Maggi
Management Engineering,
Politecnico di Milano

Alessandro Memmolo
Mathematical Engineering,
Politecnico di Torino

Andrea Mirenda
Computer Engineering,
Politecnico di Torino

Vincenzo Morano
Management Engineering,
Politecnico di Milano

Andrea Angelo Raineri
Cybersecurity Engineering,
Politecnico di Torino

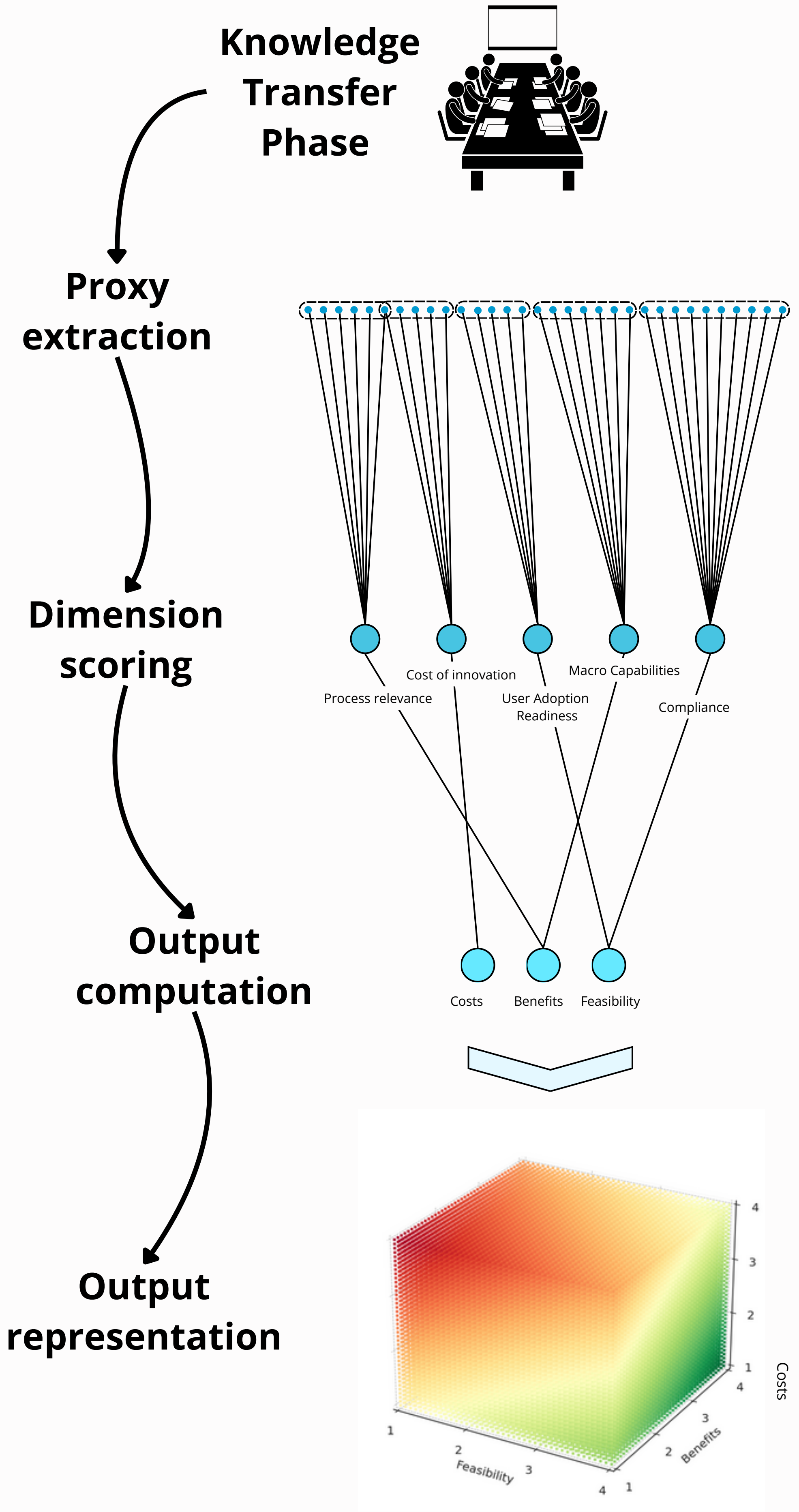
Executive Summary

The financial services industry is at a critical inflection point, confronted with both the immense promise of Generative Artificial Intelligence (GenAI) and widespread hesitation around its adoption. While McKinsey [1] estimates that GenAI could generate between $200 and $340 billion in annual value for banking, practical implementation remains slow. The lack of a structured evaluation approach, regulatory uncertainty, high costs, data privacy concerns, and a shortage of specialized skills continue to hold back large-scale deployment. This underscores the need for a structured methodology to guide financial institutions through this transformation in a strategic and responsible manner.
The GAIA-FS (Generative AI Adoption in Financial Services) project addresses this need by creating a decision-support framework for the ex-ante evaluation of GenAI use cases. Developed through industry research and expert interviews, the framework provides a systematic and repeatable process to assess whether, where, and how financial institutions should adopt GenAI, moving beyond speculative pilots toward value-driven implementation.
At its core, GAIA-FS employs a multidimensional weighted scoring model, inspired by a neural network structure. It evaluates business processes across five critical dimensions: Process Relevance, Cost of Innovation, User Adoption Readiness, Regulatory Compliance, and Macro Capabilities. Each dimension is quantified through proxy variables, ensuring objectivity and comparability across use cases. Rather than producing a single score, the framework synthesizes its analysis into three output dimensions: Benefits, Costs, and Feasibility. This explicitly allows decision-makers to evaluate trade-offs between potential value, required investment, and implementation complexity.
This visualization categorizes opportunities into actionable profiles (e.g., “No Brainer,” “Fair Risk,” or “Titanic”), helping stakeholders prioritize initiatives, validate Proof of Concept projects, or discard unviable options.
For Reply, the framework represents a scalable asset for consulting engagements: it provides a structured, evidence-based approach that supports financial institutions in making informed strategic decisions, balancing innovation potential with risk management, regulatory compliance, and long-term sustainability. Used consistently, GAIA-FS can help the financial services sector unlock the transformative value of GenAI while navigating its operational, financial, and regulatory complexities.

Key words

Generative AI; Financial Services; Regulation bodies; Technology Support.

Generative AI Adoption Evaluation Framework



Project description written by the Principal Academic Tutor

The GAIA-FS (Generative AI Adoption in Financial Services) project was created with the goal of providing a structured tool to assess, ex-ante, the potential for adopting Generative Artificial Intelligence (GenAI) solutions within the processes of financial institutions. Starting from a context characterized by strong interest but also significant uncertainty in the sector, the GAIA-FS project proposes a multidimensional framework designed to understand whether to integrate GenAI in a process of financial institutions. The project's output is a decision support model that analyzes processes across five key dimensions: regulatory compliance, innovation costs, user adoption readiness, strategic relevance of the process, and technical compatibility of the process with GenAI. The final output provides an evaluation based on three criteria - Benefits, Costs, Feasibility - which make it possible to visualize the risks and opportunities associated with adopting GenAI in a given process, thus facilitating informed decisions between immediate implementation, pilot experimentation, or abandonment of the initiative.

Team description by skill

The GAIA-FS team is composed of 7 students from different engineering majors:

• Simone Brandimarte - MSc Engineering and Management
• Stefano Fumero - MSc Computer Engineering
• Luigi Maggi - MSc Engineering and Management
• Alessandro Memmolo - MSc Mathematical Engineering
• Andrea Mirenda - MSc Computer Engineering
• Vincenco Morano - MSc Engineering and Management
• Andrea Raineri - MSc Cybersecurity

To successfully complete the project, the team brought together complementary skills, combining managerial and technical expertise to move from initial ideas to a working prototype. On the management side, Luigi, Vincenzo, and Simone took the lead in engaging with key stakeholders, conducting interviews with leaders in the financial sector to capture business needs, expectations, and current limitations. Building on these insights, they identified the key elements to be used as proxies in the framework, such as the number of people involved, the level of seniority required, and the specific parts of the organization impacted, laying the groundwork for assessing how GenAI could bring real value to the organization. In parallel, Andrea Raneri, Andrea Mirenda, Stefano, and Alessandro focused on turning this vision into reality. They translated the business inputs into technical specifications and built a functional prototype, implementing the web application and designing the neural network architecture of the framework. Beyond development, they validated the feasibility and the actual added value of the proposed business ideas, ensuring that the technical solution truly aligned with strategic objectives. Their work also enabled the framework to generalize its assessments across a wide range of banking processes, providing a robust foundation for future improvements and scalability.

Goal

The goal of the GAIA-FS project is to design a structured decision support framework that enables financial institutions to systematically evaluate the adoption of Generative AI across their processes before implementing it (ex-ante). Rather than relying on hype-driven experimentation or speculative pilots, the framework aims to provide a transparent, evidence-based methodology to assess where, how, and whether Generative AI can create tangible value. The tool is meant to support consultants at Reply in guiding clients through this evaluation, balancing opportunities for efficiency and innovation with the costs, risks, and regulatory constraints inherent to the financial sector. Ultimately, the project aims to bridge the gap between the potential of Generative AI and its responsible, large-scale deployment in financial services, ensuring that institutions commit resources only to initiatives with high strategic relevance, feasibility, compliance alignment, and with the ability to deliver financial value.

Understanding the Problem

The financial services industry is experiencing both the immense promise and the profound uncertainty of Generative AI. On one hand, studies suggest that GenAI could unlock unprecedented value across banking, insurance, and asset management by enhancing efficiency, reducing costs, and enabling new customer experiences. On the other hand, institutions face a fundamental and still unresolved question: *"Should we integrate Generative AI into our processes, and if so, how?"*

By late 2024, over 90% of financial services executives were still in the early stages of GenAI adoption, with only 11% ready for the upcoming European AI Act and just 6% of retail banks prepared to scale AI solutions [2][3]. Surprisingly, the largest banks were found among the slowest movers, despite having the most resources [3]. Surveys by BCG and Capgemini further confirmed this slow progress: two-thirds of managers rated their adoption maturity as low or nonexistent, while more than 60% of banks had yet to define metrics to measure AI's impact [4][5]. Where adoption occurred, it was mostly limited to isolated pilots like chatbots or productivity tools, rather than systemic transformation [1].

Another key reason for slow adoption is the trial-and-error approach used to assess GenAI's value. As Head of Procurement Analytics in Allianz and Head of Analytics and Artificial Intelligence–driven Products at

UniCredit noted during the interviews done, projects were often rushed by fear of missing out rather than performance, with most pilots failing to deliver measurable results as shown in MIT's finding that 95% add no value [6]. This wastes resources for large institutions and forces smaller ones into risky trade-offs between investing at the expense of other priorities or falling behind competitors. The lack of an ex-ante evaluation method creates a structural gap, leaving both large and small firms without a reliable way to assess GenAI's potential before committing resources.

At present, decision-making in this space is often fragmented, anecdotal, and reactive. Many institutions rely on small-scale pilot projects, typically focused on narrow use cases such as customer service chatbots or document generation. While these experiments demonstrate potential, they rarely evolve into scalable solutions that transform operations at the enterprise level. As a result, organizations risk either missing critical opportunities or investing heavily in initiatives that fail to deliver sustainable impact.

Compounding this challenge is the highly regulated environment in which financial institutions operate. The introduction of the EU's AI Act, DORA, and similar regulatory frameworks worldwide sets out strict requirements around transparency, explainability, data governance, and operational resilience. These rules are intended to safeguard consumers and markets but also create substantial complexity for adopters. In many cases, compliance considerations are overlooked during experimental deployments, leaving institutions vulnerable to regulatory, reputational, and financial risks.

Against this backdrop, the GAIA-FS project was conceived to fill a crucial gap: the absence of a structured, repeatable methodology for **ex-ante evaluation** of GenAI initiatives. The challenge is not merely technological but deeply strategic. Financial institutions require a framework that simultaneously accounts for potential benefits, implementation costs, user readiness, technical feasibility, and regulatory compliance.

Exploring the Opportunities

The search for effective ways to adopt Generative AI in financial services starts from the market reality. Institutions want clear value, predictable cost, and proven control, yet most available tools focus on isolated pieces such as process mining, RPA scorecards, or generic risk checklists. These approaches inform parts of the decision but do not by themselves show where GenAI is appropriate, which operating model to prefer, or how to compare use cases across business lines.

Interviews with banks and internal stakeholders consistently point to three needs that any credible solution should satisfy: outputs that executives can understand and explain, traceability and audit for compliance teams, and prioritization of processes where gains in quality or efficiency can be demonstrated.

The approach should begin with the process rather than the model. We require a formal measure of process relevance. Here, by **process relevance** we mean the material importance of a workflow to the institution, reflected in the breadth of personnel involved, the cost and time it consumes, its centrality to interconnected activities, and how its demand scales with the business. Relevance acts as a multiplier: the more relevant a process is, the more improvements amplify positive outcomes across the organization, while failures equally amplify the risks and disruptions they cause.

The solution space is naturally framed by a map of **GenAI capabilities** and their maturity. Practical options range from retrieval augmented question answering and document understanding to translation, multimodal generation and editing, speech, code generation, and synthetic data. Each capability brings different requirements for data quality, supervision, and controls, and each exhibits specific failure modes. A disciplined exploration links process tasks to those capabilities and asks what level of reliability and controllability is achievable with the available data and infrastructure.

Any approach must also respect a tight **regulatory** perimeter. The AI Act excludes certain practices at the outset, while GDPR and DORA require attention to data governance, transparency, and resilience. Rather than treating compliance as a final gate, exploration should weave it into the assessment from the first step with concrete checks on data freshness and bias, ownership and lawful basis, explainability that matches decision impact, and operational safeguards.

Finally, exploration should take into account both economics and adoption. On the economic side, cost is driven by factors such as engineering effort, the expected volume of token inputs and outputs, the frequency of use, integration with existing systems, and the ongoing effort required for monitoring and support. On the other hand, adoption depends on how clearly the solution addresses a well-defined problem, whether it delivers a tangible improvement over current practices. Taken together, these elements enable institutions to make fair like for like comparisons, choose the right scope and operating model, and move beyond pilots to scaled initiatives that deliver measurable value while staying within operational and regulatory guardrails.

Generating a solution

The model operates as a neural style Decision Support System. It ingests process level inputs, aggregates them into a set of intermediate dimensions, and then synthesizes three final outputs that guide prioritization and governance.

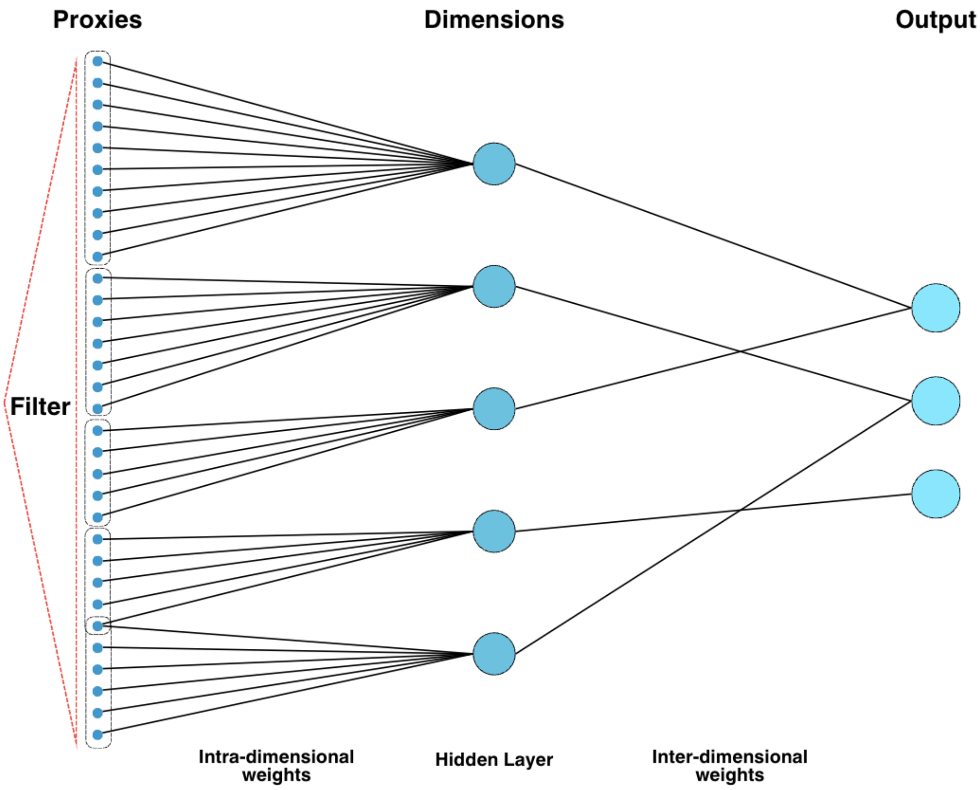


Figure 1: GAIA FS full model. GAIA FS filters out unacceptable risk use cases under the AI Act, aggregates proxies scored 1 to 4 into five dimensions that act like a hidden layer, and combines them to produce Benefits, Costs, and Feasibility; the final scores are plotted in a three dimensional decision space to classify and prioritize opportunities.

The five dimensions are Process Relevance, Cost of Innovation, User Adoption Readiness, Regulatory Compliance, and Macro Capabilities. Each dimension is measured via 1–4 proxies with intra/inter-dimension weights.
Before scoring, we apply a **strict** eligibility **filter** aligned with the EU AI Act. Prohibited use cases, including social scoring, large scale facial scraping, and real time biometric identification outside narrowly defined exceptions, are discarded. Only admissible cases advance to evaluation.

Compliance is assessed along two subdimensions through a structured set of questions. The first is *Data Quality* and Governance, which examines data freshness, bias, roles and ownership, and alignment with GDPR and the AI Act. The second is *Explainability*, which examines decision impact, the required level of justification, the appropriate granularity of explanations, and sector specific constraints. The Compliance Score is the arithmetic mean of the two subdimension scores. Scores are reported on a 1 to 4 scale, where 1 indicates high compliance and 4 indicates low compliance, and higher scores signal greater automation complexity and governance overhead.

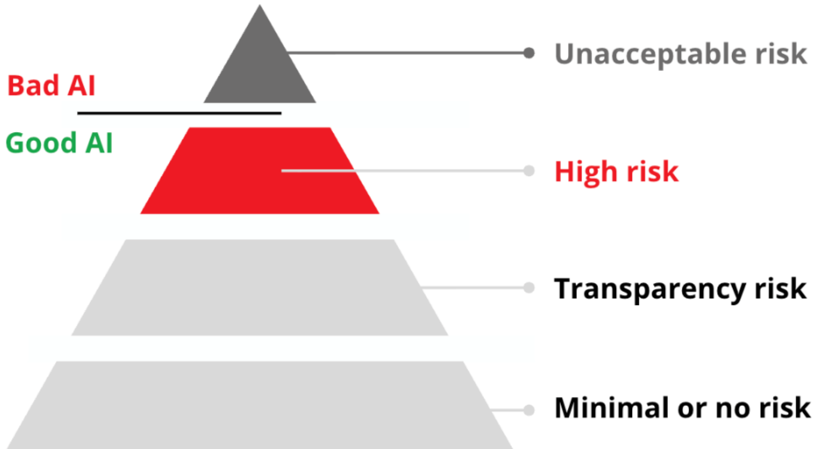


Figure 2 - AI Act risk tiers from low to high: Minimal or no risk, Transparency risk, High risk, Unacceptable risk.

Cost of Innovation estimates the total effort to build, integrate, and run the solution, on a 1–4 comparative scale where higher means lower relative cost; it combines the depth of expertise and engineering required (from simple API orchestration to fine-tuning and custom pipelines), expected token input/output multiplied by usage frequency, integration and change-management with existing systems and controls, and steady-state run, monitoring, and support costs.

Macro Capabilities describe the fit between the use case and what current GenAI can reliably do, mapping the task to capabilities such as retrieval-augmented question answering, text understanding and drafting, translation, multimodal generation/editing, speech, code generation, and synthetic data; the score reflects not just availability but also maturity, controllability, and known failure modes for the specific task and data.

User Adoption Readiness gauges whether people and context can actually absorb the solution: clarity of problem–solution fit, perceived quality versus today's baseline (ideally backed by blind preference tests), explainability acceptable for the decision risk, required supervision and hand-off design, and the presence of owners, policies, and training so that the new workflow is trusted and used.

Process Relevance assesses the strategic and operational importance of a process. It considers the breadth of personnel affected, the share of personnel expenditure and workload it absorbs, its centrality to cross functional workflows, and how demand for the process scales as the business grows. A highly relevant process amplifies outcomes, so improvements or failures given by the GenAI initiative propagate widely and translate into material business impact. The score is reported on a scale from 1 to 4, where 1 indicates low relevance with limited and local impact, 2 indicates moderate relevance, 3 indicates high relevance with organization wide implications, and 4 indicates critical relevance in which changes are likely to affect many functions and produce measurable business effects.

The outputs provided by our framework are **Benefits**, **Costs**, and **Feasibility**, where:

- Benefits is a weighted combination of Process Relevance and Macro Capabilities.
- Costs derive from the Cost of Innovation dimension.
- Feasibility is computed from User Adoption Readiness and Compliance after the ex-ante AI Act filter.

We then place these three results in a color coded three dimensional decision space to support classification and prioritization.

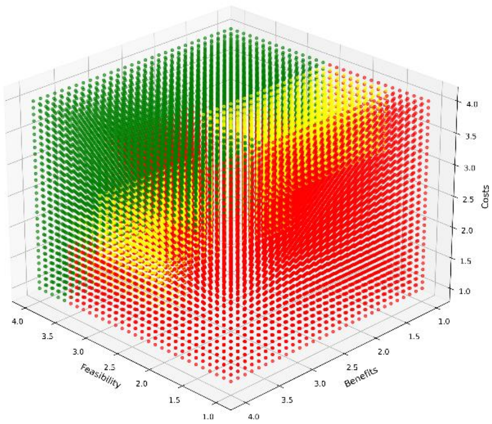


Figure 3 - Graphical representation of the colored-output provided by the framework

The $[1–4]^3$ output-space is partitioned into cost-aware decision profiles: *No Brainer, Smart Bet, Time Waste; High ROI, Fair Risk, Not Worth It; Target, Uncertain Value, Titanic.*

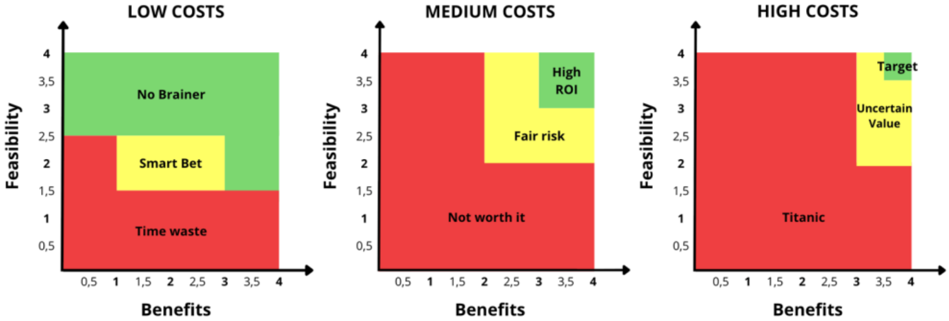


Figure 4 - Decision space by cost tier: three panels fix cost level left to right Low, Medium, High. Benefits on x and Feasibility on y, both 1 to 4. Colored zones signal actions, green proceed, yellow review, red stop. Low cost moves from Time waste to Smart Bet to No Brainer as scores rise. Medium cost is mostly Not worth it until higher thresholds, then Fair risk, with High ROI in the top right. High cost is largely Titanic unless benefits are near 4 and feasibility is high, yielding Uncertain Value and Target. The chart enables fast prioritization and gating.

Bibliography

[1] McKinsey & Company. (2023, December 5). Capturing the full value of generative AI in banking. McKinsey & Company. Retrieved from https://www.mckinsey.com/industries/financial-services/our-insights/capturing the-full-value-of-generative-ai-in-banking

[2] Ernst & Young (EY). (2024, December 10). Lack of upskilling and low regulatory preparedness is holding back deeper GenAI adoption within European financial services firms. EY. Retrieved from https://www.ey.com/en_gl/newsroom/2024/12/lack-of-upskilling-and-low-regulatory-preparedness-is-holding-back-deeper-genai-adoption-within-european-financial-services-firms

[3] Criddle, C., & Quinio, A. (2024, June 30). Financial services still shun AI over job and regulatory fears. Financial Times. Retrieved from https://www.ft.com/content/0675e4d9-62a1-4d6c-9098-a8cb0d1e32ed

[4] Boston Consulting Group, "Generative AI in Banking: Six Myths You Need to Ignore," 2025. [Online]. Available: https://cdn.jsdelivr.net/gh/abncharts/abncharts.public.1/abnasia.org/1733996415457_watermarked.pdf.

[5] Capgemini, "Only 6% of retail banks have built an enterprise roadmap to drive AI-driven transformation at scale," Press Release, 2024. [Online]. Available: https://www.capgemini.com/news/press-releases/only-6-of-retail-banks-have-built-an-enterprise-roadmap-to-drive-ai-driven-transformation-at-scale/.

[6] MIT NANDA (2025). The GenAI Divide: State of AI in Business 2025. Retrieved from https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf