# Phase 1: Foundation Building

## I. Domain Finance

I have chosen finance as the target domain due to its inherent complexity, structured and unstructured data sources, and the high stakes associated with decision-making. Finance includes corporate disclosures, such as 10-K, 10-Q, or 13F filings, regulatory reporting, and investment research, all of which require precise interpretation of technical language and numerical data. This domain also introduces practical challenges such as inconsistent terminology, nuanced accounting concepts, and evolving regulatory standards.

Moreover, finance is an ideal environment for evaluating large language models because the applications are both impactful and risk-sensitive. Financial institutions are actively deploying AI in areas such as risk management, compliance automation, and investor decision-support systems, where accuracy, explainability, and compliance are critical. Misclassification of risks or hallucinated financial metrics can have direct legal, reputational, and monetary consequences. This ensures that prompt engineering strategies are meaningfully tested under conditions that demand rigor.

Overall, finance forces LLMs to work across modalities: tables, footnotes, and narratives often appear together in filings, and regulatory documents may exceed hundreds of pages. These features make the domain especially suitable for assessing whether LLMs can handle dense, technical, and domain-specific content while maintaining fidelity to source data.

## II. Selected Tasks

**Task 1: Information Synthesis - Summarizing Regulatory Filings**

Financial professionals rely on filings such as the 10-K and 10-Q to evaluate company performance, risk exposure, and forward-looking statements. However, these reports are lengthy and filled with repetitive, technical disclosures. The task for the LLM is to condense these documents into coherent, compliance-ready summaries highlighting:

- Key financial metrics (Revenue, Net Income, EPS, Cash)
- Top risk disclosures
- Insights from Management's Discussion and Analysis (MD&A)

This task tests whether the LLM can reduce verbosity while preserving domain-critical content, balancing conciseness with accuracy.

**Task 2: Classification/Analysis - Identifying Risk Factors**

Investment analysts routinely interpret financial disclosures and market data to categorize risks into categories such as operational, regulatory, market, or strategic. The LLM will be tested on its ability to classify risks accurately while providing justification. Unlike summarization, this task emphasizes interpretive reasoning to infer the nature of a risk and explain it with direct textual or numerical evidence. This evaluates whether LLMs can not only extract content but also apply structured analytical frameworks consistently, mirroring the reasoning process of human analysts.

## III. Success Criteria

**Task 1: Information Synthesis**

- **Accuracy:** Numerical values (e.g., revenue, net income) and financial terminology must be captured without distortion.
- **Completeness:** Summaries must include performance metrics, top risk factors, and MD&A insights.
- **Concise:** Length must be reduced by ≥70% while preserving critical insights.

**Task 2: Classification/Analysis**

- **Correctness:** At least 80% alignment with domain expert categorizations of risks.
- **Justification:** Each classification must reference explicit textual or numerical evidence from the source.
- **Consistency:** Similar risks should be classified in the same way across different contexts.

**Conclusion**

The finance domain, with its blend of structured metrics and qualitative disclosures, provides a robust testbed for LLM evaluation. By focusing on information synthesis and risk classification, this framework measures not only factual accuracy but also reasoning, evidence use, and compliance-readiness for real-world adoption of LLMs in high-stakes industries.