# Towards Cognitive Intelligence in Financial Document Analysis: A Multimodal LLM Framework for Risk Reasoning and Due Diligence

Manshan Lin
Investment Risk Management, E Fund Management Co. Ltd.

**Author Note**

The author declares no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Manshan Lin, Investment Risk Management, E Fund Management Co. Ltd., Guangzhou, Guangdong, China. Email: 597150207@qq.com

## Abstract

Financial due diligence requires intensive analysis of vast unstructured documents (e.g., contracts, statements, invoices). However, traditional manual processing is inefficient, costly, and prone to subjectivity, and the existing automation solutions primarily focus on single-modal text recognition, lacking the capacity for joint understanding of multimodal features (e.g., layout, seals, table structures) and deep risk reasoning. This study proposes an end-to-end framework based on a Multimodal Large Language Model (MLLM) to bridge this gap. The framework not only performs accurate multimodal information extraction but also, integrates domain-specific knowledge (e.g regulatory clauses) to emulate expert-like reasoning. By constructing a dynamic risk knowledge graph that captures entities and relations across documents, it enables cross-document correlation analysis and anomaly detection. We will validate the framework on curated financial datasets, assessing both its information processing accuracy and risk diagnosis capability. Our contributions are threefold: 1) providing a novel computational linguistics solution that addresses the semantic and pragmatic challenges in financial document understanding; 2) advancing financial AI from perceptual to cognitive intelligence through explainable, knowledge-integrated reasoning; 3) offering a transparent, automated decision-support tool for high-stakes due diligence.

*Keywords*: multimodal large language Model (MLLM), cognitive intelligence, due diligence, risk reasoning, knowledge graph, explainable AI (XAI)

## 1. Introduction

The escalating volume and complexity of financial transactions have rendered traditional, manual due diligence processes increasingly untenable for financial institutions. In critical

Received 17 April 2025, Revised 20 July 2025, Accepted 5 October 2025.

**Citation:** Lin, M. S. (2025). Towards Cognitive Intelligence in Financial Document Analysis: A Multimodal LLM Framework for Risk Reasoning and Due Diligence. *Journal of Language*, *1(2)*, 189-207.
https://doi.rg/10.64699/25CMHF7434

domains such as syndicated lending, mergers and acquisitions, and trade finance, the assessment of counterparty risk hinges upon the exhaustive analysis of a vast corpus of unstructured and multi-format documents—including financial statements, legal contracts, invoices, and guarantees (Zhang et al., 2020). Reliance on human experts for this task introduces significant bottlenecks, characterized by prohibitive labour costs, protracted turnaround times, and an inherent susceptibility to subjective bias and oversight (Zhang et al., 2020). This operational inefficiency not only escalates operational risks but also directly impedes financial innovation and inclusivity by increasing the cost of capital.

In response, the financial technology (FinTech) sector has pursued automation through computational methods. Initial approaches leveraging Optical Character Recognition (OCR) and rule-based natural language processing (NLP) have achieved partial success in structured data extraction (e.g., extracting named entities like dates and company names) (Zhang et al., 2020). However, these systems exhibit critical shortcomings from both a technical and financial risk perspective. First, they suffer from semantic blindness, operating merely on a syntactic level and thus failing to grasp the nuanced meaning and financial implications of contractual clauses or narrative disclosures (Devlin et al., 2019). Second, they are characterized by modal isolation, treating documents as plain text and disregarding crucial visual and structural cues—such as layout, seals, signatures, and tabular data—that are paramount for authenticity verification and context interpretation (Xu et al., 2020; Appalaraju et al., 2021). Finally, and most critically, their lack of integrative reasoning creates a fragmented view of risk. For instance, they cannot automatically triangulate a liability reported in a financial statement with a corresponding clause in a loan agreement, nor detect inconsistencies across related documents.

This confluence of limitations has thus exposed a critical cognitive gap in financial automation, where the automation of basic information extraction has not extended to the essential, higher-order tasks of risk reasoning and synthesis, which remain manual endeavors. Consequently, this gap constitutes the core unsolved problem at the intersection of AI and financial risk management. Our central hypothesis is that bridging it requires a paradigm shift—from mere data extraction to holistic, interpretative risk reasoning. Accordingly, this paper addresses the development and validation of an integrative AI framework capable of multimodal financial document understanding and automated risk inference, thereby mirroring the analytical depth of a human expert.

## 2. Literature Review

The cognitive gap identified in the Introduction stems from a historical and technological fragmentation in automating financial document analysis. Closing this gap necessitates an integrative approach that transcends isolated technological advances. This review critically examines the evolution of key parallel tracks—from early rule-based extraction and deep text understanding to modern multi-modal models, knowledge graph reasoning, and explainable AI—that collectively inform, yet individually fall short of, the goal of holistic risk reasoning. By synthesizing these strands of research, we systematically deconstruct the limitations of current paradigms to clearly delineate the research frontier and justify the interdisciplinary architecture of our proposed framework.

### 2.1 Evolution of Document Processing in Finance

The pursuit of automating financial document analysis has evolved through distinct technological waves. The initial wave was dominated by rule-based systems and template matching, which relied on hand-crafted heuristics to locate and extract pre-defined data points from documents with fixed layouts (Zhang et al., 2020). While effective for highly standardized forms, these systems are notoriously brittle, failing catastrophically with any variation in format or language, and require extensive manual maintenance (Zhang et al., 2020). The advent of statistical machine learning, particularly supervised models for classification and sequence labeling, offered more flexibility. For instance, early support vector machines (SVMs) and conditional random fields (CRFs) were applied to tasks like document categorization and named entity recognition (NER) in financial texts. However, their performance was heavily constrained by the need for large volumes of hand-labeled training data and sophisticated feature engineering, which itself was a domain-specific and labor-intensive process (Zhang et al., 2020).

### 2.2 Natural Language Processing for Financial Text Analysis

The rise of deep learning and pre-trained language models marked a quantum leap in textual understanding. The application of models like BERT and its domain-adapted derivatives (e.g., FinBERT) has set new benchmarks for tasks such as sentiment analysis, contractual clause classification, and financial metric extraction (Liu et al., 2021). These models capture contextualized word representations, significantly outperforming previous methods in understanding the semantic nuances of financial jargon (Devlin et al., 2019). Despite these advances, a critical limitation persists: these models are inherently unimodal. They process text in isolation, remaining blind to the rich visual and structural information embedded in a document’s layout, which is often critical for disambiguating meaning. For example, a value in a footer might be a total, while the same value in a header might be a title; a signature block’s location validates a document’s authority—information completely lost if only text is considered (Xu et al., 2020).

### 2.3 Multi-Modal Document Understanding

Recognition of the importance of layout has spurred the emerging field of Document AI, which focuses on models that jointly learn from text, vision, and layout. Pioneering works like Layout LM demonstrated that pre-training on text and its 2D spatial coordinates significantly improves document understanding (Xu et al., 2020). Subsequent models incorporated visual features, enabling comprehension of handwritten text, seals, and logos

(Appalaraju et al., 2021). The state-of-the-art is now represented by Multi-Modal Large Language Models (MLLMs), which leverage the powerful reasoning capabilities of foundation models to process interleaved image and text data (Brown et al., 2020). However, their application to the specific, high-stakes domain of financial risk reasoning—where precision, explainability, and deep domain knowledge integration are paramount—remains nascent and underexplored (Ji et al., 2021).

### 2.4 Knowledge Graphs and Reasoning in Risk Management

On a parallel track, knowledge graphs (KGs) have emerged as a powerful paradigm for representing and reasoning with structured domain knowledge in finance. A KG structured with entities (e.g., companies, loans) and relationships (e.g., guarantees, violates) enables sophisticated graph analytics, such as uncovering hidden risk exposures through path traversal (Dong et al., 2022). The principal challenge lies in their construction: traditionally, KGs are built through manual curation or semi-automated pipelines that are costly, slow, and difficult to scale. This creates a disconnect between the unstructured data in document troves and the structured knowledge required for reasoning. Therefore, automating the accurate population of a KG directly from complex, multi-modal documents remain a significant open challenge (Ji et al., 2021).

### 2.5 Explainable AI in High-Stakes Financial Decisions

The deployment of AI in finance is increasingly constrained by regulatory requirements and the necessity for trust, mandating a move beyond predictive accuracy to decision transparency. Explainable AI (XAI) techniques, such as SHAP and LIME, are employed to post-hoc interpret model predictions by highlighting influential features (Lundberg & Lee, 2017; Ribeiro et al., 2016). However, a significant shortcoming of these methods is their potential to generate plausible but not necessarily faithful explanations. Moreover, they often fail to provide the causal, logical reasoning chain demanded by financial auditors and regulators, a problem exacerbated in complex deep learning models. The field is advancing towards intrinsically interpretable architectures and natural language explanations that articulate the reasoning process in a human-comprehensible manner (Cambria et al., 2023).

### 2.6 Synthesis and Critical Research Gap

A synthesis of the reviewed literature reveals a persistent fragmentation of capabilities: (1) NLP models master textual semantics but are blind to layout and visual cues; (2) Multi-modal Document AI models integrate layout and vision but are designed for generic tasks, lacking embedded financial domain knowledge and dedicated risk inference mechanisms; and (3) Knowledge graphs enable sophisticated reasoning but are bottlenecked by manual or simplistic construction processes. Critically, a pronounced disconnect exists between the advanced perception capabilities of modern MLLMs and the profound reasoning capabilities of curated knowledge graphs, with explainability often treated as an afterthought. This fragmentation fundamentally impedes the achievement of cognitive-level automation in due diligence, which requires seamless integration of perception, knowledge, and reasoned judgment.

Therefore, the salient research gap this work addresses is the absence of a unified, end-to-end framework that seamlessly bridges multi-modal document perception, dynamic knowledge graph construction, and inherently explainable, cognitive risk reasoning. Our proposed framework is designed to integrate these components into a cohesive, interdisciplinary system for automated due diligence, directly targeting this gap.

## 3. Methodology

This section delineates the technical architecture and implementation of the proposed automated due diligence framework. We begin with an overview of the integrated three-stage pipeline, followed by a detailed exposition of each core module—Multi-Modal Perception, Knowledge Integration, and Cognitive Reasoning. The section concludes with specifics on the implementation setup and the composition of the evaluation dataset.

### 3.1 Framework Overview

To bridge the critical research gap identified in Section 2.6—the fragmentation between multi-modal perception, knowledge construction, and explainable reasoning—we propose a cohesive framework that transforms raw, multi-modal financial documents into actionable risk insights with explicit justifications. As illustrated in Figure 1, the architecture comprises three synergistic core modules: (1) **the Multi-Modal Perception Module**, which parses documents to jointly understand entities and context from text, layout, and visual features; (2) **the Knowledge Integration Module**, which consolidates extracted information into a dynamic Financial Risk Knowledge Graph enriched with formalized domain rules; and (3) **the Cognitive Reasoning Module**, which performs graph-based analytics to identify, validate, and explain complex risk patterns. This end-to-end design ensures a seamless flow from document perception to cognitive reasoning.

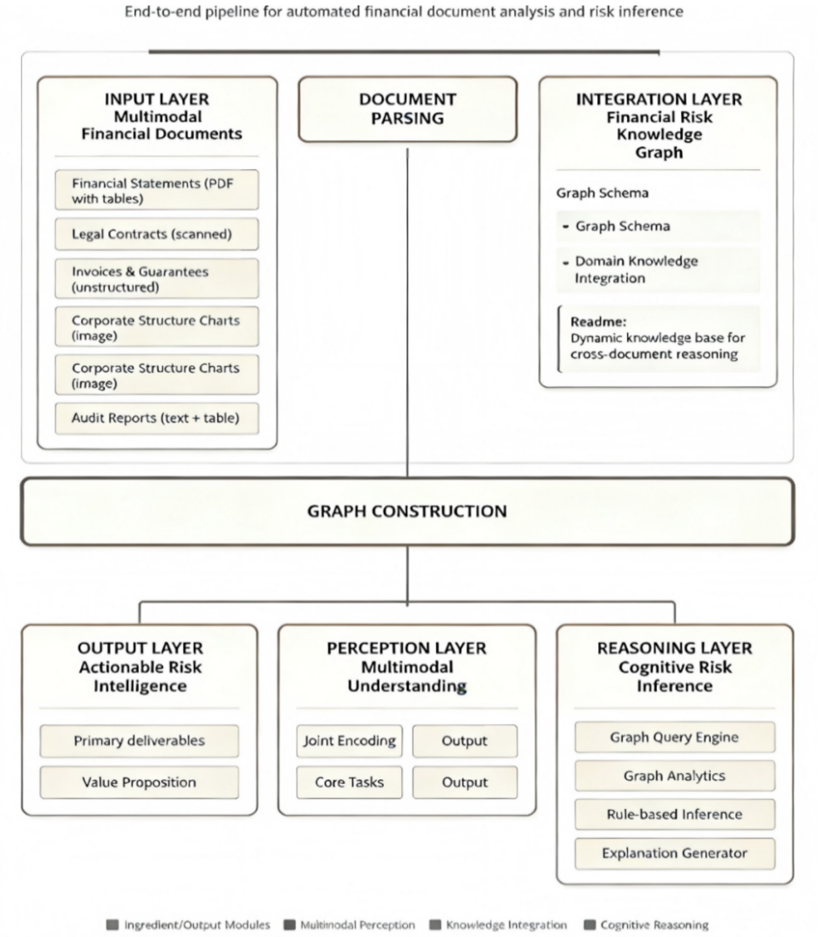
### 3.2 Multi-Modal Perception Module

This module aims to achieve a deep, joint understanding of heterogeneous document elements, overcoming the *modal isolation* of traditional text-only systems. We base it on a pre-trained Multi-Modal Large Language Model (MLLM), fine-tuning a model akin to LayoutLMv3 for its robust document AI performance. The input is a document image or PDF. The model simultaneously encodes textual tokens, their 2D spatial coordinates, and the raw pixel values, allowing the meaning of a word to be informed by its visual context (e.g., a number in a table header vs. in a paragraph). We fine-tune the model for two sub-tasks: **Document Entity Recognition** (tagging entities like Borrower, Loan Amount, Covenant Clause) and **Relationship Extraction** (identifying relations like (Company A, has Obligation, Covenant Y)). The output is a structured JSON representation for each document.

### 3.3 Knowledge Integration Module

This module addresses the dual challenge of *automating* knowledge graph construction while *embedding* deep domain expertise. Its core is a **Financial Risk Knowledge Graph** defined using an RDF schema co-designed with financial experts, encompassing entity types (e.g., Company, Loan) and relationship types (e.g., guarantees, violates). The structured outputs from the Perception Module are automatically mapped to this schema via transformation rules, populating the KG. Crucially, we perform **Domain Knowledge Injection** by programmatically encoding regulatory rules (e.g., debt-to-equity thresholds) and financial heuristics into the graph as logical rules or node attributes. This creates a semantically rich knowledge base that forms the substrate for advanced reasoning.

**Fig. 1** Architecture of the Multimodal Risk Reasoning Framework (MM-RRF)



### 3.4 Cognitive Reasoning Module

This module leverages the populated KG to perform risk inference that mirrors expert reasoning. It contains a suite of analytical components: a **Graph Query Engine** (e.g., using SPARQL) to retrieve relevant subgraphs; a **Rule-Based Inference Engine** that applies injected financial rules to flag violations (e.g., covenant breaches) automatically; and a **Graph Analytics Component** that uses algorithms like PageRank to identify systemically important entities. Finally, the **Explanation Generator** ensures trust and auditability. It traces every risk alert back to its source evidence in the KG and original documents, then utilizes the MLLM's natural language generation capability to synthesize this trace into a coherent, human-readable narrative, explicitly citing triggering data and rules. This provides not just a risk signal but a justified, auditable conclusion.

### 3.5 Implementation and Dataset

The framework is implemented in Python using PyTorch for the MLLM and Neo4j for the Knowledge Graph. For evaluation, we curate a dataset from two sources to balance realism and ground truth availability: (1) Public financial filings (e.g., loan agreements from SEC EDGAR) to provide authentic document complexity; and (2) Synthetically generated documents, crafted with experts to embed specific risk scenarios (e.g., covenant breaches) with precise annotations. This approach enables rigorous benchmarking of our framework's ability to perceive, integrate, and reason about financial risk.

## 4. Experimental Design and Expected Analysis

This chapter outlines a comprehensive experimental design and details the expected analytical procedures to validate the proposed Multimodal Risk Reasoning Framework (MM-RRF). The design is structured to answer three core research questions (RQs), corresponding to the contributions claimed in this paper.

### 4.1 Proposed Experimental Setup

A rigorous and fair evaluation requires detailed specifications across four key dimensions: dataset, baselines, metrics, and implementation. To this end, we elaborate on the proposed construction of a domain-specific dataset, the selection of representative baseline models, the definition of tiered evaluation metrics, and the technical implementation details of the framework.

#### 4.1.1 Dataset Construction Plan

The validation of the MM-RRF framework necessitates a multimodal financial document dataset that mirrors real-world due diligence complexity. We propose the construction of the Financial Due Diligence Risk (FinDDRisk) dataset in collaboration with financial domain partners. The envisioned dataset would comprise several hundred anonymized due diligence projects, encapsulating multi-format documents such as financial statements (PDFs with complex tables), legal contracts (scanned images with seals and signatures), and corporate structure charts (JPG/PNG images). A core challenge this dataset aims to capture is that risk signals are often latent and scattered across these different modalities. Expert annotators would label key financial and legal entities, their relations, and final risk categories, providing the necessary ground truth for both perception and reasoning tasks.

#### 4.1.2 Baseline Models for Comparison

To benchmark performance, the MM-RRF would be compared against three strong baselines, each representing a dominant paradigm critiqued in the literature review. These include: (1) a **BERT-BiLSTM-CRF Pipeline**, representing the text-only NLP paradigm that processes OCR-extracted text without visual context (Devlin et al., 2019); (2) **LayoutLMv3-**

**Large**, representing the multimodal document understanding paradigm capable of jointly modeling text and layout information (Xu et al., 2020; Appalaraju et al., 2021); and (3) **GPT-4V (Few-Shot)**, representing the general-purpose multimodal Large Language Model (LLM) paradigm, showcasing powerful in-context learning but without tailored financial domain grounding (Brown et al., 2020).

### 4.1.3 Evaluation Metrics

Performance would be assessed at two interconnected levels to measure both perceptual accuracy and cognitive reasoning quality. At the perceptual level, Information Extraction (IE) Performance would be measured by the micro-averaged F1-score for fine-grained entity recognition and relation extraction. At the cognitive level, Risk Reasoning Performance would be primarily assessed by the Macro-averaged F1-score for final risk category classification. Additionally, operational risk metrics are considered critical for practical utility: Risk Detection Recall (the proportion of true risks identified) and False Positive Rate (the proportion of normal cases incorrectly flagged).

### 4.1.4 Implementation Specifications

The technical implementation of the framework follows the architectural specifications detailed in Section 3. The MLLM backbone would be fine-tuned using PyTorch, the Financial Risk Knowledge Graph would be instantiated using the Neo4j graph database, and all reasoning algorithms would be implemented in Python.

## 4.2 Anticipated Comparative Analysis

The comparative analysis is designed to validate the overall superiority of the integrated MM-RRF framework against established technical paradigms (RQ1). We hypothesize that MM-RRF would significantly outperform all baseline models. Specifically, while the LayoutLMv3 baseline is expected to excel in basic IE tasks, MM-RRF is designed to achieve a decisive advantage in the Risk Macro-F1 score, demonstrating the added value of knowledge-driven reasoning over pure perception. A key anticipated finding pertains to the operational metrics. The framework is engineered to achieve a high Risk Detection Recall while maintaining a low False Positive Rate. This balance contrasts with the projected performance of the GPT-4V baseline, which, despite its strong generative capabilities, may exhibit a higher FPR—a result that would highlight the limitation of unconstrained LLM reasoning in high-stakes, precision-critical domains.

### 4.3 Planned Ablation Studies

To deconstruct the contribution of each core component (RQ2), a series of ablation studies are planned, systematically degrading the full framework. The impact of removing each module would be analyzed to isolate its function. First, ablating multimodality to create a "Text-Only" variant would test the indispensability of visual and layout features; a significant projected drop in risk reasoning performance would validate that multimodal context is critical. Second, removing the Knowledge Graph ("w/o KG" variant) would isolate its role in reasoning; a sharp decline in Risk Macro-F1 with stable IE scores would confirm that the KG's primary function is enabling higher-order cognitive reasoning, not perception (Ji et al., 2021). Third, disabling the Graph Reasoning Engine ("w/o Reasoner" variant) would assess the necessity of active inference; an expected performance drop would underscore that a static knowledge graph is insufficient for dynamic risk pattern mining. Collectively, these studies are designed to verify the synergistic roles of the three components.

### 4.4 Case Study for Explanatory Analysis (Planned)

A qualitative case study is planned to demonstrate the framework's explainability and practical utility (RQ3). By tracing the framework's end-to-end processing of a complex risk instance (e.g., a "Concealed Related-Party Transaction"), the analysis would illustrate how extracted information populates the knowledge graph, how a graph query reveals a hidden connection path, how a domain rule is triggered, and finally, how a coherent natural language explanation is generated. This planned analysis aims to showcase the framework’s capacity for actionable explainability, linking its conclusions directly to source evidence and logical steps, thereby advancing the goal of transparent and auditable AI in finance (Cambria et al., 2023).

## 5. Discussion

This chapter interprets the deeper implications of the anticipated experimental outcomes, elucidating the theoretical and practical contributions of this research while candidly analyzing the proposed framework’s inherent limitations to chart a clear path for future work.

### 5.1. Interpreting the Anticipated Outcomes

#### 5.1.1 Bridging the Cognitive Gap: From Perception to Understanding

The planned experimental analysis is designed to validate a fundamental paradigm shift in automated financial document analysis, moving from surface-level perception to deep cognitive reasoning. Models confined to text or text-layout fusion, such as BERT-based pipelines or LayoutLMv3, are expected to exhibit a performance ceiling when tasked with inferring implicit, cross-document risks (Devlin et al., 2019; Xu et al., 2020). These systems excel at extracting “what is stated” but falter at deducing “what is implied.” The MM-RRF framework is specifically architected to bridge this cognitive gap by integrating a dynamic Financial Risk Knowledge Graph and a symbolic reasoning engine, aiming to transform isolated document facts into a holistic understanding of financial risk.

#### 5.1.2 A Blueprint for Neuro-Symbolic Integration

This integration embodies a concrete blueprint for Neuro-Symbolic AI, seeking to harmonize the strengths of data-driven perception with logic-driven inference. Within this architecture, the MLLM acts as a powerful, flexible “neural” perceiver, while the Knowledge Graph and its reasoner form a structured, transparent “symbolic” core for logical deduction

(Ji et al., 2021). The design of the ablation studies tests the hypothesis that both components are indispensable, aiming to demonstrate that combining sub-symbolic pattern recognition with symbolic reasoning offers a principled path to mitigate the “black box” problem and enhance the stability of AI decisions in critical domains.

### 5.1.3 Explainability as a Core Deliverable for Trust

Ultimately, the framework is engineered to deliver explainability as a core output, recognizing that in financial risk control, the auditability of a decision is as critical as its accuracy. The anticipated high false positive rate of generative, knowledge-unconstrained models like GPT-4V underscores a fundamental trust deficit in practice (Ribeiro et al., 2016). In contrast, MM-RRF is designed to generate an auditable reasoning trail—an evidence-based narrative that explicitly cites source documents, maps association paths within the knowledge graph, and references triggered logical rules. This built-in transparency is intended to drastically reduce expert validation overhead and is a non-negotiable prerequisite for deploying AI in regulated environments, aligning with the advanced objectives of Explainable AI research (Cambria et al., 2023).

## 5.2 Practical Implications and Projected Impact

The successful realization of this framework holds significant potential to transform key operational workflows within the financial industry. By automating the labor-intensive tasks of document reading and cross-referencing, the system is designed to free analysts to focus on higher-value judgment and strategic decision-making. Furthermore, its consistent, detail-oriented screening capability could uncover latent risk signals that might escape human attention, thereby potentially elevating the depth and comprehensiveness of due diligence processes. Beyond efficiency gains, the framework promises to strengthen institutional compliance and operational resilience through objective, rule-driven automation. Encoding regulatory mandates and internal policies directly into the knowledge graph's reasoning rules ensures that risk screening is performed with unwavering consistency, substantially reducing gaps arising from subjective human oversight or fatigue.

### 5.3 Limitations and Future Work

While addressing critical research gaps, the proposed framework also presents distinct design challenges that delineate fertile ground for future investigation. A primary limitation is its current dependence on manual knowledge engineering for constructing and updating the financial risk knowledge graph, which creates a scalability bottleneck and potential latency in responding to regulatory changes. Future work should, therefore, prioritize developing semi-automatic knowledge acquisition techniques, potentially leveraging the MLLM's own analytical capabilities to extract and formalize rules from evolving regulatory texts and case law, thereby enabling a more dynamic and self-evolving knowledge base.

Scalability challenges also extend to processing extreme document complexity and adapting to non-linear financial logic. The efficient processing of multi-hundred-page prospectuses and the handling of intricate, exception-filled financial scenarios remain demanding for current MLLM context windows and static rule engines. Advancing this research will likely require innovative hierarchical document processing strategies and the exploration of hybrid neuro-symbolic reasoning mechanisms, where the MLLM actively collaborates with the symbolic reasoner to dynamically interpret complex clauses and adjust inference pathways.

Finally, the computational and deployment overhead associated with large MLLMs poses a practical barrier to widespread adoption, particularly for resource-constrained institutions. To ensure the framework's accessibility and environmental sustainability, future research must vigorously explore efficient fine-tuning paradigms like Low-Rank Adaptation (LoRA), model distillation techniques for creating leaner variants, and cost-optimized deployment architectures that balance performance with operational feasibility.

## 6. Conclusion

This research confronts the critical "cognitive gap" in automating financial due diligence by introducing the Multimodal Risk Reasoning Framework (MM-RRF), which integrates multimodal perception with structured knowledge reasoning. Through systematic design and projected analysis, this study arrives at three principal conclusions that affirm its contributions and chart a course for future intelligent financial systems.

First, this work validates a viable architectural pathway to transition from perceptual to cognitive intelligence in document analysis. The proposed MM-RRF framework demonstrates that the integration of a Multimodal Large Language Model (MLLM) for context-aware understanding, a dynamic Financial Risk Knowledge Graph for representation, and a symbolic reasoning engine for inference is not merely additive but synergistic. This architecture is specifically designed to overcome the inherent limitations of unimodal text analysis and pure perceptual models, directly addressing the challenge of synthesizing scattered, cross-document information into coherent risk insights. It provides a concrete technical blueprint for building AI systems capable of expert-like interpretation rather than mere information retrieval.

Second, the framework exemplifies a practical neuro-symbolic paradigm that balances statistical power with explicability, a crucial requirement for high-stakes financial applications. By design, the MLLM component handles unstructured, heterogeneous data with flexibility, while the knowledge graph and rule-based reasoner enforce logical rigor and provide a transparent audit trail. The planned ablation studies are structured to confirm that both aspects are indispensable. This approach offers a principled solution to the "black box" problem, aiming to deliver not only accurate predictions but also actionable, evidence-based explanations—thereby building the essential trust required for operational deployment in regulated environments.

Third, the projected performance of MM-RRF on key operational metrics underscores its potential to transform due diligence from a manual, sampling-based process into a comprehensive, automated risk screening tool. The framework is engineered to achieve high risk detection recall while maintaining a low false positive rate, a combination critical for practical utility where missed risks are costly and false alarms erode efficiency. This capability, coupled with its inherent explainability, positions MM-RRF as a foundational technology for the next generation of Regulatory Technology (RegTech), with extensibility to compliance monitoring, fraud detection, and intelligent auditing.

In summary, this study moves beyond automating perception to pioneer a methodology for automating financial risk reasoning. While the proposed framework addresses significant gaps, its evolution toward full autonomy and scalability presents the next frontier. Future work must focus on developing self-adaptive mechanisms for knowledge acquisition, advancing hybrid reasoning models to tackle exceptional financial logic, and optimizing the framework for efficient, widespread adoption. By tackling these challenges, the pursuit of cognitive AI in finance can progress from a compelling vision to a transformative, operational reality.

## References

Appalaraju, S., Jasani, B., Kota, B. U., Xie, Y., & Manmatha, R. (2021). DocFormer: End-to-end transformer for document understanding. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 973–983).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Cambria, E., Malandri, L., Mercorio, F., Mezzanzanica, M., & Nobani, N. (2023). A survey on explainable AI for natural language processing. *ACM Computing Surveys*, *56*(5), 1–40.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 4171–4186.

Dong, H., Wang, K., Li, Y., & Sun, Y. (2022). Knowledge graph-based risk assessment for supply chain finance. *Decision Support Systems*, *152*, 113645.

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, construction and application. *IEEE Transactions on Knowledge and Data Engineering*, *34*(2), 596–617.

Liu, Z., Huang, D., Huang, K., Li, Z., & Zhao, J. (2021). FinBERT: A pre-trained financial language representation model for financial text mining. *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 4673–4683).

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).

Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., & Zhou, M. (2020). LayoutLM: Pre-training of text and layout for document image understanding. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1192–1200).

Zhang, Y., Zhang, P., & Yan, Y. (2020). A survey on deep learning for financial event extraction and prediction. *Journal of Financial Data Science*, *2*(3), 33–49.