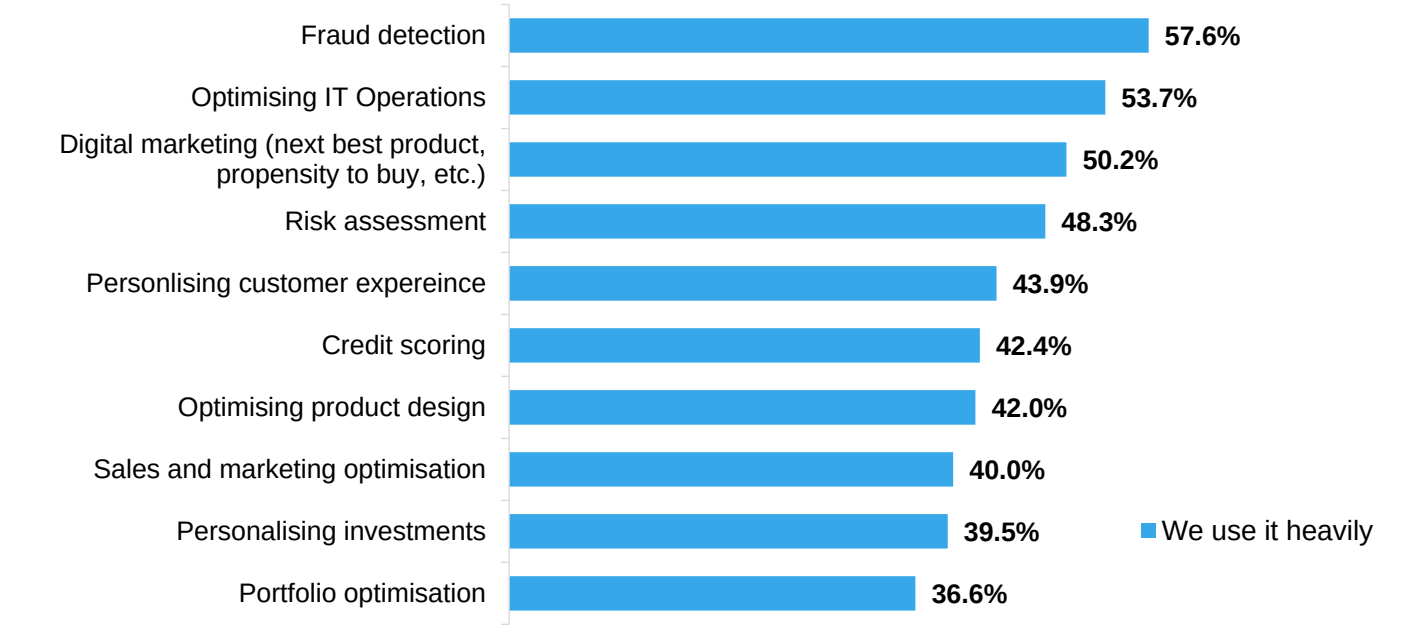
Beyond the Hype:

How to Create a Competitive Edge with AI

reference.point

Total spending on AI in financial services has steadily increased over the last several years and that trend is expected to accelerate at an ~30% CAGR, boosting total spending from an estimated $35 billion in 2023 to a projected ~$100 billion by 2027 (Statista[1]). Broadly speaking, most financial services firms have at a minimum begun experimenting with artificial intelligence, with over 50% leveraging Artificial Intelligence (AI) in areas such as fraud detection, IT operations, and digital marketing (Economist[2]). In an environment where all firms are racing to utilize AI to provide clients with innovative products, improve bottom line, and gain market share, the winners will be those that understand the requirements for AI modeling and can harness their existing data and technology environments to support a steady pipeline for the increasing list of applicable AI use cases.

Figure 1. To what extent does your organisation use artificial intelligence for the following business uses?



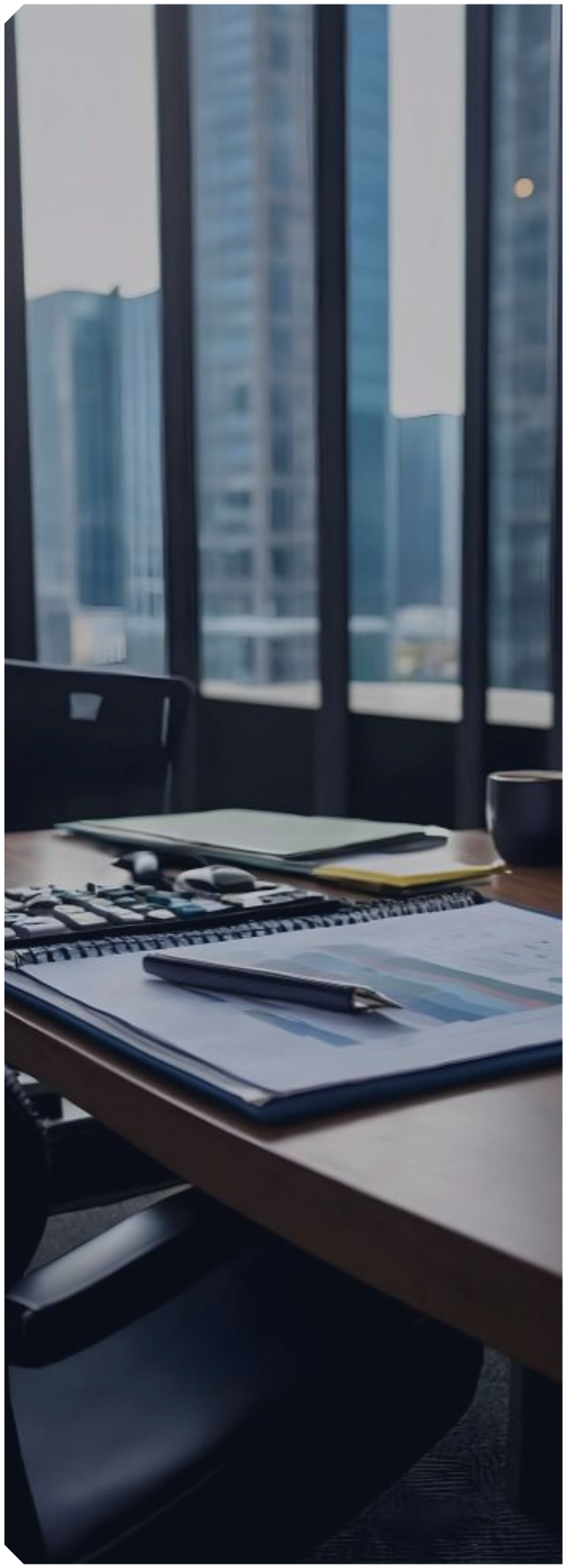
Key Investment: AI Assessment Framework

70% of all projects regardless of type fail and, as of year-end 2023, up to 85% of AI-related projects failed (70% of projects fail, 85% of projects fail[3]). While all-in cost to build an AI model will vary depending on several factors, such as dataset size, model complexity, computing power needed, and training requirements, OpenAI predicts the overall cost of an AI model to increase between 4-5x by 2030. With costs high and projected to rise, successful firms will be those who intimately understand the cost structures of an AI model, how to build reusable assets, and how to extract value. Thankfully, most AI projects follow a similar approach that can be deconstructed to create a reusable framework to assess a use case's costs, reusable components, potential benefits, and ultimately ROI. Throughout this piece, we will discuss several relevant topics to consider when creating an AI Assessment Framework and building out an AI pipeline, such as:

1	2	3	4	5
Compliance, Risk, and Regulatory Concerns	Harness Reusability	Scrutinize Your Data Assets	Understand Model Fit	Importance of Deployment/ Usability

[1] Financial sector AI spending forecast 2023 | Statista
[2] aiinfinancialservices.pdf (economist.com)
[3] Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence

For the first time, the Financial Stability Oversight Council (Council) included a section outlining the risks posed by AI in its 2023 annual report, published in December. The Council, which is chaired by Treasury Secretary Yellen and includes all major US regulators, recommends that all financial institutions “ensure that oversight structures [in relation to AI] keep up with or stay ahead of emerging risks to the financial system.” The Council also references a recommendation, and perhaps a hint, that supervisory authorities further “build out expertise and capacity to monitor AI innovation and usage and identify emerging risks.”



As AI continues to become more prevalent throughout the Financial Services sector, regulators will undoubtedly ramp-up oversight, which could significantly slow down progress. Over time, we foresee a ‘use case dependent’ sliding scale approach being adopted, where certain use cases require more scrutiny than others. “Scrutiny levers” will likely include sensitivity of data, explainability needs, and risk of human fault. For instance, using an AI model to make loan decisions will require a very high level of scrutiny and oversight, while using AI for process automation in middle-office operations will have significantly less. Additionally, a large language model trained on sensitive data, such as PII, is inherently at risk of inappropriately revealing that data and, therefore, will and should have additional infosec requirements. This risk can be partially mitigated by obfuscating data prior to modeling.

While regulators will likely slow the pace of innovation in an effort to reduce risk, the best way firms will be able to maintain innovation velocity as well as regulator confidence will be through reusability of assets. When assessing a library of AI use cases, heavy weight should be given to those use cases that will enable reusable asset creation across the data itself, the data pipeline, modeling, and/or integration. Top firms will utilize continued investments in AI infrastructure to bring their cost-basis on future projects down and, therefore, open up their ability to chase additional use cases. For a deeper dive into Wealth Management-specific use cases, see our in-depth article "Prioritizing Pivotal AI Use Cases within Wealth Management."

Once a problem statement has been defined, it is important to gauge what data sets will be required to build the model and evaluate the state of each data set, including existing and net new assets. Evaluating existing data assets requires additional scrutiny as the “garbage in, garbage out” concept becomes exponentially more problematic when compounded with the current inability to fully solve interpretability and explainability of complex models. While solving the black box paradox is a larger problem, organizations can increase the explainability of their data assets, which will in turn increase confidence in the models. Net new assets, which are completely new data sets or existing data with alterations, will require significant effort to ensure quality, but also to ensure the new asset doesn’t cause friction with the existing pipeline. Additionally, it is highly likely that there will be a core set of data that will be reused consistently across multiple models, and therefore, if built appropriately can be reused (e.g., core reference data, common transaction data). With that in mind, below are some leverageable questions to consider when building out a framework:

Example: Would risk, compliance, and potentially regulatory counterparts approve the quality of the data? If not, what work is required to achieve the desired level of quality?

1. **Data Origination:** Are there any inherent biases in the data collection process?
   - E.g., racial biases in loan approval processes (Federal Reserve[4])
2. **Data Loss/Leakage:** Is there any data loss during movement from origination source to model source?
3. **Data Cleansing/ Normalization:** Are all transformations and normalizations of the data documented or at a minimum auditable?
4. **Data Distribution:** Does the distribution of data reflect the real world?

[4] 2022 Report on Firms Owned by People of Color: Based on the 2021 Small Business Credit Survey (fedsmallbusiness.org)

A typical data scientist will explore multiple algorithm options while testing the viability of a single AI use case; this may make it seem unpredictable to gauge the ongoing costs prior to some level of implementation, but model algorithm types can be assessed in the aggregate to at a minimum provide a comparative cost in relation to each other. By creating and applying a flexible framework that teases out the underlying requirements of the use case in question and mapping it against the most applicable (or top x) model algorithm types, it becomes possible to estimate costs within a reasonable margin of error. Some example questions to tease out the correct algorithm include:

- Is the use case generally focused on predictive capabilities, processing power, or automation?
- What level of risk and regulatory scrutiny is expected? Is high transparency and/or interpretability a requirement?
- What is the size of the available data set? What is available in terms of compute resources?

Table 1. Artificial intelligence and machine learning model selection considerations

Label	Model Type	Transparency	Precision	Compute Power	Data Requirements	Complexity	Interpretability	Estimated Level of Effort
1	Artificial Intelligence	L	L/M/H	H	H	H	H	H
1.1	Machine Learning	L/M/H	H	L/M/H	M/H	M/H	L/M/H	L/M/H
1.1.1	Supervised Machine Learning	L/M	H	L/M/H	M/H	L/M/H	M	L/M/H
1.1.2	Unsupervised Machine Learning	L	N/A	M/H	M/H	M/H	L	M/H
1.1.3	Reinforcement Learning	L	H	M/H	M/H	H	L	M/H
1.2	Expert System	H	H	L/M	L	M/H	H	L/M
1.3	Natural Language Processing	M	M/H	M/H	H	M/H	M	M
1.4	Robotics Process Automation	H	H	L	L	L	N/A	L

- **Transparency** refers to how understandable and explainable the system's decisions are.
- **Precision** is the accuracy and reliability of the system's output.
- **Compute Power** indicates the computational resources required to effectively utilize the model.
- **Data Requirements** measure the amount and quality of data needed for effective functioning.
- **Complexity** evaluates the intricacy of implementing and managing the system.
- **Interpretability** reflects the ease of understanding the system's outputs and reasoning.
- **Estimated Level of Effort** is an aggregated measure to understand the required costs and resources to effectively develop and deliver this solution.

If the organization has a low risk profile and limited resources, the range of solutions may be limited to supervised machine learning and robotics process automation opportunities. Alternatively, more established data science organizations may have a wider range that includes the full scope of machine learning tools along with natural language processing techniques.

Importance of Deployment/Usability

The perfect AI model is only useful when it is implemented and deployed where it can be executed as part of the process, user experience or customer experience for which it is intended. The environment where a model is required to be deployed often has more stringent performance, resiliency, and security requirements. Most likely, the model will need to be integrated with existing Customer and/or Employee UI systems and databases. This integration is often underestimated and perhaps left until late in the lifecycle leading to additional costs (development and runtime) and project delays impacting the business outcome.

Key questions to ask early in the development process:

- *In what systems/channels will the model be deployed?*
- *What are the response time requirements?*
- *What data is required for the model execution and is that data required to execute the model available in those channels?*
- *Will an API set need to be developed?*
- *Will execution of the model be required or optional? If the model and data are not available with the system/channel continue to function or is it 100% mission critical?*
- *Must the model execute in real-time? Or can it execute 'off-line' with the results made available in the real-time experience?*

In addition to the technical aspects, how the output of the AI model will be manifested in a meaningful and clear way to the users must be considered. There is often a need to incorporate a translation layer to facilitate user adoption. When building out the mechanism(s) to manifest the model outputs, such as an API set, it is important to account for the translations required to make the output understandable.

Will users be capable of receiving the model output, digest it, and action it according to how the model expects them to?

Will there be a feedback loop? How will the translation layer prevent bad feedback from being self-reinforced?

Has change management been accounted for? Particularly, training and adoption needs?

- Is there a need to re-train employees?
- What are the testing and change timeframes?

While AI projects are not and should not be treated the same as app dev projects, it is worthwhile to conduct upfront analysis to understand the costs required to go from AI to useful AI. Useful AI aims to increase adoptability of the model and decrease risk through well thought out human-computer interaction.

Final Thoughts

Many artificial intelligence and machine learning algorithms have been around since the 1950s, but because of the advent of LLMs, such as ChatGPT (which have only recently become viable due to extremely large data sets, rapid improvements in GPUs, and cloud computing), many firms will feel the need to green light their own AI models to keep pace. This is classic action bias that led to previous AI winters in the mid-1970s, late-1980s and early 2000s.

The ultimate winners will be those firms who can:

1. Create a reusable pipeline to build quality AI models quickly.
2. Minimize sunk costs by focusing on the highest ROI use cases.
3. Properly plan and account for deployment needs to ensure the AI is useful to the business.

How RP Can Help

Reference Point can assist firms in the following ways:

If you are just getting started on the AI journey, RP can help you get going with:

- Framework Design & Initial Use Case Ideation
- Organizational Readiness Assessment, including assessment of:
  - Organizational Design (e.g., governance, resourcing)
  - Data Pipeline
  - Tooling & Vendor Selection
- Product Backlog Creation

If you are starting to implement AI and are worried about risk, regulation and compliance, RP can provide:

- Model Management & Model Risk Governance Best Practices (incl. AI)
- Data Risk Management Best Practices
- Business-line and technical regulatory expertise.

If you have experience and are trying to scale to manage costs and get better speed to market, RP can help:

- Design reusable technical & data architectures.
- Enhance AI use case framework to focus on program scalability.
- Provide both deep domain expertise and AI expertise.

If you are looking at AI to address a specific set of business needs, RP can help:

- Conduct third-party ROI and risk assessments.
- Design reusable technical & data architectures.
- Author patterns
- Provide both former c-suite domain expertise and AI expertise.

About Reference Point:

Reference Point is a strategy, management, and technology consulting firm focused solely on the financial services industry. Since 2002, we have been a trusted advisor on a wide range of strategic initiatives, helping clients implement leading-edge solutions that quickly and significantly solve challenges and position financial institutions for the future. Our unique approach of combining former industry practitioners with top-tier management consultants provides clients with unrivaled experience and practical, implementable future.