Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Prompt Engineering Frameworks for Generative AI in Credit Analysis

Deepak Pai
Information Architecture, Pittsburgh, PA
MS Information Tech Management, Illinois Tech, Chicago
write2deepakpai@gmail.com

ARTICLE INFO	ABSTRACT

Received: 25 Mar 2025

Revised: 19 Apr 2025

Accepted: 09 May 2025

The financial services industry is currently in the midst of a significant transformation, largely sparked by the rise and widespread integration of generative artificial intelligence (GenAI) technologies. These tools and technologies available have the potential to fundamentally change how we operate and make strategic decisions. In this fast-paced environment, generative AI's knack for synthesizing, analyzing, and extracting insights from extensive and varied datasets is reshaping credit analysis. Traditionally, this field has leaned on careful data examination, statistical modeling, and the judgment of experts to evaluate the financial reliability of individuals or organizations while also proactively tackling the uncertainties that accompany lending. This paper captures prompt engineering techniques for credit risk analysis while leveraging GenAI.

Keywords: Generative artificial intelligence, Prompt engineering, finance, large language models.

1. Introduction:

GenAI assists the user to create content in different formats. It uses various unsupervised or semi-supervised machine-learning algorithms to generate the outputs. This has helped organizations rapidly build foundation models from a vast quantity of unlabeled data. These foundation models can serve as the basis for AI systems that can be used for different implementations [1]. There are several outstanding publicly available GPTs (Figure 1) that make content creation easy and provide path to enhance productivity and finer decision making.

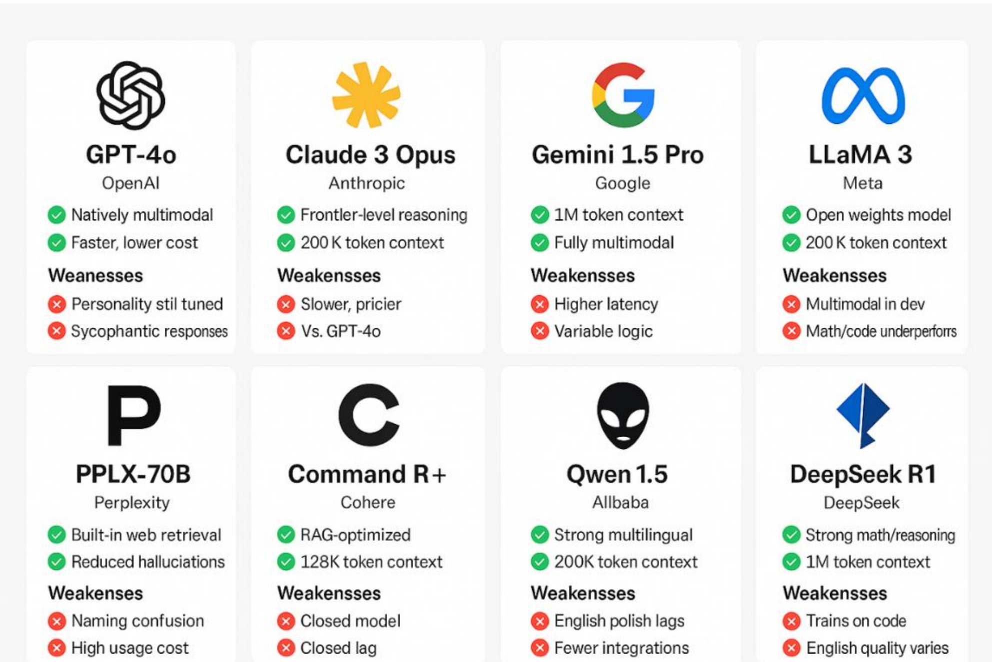

Figure 1: LLM Benchmarking (As of Apr. 2025)

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Imperative of Generative Artificial Intelligence in Financial Services Sector

GenAI in financial context is heavily reliant on deep learning models and its training data. Its operational model revolves around Generative Adversarial Network (GAN) and variation auto encoder (VAEs). [5]. There are several advantages of making the most of GenAI in financial decisioning process (figure 2).

Figure 2: Generative AI Advantages

Focus Area: Credit Risk Analysis

Credit risk analysis is a complex process that involves a wide range of abilities, including contextual understanding, logical reasoning, the application of domain-specific knowledge, and implicit and causal reasoning. Evaluating the performance of large language models (LLMs) in credit risk assessment provides crucial insights into their practical utility in real-world scenarios. This work introduces the application of LLMs for generating comprehensive credit risk reports - a critical task in financial decision-making. More specifically, we explore novel prompt-engineering approaches designed to enhance the quality and fidelity of credit risk assessments; we compare the credit risk assessments of the LLM against human analysts through a user-centered, human-based evaluation, demonstrating the proposed procedure's efficacy in dealing with the credit risk assessment task.

Using LLMs for credit risk analysis is quite a challenging task for a few reasons. First off, credit risk analysis is a complex field that involves looking at a variety of factors, from an individual's credit history and current financial situation to wider economic trends and specific industry conditions. Secondly, financial data is always changing and can be quite volatile, which means that the relationships between different variables are constantly evolving. This adds another layer of complexity. Lastly, there are significant challenges when it comes to using a language model like GPT-4 for tasks that are typically handled by credit analysts—those professionals who have a deep understanding of financial dynamics and risk. This creates a need to bridge the gap between the broad data analysis capabilities of GPT and the specialized knowledge and intuition that credit analysts possess. When it comes to generating detailed

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

credit risk reports, it really tests the limits of current language models like GPT as it requires a mix of domain expertise, contextual awareness, and various reasoning skills. Some components may be manageable with the language models we currently use, but the full scope of the task is much more challenging. The success of GPT (LLM) in crafting these credit risk reports is tied to its ability to comprehend and integrate this nuanced knowledge and intuition, which calls for a strategic approach to prompt engineering.

Currently, the biggest challenge with leveraging GPT and its existing strategies for prompt engineering is the unpredictable responses, especially when it encounters unexpected anomalies, newer data, and changing requirements. This inconsistency poses a significant risk in financial decision-making, where it's pivotal to have quality at the forefront in providing insights.

In this paper, we explore prompt engineering strategy that standardizes the content and quality of GPT-4's output, making it more predictable and insightful, even when handling complex, dynamic tasks like credit risk analysis. The unique strength of this approach lies in its potential to act as a 'missing piece' in prompt engineering; it changes how we interact with LLMs by ensuring that the quality of insights and analyses remains consistent, even when dealing with complex, dynamic tasks like credit risk analysis. Moving beyond traditional applications of LLMs, our research exposes these models to the financial industry's requirements, successfully meeting the pragmatic needs of credit analysis.

2. Background

As the application of Large Language Models (LLMs) in the ever more advanced and complex business of credit risk analysis continues to evolve itself in profound ways, it is not just important but increasingly imperative to report not just about the intricate technical abilities and attributes of such sophisticated models but also to thoroughly explore their multifaceted ethical implications along with the biases all too common in the automated results of such computerized intelligence systems. The application of historical data, an integral part of the majority of machine learning models, can automatically be used to perpetuate and enhance existing biases and inequalities, and therefore it is of the highest priority that timely engineering interventions include and apply bias mitigation mechanisms that ensure the provision of equitable and just treatment to a wide array of diverse demographic groups, such as but not limited to race, gender, and socioeconomic status. By virtue of the diligent work of refining prompts to specifically incorporate fairness measures and context sensitivity, we can significantly enhance the capability of the model to generate reviews that not only are fair but also highly accurate and relevant in their applications. Also, as fiscal regulations become tougher and more complicated in an age of globalization, the practice of including checks for compliance in the context of prompt engineering could have the effect of simplifying complexity and improving conformity with the extensive array of legislative requirements, therefore creating a higher level of trust and confidence in automated decision-making processes in finance. This simultaneous focus on enhancing performance and solving ethical problems will be of extreme importance in building a robust foundation for the future of AI-driven credit risk assessment methods.

3. Methodologies

Let's look closely at the widely used motivating concepts and fundamental models. [Table 1] We will also closely review improving these elements so we may fairly assess how these strategies influence credit risk assessment in various financial situations. This helps us to ensure that our methods not only meet the high regulatory criteria set by the industry but also solve any possible bias in the basic data. Closely analyzing these motivating strategies will help us to understand their effectiveness in reducing different types of risk and promoting justice. In the financial scene, this finally leads to more accurate credit evaluations, benefiting both borrowers and lenders.

Table 1: Prompt Engineering Techniques

Prompting Technique	**Focus Area**	**Prompt Example**	**Use Case**
Role Based	User specifies a Role for LLM to adapt.	You are a financial advisor. Explain the best possible steps for investment for retirement for a young student or a professional starting his career.	Create educational content. Financial advice

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Chain-of-thought	Encourages LLM to think step by step and provide logical reasoning, thereby enabling sharper responses for complex problems.	I am looking to invest 15000 dollars into stocks with 5% dividend and 6% growth YoY. Help me with investment plan with proper reasoning	Budgeting, Financial modeling, Analyze Investments.
Zero-shot	LLM relies solely on task description and its general knowledge	I am new to finance. What does Investment portfolio mean	Quick explanation of terminology or jargon
Retrieval Augmented Generation (RAG)	LLM fetches relevant docs, gathers context, and answers only based on evidence/citations – Reduces hallucinations.	You are senior Risk Analyst. Answer strictly based on Context, Cite Source with #, keep output in JSON.	Instant loan score, Real time fraud, Early warning monitoring
Contextual	LLM relies on contextual information to generate the output aligned with user's needs.	I am a 40 yr. old single parent earning 75000 per annum with 1800 monthly rent and 500 in other expenses. I plan to buy a house next year with a 5% down payment. Create a monthly budget plan to achieve this.	Scenario-based requirements.
Constraint	LLM is provided with a constraint or limitation to work with.	Summarize the contents of this pdf in less than 300 words	Summarize requirements. concise output needs.
Few-shot	LLM is fed few examples in prompt demonstrating the expected output format.	As my Financial advisor craft an email showcasing the strength of my investment portfolio in the following format: 1. Portfolio Monthly Review: Dear User, Your portfolio grew by 1% this month with growth showing in the investment products :a,b,c 2. Portfolio Yearly Review: Dear User, your portfolio grew by 3% this year with growth observed in the following sectors: x,y,z	Tasks with standardized output, Reports.
Tree of thought	LLM has multiple options to navigate before generating the output	I am planning on visiting New York City on a tight budget for 3 days. I would like to visit the top 10 attractions that can be covered, using either of public transportation, walk or ride sharing. I would prefer to use one of the mode of commuting at a given time or it is also acceptable to combine two modes of commute provided it saves me more money and time. I want you to craft a detailed plan and highlight the pros and cons of each option considering my budget.	Risk Assessment, Portfolio allocation, Strategic decision requirements.
Prompt Chaining	LLM has the need to build on top of the previous prompt(s) - Complex task is split into multiple prompts.	As my financial advisory I want you to recommend a plan to calculate monthly saving so that i have 20 million in retirement in 30 years from now assuming 6% annual return. Using my monthly savings from previous step allocate 50% to stocks and 40% to bonds and 10% to other.	Portfolio allocation, Investment strategy, Financial planning, Budgeting.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

		Recommend a plan that is easy to understand and adapt for a 40 yr. old single parent earning 75000 per annum with 1800 monthly rent and 500 in other expenses	

4. Literature Survey

Let's take a closer look at the journal that tries to establish enhancing financial market integrity and risk management [6]. The study explores the importance of prompt engineering in managing credit risk, particularly in risk detection and credit scoring. This raises the relevance and accuracy of existing AI predictions, which will be helpful in making better decisions. The advantages include a significant improvement in accuracy that complies with regulations and a decrease in error rates by around 20% for more complex problems. In contrast, the processes are limited, because they are based on the user's configuration and may not be as successful based on the model they choose, as seen in the comparison of ChatGPT-4 and Google Gemini.

The paper investigates the role of prompt engineering in optimizing large language models (LLMs) like ChatGPT-4 and Google Gemini for applications in financial market integrity and risk management, emphasizing the importance of prompt configurations in enhancing the relevance and accuracy of AI-generated outputs in financial decision-making.

Through experiments comparing the performance of ChatGPT-4 and Google Gemini, the study finds that ChatGPT-4 significantly outperforms Google Gemini in generating accurate financial insights, with a noted 30% improvement, and demonstrates that optimized prompt strategies can reduce error rates by approximately 20% when dealing with complex financial queries

Research Gap:

The paper does not address the long-term implications of integrating prompt engineering into financial services, particularly how these AI tools may evolve and adapt to changing regulatory environments and market conditions over time. This gap suggests a need for further research on the sustainability and adaptability of prompt engineering strategies in dynamic financial landscapes.

There is a lack of exploration into the comparative effectiveness of prompt engineering across different financial sectors beyond credit and market risk analysis. Future studies could investigate how prompt engineering impacts other areas such as investment strategies, fraud detection, and compliance monitoring, which may reveal additional insights into its overall utility in the financial industry.

5. Case Studies and Application

Industry Use Case:

Use Case 1: McKinsey's Embrace GenAI in Credit Risk.

A recent McKinsey survey of banks and lenders revealed that many institutions are experimenting with generative AI in portfolio monitoring and credit review processes One bank developed a GenAI tool to automatically draft climate risk assessment reports for corporate clients. The system uses an LLM that is fed the client's financial reports and sustainability disclosures. To ensure accuracy, the relevant sections of these lengthy documents are first retrieved and provided in the prompt, along with *carefully designed instructions asking the model to extract and summarize key information*. The output is a summary of the client's climate risk profile, complete with references to the source data [3]. This approach showcases how prompt engineering goes hand-in-hand with data preparation: the AI was not turned loose on an entire annual report; instead, the prompt was tightly scoped to sections identified by an upstream process, and phrased to ask for specific summaries. The reported outcome was that the LLM could synthesize a coherent analysis with supporting evidence in much less time than a human analyst would take, albeit with the need for a human to verify the result. Such use cases indicate that prompt engineering for credit analysis often involves *structuring the task into parts*: first retrieving or highlighting relevant data, then instructing the model clearly on what to do with it (e.g., "Summarize X focusing on Y"), and even specifying the format of the answer (like

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

including source citations or a risk rating). This structured prompting is effectively a framework in itself, guiding the generative model through a complex workflow. The success of these trials has led to about 60% of surveyed institutions planning GenAI projects in credit risk monitoring, expecting efficiency and insight gains.

Use Case 2: Moody's GenAI-powered Loan-Covenant & Early-Warning Monitoring.

When a commercial loan officer brings up "covenant monitoring," they're really talking about ensuring that every borrower sticks to the commitments—the covenants—they agreed to in the credit contract. This includes things like "keep your debt-service-coverage ratio above 1.25," "submit audited financials within 90 days," and "no additional senior liens," among others. Traditionally, someone would have to sift through Excel files, PDFs, company news, and emails to verify if those commitments are still being honored. However, Moody's new GenAI-driven monitoring system is designed to eliminate that tedious process and, even more importantly, to identify potential issues before they escalate into defaults.

2.a. Let's dive into how the platform processes and comprehends covenant language.

Document capture:

Whenever a lender secures a deal, all the legal documents—like the loan agreement, amendments, and waivers—are either uploaded or automatically gathered from the deal room.

Clause extraction using a large language model:

An LLM, fine-tuned on thousands of historical agreements, meticulously reads each paragraph, pinpoints covenant clauses (for example, "Maintain DSCR ≥ 1.25"), and converts them into structured data: metric = DSCR, threshold = 1.25, frequency = quarterly, source = borrower's consolidated income statement.

Development of Prompt Templates:

The engineering team quickly realized that using free-form prompts led to some pretty wild and inaccurate metrics. So, they decided to wrap every extraction call in a strict JSON schema prompt—something like "return only objects with keys …"—and then follow it up with a second prompt that double-checks the math on any numerical covenant. This two-step approach really helped cut down parsing errors to less than 2% during pilot tests, as noted in Moody's internal case files.[4]

2.b. Evidence collection:

The system keeps track of a calendar for when borrowers are expected to submit their documents, like financial statements and compliance certificates or other credible documents. It automatically tags incoming files to the right covenant line item using a special "document-type" prompt. If a statement is submitted late or a ratio doesn't pass the test then the relationship managers get an exception ticket in near real time instead of waiting until the end of the quarter to conclude investigation and take corrective action.

2c. What does prompt engineering really bring to the table?

Well, to start with, it promotes precision over fluff. The extraction prompt ensures the model provides exactly the required JSON format, cutting out all the unnecessary legal jargon.
Then there's explainability. Leveraging the chain-of-thought prompting reasoning captured in the logs, auditors leverage findings that is easy to follow and trace as to why the model flagged a potential covenant breach. For instance, if it notes, "EBITDA for Q2 2025 = $8.9 M; required ≥ $9.5 M," you can see exactly how it arrived at that conclusion. And let's not forget about noise reduction. The news prompt features a "materiality" section that sets thresholds for things like revenue impact, credit-rating outlook, or vacancy levels. This means analysts can expect maybe five high-quality alerts each week instead of being bombarded with 500 RSS items.

In Essence - Impact lenders are sharing some exciting news! According to pilot banks speaking with Moody's, they can now detect covenant breaches in "minutes, not days." This means analysts can focus more on creating action plans instead of chasing down files. Portfolio managers appreciate that every alert connects back to the specific covenant paragraph, providing them with a reliable reference when they reach out to borrowers. In essence, this system functions like a dedicated junior analyst who never misses a filing deadline and is always on top of breaking news—allowing humans to concentrate on making those crucial judgment calls

6. Conclusion

Based on Industry use cases, Literature review and prompt engineering strategies available – there is no single umbrella strategy that can serve all financial tasks leveraging same strategy. For Example - The approach taken for KYC may not be the same required for Fraud detection or Credit Risk Analysis. Prompt engineering techniques evaluation study for credit risk analysis suggests that :

Chain-of-thought (step-by-step) prompts raise GPT-4 credit-risk-report accuracy by upto 11-15 pp, giving analysts clearer, rationale-rich outputs [7]

Retrieval-augmented generation (RAG) injects live regulatory and financial snippets, cutting hallucinations and boosting exact-match Q&A scores on bank reports from 0.73 to 0.92 [8]

Few-shot templates keep GPT-4's memos schema-compliant and 60-90 % preferred over free-text outputs by human reviewers [7]

References:

[1] Sai, S., Arunakar, K., Chamola, V., Hussain, A., Bisht, P., & Kumar, S. (2025). Generative AI for Finance: Applications, Case Studies and Challenges. Expert Systems, 42(3), e70018.

[2] Staegemann, Daniel, Christian Haertel, Christian Daase, Matthias Pohl, Mohammad Abdallah, and Klaus Turowski. "A Review on Large Language Models and Generative AI in Banking."

[3] Kremer, A. (2024, July). McKinsey's Risk & Resilience Practice. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/embracing-generative-ai-in-credit-risk

[4] Pieretti, C. (2025). Moody's GenAI-powered Research Assistant drives significant efficiency gains for financial services industry. In*AI and gen AI*. Moody's. https://www.moodys.com/web/en/us/site-assets/genai-research-assistant-financial-services.pdf

[5] Liu, Y., and J. Wang. 2024. "Analysis of Financial Market Using Generative Artificial Intelligence." Academic Journal of Science and Technology 11,no. 1: 21–25. https://drpress.org/ojs/index.php/ajst/article/view/21745

[6] Joshi, S. (2025). Leveraging prompt engineering to enhance financial market integrity and risk management. *World Journal Of Advanced Research and Reviews*, *25*(1), 1775–1785. https://doi.org/10.30574/wjarr.2025.25.1.0279

[7] Teixeira et.al.(2023) Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting https://dl.acm.org/doi/fullHtml/10.1145/3604237.3626902

[8] Iaroshev I, Pillai R, Vaglietti L, Hanne T. Evaluating Retrieval-Augmented Generation Models for Financial Report Question and Answering. Applied Sciences. 2024; 14(20):9318. https://doi.org/10.3390/app14209318