VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Project Report on

FINANCIAL RISK ANALYSIS USING LLM

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2024-25

Submitted by

Tasmiya Sarfaraz Khan (D17 - A , Roll no - 30)

Purtee Santosh Mahajan (D17 - A , Roll no - 39)

Ketaki Dhananjay Nalawade (D17 - A , Roll no - 44)

Srushti Satish Sambare (D17 - A , Roll no - 54)

Project Mentor
Dr. (Mrs.) Sujata Khedkar
Associate Professor

(2024-25)

Index

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Certificate

This is to certify that **Tasmiya Sarfaraz Khan (D17 - A , Roll no - 30) , Purtee Santosh Mahajan (D17 - A , Roll no - 39) , Ketaki Dhananjay Nalawade (D17 - A , Roll no - 44) , Srushti Satish Sambare (D17 - A , Roll no - 54)** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "**Financial Risk Analysis Using LLM**" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Dr.(Mrs.) Sujata Khedkar** in the year 2024-25 .

This project report entitled **Financial Risk Analysis Using LLM** by **Tasmiya Sarfaraz Khan (30), Purtee Santosh Mahajan (39), Ketaki Dhananjay Nalawade (44), Srushti Satish Sambare (54)** is approved for the degree of **BE Computer Engineering**.

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide: Dr. (Mrs.) Sujata Khedkar

Project Report Approval
For
B. E (Computer Engineering)

This thesis/dissertation/project report entitled **Financial Risk Analysis Using LLM** by **Tasmiya Sarfaraz Khan (30) , Purtee Santosh Mahajan (39) , Ketaki Dhananjay Nalawade (44) , Srushti Satish Sambare (54)** is approved for the degree of **BE Computer Engineering**.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Tasmiya Sarfaraz Khan (30)

Purtee Santosh Mahajan (39)

Ketaki Dhananjay Nalawade (44)

Srushti Satish Sambare (54)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Dr. (Mrs.) Sujata Khedkar** for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

Chapter Index

List of Figures

List of Tables

Abstract

In an increasingly volatile and complex financial landscape, identifying, assessing, and managing risks has become more critical than ever for investors, analysts, and corporate decision-makers. Traditional risk analysis relies heavily on structured financial data and numerical indicators, often overlooking the wealth of insights hidden within unstructured content such as earnings call recordings, corporate videos, ESG reports, financial news, and investor presentations. These sources often contain early warning signals, implicit sentiment, or contextual risk indicators that are challenging to quantify using conventional methods.

This project, **"**Financial Risk Analysis Using LLM**"**, introduces a multi-modal, AI-powered framework for comprehensive risk analysis by leveraging the capabilities of advanced Large Language Models (LLMs). It aims to extract and interpret meaningful insights from various formats—audio, video, PDFs, and online news articles—to generate structured risk summaries and actionable intelligence.

The system includes five core modules: Earnings Call Analysis (tone and timestamp-based risk signals), YouTube Video Analysis (investment recommendations and risk breakdowns), Real-Time News and Stock Data Analysis (risk summary tables with mitigation strategies), PDF Graph and Chart Interpretation (extraction of visual insights), and ESG Report Risk Evaluation. Each module applies natural language understanding to transform raw, unstructured data into coherent, categorized risk profiles that help stakeholders make informed decisions.

By automating the identification of risk types, potential impacts, and mitigation strategies, this project bridges the gap between traditional quantitative models and modern-day qualitative data sources. The result is a robust and scalable financial risk intelligence solution capable of supporting analysts, regulators, and investors in a data-driven yet context-aware manner. This not only enhances the timeliness and depth of risk assessments but also aligns with the growing demand for transparency and sustainability in corporate risk disclosures.

Chapter 1: Introduction

1.1 Introduction

In the digital age, however, financial decisions are no longer based solely on structured data such as balance sheets, income statements, or market ratios. Instead, stakeholders now turn to a multitude of unstructured data sources—including earnings call recordings, YouTube interviews with executives, social media sentiment, ESG disclosures, and financial news articles—to understand a company's true risk profile. These data streams offer critical context that can signal impending risks or uncover hidden vulnerabilities. Yet, most traditional risk analysis tools are ill-equipped to interpret them.

In an era marked by increasing economic globalization, businesses present tremendous opportunities to expand but also face increasing pressures from competitive markets [3]. As organizations seek to capture this environment actively, risk management and compliance have become key components of corporate governance, especially in financial services—ensuring that companies can maintain trust and stability [4]. Today's complex corporate environment calls for a comprehensive risk management strategy to build organizational resilience and sustainability [5]. Financial pitfalls, in particular, are receiving accelerating engrossment due to their potentially eloquent impact on the organization's future. These risks can arise from a variety of sources, such as market fluctuations, regulatory changes, or faulty internal controls, making it necessary for companies to take a holistic approach to risk assessment and management [6].

The field of risk assessment and management has been firmly established over the last 30–40 years, during which time it has evolved into an organized scientific discipline [7]. The decision-makers need to see beyond the risk evaluation; they need to combine the risk information they have received with information from other sources and on other topics [7].

This project presents an AI-driven, multi-modal system that integrates advanced natural language processing through Large Language Models (LLMs) to analyze financial content from diverse formats. It is designed to identify various risk types—such as operational, financial, reputational, and ESG-related—and evaluate their potential impacts, likelihoods, and mitigation strategies. By applying tone analysis, risk mapping, timestamp insights, and contextual understanding across audio, video, text, and graphical content, the system transforms raw, unstructured data into actionable, structured insights.

By unifying insights from earnings calls, corporate media, public news, ESG reports, and financial documents, this project aims to provide a 360-degree, real-time view of corporate risk exposure—offering a modernized framework for financial risk analysis that is scalable, adaptive, and truly insightful.

1.2 Motivation

Modern financial risk assessment is no longer confined to numbers on a balance sheet—it must account for insights hidden in tone, language, visual cues, and real-time market signals. With the increasing complexity and interconnectedness of financial systems, decision-makers require a multi-dimensional, real-time understanding of risk that goes beyond structured financial data. This project is motivated by the realization that critical risk indicators are often buried within unstructured sources like earnings call recordings, executive interviews, ESG reports, and financial news—making them difficult to analyze through conventional methods.

Earnings call audio, for instance, is not just about what is said, but how it is said. Subtle shifts in tone, pauses, and emphasis often reveal a company's confidence—or lack thereof. By applying tone analysis, timestamped insights, and risk classification to these calls, we can surface hidden signals that might otherwise be overlooked. Similarly, executive interviews and YouTube earnings videos offer rich visual and verbal information. Extracting company overviews, risk types, and investment cues from such content gives investors and analysts a deeper, more human interpretation of financial communication.

Beyond executive commentary, real-time financial news and stock movements offer a pulse on the market's perception and external risks. Summarizing such data into structured risk tables—categorized by type, impact, likelihood, and mitigation—helps in identifying emerging threats and vulnerabilities as they unfold. Furthermore, many organizations present key insights through charts and graphs embedded in PDFs. These visuals are powerful but often neglected in automated systems. Extracting data from these sources and converting them into actionable risk summaries enhances the granularity of analysis.

Equally important are ESG disclosures, which are now pivotal in investment decisions. ESG risks are often complex, qualitative, and distributed across lengthy documents. Automating the extraction of risk types, descriptions, intensity, and mitigation strategies from ESG reports ensures that sustainability and governance risks are part of the broader financial narrative.

The idea of bringing all these capabilities together on a single AI-powered platform was inspired by the fragmented nature of current risk analysis workflows. Today, analysts juggle multiple tools to gain partial insights, often missing crucial interconnections. With the rise of Large Language Models (LLMs) like GPT-4 and Google's Gemini, it is now feasible to process and unify insights from multi-modal, unstructured data at scale. This project leverages that potential to create a unified, intelligent system that delivers a 360-degree view of financial risk—structured, contextual, and ready for informed decision-making.

1.3 Problem Definition

Traditional financial risk analysis tools are primarily built for structured data and fail to account for the growing volume of unstructured and multi-modal information—such as speech, video, charts, and ESG reports—that often contains early indicators of risk. There is currently no unified platform capable of analyzing and extracting risk-related insights from these diverse sources in an automated and intelligent manner.

This gap leads to fragmented analysis, delayed risk identification, and missed strategic signals. Analysts are forced to use multiple disconnected tools, resulting in inefficiencies and incomplete understanding of a company's risk exposure. The absence of integrated tone analysis, visual data interpretation, and contextual risk summarization further limits decision-making accuracy.

The core problem is the lack of a centralized system that leverages Large Language Models (LLMs) to convert unstructured financial data across various formats into a structured, actionable risk profile—covering types of risk, impact, likelihood, and mitigation strategies. This project addresses that need through a single AI-driven solution that brings together all these capabilities on one platform.

1.4 Existing Systems

Existing financial risk analysis systems are largely built around structured data, focusing on key performance indicators such as profit margins, debt ratios, liquidity measures, and market volatility. These systems typically rely on statistical models, dashboards, or rule-based engines to monitor financial health. Some of the popular tools and platforms include:

- Bloomberg Terminal, Refinitiv Eikon, and S&P Capital IQ: These offer financial data analytics, news integration, and performance monitoring, but primarily deal with numerical and structured data.
- Risk Management Information Systems (RMIS): Used by financial institutions to track operational, credit, or market risk.
- Basic NLP-based tools: Some tools use keyword extraction or sentiment analysis on earnings call transcripts or news articles, but with limited depth and context-awareness.
- ESG data providers: Platforms like MSCI and Sustainalytics assess ESG risks, but their methods are often proprietary, and reports are not user-customizable or based on real-time extraction from uploaded documents.

Though powerful in isolation, these systems are domain-specific, have limited support for multi-modal unstructured data, and often lack the intelligence to unify insights from disparate formats like audio, video, PDFs, and raw ESG reports.

1.5 Lacuna of the existing systems

Despite the availability of sophisticated tools, several major gaps persist:

- Lack of Multi-Modal Input Support: Most existing systems do not analyze content across audio, video, graphs, and documents. For example, earnings call audio tone and visual body language cues from interviews are overlooked.
- No Unified Platform: Current solutions are siloed. One tool handles transcripts, another ESG, another charts—leading to scattered workflows and inconsistent analysis.
- Minimal Use of LLMs for Contextual Understanding: Existing systems use traditional NLP or manual tagging, which miss the nuance in tone, risk category interlinkages, and strategic implications.
- Static Risk Models: Risk scoring is often rule-based and lacks dynamic context sensitivity. These models can't understand newly emerging risks from breaking news or real-time market sentiment.
- Non-Customizable for User-Uploaded Content: Analysts cannot upload custom reports, presentations, or PDFs for instant risk breakdown using existing tools.

1.6 Relevance of the Project

This project addresses the gaps mentioned above by introducing a unified, LLM-powered financial risk analysis system capable of interpreting and synthesizing risks from various unstructured sources—all on a single platform. The relevance lies in several transformative aspects:

- End-to-End Risk Intelligence: From real-time tone analysis in earnings calls to extracting risk tables from ESG and PDF visuals, the system provides comprehensive coverage.
- LLM-Driven Contextual Insights: By leveraging models like Gemini, the platform understands context, sentiment, risk relationships, and implications across formats.
- Real-Time, Analyst-Friendly Tool: Users can upload custom content (YouTube videos, PDFs, ESG reports) and receive instantly structured, actionable outputs.
- Strategic Financial Decision-Making: Investors, analysts, and compliance officers benefit from a 360-degree view of risk—integrated from media, speech, reports, and market data—facilitating more informed, proactive decisions.
- Customizable, Scalable, and Transparent: Unlike black-box proprietary systems, this solution can be customized and scaled across sectors, giving users control over data inputs and interpretations.

Chapter 2: Literature Survey

A. Brief Overview of Literature Survey

In the domain of financial risk analysis, traditional models have primarily relied on structured data and statistical methods, often falling short in capturing qualitative insights from unstructured sources such as earnings calls, financial news, and ESG reports. With the rise of Large Language Models (LLMs), there has been a paradigm shift towards more contextual and interpretive approaches to risk evaluation.

Recent literature emphasizes the role of LLMs in augmenting financial analysis through sentiment detection, semantic comprehension, and argument mining. Studies explore LLMs' capability in extracting risk indicators, performing credit assessments, and generating summaries across diverse data modalities. Moreover, the integration of LLMs with time-series models, knowledge graphs, and advanced transcription tools showcases their versatility in real-world financial applications.

B. Related Works

2.1 Research Papers Referred

Title of Paper	Journal	Year	Methodology Used	Merits
RiskLabs: Predicting Financial Risk Using Large Language Model Based on Multi-Sources Data [8]	eprint arXiv:2404.07452	April 2024	RiskLab framework comprises four main modules: 1) Earnings Conference Call Encoder; 2) Time-Series Encoder; 3) Relevant News Encoder; and 4) Multi-Task Prediction Block. It combines self-attention, Bayes-VaR forecasting, and dynamic time windows with news filtering and contextual compression for flexible training.	It integrates diverse information sources for a holistic market view and enables multi-task predictions, offering investors nuanced insights.

Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting [9]	Conference: ICAIF '23: 4th ACM International Conference on AI in Finance	November 2023	A novel prompt engineering technique designed to enhance the quality of responses from GPT-4 by ensuring it addresses specific aspects of credit risk analysis.	The application of LGP and Bayesian networks led to the generation of high-quality credit risk reports that were statistically preferred by evaluators over traditional human-generated reports. The use of LLMs like GPT-4 can increase the efficiency and scalability of credit risk assessments.
Fusing LLMs and KGs for Formal Causal Reasoning behind Financial Risk Contagion [10]	eprint arXiv:2407.17190	July 2024	Risk Contagion Causal Reasoning Model is used to understand how financial risks spread. Financial Knowledge Graphs (KGs) provide important information to help LLMs reason about risks. Fusion Module helps combine information from LLMs and KGs effectively. Uses Sankey diagrams to show how risks spread and their intensity.	The model helps uncover the reasons behind how risks spread. Identifying risk pathways can lead to better strategies to prevent financial crises.
Identifying Corporate Credit Risk Sentiments from Financial News[11]	NAACL-HLT 2022: Industry Track Papers.	2022	Credit Relevance model - used to filter out irrelevant news articles that	Automates the analysis of financial news Provides insights into credit risk

			do not pertain to credit risk Targeted entity sentiment Model - It classifies the sentiment of each entity as Positive, Negative, or Neutral based on the context in which they are mentioned Risk Categorization Model - categorizes sentences or paragraphs into specific risk categories related to credit events. Custom scoring mechanism (Credit Sentiment Score - CSS)	Effectively distinguishes between defaulters (companies that have experienced severe credit events) and non-defaulters (companies that have not)
Predicting Companies' ESG Ratings from News Articles Using Multivariate Time Series Analysis.[12]	arXiv:2212.11765 [q-fin.GN]	2023	Multivariate time series analysis Sentiment analysis Semantic and topic analysis CNN and transformer-based models	Creation of a large and diverse ESG-related news dataset Accurate predictions of ESG ratings Comprehensive analysis of model capabilities
ChatGraph: Chat with Your Graphs[13]	arXiv, 2024. [Online]. Available: https://arxiv.org/abs/2401.12672v1	2024	The ChatGraph methodology combines API retrieval, graph-aware LLM, and fine tuning modules to generate API chains for graph analysis from natural language input, enhancing	ChatGraph simplifies graph analysis by enabling natural language interactions and integrates advanced modules like API retrieval and graph-aware LLMs for diverse

			user interaction with graphs.	real-world applications.
From Pixels to Insights: A Survey on Automatic Chart Understanding in the Era of Large Foundation Models[14]	arXiv preprint arXiv:2403.12027 (2024).	2024	The methodology for automatic chart understanding encompasses several key components aimed at enhancing the interpretation of charts. It begins with classification-based models that leverage visual and textual features, often using CNNs for chart encoding and LSTMs for question processing	Chart-to-Table Conversion: Extracts structured data from visual representations for better analysis. Classification-base d Models: Utilize visual and textual features, often employing CNNs for chart encoding and LSTMs for question encoding
ChartQA: A Benchmark for Question Answering about Charts with Visual and Logical Reasoning[15]	arXiv preprint arXiv:2203.10244.	2022	The study uses a dataset of 9.6K human-written and 23.1K generated questions, with ChartOCR and neural models to extract data and visual features from charts. Transformer models like T5 and VL-T5 process and integrate this information, are trained for accuracy, and evaluated for performance improvement.	Complex Reasoning: Handles complex questions needing logical and arithmetic operations. Data Integration: Combines visual and tabular data for better chart understanding. Human-Centric Design: Reflects natural language and practical question types.
Chart-based Reasoning: Transferring Capabilities from LLMs to VLMs [16]	eprint arXiv:2403.12596	March 2024	The study enhances core image representations to improve reasoning in VLMs. Fine-tuning is done using synthetic datasets with	Synthetic data generation creates diverse training examples, improving model robustness. This approach transfers reasoning from

			reasoning traces from advanced LLMs. The hybrid online setup refines numerical reasoning, and the methods are evaluated on the ChartQA benchmark for visual question answering on charts.	larger LLMs to smaller VLMs, boosting performance with minimal computational resources.
Transcribing in the digital age: qualitative research practice utilizing intelligent speech recognition technology[17]	European Journal of Cardiovascular Nursing, Volume 23, Issue 5, Pages 553–560	July 2024	The study employed intelligent speech recognition within Microsoft Teams for simultaneous transcription, algorithmic processing, accuracy checking, and secure data management to enhance transcription efficiency and accuracy.	Speech recognition technology enhances efficiency, cost savings, data immersion, reduces manual effort, and improves accessibility.
From voice to ink (Vink): development and assessment of an automated, free-of charge transcription tool[18]	BMC Research Notes, vol. 17, no. 95	2024	The transcription methodology includes manual transcription, professional services, and software-based programs, each offering trade-offs in time efficiency, data quality, and potential researcher bias.	Manual transcription ensures deep engagement with data, professional services save time, and software programs offer speed and cost-effectiveness.
Automatic speech recognition and the transcription of indistinct forensic audio: how do the	Frontiers in Communication. 9. 1-9. 10.3389/fcomm.2024.1281407/full.	2024	The study compared the accuracy of 12 ASR systems, including Whisper,	The study offers a thorough comparison of ASR systems, highlighting

new generation of systems fare?[19]			Descript, and Sonix, against human transcribers on forensic-like audio, focusing on word error rates and audio quality.	Whisper's accuracy and relevance to forensic applications.
Transcription and Qualitative Methods: Implications for Third Sector Research[20]	Voluntas 34, 140–153	2023	The study analyzed 212 qualitative research articles to examine how transcription is discussed, finding that 41% omitted it, while others varied in detail, highlighting the impact of theoretical background on transcription style choices.	The study underscores the need for detailed reporting and an interpretivist approach to transcription in qualitative research.
Advanced Search and Summarization of Educational Documents Using Machine Learning [21]	Journal of Nonlinear Analysis and Optimization Vol. 15, Issue. 1, No.6 : 2024 ISSN : 1906-9685	2024	The study focuses on fine-tuning language models to better extract key information from various texts, particularly in the fields of science and literature.	The methodology uses embeddings to capture semantic meaning, which helps in generating contextually rich summaries. Advanced pre-trained LLM (Flan-T5, BART) provides high-quality summarization.
LaMSUM: A Novel Framework for Extractive Summarization of User Generated Content using LLMs [22]	Arxiv. Computation and Language (cs.CL); Machine Learning (cs.LG)	2024	This innovative work introduces LaMSUM, a framework that harnesses the power of LLMs to create extractive summaries from user-generated content.	This addresses the challenge of summarizing extensive content that exceeds the context window of LLMs by employing a multi-level

				summarization approach.
A Comprehensive Survey on Process-Oriented Automatic Text Summarization with Exploration of LLM-Based Methods [23]	Arxiv Artificial Intelligence (cs.AI)	2024	This study delves into the significance of ATS in reducing human effort in processing large texts, while also addressing the practical applications of these methods in real-world scenarios. It highlights the evolution of ATS techniques, particularly in light of the transformative impact of Large Language Models (LLMs).	The introduction of a "Process-Oriented Schema" offers a structured and practical framework for understanding ATS, aligning theoretical concepts with real-world implementations.
Mining Both Commonality and Specificity From Multiple Documents for MultiDocument Summarization [24]	IEEE Access (Volume: 12)	2023	This study presents an innovative method that balances the need for coverage and content diversity by utilizing a class tree derived from hierarchical clustering of documents. By selecting sentences based on their relevance to both common themes and unique specifics, this approach ensures that summaries are both informative and varied. .	The method captures the overall common information across all documents while also highlighting the unique characteristics of different subclasses.

From Moments to Milestones: Incremental Timeline Summarization Leveraging Large Language Model [25]	Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)	2024	1.LLM-TLS approach collects and preprocesses text data, detects events using large language models (LLMs), clusters similar events, and constructs a dynamic timeline, updating it in real-time. 2. Summarizes the clustered events, evaluates performance using metrics like precision and ROUGE, and refines the process iteratively based on comparative analysis with baseline models.	1.Offers real-time, scalable timeline construction with dynamic event clustering, 2.Effectively handles noisy and large data streams 3. Generates high-quality abstractive summaries
Temporal analysis of topic modeling output by machine learning techniques[26]	International Journal of Data Science and Analytics	2024	Data is collected, preprocessed, and topics are modeled using LDA/NMF, with trends tracked through feature vectors. Dimensionality reduction and clustering identify patterns, which are evaluated, visualized, and interpreted for reporting.	Flexibility, applicability, various datasets. Robust evaluation, validation. Dynamic Topic Analysis
Noun and Verb Phrase Extraction in Natural Language Processing: A Comparative Study of Approaches[27]	Procedia Computer Science Volume 235 , 2024, Pages 2876-2885	2024	Preprocessing text (tokenization, stemming) using POS tagging with models to label nouns and verbs, incorporating rule-based,	1. High accuracy: 96% POS tagging, 95% SpaCy. 2. Versatile, adaptable across domains/languages 3. User-friendly

			statistical, and machine learning techniques (SpaCy, NLTK) for extraction. Performance evaluation shows high accuracy (SpaCy 95%, POS tagging 96%)	
Visualizing Parts of Speech Tags by Analyzing English Language Text [28]	IEEE	2024	CRFs and Spacy Integration Evaluates CRFs with Spacy against traditional methods, showing superior tagging accuracy. Includes graphical or annotated text to visualize POS tags	Accuracy: Combines CRFs and Spacy for better POS tagging accuracy. Real-time Processing, Improved Interpretability. Easily scalable and adaptable.

Table 2.1 Literature survey

Chapter 3: Requirement Gathering for the Proposed System

3.1 Introduction to requirement gathering

Requirement gathering is a crucial initial step in the development of the proposed Financial Risk Analysis system, as it ensures a clear understanding of the user needs, technical goals, and data requirements. Given the system's complexity integrating earnings calls, financial reports, news articles, and video content for AI-driven risk analysis, accurate and well-defined requirements help in aligning the development process with the intended functionality. This phase lays the groundwork for designing a robust, scalable, and user-friendly platform by identifying what the system must do (functional requirements) and how it should perform (non-functional requirements), ultimately ensuring the delivery of a reliable and insightful solution for investors and analysts.

3.2 Functional Requirements

1. **Data Upload & Processing**:
    a. Allow users to upload corporate earnings call audio, YouTube video URLs, transcripts in PDF format, and financial reports.
    b. Extract audio from video URLs and perform speech-to-text to generate a transcript.
    c. Generate a summary and perform timeline analysis on the earnings call transcript.
    d. Perform tone and semantic analysis on earnings call transcripts for risk analysis.
    e. Implement argument mining to extract key points and assess risks from the transcripts and news.
    f. Integrate a search bar to fetch company financial reports based on company name and year.
    g. Extract relevant company news articles based on the provided time frame.
    h. Combine data from earnings calls, transcripts, company news, and financial reports to generate a comprehensive risk report.
2. **Risk Table Generation**:
    a. Perform a detailed risk analysis based on financial performance metrics, corporate strategy, governance, capital allocation, market focus, ESG, and forward-looking statements.
    b. Provide users the ability to download the final risk table in CSV format.
3. **Graph Interpretation**:
    a. Dynamically generate and interpret visual financial data, including charts and graphs, with an emphasis on identifying potential risks or insights.
4. **Data Validation**:
    a. Ensure uploaded data belongs to the surrounding relevant timeframe, such as matching earnings call quarters with financial report years and news periods.

3.3 Non-Functional Requirements

1. **Performance**:
    a. System should process uploaded audio, video URLs, and PDF transcripts efficiently with minimal latency.
    b. Real-time risk analysis and report generation should be optimized for fast turnaround.
2. **Scalability**:
    a. The system should scale to handle multiple users concurrently uploading and analyzing large datasets, ensuring high availability during peak loads.
3. **Accuracy**:
    a. The tone, semantic analysis, argument mining, and risk assessment algorithms should be highly accurate and robust to ensure reliable decision-making for investors.
4. **Security**:
    a. Ensure that all uploaded data is securely stored and processed.
    b. Implement access controls, encryption, and secure data transmission.
5. **User Experience**:
    a. The system should have an intuitive interface that allows users to easily upload files, analyze data, and download reports.
    b. Provide clear instructions and feedback during the process, with visual aids for graph interpretation.
6. **Compatibility**:
    a. Support for multiple formats, including audio, video URL, PDF, and different financial report formats, ensuring compatibility with a wide range of data sources.

3.4 Hardware, Software , Technology and tools utilized

1. **Frontend:**

   The frontend will be done using React.js - a JavaScript Frontend Framework and Streamlit which will provide a user interface for our project.

2. **Transcribing the source earning call:**

   Google Text-to-Speech: We have used Assembly AI tool to transcribe the uploaded source file.

3. **Summarization of transcript:**

   We will use argument mining to extract and summarize key arguments from the transcribed text. This method identifies core arguments and supporting evidence. Advanced models like BERT or Gemini will be employed to enhance the summarization, with BERT providing bidirectional context and Gemini offering deeper argument analysis. This approach ensures a focused and insightful summary of complex discussions.

4. **Extracting Financial Data:**

   NLP Model and Named Entity Recognition (NER): For the extraction of financial data from our transcripts, we will require NER techniques. The pre-trained model can be used which has NER techniques, but it should be trained specifically for the extraction of financial data. We need to train the model for extracting quantitative measures, keywords related to revenue, earnings per share, revenue generated by the company and a lot more related parameters. This model will identify all the texts that match financial measures and parameters from the generated transcripts.
5. **Data interpretation of charts :** Users can chat with charts or graphs, which the Gemini model will analyze to extract meaningful insights and trends, enabling a deeper understanding of complex data and supporting informed decision-making.
6. **Risk assessment and Investment prediction:** Using either the Gemini or PaLM model, we will analyze argumentative structures in financial reports, news to assess risks and predict investment outcomes, helping users make more informed decisions based on market insights.
7. **News as input :** Integrate media news using a news API to evaluate the current economic, political, and social climate impacting market sentiment and investor behavior. This helps gauge how events and trends influence market dynamics and investment decisions.
8. **Financial Report Retrieval and Risk Analysis**: A search bar will be available for users to input a company name and year, allowing them to fetch and download the company's financial reports using the Google Custom Search API. Once the reports are uploaded, the system will provide a risk analysis, offering insights into potential financial risks. This feature streamlines the process of accessing and evaluating financial data, supporting informed decision-making and thorough risk assessment.

3.5 Constraints

1. **Data Dependency**:
   a. The system is reliant on the availability of corporate earnings calls, financial reports, and news articles from third-party sources (company websites, YouTube, and news outlets).
2. **Time Frame Alignment**:
   a. Data sources must align in terms of the timeframe (e.g., earnings calls, financial reports, and news must be from the same quarter or year), which imposes constraints on the processing of mismatched data.
3. **Processing Power**:
   a. High computational resources may be required for processing audio, video, large transcripts, and conducting detailed risk analysis.
4. **Legal and Compliance**:

a. The system must comply with legal regulations regarding data usage and storage, especially when handling financial information and corporate documents.

5. **Data Quality**:
    a. The accuracy of the analysis is highly dependent on the quality of the uploaded audio, transcript, and financial reports. Poor data quality could negatively impact risk assessments.

Chapter 4: Proposed Design

4.1 Block diagram of the system

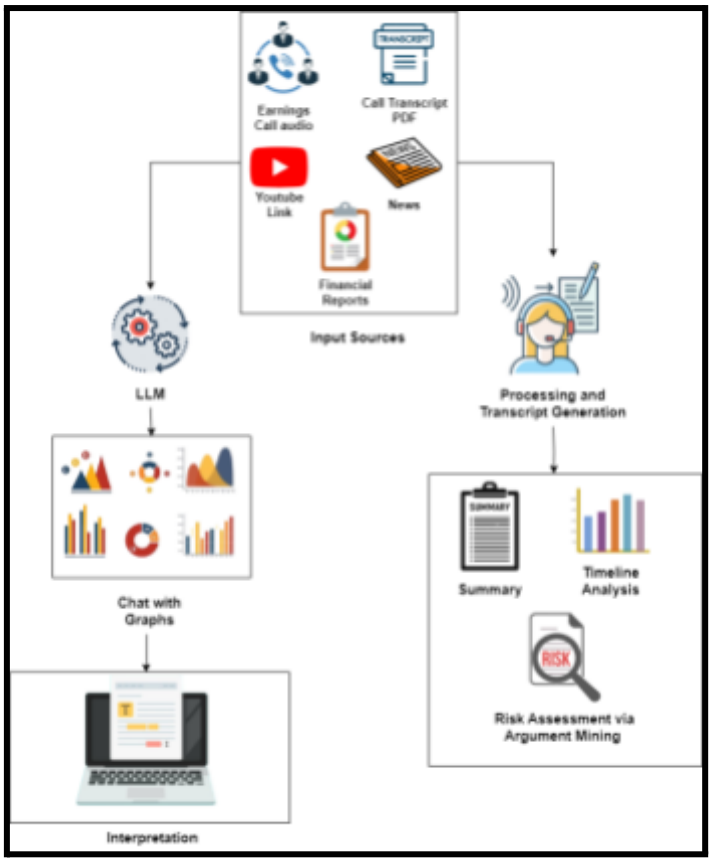


Fig. 4.1 Block diagram of the proposed system

This will be the whole block diagram of what the user can give as input to the system and what he will get as output.

The proposed system - 'Fincalls - Risk Analyzer' acts as an effective solution for the economic development of the company as well as provide an ease of analysis to the investors.

Following sources of data are taken into consideration.

1. Corporate Earnings Calls Audio
    a. The corporate earnings call audio is available on the company websites. It can be uploaded. The tone and semantic analysis of the call will be done and a risk analysis will be given.
    b. Also, a transcript of the call will be generated.
    c. Using this generated transcript, the summary of the call as well as the timeline analysis of the call will be given.
2. Corporate Earnings Calls Youtube Video URL
    a. Along with the youtube channels owned by the companies, there are multiple channels that post earnings calls video recordings on youtube.
    b. Again, after extracting the audio, the tone and semantic analysis of the call will be done and a risk analysis will be given.
    c. Using this generated transcript, the summary of the call as well as the timeline analysis of the call will be given.
3. Corporate Earnings Calls Transcript PDF
    a. Sometimes, even the direct transcript is available instead of the call.

   b. From this transcript, the summary of the call as well as the timeline analysis of the call will be given.
   c. Also, via argument mining, the risk analysis will be given.
4. Company News
   a. The company name and the duration of the release of the news articles will be given.
   b. And from this, the relevant news will be extracted.
   c. A risk analysis will be given corresponding to the extracted news.
5. Financial Report and ESG Report
   a. A search bar will be provided, where, after giving the company name and the year as an input, the financial reports of the company will be fetched.
   b. They can be downloaded and used further.
   c. Once uploaded, a risk analysis of the reports will be provided, too.

Note that, the uploaded data should belong to surrounding durations. That is, the quarter to which the earnings call belongs should be a part of the year to which the financial report belongs. Also, the news duration should lie in the same duration.

Now, the risk analysis from all the above mentioned sources will be combined to form a Risk Report. This report can be downloaded in a PDF format.

The Risk Report will be created based on the following aspects:

1. Financial Performance Metrics: To analyze metrics focusing on revenue growth, profitability, cash flow, and debt levels and identify any potential risks associated.
2. Corporate Strategy and Governance: Review recent changes in leadership or corporate structure to identify potential risks associated with these changes.
3. Capital Allocation and Dividends: It looks at the company's strategies for reinvesting profits, paying dividends, conducting share buybacks, and managing debt.
4. Product and Market Focus: Evaluate how well the company’s products and market strategy align with current and future market trends.
5. ESG and Social Impact: Examine the company’s ESG reports, CSR initiatives, and sustainability goals to identify any gaps or areas where the company is underperforming.
6. Forward-Looking Statements and Guidance: Identify the company's future expectations by reviewing management’s guidance on growth prospects, potential risks, and strategic priorities highlighted during the earnings call.

4.2 Modular design of the system

Fig. 4.2 Modular Diagram - Risk Assessment

The modular diagram of different inputs the user will feed the system and the many outputs he will get for risk assessment.

The user will have to upload either or all of the supported input sources which comprise of the Earnings Call audio, Earnings Call Transcript, Earnings Call YouTube Video URL, News and financial reports. Then, the user can either go for 'Chat with Graphs' or 'Risk Analysis' module.

If the user chooses the former, then, the numerical and the statistical data would be extracted from the given sources and would be represented in a graphical format. The interpretation of the formed graphs will be given and the user can interact and chat with those visualisations.

However, if the user chooses the latter, then, a separate risk analysis will be done on each source and the combined results will be used to form a risk report. Along with that, some other features include summary and timeline analysis of the Earnings Call.

4.3 Detailed Design

1) Earning Call Audio

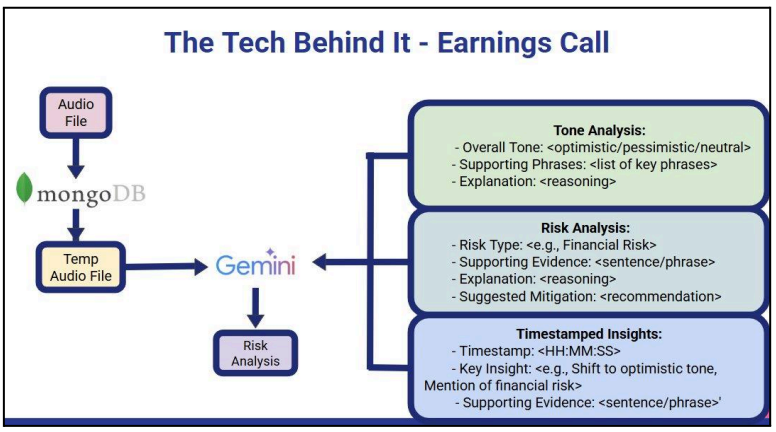


Fig. 4.3.1 Process flow diagram of Earnings call module

1. Audio Input & Temporary Storage : The process begins with the user uploading an earnings call audio file. This file is temporarily stored and registered in a MongoDB database, which helps manage file handling, track progress, and store references for future use.

2. Gemini – Large Language Model Processing : Once stored, the audio is passed to Gemini, a powerful Large Language Model (LLM) that transcribes the content and performs intelligent analysis. Gemini is responsible for generating three major outputs:

A. Tone Analysis

- Overall Tone: Detects whether the tone is optimistic, pessimistic, or neutral.
- Supporting Phrases: Lists key phrases spoken during the call that influenced the tone decision.
- Explanation: Provides reasoning behind the detected tone, based on the content and context.

B. Risk Analysis

- Risk Type: Identifies specific types of risks mentioned (e.g., financial, operational, regulatory).
- Supporting Evidence: Shows the exact sentence or phrase indicating the risk.
- Explanation: Explains how the phrase represents a potential risk.
- Suggested Mitigation: Provides mitigation strategies or recommendations generated by the model.

C. Timestamped Insights

- Timestamp: Indicates the exact time (HH:MM:SS) in the audio where an important insight occurred.
- Key Insight: Describes what changed or was revealed (e.g., shift in tone or mention of a major risk).
- Supporting Evidence: The specific sentence or phrase spoken at that moment.

2) Earning Call Youtube Video Url

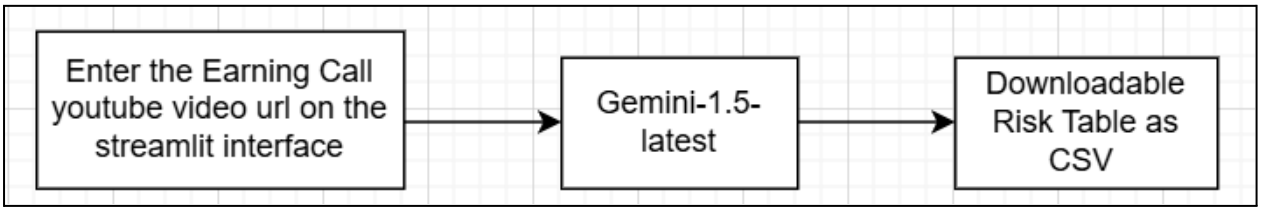


Fig 4.3.2 Earning Call Youtube Video Url Modular Diagram

Input via Streamlit Interface:

- The user enters the YouTube URL of a company's earnings call video.
- The interface is built using Streamlit, providing a simple and user-friendly experience.

Processing with Gemini-1.5 (LLM):

- The system extracts the audio and generates a transcript from the video.
- The transcript is passed to Gemini-1.5, a powerful Large Language Model.
- Gemini performs the following analyses:
    - Risk Identification – Categorizes types of risk (e.g., financial, operational, strategic).
    - Summary Generation – Produces concise risk descriptions
    - Impact Assessment – Evaluates the potential effect of each risk.
    - Likelihood Estimation – Assigns a probability or likelihood rating.
    - Mitigation Strategy – Suggests how each risk can be addressed.
    - Investment Recommendation – Provides insights like Buy/Hold/Sell based on overall tone and content.

Output – Downloadable Risk Table:

- The analyzed data is structured into a risk summary table.
- The table includes: Risk Category , Summary , Potential Impact , Likelihood , Mitigation Strategy , Investment Recommendation
- The user can download the table as a CSV file for further analysis or reporting.

3) Latest News and Stock Data Risk Analysis

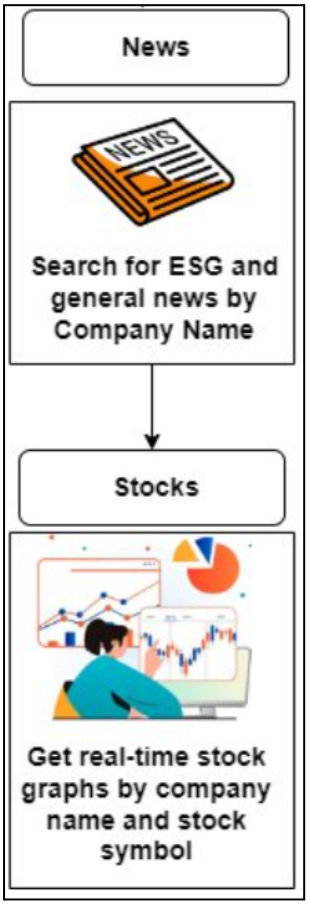


Fig4.3.3. Latest News & Stock Data

Input – Company Name or Stock Symbol: The user provides the company name or stock ticker symbol on the interface.

News Section: The system searches and retrieves real-time news articles related to the specified company.News sources include : General financial news and ESG-related news (Environmental, Social, and Governance). This helps to detect:
Emerging risks , Controversies or lawsuits , Environmental or reputational issues
and Regulatory or compliance challenges.

Stocks Section: The system fetches real-time stock market data for the given company. It includes: Live stock price charts , Historical performance and Key financial indicators. This helps analyze: Volatility trends , Market sentiment and Correlation of news events with stock movements

Combined Analysis: The integration of news and stock data offers a comprehensive view of risk: News provides qualitative risk indicators and Stock graphs show quantitative impact. It helps identify the potential market response to specific risk factors.

Output: The system compiles this information into a summarized risk table (in other modules). Users can visually correlate news sentiment and stock volatility to assess financial stability.

4) Annual Public reports for Chart and Graph Risk Analysis & Environmental , Social and Governance (ESG) Reports for ESG Risk Analysis

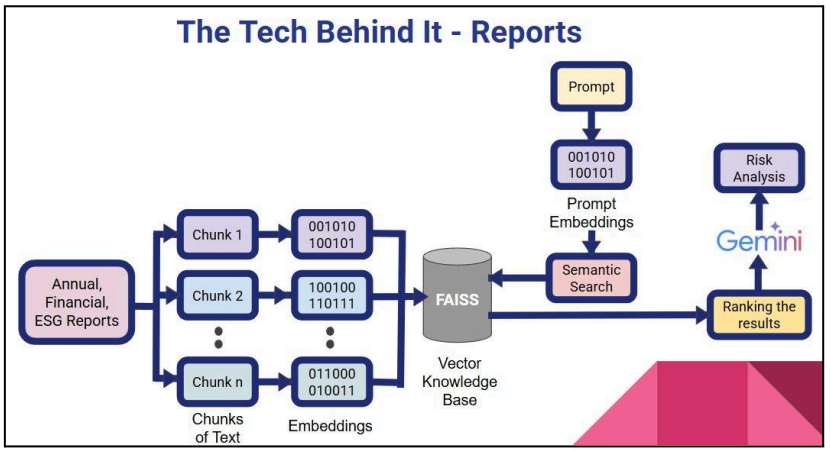


Fig4.3.4. Process flow diagram of report processing

1) Report Input:

- Accepts Annual, Financial, and ESG reports as input.
- These reports contain detailed risk-related information.

2) Chunking the Reports:

- The input document is split into smaller, manageable chunks.
- This ensures better processing and analysis by the system.

3) Text Embeddings:

- Each chunk is converted into a numerical vector (embedding).
- These embeddings capture the semantic meaning of the text.

4) FAISS Vector Knowledge Base:

- All embeddings are stored in a FAISS vector database.
- FAISS enables fast and efficient semantic search across all chunks.

5) Prompt and Semantic Search:

- A user's question or prompt is also converted into an embedding.
- FAISS performs semantic search to find the most relevant chunks based on the prompt.

6) Ranking the Results:

- The retrieved chunks are ranked in order of relevance.
- Only the top-ranked chunks are sent to the next stage.

7) Gemini-Powered Risk Analysis:

- Gemini processes the selected chunks and performs deep risk analysis.
- It identifies types of risks, explains their relevance, and suggests mitigation strategies.

8) Final Output:

- The system returns a structured risk summary based on the report.
- This makes complex reports easier to interpret for financial decision-making.

4.4 Project Scheduling & Tracking using Timeline / Gantt Chart

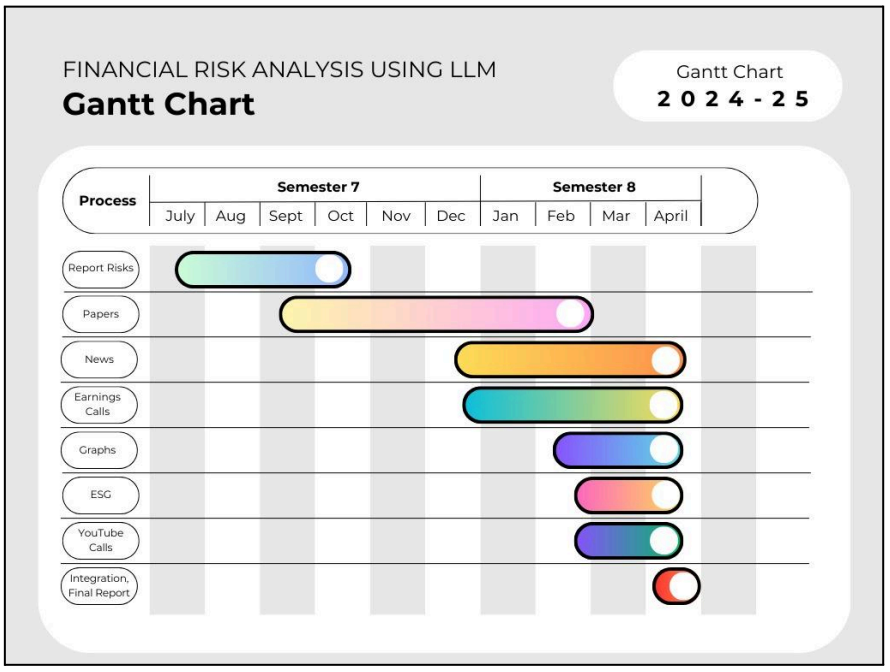


Fig4.4.1. Gantt chart of the project

This is the Gantt chart for our project. It has timelines of our project.

Chapter 5: Implementation of the Proposed System

5.1. Methodology employed for development

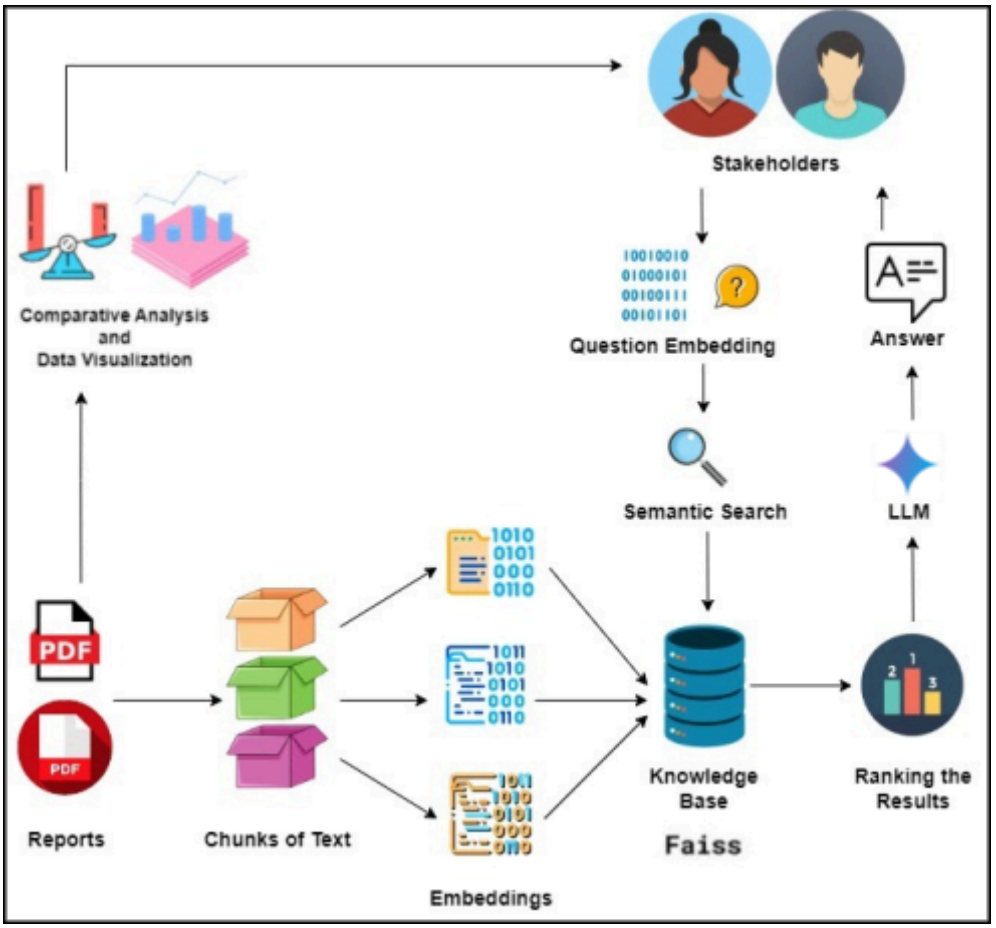


Fig 5.1. System Architecture of Risk Assessment Architecture

1. Transcription Module

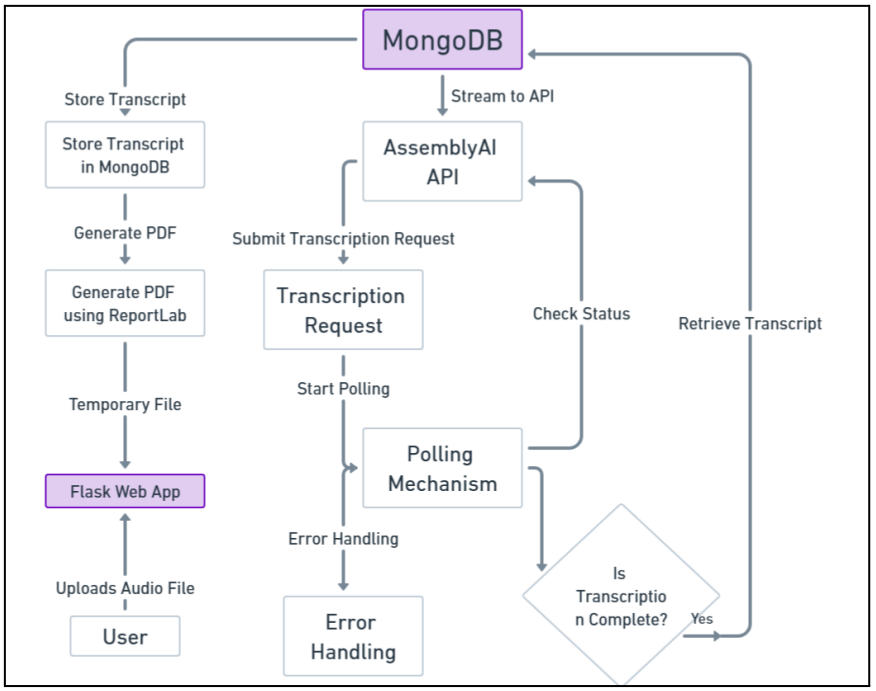


Fig 5.1.1.Transcription Data flow diagram

The transcription module begins with a user uploading an audio file via the Flask web app, which is then sent to AssemblyAI for transcription. The completed transcription is stored in MongoDB, converted to a PDF, and returned to the user.

2. Summarization

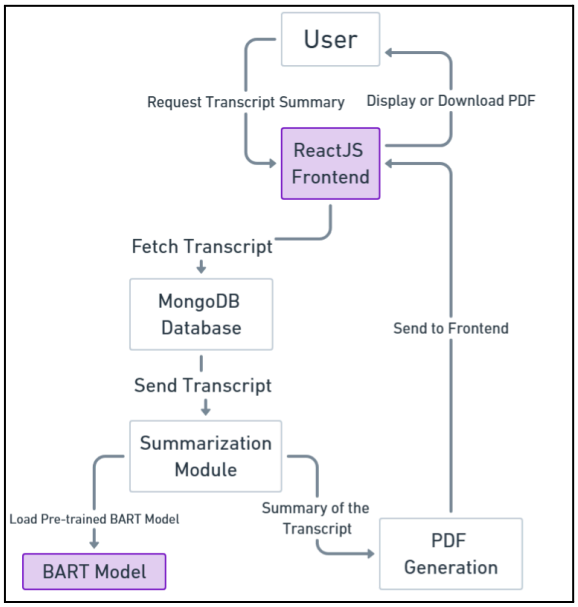


Fig 5.1.2. Summarization Data flow diagram

The summarization process begins with a ReactJS frontend requesting a transcript from MongoDB. The transcript is summarized using a BART model, then converted to a PDF, and displayed or made available for download to the user.

3. Timeline Analysis Module

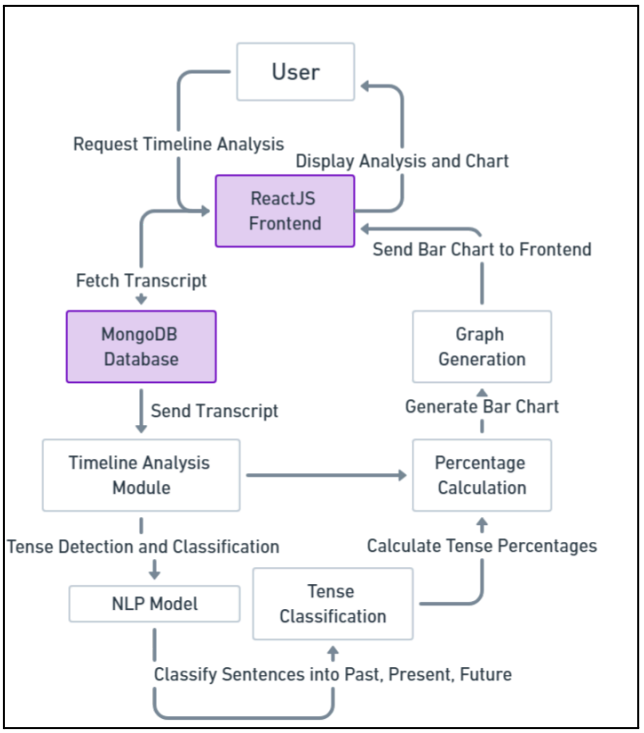


Fig 5.1.3. Timeline Analysis Data flow diagram

The timeline analysis system fetches a transcript from MongoDB via the ReactJS frontend, processes it with an NLP model for tense detection, and categorizes sentences into past, present, or future. The results are visualized as a bar chart and displayed on the frontend.

4. Risk Assessment Module

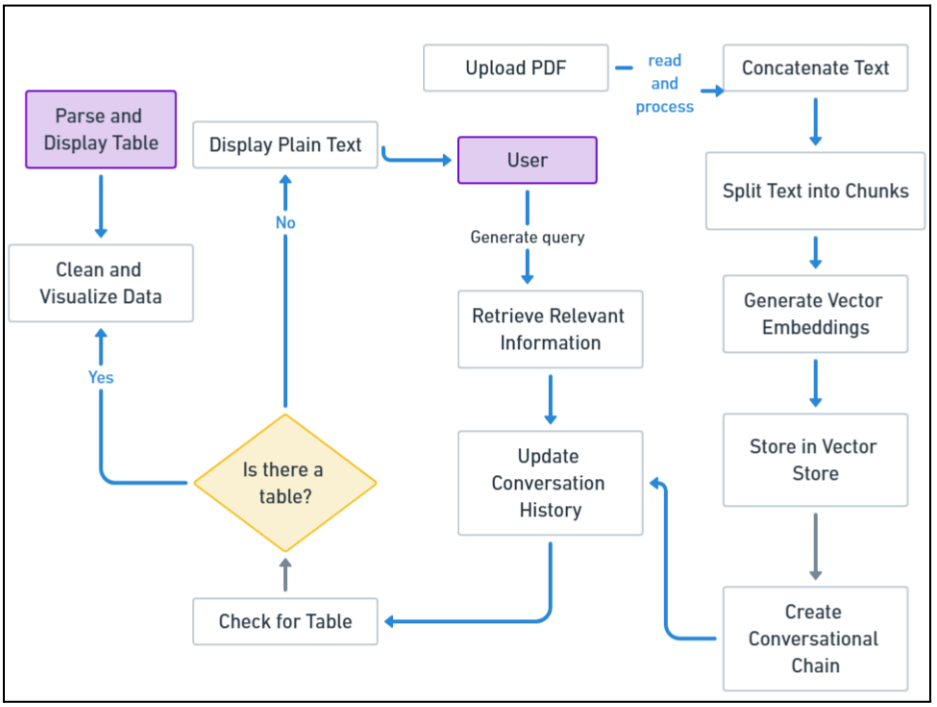


Fig 5.1.4. Risk Assessment Data flow diagram

The Risk Assessment Module begins with financial data being uploaded and preprocessed to extract key information.

Block Diagram for Transcript Generation and Summarization

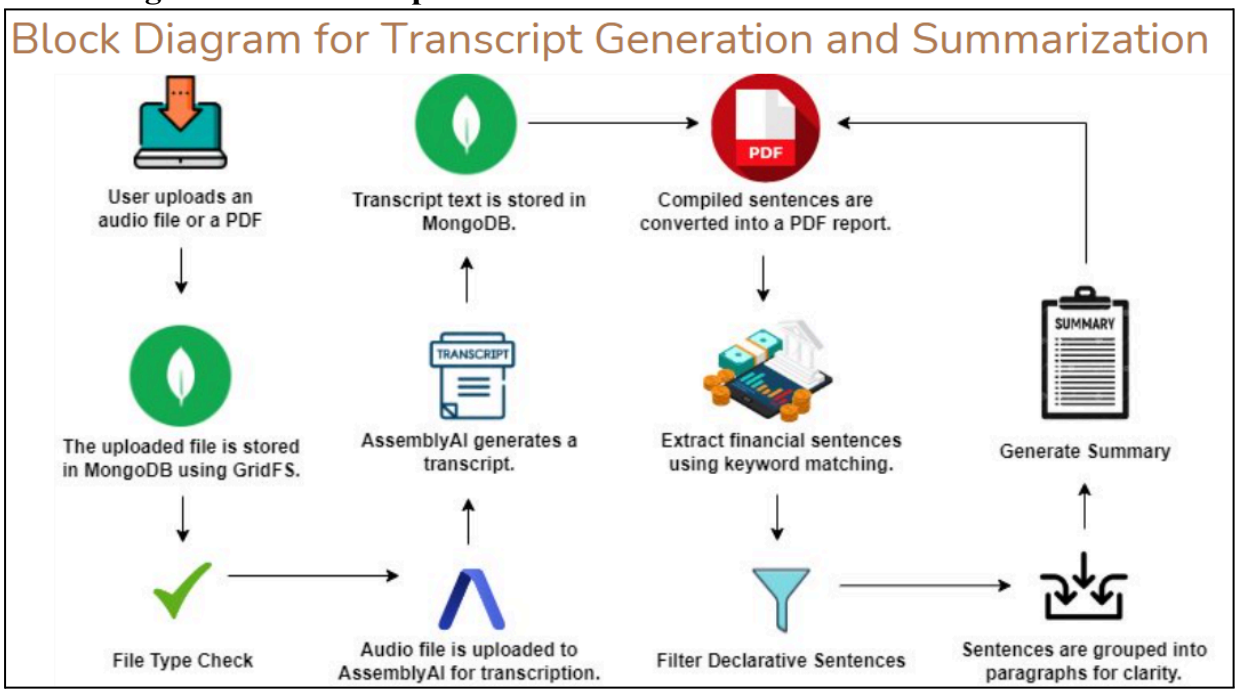


Fig 5.1.5. Block diagram for Transcript generation and Summarization

User uploads an audio or PDF file. Audio is transcribed via AssemblyAI and stored in MongoDB. Financial sentences are extracted using keywords, summarized, and compiled into a downloadable PDF.

Block Diagram for Timeline Analysis

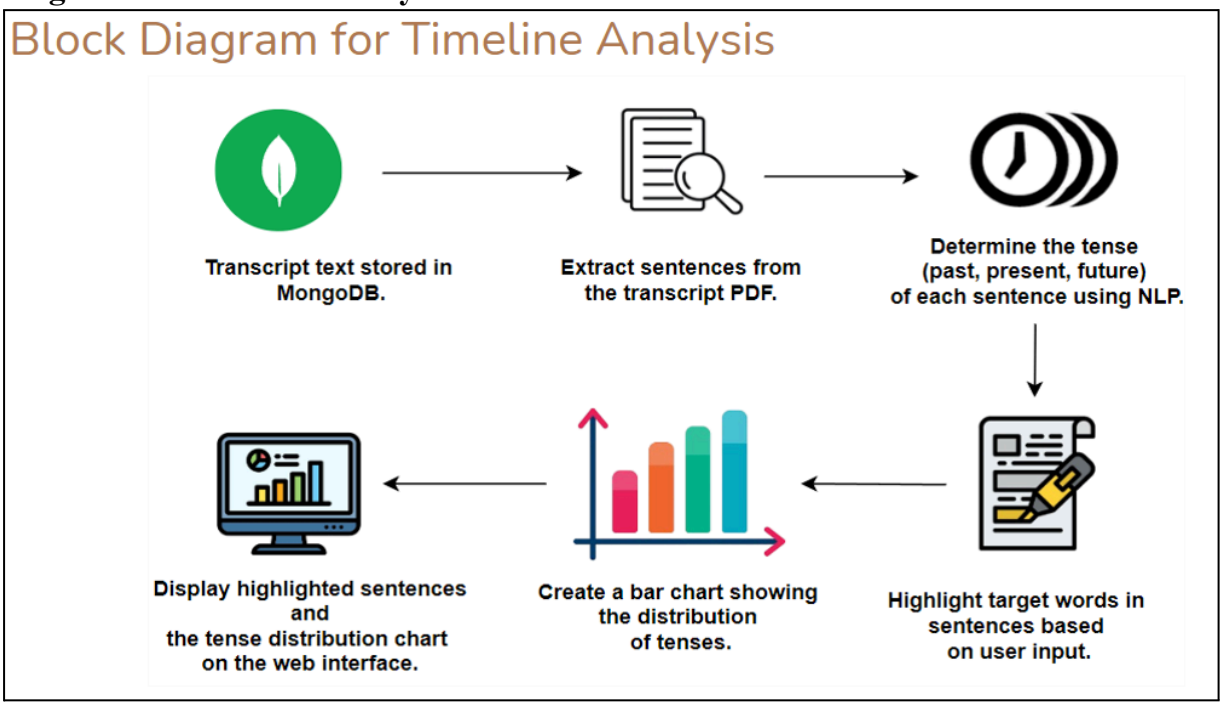


Fig 5.1.6. Block diagram for Timeline Analysis

Transcript sentences are extracted, their tense is detected using NLP, and target words are highlighted. A bar chart shows tense distribution, all displayed on the web interface.

5.2 Algorithms and flowcharts for the respective modules developed

i. Transcript Generation

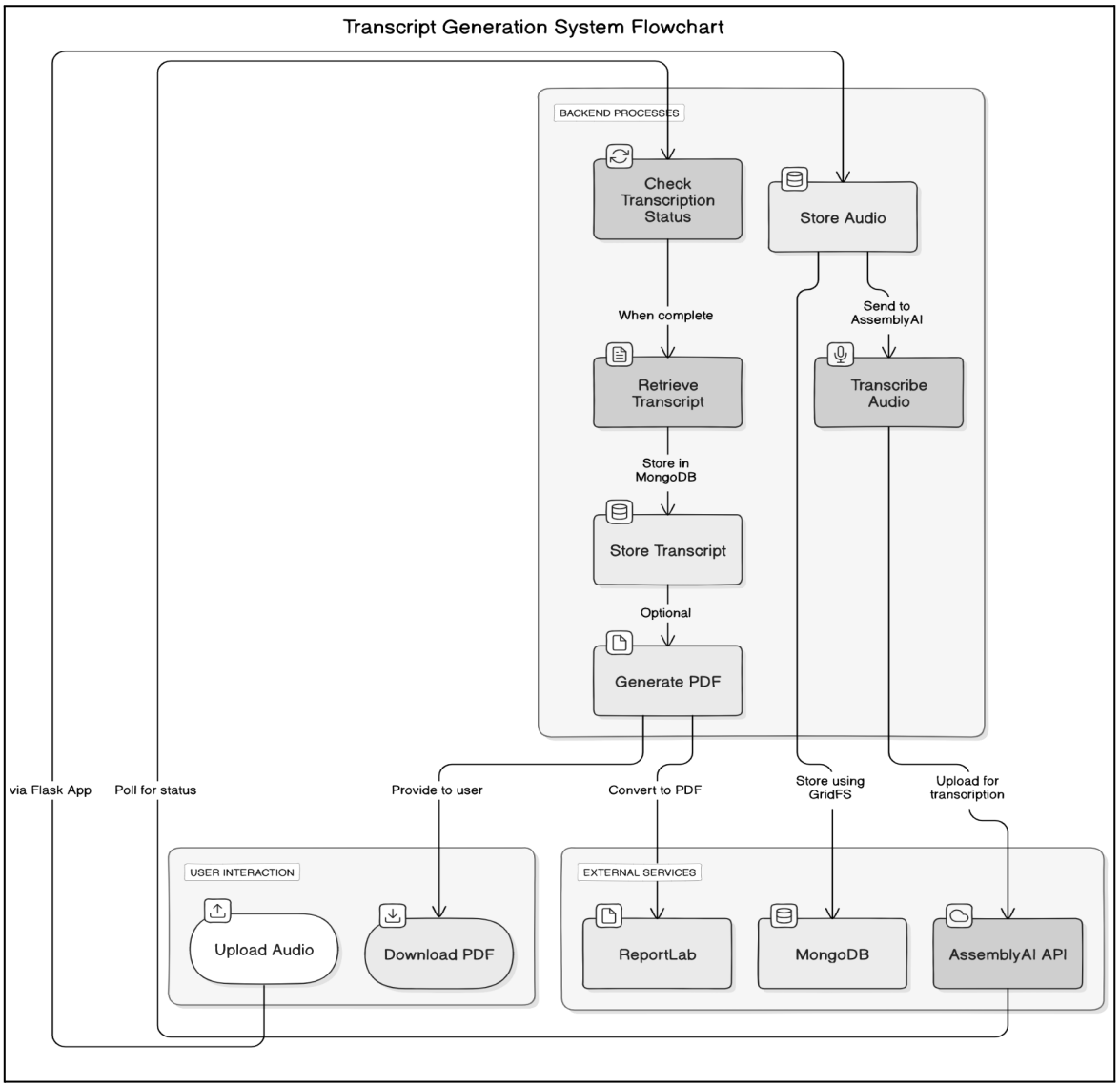


Fig 5.2.1. Flowchart of Transcript Generation

The flowchart depicts the transcript generation system's process. Users upload audio files, which are stored using GridFS and sent to the AssemblyAI API for transcription. The system monitors the transcription status and retrieves the transcript upon completion. The retrieved transcript is stored in a MongoDB database, and an optional step allows the generation of a PDF using ReportLab. Users can then download the PDF. The system operates via a Flask app, handling user interactions and polling for transcription status updates. External services, including MongoDB, ReportLab, and AssemblyAI, are integrated into the process.

ii. Transcript Summarization

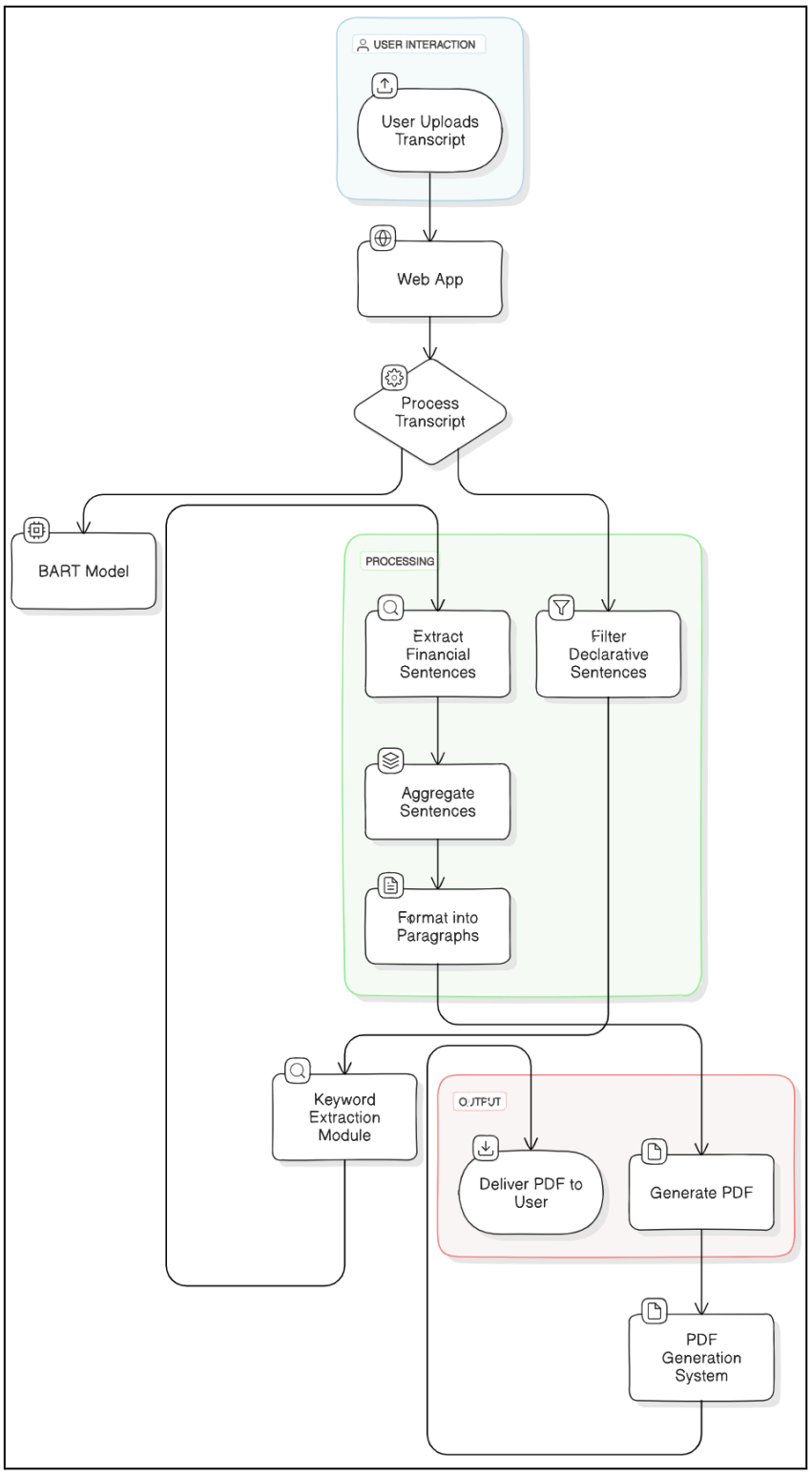


Fig 5.2.2. Flowchart of Transcript Summarization

This is the Flowchart of Transcript Summarization module. The transcript in the above process is stored in mongoDB. Then it is processed using a BART model to extract financial and declarative sentences, aggregate them into paragraphs, and extract keywords. Finally, the system generates a PDF and delivers it to the user.

iii. Timeline Analysis

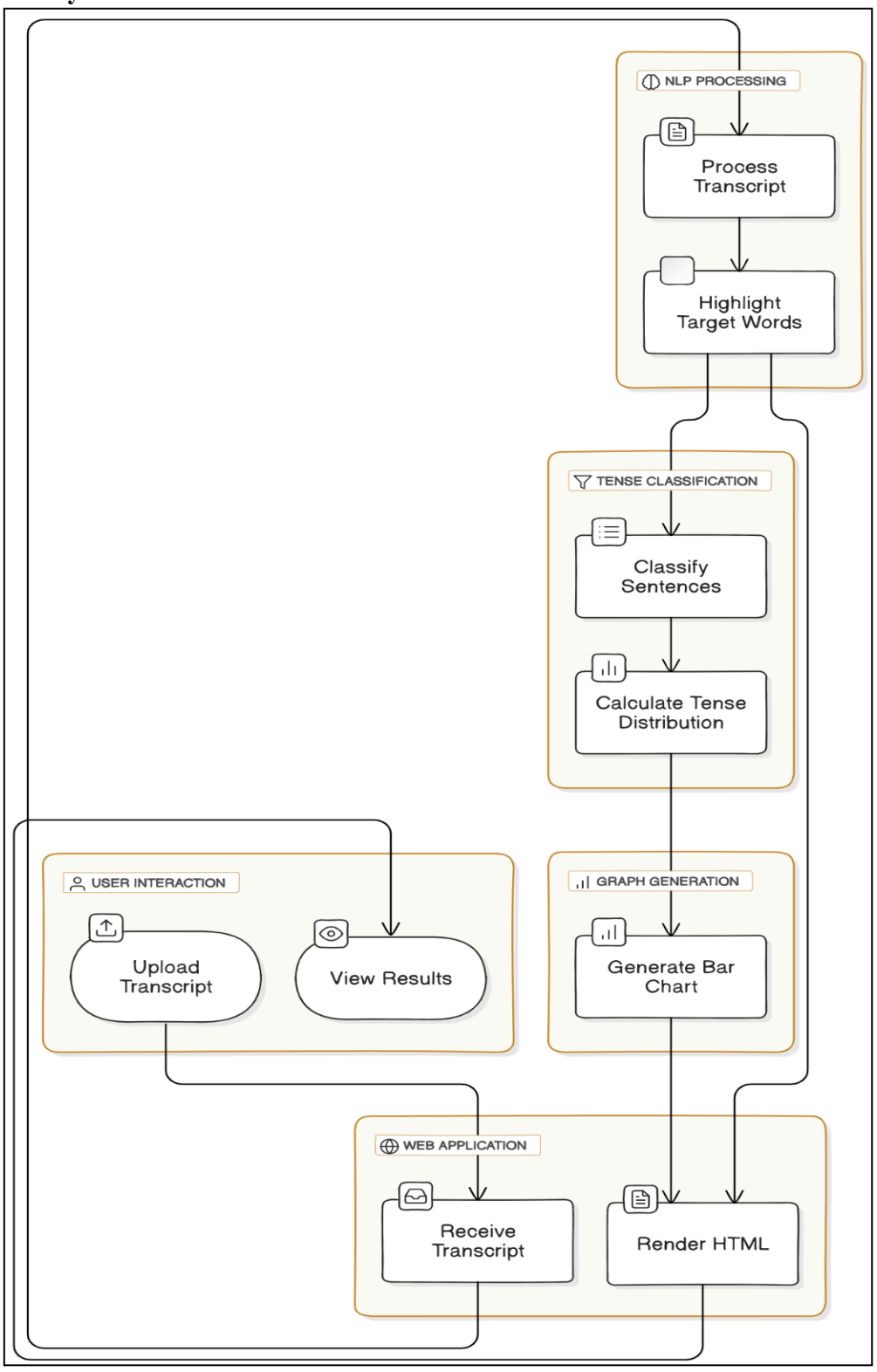


Fig 5.2.3. Flow Chart of Tense distribution

The flowchart shows a timeline analysis module where the transcript is saved in mongoDB, the system processes it to highlight words, classifies sentences by tense, calculates tense distribution, generates a bar chart, and displays the results in HTML.

iv. Risk Assessment

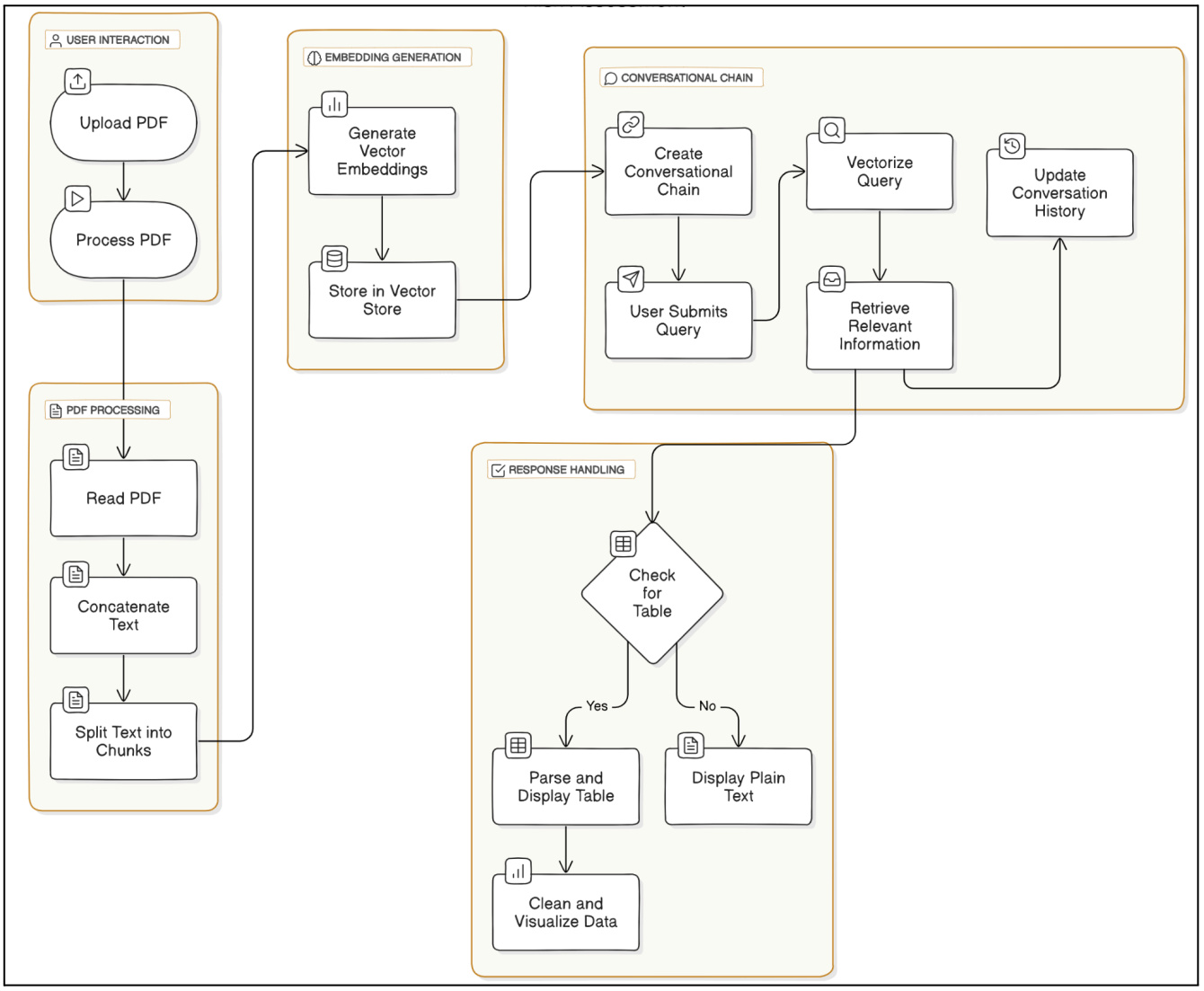


Fig 5.2.4. Flow Chart of Risk Assessment

The flowchart shows a system where a user uploads a PDF for risk assessment, which is processed into text chunks and converted into vector embeddings stored in a vector store. When the user submits a query, the system retrieves relevant information, checks if the response contains a table, and either parses and visualizes the table or displays plain text.

5.3 Datasets source and utilization

The dataset used in this project is a combination of multiple real-world financial data sources that are integrated to provide a comprehensive analysis of corporate financial risk. The major sources and their utilization are as follows:

1. Corporate Earnings Calls (Audio and Transcripts)

- Source: Official company websites and YouTube channels.
- Utilization:
    - Audio files are transcribed using AI-based speech-to-text tools.
    - Transcripts are summarized to extract key insights.

- Tone and semantic analysis is performed to detect sentiment, emotional cues, and potential financial risks.
- These insights feed into the overall risk assessment and investment prediction module.

2. News Articles and Financial Reports

- Source: Fetched dynamically using financial and news APIs.
- Utilization:
  - Articles and reports are filtered based on company name and financial year.
  - Natural Language Processing (NLP) techniques are applied for sentiment analysis.
  - News sentiment is factored into the risk score and investment outlook for the company.

3. Charts and Financial Visualizations

- Source: Generated internally from financial report data.
- Utilization:
  - Used to identify trends such as revenue growth, expenses, and profit margins.
  - Visual aids help in interpreting the risk profile of companies.
  - Supports decision-making for investors through intuitive insights.

Together, these sources enable a multi-modal approach to financial risk analysis, leveraging audio, text, and visual data to provide accurate, timely, and explainable assessments for investors.

Chapter 6: Testing of the Proposed System

6.1 . Introduction to testing

In this project, testing was carried out to evaluate the effectiveness of the LLM in extracting and identifying risk-related statements from company annual reports. Given the subjective and qualitative nature of risk analysis, a manual evaluation approach was adopted. The aim was to assess whether the generated outputs meaningfully captured potential risks and aligned with the original context from the reports.

6.2. Types of tests Considered

As the testing was entirely manual, the following evaluation type was considered:

- Manual Evaluation: Each model-generated risk statement was manually compared against the corresponding excerpt from the annual report to assess:
    a. Relevance to actual risks.
    b. Correctness of interpretation.
    c. Clarity and usefulness of the generated insight.

This approach helped ensure that the extracted information retained its contextual integrity and practical value in risk assessment.

6.3 Various test case scenarios considered

1. Annual Report Risk Testing

Output from Risk Report	Reference from Actual Annual report
Foreign Exchange Fluctuations: The report extensively details the company's exposure to foreign currency exchange rate risk, stating that fluctuations "may adversely impact the fair value of its financial instruments" and affect profit. This is a significant financial risk.	**Financial risk management** The Group is exposed primarily to fluctuations in foreign currency exchange rates, credit, liquidity and interest rate risks, which may adversely impact the fair value of its financial instruments. The Group has a risk management policy which covers risks associated with the financial assets and liabilities. The risk management policy is approved by the Board of Directors. The focus of the risk management committee is to assess the unpredictability of the financial environment and to mitigate potential adverse effects on the financial performance of the Group.
Foreign Exchange Losses: The report notes an exchange loss of ₹1,162 crore in FY2023 on foreign exchange contracts not qualifying for hedge accounting. This represents a substantial negative impact. Conversely, there was a gain of ₹109 crore in FY2024, but the overall risk remains.	Exchange gain of ₹109 crore and loss of ₹1,162 crore on foreign exchange forward, currency options and futures contracts that do not qualify for hedge accounting have been recognised in the consolidated statement of profit and loss for the years ended March 31, 2024 and 2023, respectively.
Potential Profit Impact from Exchange Rates: A 10% appreciation/depreciation of functional currencies against various foreign currencies could increase/decrease profit before taxes by approximately ₹338 crore (FY2024) and ₹713 crore (FY2023). This shows significant vulnerability to exchange rate shifts.	10% appreciation / depreciation of the respective functional currency of Tata Consultancy Services Limited and its subsidiaries with respect to various foreign currencies would result in increase / decrease in the Group's profit before taxes by approximately ₹338 crore for the year ended March 31, 2024.
Employee Turnover: 1.2% for permanent employees in FY2023-24. This is relatively low, suggesting a positive aspect in terms of employee retention. However, the data for FY2022-23 shows a significantly lower turnover rate (0.01%), which could indicate a potential negative trend.	**Turnover rate for permanent employees[a]** FY 2023-24: Male 12.5%, Female 12.5%, **Total 12.5%** FY 2022-23: Male 20.2%, Female 20.1%, **Total 20.2%** FY 2021-22: Male 17.3%, Female 17.7%, **Total 17.4%** Permanent Employees

Table 6.3.1 Annual report risk testing

2. ESG Risk Analysis Testing

ESG Report Output	References from actual annual report
1. Environmental Performance: Carbon Footprint: Accenture has committed to achieving net-zero emissions by 2025, focusing on reducing Scope 1, 2, and 3 emissions and offsetting remaining emissions through nature-based solutions. They aim for 100% renewable electricity by 2023 and have operated 85% in FY2022.	**Net-Zero Emissions by 2025** To meet these commitments, we set a goal to achieve net-zero emissions by 2025 by first focusing on actual reductions across our Scope 1, 2 and 3 emissions and then removing any remaining emissions through nature-based carbon removal offsets.
Labor Practices: Accenture highlights its commitment to employee well-being, offering programs for physical, mental, and financial health. An 85% employee satisfaction rate regarding flexible work arrangements is reported. However, the document lacks detail on fair wages, working conditions, and employee rights.	**Our Health, Safety and Well-Being** We are committed to creating a place where people can be successful both professionally and personally. We take a holistic view of well-being—including physical, mental, emotional and financial well-being—providing specially defined programs and practices to meet our people's fundamental human needs. During fiscal 2022, our people have embraced omni-connected ways of working. According to a survey, 85% of our global respondents feel empowered to work flexibly within their teams.

Table 6.3.2 ESG Risk Analysis Testing

6.4. Inference drawn from the test cases

Based on manual testing across diverse report excerpts, the following inferences were drawn:

- The system was generally successful in identifying explicitly mentioned risks.
- Performance was slightly less accurate for implicit or indirect risk references, where contextual understanding is deeper.
- False positives (non-risk statements flagged as risks) were minimal, indicating good baseline prompt quality.
- In a few cases, multi-factor risks were partially captured, suggesting scope for refining how compound sentences are parsed and interpreted.
- Overall, manual testing confirmed that the system provides meaningful and contextually accurate risk summaries, with room for improvement in nuanced cases.

Chapter 7: Results and Discussion

7.1. Screenshots of User Interface (UI) for the respective module

1. Earnings Call Analysis Module

Fig.7.1.1 Landing page of Earnings Call module

This is the home page of Earnings Call Analyzer module

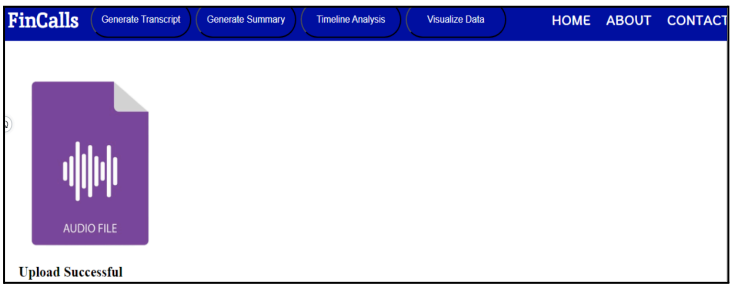


Fig. 7.1.2 User successfully uploaded the earnings call

The earnings call was successfully uploaded by the user.

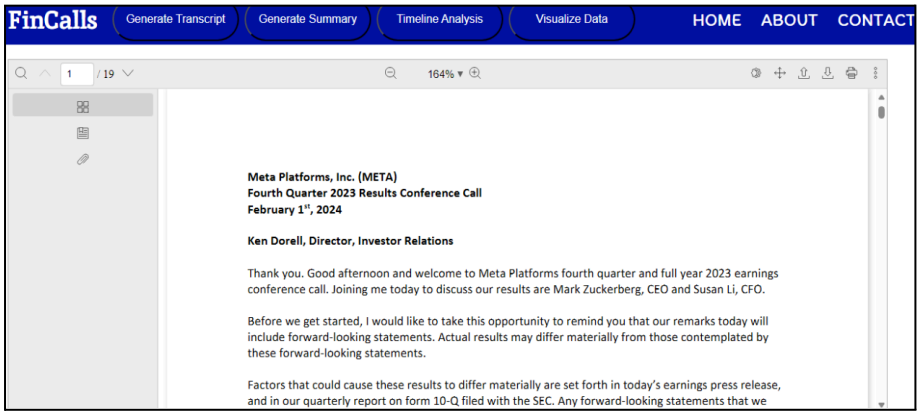


Fig. 7.1.3 Transcript generation result

The transcript of the Earnings call was generated (19 pages)

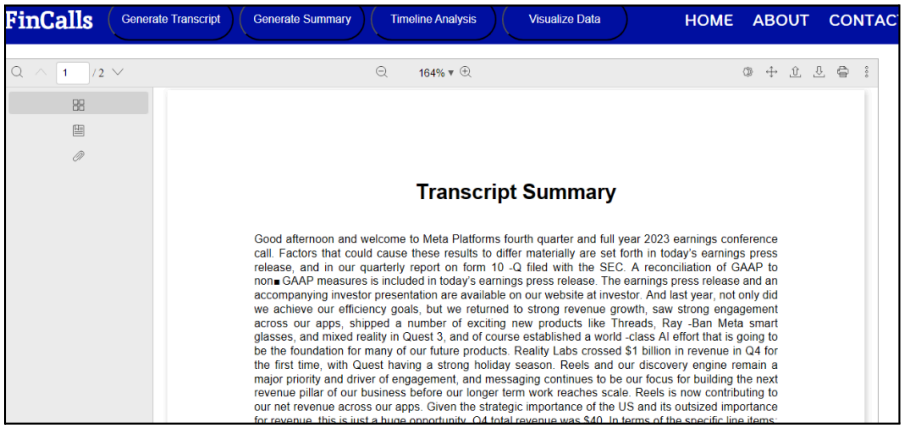


Fig 7.1.4 Transcript summary generated

Here, the summary of transcript is generated (2 pages)

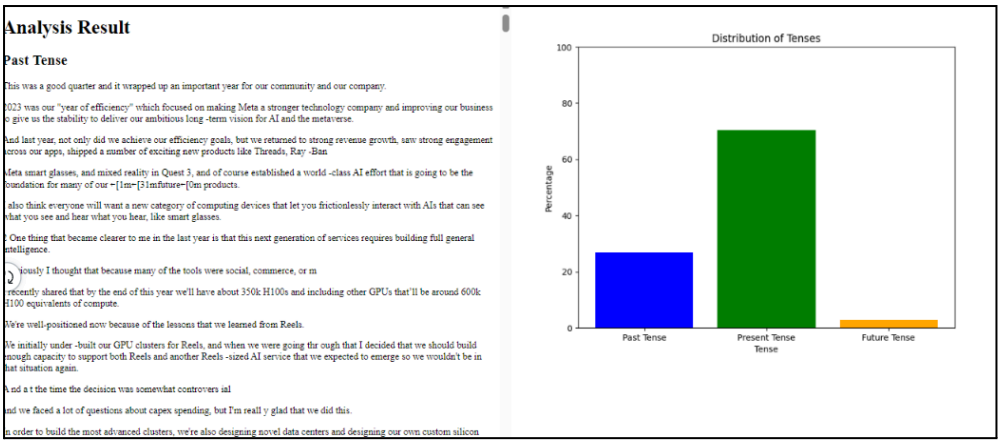


Fig 7.1.5 Timeline Analysis generated

The timeline analysis of the Earnings call generated, where the user can study how much of the whole call was discussed about the past events, the present condition and the future plannings.

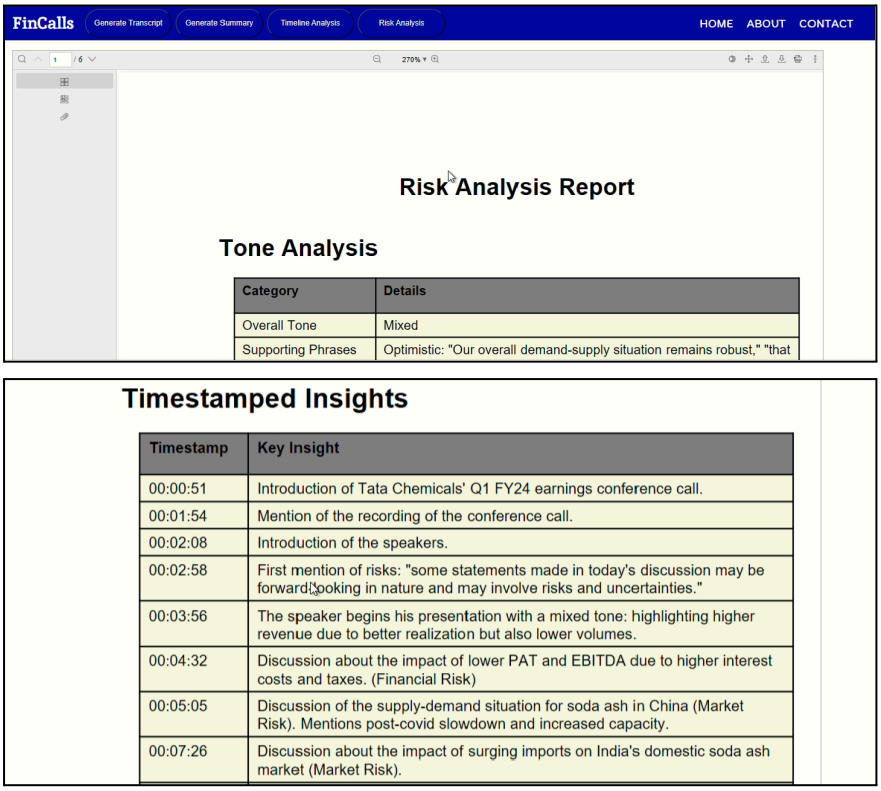


Fig 7.1.6 Earning Calls Risk Assessment

We have done tone analysis, timestamped based insights, and potential risks

- **Annual Report Risk Assessment**
  The input taken here is a sample document which contains some information about the company.

XYZ Corporation Financial Overview

In the fiscal year ending December 31, 2023, XYZ Corporation reported a total revenue of $1.2 billion, reflecting a modest increase of 5% compared to the previous year. However, the company's net profit experienced a significant decline, dropping by 15% to $150 million. This downturn is attributed to rising operational costs and increased competition in the market.

XYZ Corporation has implemented several strategic initiatives to address these challenges. The company has invested heavily in technology upgrades, allocating $100 million towards enhancing its digital infrastructure and improving operational efficiencies. This strategic pivot aims to streamline processes and reduce costs in the long run.

Despite these efforts, the company acknowledges the inherent risks associated with its expansion strategy. The volatility in raw material prices poses a threat to profit margins, as indicated by a 10% increase in material costs over the past year. Additionally, the ongoing geopolitical tensions in key markets have raised concerns regarding supply chain stability, potentially impacting product availability and pricing.

Looking forward, XYZ Corporation remains committed to diversifying its product offerings and exploring new market opportunities. The management team is optimistic about achieving a revenue target of $1.5 billion by the end of 2024. However, they also caution that achieving this goal will require navigating significant market uncertainties, including regulatory changes and economic fluctuations.

In summary, while XYZ Corporation has laid out a comprehensive strategy to enhance profitability and growth, the company remains vigilant of the risks that could affect its financial health and operational success.

Fig. 7.1.7 Input Document of Risk Assessment module

The input here is a sample document that provides details about the company.

Statements Depicting Risk:

- "The company's net profit experienced a significant decline, dropping by 15% to $150 million."
- "This downturn is attributed to rising operational costs and increased competition in the market."
- "The volatility in raw material prices poses a threat to profit margins, as indicated by a 10% increase in material costs over the past year."
- "Additionally, the ongoing geopolitical tensions in key markets have raised concerns regarding supply chain stability, potentially impacting product availability and pricing."
- "Despite these efforts, the company acknowledges the inherent risks associated with its expansion strategy."
- "However, they also caution that achieving this goal will require navigating significant market uncertainties, including regulatory changes and economic fluctuations."

Fig. 7.1.8 Statements Depicting Risk

Here we extract the statements that depict risk for the stakeholders.

Negative Impact Statements:

- "The company's net profit experienced a significant decline, dropping by 15% to $150 million."
- "Rising operational costs"
- "Increased competition in the market"
- "10% increase in material costs over the past year"
- "Concerns regarding supply chain stability, potentially impacting product availability and pricing"
- "Significant market uncertainties, including regulatory changes and economic fluctuations"

Fig. 7.1.9 Negative Impact Statements

The sentences that show negative impact for the stakeholders.

- **Negative:** Net profit decreased by 15% to $150 million.
- **Negative:** Material costs increased by 10% over the past year.
- **Positive:** Revenue increased by 5% to $1.2 billion.
- **Positive:** Investment in technology upgrades: $100 million.
- **Target:** Revenue target of $1.5 billion by the end of 2024.

Fig. 7.1.10 Quantifiable data classified into positive and negative trends

We classify the quantifiable data into positive and negative trends.

Assessment Summary:

XYZ Corporation faces several significant risks that could impact its financial health and operational success.

- **Profitability Decline:** The company experienced a substantial decline in net profit, attributed to rising operational costs and increased competition. This trend could continue if not addressed effectively.
- **Raw Material Price Volatility:** Fluctuations in raw material prices pose a threat to profit margins, as evidenced by the 10% increase in material costs. This could further impact profitability if not mitigated.
- **Supply Chain Disruptions:** Geopolitical tensions and potential supply chain disruptions could negatively impact product availability and pricing, affecting both revenue and customer satisfaction.
- **Market Uncertainties:** Regulatory changes and economic fluctuations add to the existing challenges, requiring careful navigation and strategic adjustments to achieve the targeted revenue growth.

Fig. 7.1.11 Risk Assessment of all the extracted data

We get the risk assessment of all the data that is extracted.

Risk Mitigation Suggestions:

- **Cost Optimization:** Implement cost reduction strategies across all operations to address the rising operational costs. This could include streamlining processes, negotiating better deals with suppliers, and exploring alternative sourcing options.
- **Competitive Differentiation:** Develop a strong competitive strategy to address the increasing market competition. This could involve product innovation, targeted marketing campaigns, and enhancing customer service.
- **Raw Material Hedging:** Explore hedging strategies to mitigate the impact of raw material price volatility. This could involve using forward contracts or other financial instruments to lock in prices.
- **Supply Chain Diversification:** Diversify supply chains to reduce reliance on specific regions and minimize the impact of geopolitical tensions. This could involve sourcing materials from multiple locations or establishing alternative manufacturing facilities.
- **Strategic Market Analysis:** Continuously monitor regulatory changes and economic fluctuations to proactively adapt the business strategy. This could involve conducting regular market research, engaging with industry experts, and developing contingency plans.

By actively addressing these risks and implementing mitigation strategies, XYZ Corporation can improve its resilience and enhance its chances of achieving its ambitious growth targets.

Fig. 7.1.12 Risk Mitigation Suggestions

Suggestions about Risk Mitigation from the document given as input is provided.

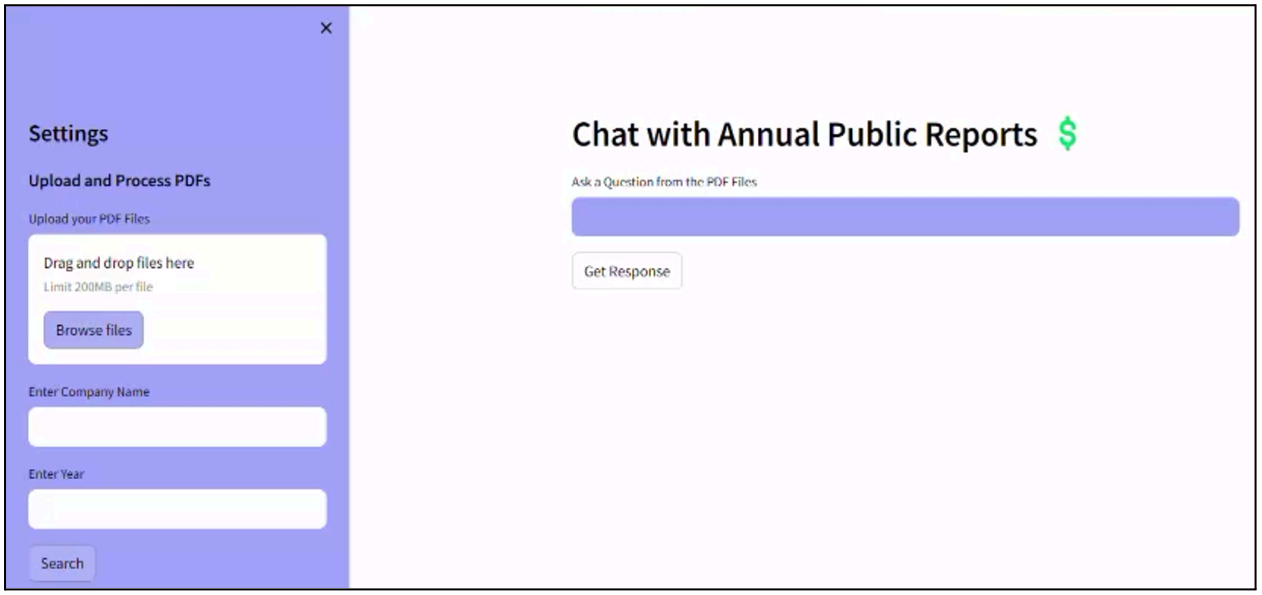


Fig. 7.1.13 ChatBot Home Page

This is the home page of the Chatbot.

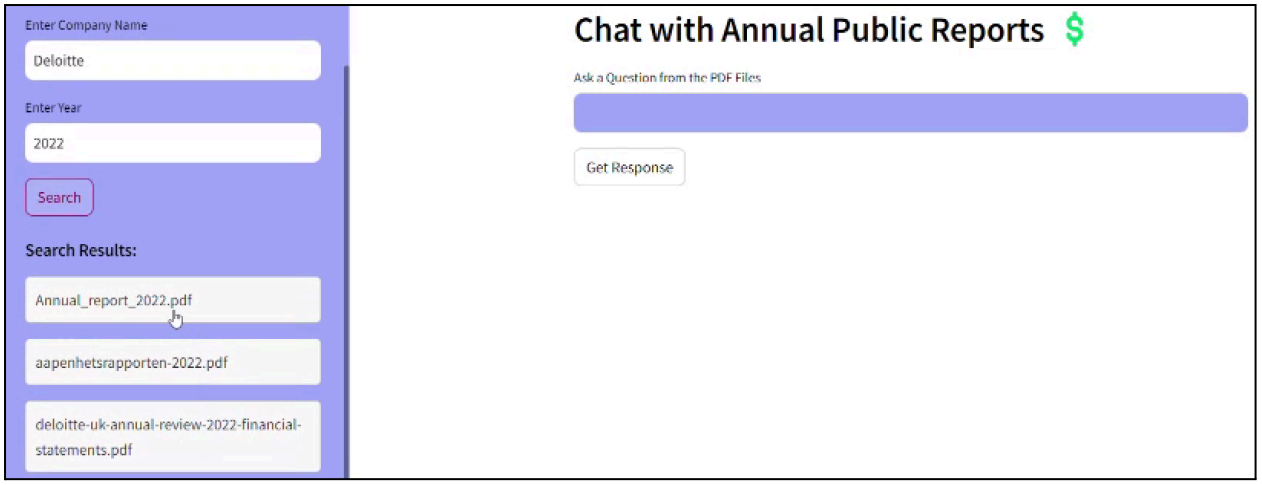


Fig. 7.1.14 Search company name and year

We get the results after entering the name of any company and particular year (Deloitte and 2022 in this case), that is all the documents available over the net related to Deloitte and 2022.

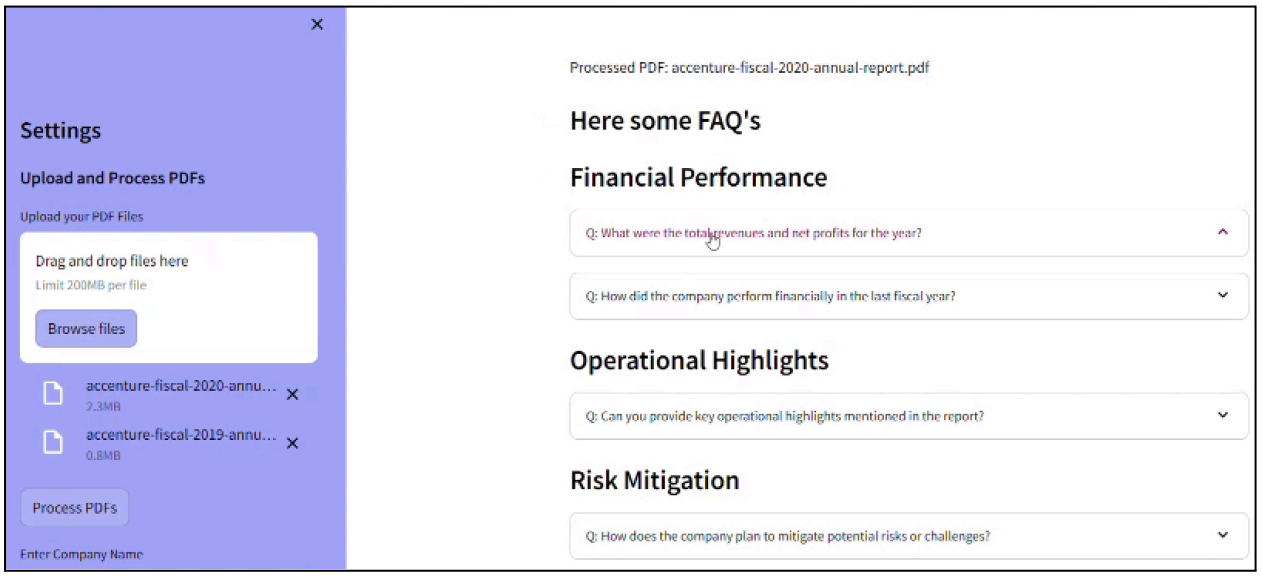


Fig. 7.1.15 FAQs generated

FAQs generated category wise with answers taken from the 2 uploaded PDFs

Fig. 7.1.16 Answers of the FAQs

Few answers retrieved from the two input documents in FAQ section

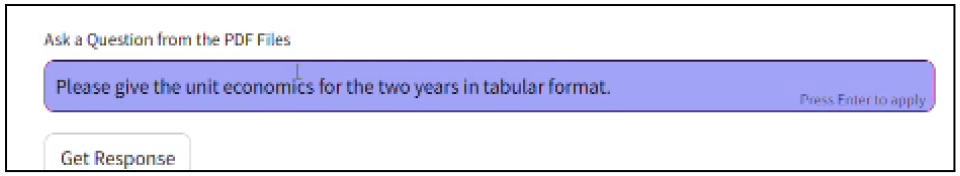


Fig. 7.1.17 Question entered by the user

The user can enter any query in the query box.

Fig.7.1.18 Answer to the given question

The chatbot will get the asked data from the pdf and represent the data in tabular form with visual representation (graphs), if visualization is possible.

- **News Risk Report**

Risk & Opportunity Report

Risk Analysis Table

Risk Category	Summary	Potential Impact	Likelihood	Mitigation Strategy
Risk Category	Summary	Potential Impact	Likelihood	Mitigation Strategy
Financial	Insider selling of a significant number of shares by CEO and SVP may signal negative outlook or potential financial difficulties.	Medium	Medium	Conduct thorough internal review of the company's financial health and communicate transparently with investors.
Reputational	Negative media coverage related to insider selling could damage Apple's reputation and investor confidence.	Medium	Medium	Proactive communication strategy addressing the insider selling, emphasizing positive aspects of the company's financial health and future prospects.
Financial	Analyst downgrades of price targets could negatively impact Apple's stock price and valuation.	Medium	Medium	Closely monitor analyst reports and engage with analysts to address concerns and clarify company strategies.
Regulatory	Potential future antitrust investigations similar to those faced by other tech companies could lead to fines or changes in business practices.	High	Medium	Proactively comply with all existing regulations and implement robust compliance programs to minimize the risk of future regulatory actions.

Fig. 7.1.19 News Risk and Opportunity Report

This table will show potential risk, its impact and mitigation strategy.

Strength & Opportunity Matrix

Category	Positive Indicator	Strategic Impact
Category	Positive Indicator	Strategic Impact
Financial Performance	EPS exceeded consensus estimate by $0.04	Strong financial performance indicates efficient operations and market demand, leading to increased profitability and investor confidence.
Financial Performance	Return on equity of 160.83%	Exceptional return on equity signifies high profitability relative to shareholder investment, enhancing shareholder value and attracting investors.
Financial Performance	Net margin of 24.30%	High net margin indicates strong pricing power and cost control, contributing to greater profitability and financial stability.
Financial Performance	Analysts anticipate 7.28 EPS for the current year	Positive earnings forecast suggests continued strong financial performance and growth prospects, boosting investor confidence.
Financial Performance	Quarterly dividend of $0.25	Dividend payments demonstrate financial strength and commitment to shareholder returns, enhancing investor appeal.
Financial Performance	Market cap of $2.96 trillion	Large market capitalization signifies strong brand recognition, market dominance, and significant financial resources, enhancing stability and growth potential.

Fig. 7.1.20 News Strength and Opportunity Matrix

This table will show potential risk, its positive indicator and mitigation strategy.

- **Risk assessment of graphs from annual public report**

Risk Analysis Table

Risk Category	Summary	Potential Impact	Likelihood	Mitigation Strategy
Financial Risk (Credit Risk)	Increased loan defaults across client segments (particularly noticeable if economic downturn occurs).	High	Medium	Diversify loan portfolio, strengthen credit scoring models, implement stricter lending criteria, increase reserves for potential loan losses.
Market Risk (Economic Downturn)	A significant economic recession could severely reduce demand for credit and capital, impacting overall loan volume and profitability.	High	Medium	Develop robust contingency plans for economic downturns, explore alternative revenue streams, strengthen relationships with key clients to ensure continued business.
Operational Risk (Data Security)	Potential for data breaches impacting customer information and financial data.	Medium	Medium	Invest heavily in cybersecurity infrastructure, employee training, regular security audits, and robust incident response plan.

Fig. 7.1.21 Annual report Risk Analysis Matrix

Positive Indicators Table

Indicator	Value	Strategic Impact
Overall Growth in Credit and Capital	Consistent growth from ~$1.9B in 2005 to ~$3.18B in 2023	Demonstrates strong market position and ability to attract clients. Supports expansion and increased profitability.
Diversified Client Base	Significant lending across corporate, small business, and consumer segments.	Reduces risk exposure; strong diversification across different market sectors.
Increase in Consumer Lending	Growth in the consumer segment from 2019-2023.	Indicates success in expanding into new market segments.

Fig. 7.1.22 Annual Reports Strength and Opportunity Matrix

Negative Indicators Table

Indicator	Value	Strategic Impact
Volatility in Certain Sectors	Fluctuations in lending to small businesses and commercial clients between years.	Indicates sensitivity to economic cycles. Requires improved risk management in these segments.
Potential for Concentration Risk	Significant proportion of lending may be concentrated in certain geographic locations or industries. (Data not provided, but a potential risk identified from chart trend)	Increased risk exposure to local economic downturns or sector-specific challenges. Diversification efforts are necessary.
Limited Data Pre-2019 on Government Lending	Lack of detailed breakdown of government, government-related, and nonprofit lending before 2019.	Prevents full assessment of historical trends and opportunities in this segment.

Fig. 7.1.23 Annual Report Negative Indicator Matrix

- **ESG risk assessment**

Risk Analysis Table

Risk Category	Summary of Risk	Potential Impact	Likelihood (Low/Medium/High)	Mitigation Strategy
Environmental: Greenhouse Gas Emissions	The report mentions Scope 1, 2, and 3 GHG emissions but provides no specific data on their magnitude or reduction targets.	Reputational damage, regulatory fines, increased operational costs, stranded assets.	Medium	Requires further investigation into the company's GHG emissions data, setting reduction targets, and implementing mitigation strategies.
Environmental: Water Usage	The report mentions water withdrawal but lacks specific data on usage and impact.	Water scarcity risks, negative impact on local communities, regulatory non-compliance.	Medium	Needs detailed information on water usage, sourcing, and management practices. Implementation of water conservation measures is necessary.
Environmental: Waste Disposal	The report mentions waste diverted from disposal but lacks quantitative data.	Environmental pollution, regulatory fines, reputational damage.	Medium	Requires detailed data on waste generation, recycling rates, and waste management strategies. Improved waste reduction and recycling programs are needed.
Environmental: Biodiversity	The report mentions significant impacts of activities on biodiversity but lacks specific details.	Loss of biodiversity, reputational damage, regulatory non-compliance.	Medium	Requires a comprehensive assessment of the company's impact on biodiversity and implementation of conservation measures.
Social: Wage Equality	The report mentions a ratio of basic salary and remuneration of women to men but doesn't provide the actual ratio.	Reputational damage, legal challenges, decreased employee morale and productivity.	Medium	Requires disclosure of the actual gender pay gap and implementation of strategies to address any inequalities.

Fig. 7.1.24 ESG Report Risk Analysis Matrix

Risk Category	Summary of Risk	Potential Impact	Likelihood (Low/Medium/High)	Mitigation Strategy
Social: Supplier and Vendor Labor Practices	The report mentions forced or compulsory labor risks and the existence of a Supplier Code of Conduct, but lacks specific details on monitoring and enforcement.	Reputational damage, legal challenges, human rights violations.	Medium	Requires robust due diligence processes to ensure compliance with labor standards throughout the supply chain. Regular audits and transparent reporting are crucial.
Governance: Transparent Communications	The report mentions fair and transparent communications as important but doesn't provide specific examples or metrics.	Loss of investor confidence, reputational damage, regulatory non-compliance.	Medium	Requires clear and consistent communication of ESG performance and risks to stakeholders. Independent verification of reported data is beneficial.
Governance: ESG Disclosures and Reporting	The report mentions mandatory regulatory reporting but doesn't specify the extent of voluntary disclosures.	Loss of investor confidence, reputational damage, regulatory non-compliance.	Medium	Requires comprehensive and transparent ESG reporting aligned with leading frameworks (e.g., GRI, SASB). Third-party assurance of reported data is recommended.

Positive Indicators Table

Positive Factor	Current Status	Strategic Impact
Board Diversity	Mentioned in the report (2024 Proxy Statement, p. 10), but specific data is missing.	Enhanced decision-making, improved reputation, better alignment with stakeholder expectations.
Anti-Corruption Policies and Procedures	Employees are required to complete anti-corruption training.	Reduced risk of corruption, improved ethical conduct, enhanced reputation.
ESG Integration in Investment Management	The report mentions the incorporation of ESG factors in investment management and advisory, but lacks specific details on the scope and effectiveness.	Improved investment performance, alignment with investor preferences, positive environmental and social impact.

Fig. 7.1.25 ESG Report Positive Indicators Matrix

Negative Factor	Current Status	Strategic Impact
Lack of Quantitative ESG Data	The report frequently mentions ESG topics but lacks specific quantitative data on key metrics.	Difficulty in assessing material risks and opportunities, limited transparency, hindering effective stakeholder engagement.
Limited Information on Mitigation Strategies	The report mentions several risks but provides limited detail on specific mitigation strategies.	Increased vulnerability to ESG-related risks, potential for reputational damage and financial losses.
Absence of Specific Targets and Timelines	The report lacks specific targets and timelines for achieving ESG goals.	Difficulty in measuring progress, lack of accountability, hindering effective management of ESG risks and opportunities.

Fig. 7.1.26 ESG Report Negative Indicators Matrix

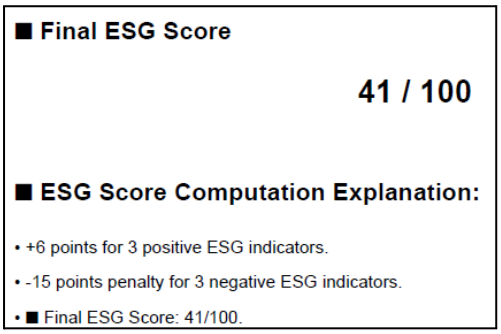


Fig. 7.1.27 ESG Report Score

7.2. Performance Evaluation measures

Metric	Description
Relevance	Whether extracted risk statements are genuinely related to risk.
Context Accuracy	Whether the risk statement retains the meaning from the original text.
False Positives Rate	Frequency of non-risk statements being incorrectly tagged.
Coverage	Proportion of actual risks in the document that were successfully identified.
User Satisfaction	Manual validation feedback and expert scoring (1–5 scale).

Table 7.2.1 Performance Evaluation Measures

7.3. Input Parameters / Features considered

Module	Input Format	Key Features Extracted
Annual Reports	. pdf/search name of company	- Risk section parsing (MD&A, Risk Factors) - Statements depicting risks - Financial risk terms
Earnings Calls	Transcript .pdf, audio (.mp3, avi)	- Tone analysis - Supporting risk statements - Timestamp based analysis

News Articles	Archived/live textual content	- Sentiment polarity - Company & entity mentions - Crisis event patterns
ESG Reports	PDF ESG disclosures, annual report (.pdf)	- ESG dimension classification - Keywords like “emissions”, “inclusion”, “governance”
Charts/Graphs	Image files (PNG/JPG/PDF)	- OCR + LLM summary - Visual trend signal extraction - Anomalies as indicators

Table 7.3.1 Input Parameters / Features considered

7.4. Comparison of results with existing systems

Parameter	Fincalls - Risk Analyzer	Legacy/Manual Methods
Context-Aware Extraction	Yes	No
Multi-Source Aggregation	Yes	No
Custom Risk Dimensions	Defined (6 Key Axes)	Generic/Manual Tagging
Efficiency	Automated Summarization	Time-consuming
Real-time Usability	LLM-powered instant outputs	Manual consolidation needed

Table 7.4.1 Comparison of results with existing systems

7.5. Inference drawn

1. The system effectively streamlines multi-source document analysis into a single cohesive Risk Report.
2. Manual testing confirms the model identifies highly relevant risks with semantic accuracy.
3. Investors and analysts can gain a 360° view of risk without navigating multiple documents.
4. Future scope includes integration of automated benchmarking, risk scoring, and real-time alerting.

Chapter 8: Conclusion

8.1 Limitations

1) Dependency on Pre-trained Models: The system relies heavily on Gemini-1.5's performance. Any inaccuracies in the model's understanding or summarization can affect the quality of risk analysis.

2) Lack of Real-time Data Sync: While the system fetches stock and news data, it may not reflect instantaneous market changes, which can be critical for time-sensitive financial decisions.

3) Audio/Video Quality Constraints: Earnings call audio and YouTube video transcripts may lose accuracy if the source has noise, overlapping speakers, or unclear speech.

4) Limited Multilingual Support: The current setup is optimized for English. Reports or calls in other languages may not be effectively processed or analyzed.

5) Static Risk Categorization: Risk categories and mitigation strategies are based on predefined patterns. This may limit adaptability to new, emerging risk types.

6) Chart and Graph Analysis Limitations: Chart extraction from PDFs may fail if visuals are too complex or non-standard in formatting, limiting the effectiveness of automated graph risk analysis.

7) Data Privacy Concerns: Sensitive documents like ESG and financial reports may raise data security concerns, especially when processed on third-party APIs or cloud platforms.

8.2 Conclusion

This project demonstrates the potential of using cutting-edge large language models like Gemini-1.5 to automate and streamline financial risk analysis from diverse data sources such as earnings calls (audio/video), financial reports, real-time news, stock data, and ESG reports.

By integrating natural language understanding, semantic search, and vector-based document retrieval, the system:

- Provides a holistic view of a company's risk profile
- Automated insights that typically require manual financial analysis
- Enhances decision-making for stakeholders such as analysts, investors, and company executives

The modular architecture allows scalable extension for different types of financial documents and dynamic market inputs, creating a powerful, AI-assisted risk analysis platform.

8.3 Future Scope

1) Cross-Company Risk Benchmarking: Compare risk profiles of multiple companies side-by-side using common KPIs and generate comparative summaries (e.g., L'Oréal vs. Nivea risk trends).

2) Automated Investment Recommendation System: Train a module using historical risk trends and stock performance to generate buy/sell/hold suggestions based on extracted risk summaries and tone.

3) Knowledge Graph Generation: Automatically create risk-centric knowledge graphs connecting entities (companies, risk types, mitigation strategies) to visualize relationships.

4) Risk Impact Simulator: Allow users to simulate hypothetical scenarios (e.g., increase in regulatory risk) and predict potential outcomes on financial health or stock price.

References

[1] Vivien E. Jancenelle Susan Storrud-barnes Rajshekhar Javalgi , (2017),"Corporate Entrepreneurship and Market Performance: A Content Analysis of Earnings Conference Calls ", Management Research Review, Vol. 40 Iss 3 pp. -

[2] Kimbrough, M.D. (2005), "The effect of conference calls on analyst and market underreaction to earnings announcements", The Accounting Review, Vol. 80 No. 1, pp. 189-219.

[3] Environmental and Public Health, Journal of. "Retracted: Prediction and Analysis of Corporate Financial Risk Assessment Using Logistic Regression Algorithm in Multiple Uncertainty Environment." Journal of Environmental and Public Health 2023 (2023): n. pag.

[4] Judijanto, Loso, Sitti Hartati Hairuddin, Subhan Subhan and Baren Sipayung. "Analysis of the Effect of Risk Management and Compliance Practices on Financial Performance and Corporate Reputation in the Financial Industry in Indonesia." The Es Accounting And Finance (2024): n. pag.

[5] Balaji, Saradha & Shreshta, Lolakpuri & Sujatha, K.. (2024). A Study on Risk Management in Corporate Business. Involvement International Journal of Business. 1. 197-209. 10.62569/iijb.v1i3.26.

[6] Sun, Weihua. (2024). Research on Corporate Financial Risk Management and Countermeasures. International Journal of Global Economics and Management. 3. 59-65. 10.62051/IJGEM.v3n1.07.

[7] Aven, T. (2015). Risk assessment and risk management: Review of recent advances on their foundation. European Journal of Operational Research, 253(1), 1–13. https://doi.org/10.1016/j.ejor.2015.12.023

[8]Cao Y, Chen Z, Pei Q, Dimino F, Ausiello L, Kumar P, Subbalakshmi KP, Ndiaye PM. RiskLabs: Predicting Financial Risk Using a Large Language Model Based on Multi-Sources Data. arXiv preprint arXiv:2404.07452. 2024 Apr 11

[9] Teixeira, Ana & Marar, Vaishali & Yazdanpanah, Hamed & Pezente, Aline & Ghassemi, Mohammad. (2023). Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting. 340-348. 10.1145/3604237.3626902.

[10] Yu G, Wang X, Li Q, Zhao Y. Fusing LLMs and KGs for Formal Causal Reasoning behind Financial Risk Contagion.arXiv preprint arXiv:2407.17190. 2024 July 24.

[11] Ahbali, N., Liu, X., Nanda, A. A., Stark, J., Talukder, A., & Khandpur, R. P. (2022). Identifying corporate credit risk sentiments from financial news. In *Proceedings of NAACL-HLT 2022: Industry Track Papers* (pp. 362-370). Association for Computational Linguistics.

[12] Tanja Aue, Adam Jatowt, Michael Färber Predicting Companies' ESG Ratings from News Articles Using Multivariate Time Series Analysis arXiv:2212.11765 **[q-fin.GN]** https://doi.org/10.48550/arXiv.2212.11765

[13] X. Li and Y. Wang, "ChatGraph: Enhancing Graph Analysis with Natural Language Interaction and Advanced Modules," arXiv, 2024. [Online]. Available: https://arxiv.org/abs/2401.12672v1

[14] Huang, Kung-Hsiang, et al. "From Pixels to Insights: A Survey on Automatic Chart Understanding in the Era of Large Foundation Models." arXiv preprint arXiv:2403.12027 (2024).

[15] Masry, A., Long, D.X., Tan, J.Q., Joty, S. and Hoque, E., 2022. Chartqa: A benchmark for question answering about charts with visual and logical reasoning. arXiv preprint arXiv:2203.10244.

[16] Carbune V, Mansoor H, Liu F, Aralikatte R, Baechler G, Chen J, Sharma A. Chart-based reasoning: Transferring capabilities from llms to vlms. arXiv preprint arXiv:2403.12596. 2024 Mar 19.

[17] Helen Eftekhari, Transcribing in the digital age: qualitative research practice utilizing intelligent speech recognition technology, European Journal of Cardiovascular Nursing, Volume 23, Issue 5, July 2024, Pages 553–560, https://doi.org/10.1093/eurjcn/zvae013

[18]H. Tolle, M. del Mar Castro, C. M. Denkinger, J. Wachinger, S. A. McMahon, A. Z. Putri, and D. Kempf, "From voice to ink (Vink): development and assessment of an automated, free-of-charge transcription tool," BMC Research Notes, vol. 17, no. 95, 2024, doi: 10.1186/s13104-024-06749-0.

[19] Loakes, Debbie. (2024). Automatic speech recognition and the transcription of indistinct forensic audio: how do the new generation of systems fare?. Frontiers in Communication. 9. 1-9. 10.3389/fcomm.2024.1281407/full.

[20] McMullin, C. Transcription and Qualitative Methods: Implications for Third Sector Research. Voluntas 34, 140–153 (2023). https://doi.org/10.1007/s11266-021-00400-3

[21] Gaddam, Shreyas, et al. "ADVANCED SEARCH AND SUMMARIZATION OF EDUCATIONAL DOCUMENTS USING MACHINE LEARNING."

[22] Chhikara, Garima, et al. "LaMSUM: A Novel Framework for Extractive Summarization of User Generated Content using LLMs." arXiv preprint arXiv:2406.15809 (2024).

[23] Jin, Hanlei, et al. "A comprehensive survey on process-oriented automatic text summarization with exploration of llm-based methods." arXiv preprint arXiv:2403.02901 (2024).

[24]Ma, Bing. "Mining both Commonality and Specificity from Multiple Documents for Multi-Document Summarization." IEEE Access (2024).

[25] Hu, Q., Moon, G. and Ng, H.T., 2024, August. From Moments to Milestones: Incremental Timeline Summarization Leveraging Large Language Models. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 7232-7246).

[26] Azizi, F., Vahdat-Nejad, H. and Hajiabadi, H., 2024. Temporal analysis of topic modeling output by machine learning techniques. International Journal of Data Science and Analytics, pp.1-51.

[27] Nair, R.P. and Thushara, M.G., 2024. Investigating Natural Language Techniques for Accurate Noun and Verb Extraction. Procedia Computer Science, 235, pp.2876-2885.

[28] P. K B, R. Sony Pinto, L. V S and S. Vrajesh, "Visualizing Parts of Speech Tags by Analyzing English Language Text," 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 2024, pp. 01-06, doi: 10.1109/ICDCECE60827.2024.10548901.

[29] Jeet Singh & Preeti Yadav. (2016), “A Study on the Factors Influencing Investors Decision in Investing in Equity Shares in Jaipur and Moradabad with Special Reference to Gender”, Amity Journal of Finance 1(1), (117-130) ©2016 ADMAA

Appendix

1. Paper Details

a. Paper published

Paper 1 :

Financial and Corporate Risk Analysis Using Large Language Models (LLMs)

Mrs. Sujata Khedkar[#1], Srushti Satish Sambare[#2], Tasmiya Sarfaraz Khan[#3],

Purtee Santosh Mahajan[#4], Ketaki Dhananjay Nalawade[#5]

[#]*Computer Engineering Department, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai - 400071, India*

[1]sujata.khedkar@ves.ac.in

[2]2021.srushti.sambare@ves.ac.in

[3]2021.tasmiya.khan@ves.ac.in

[4]2021.purtee.mahajan@ves.ac.in

[5]2021.ketaki.nalawade@ves.ac.in

***Abstract*— The purpose of this paper is to examine how Large Language Models (LLMs), specifically Gemini, may be used in financial and corporate risk analysis. In particular, we utilize LLMs to perform risk assessment activities using corporate data in the form of Corporate Earnings Calls, Annual Public Reports and Environmental, Social, and Governance (ESG) reports. Our objective is to understand the potential risks of doing business by undertaking individual risk assessment in these data sources. All these risk assessments are subsequently combined into one risk report, reflecting all possible risk exposures associated with the company's activities. This technique improves the overall efficiency of risk assessment due to the decrease of time needed for its completion while increasing the quality of available information for decision-making.**

Keywords— Large Language Models (LLMs), Gemini, Financial Risk Analysis, Corporate Risk Assessment, Corporate Earnings Calls, Annual Public Reports, ESG Reports, Generative AI, Risk Reporting, Corporate Governance

I. Introduction

In an era marked by increasing economic globalization, businesses present tremendous opportunities to expand but also face increasing pressures from competitive markets [1]. As organizations seek to capture this environment actively, risk management and compliance have become key components of corporate governance, especially in financial services ensuring that companies can maintain trust and stability [2].

Today's complex corporate environment calls for a comprehensive risk management strategy to build organizational resilience and sustainability [3]. Financial pitfalls, in particular, are receiving accelerating engrossment due to their potentially eloquent impact on the organization's future. These risks can arise from a variety of sources, such as market fluctuations, regulatory changes, or faulty internal controls, making it necessary for companies to take a holistic approach to risk assessment and management [4].

The field of risk assessment and management has been firmly established over the last 30–40 years, during which time it has evolved into an organized scientific discipline [5]. The decision-makers need to see beyond the risk evaluation; they need to combine the risk information they have received with information from other sources and on other topics [5].

The objective of this paper is to explore Large Language Models (LLMs) like Google's Gemini, which can enhance the risk assessment and management process by analyzing unstructured data from sources such as corporate earnings calls, annual public reports, and ESG reports. These insights, when combined, provide a solid foundation for more accurate and comprehensive risk assessments, ultimately contributing to better corporate decision-making.

II. Related work

The exploration of financial risk prediction and management using advanced machine learning models has been expanding, offering innovative techniques for understanding and mitigating financial risks.

In the paper RiskLabs: Predicting Financial Risk Using Large Language Models Based on Multi-Sources Data [6], the authors present the RiskLab framework, which integrates multiple modules to handle diverse sources of information such as earnings conference calls and time-series data. By

employing techniques like self-attention and Bayes-Value at Risk (VaR) forecasting, the framework demonstrates a sophisticated approach to combining news filtering and contextual compression to achieve more flexible training. A major merit of this framework is its ability to generate nuanced, multi-task predictions that provide investors with comprehensive insights into market conditions. However, a significant limitation is the reliance on large language models (LLMs), which can sometimes produce inaccurate or misleading responses, such as hallucinations, and struggle with real-time market updates due to outdated data sources.

In Enhancing Credit Risk Reports Generation Using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting [7], the authors focus on generating high-quality credit risk reports through a combination of Labeled Guide Prompting (LGP) and Bayesian networks. This methodology ensures the outputs from GPT-4 align closely with the specific requirements of credit risk analysis. A notable merit is that the generated reports were statistically preferred over traditional human-generated reports. The integration of LLMs into the credit risk process improves efficiency and scalability, making it an attractive solution for large-scale financial assessments. However, GPT-4's tendency to equally weigh all features poses a limitation, potentially failing to account for the importance of various risk factors in real-world assessments. Additionally, the model's occasional inaccuracies raise concerns about the reliability of its output.

The third paper, Fusing LLMs and Knowledge Graphs for Formal Causal Reasoning Behind Financial Risk Contagion [8], offers a novel approach to understanding how financial risks spread across interconnected entities. The Risk Contagion Causal Reasoning Model integrates Financial Knowledge Graphs (KGs) with LLMs to provide a formal causal understanding of financial contagion. By employing a fusion module and Sankey diagrams, the authors successfully show the pathways through which risks propagate. This model enhances the ability to develop targeted strategies for preventing financial crises, identifying critical nodes in the risk network. However, the challenge lies in integrating unstructured language data with structured graph data, and the model's success is dependent on the quality of the underlying financial KGs, which may not always be available or accurate.

Finally, Deep Learning Model-Driven Financial Risk Prediction and Analysis [9] explores the application of generative deep learning models for simulating financial time series data and predicting Value at Risk (VaR). By comparing the performance of different generative models, such as GANs (Generative Adversarial Networks), the paper highlights significant improvements in VaR prediction accuracy. This approach is particularly effective at capturing the complex distribution patterns in financial data. However, a notable limitation is the model’s inability to quickly adapt to sudden market changes, which can lead to inaccurate predictions during crises, where rapid responses are critical.

III. Input Sources

The system deals with three input sources. They are elaborated as follows:

A. Corporate Earnings Call

Earnings calls play a significant role in increasing investment in the company by providing economic communication. Earnings conference calls are held following the quarterly release of a company's earnings and have grown in popularity in recent years, owing to their ease of access via modern communication media (e.g., programs like EarningsCast, interactive investor-relation websites)[10][11]. The goal of these calls is to inform the market about the firm's future strategy and tactics, as well as remark on the previous quarter's revenue streams and costs [10][11].

B. Annual Public Report

An annual public report is a detailed report that public companies are required to present annually to their shareholders and the local tax office which can prepare auditor’s reports. Annual report of a company contains the directors’ report, the auditor’s report, the financial statements and the schedules and notes to the accounts [12]. For all the companies under the Ministry of Corporate Affairs (under Section 217 of Companies Act, 2013), (approx. 2263265 companies), it is mandatory to publish an annual report.

C. Environmental, Social and Governance Report (ESG)

ESG reporting is the disclosure of environmental, social and corporate governance data. As with all disclosures, its purpose is to shed light on a company’s ESG activities while improving investor transparency and inspiring other organizations to do the same. Since ESG reports summarize the qualitative and quantitative benefits of a company’s ESG activities, investors can screen investments, align investments to their values, and avoid companies with the risk of environmental damage, social missteps or corruption [13].

IV. Proposed solution

The user can upload either one or two or all of the three input sources i.e. corporate earning call, annual public report, ESG report. A risk assessment will be done on each of these sources. Risk analysis will be done based on the following parameters:

1. Highlight the statements depicting risk
2. Highlight the statements with negative impact
3. For numerical data, the system will distinguish between positive and negative trends. For example, a decrease in profits by x% will be classified as negative, whereas a decrease in losses by x% will be considered as positive. This nuanced approach will

ensure that negative statements are accurately identified while recognizing any improvements.

The assessment results are further analyzed and suggestions to lower the risk are given. Finally, these assessment results would be combined to form a risk report.

V. Methodology

Step 1: Uploading Files

1. The document is given as input using steamlit's file uploader component.
2. The user clicks on the 'Process PDF' button.

Step 2: Processing the Files

1. The process_pdf() function is called. This function reads the text content from each page of the PDF file using the 'PdfReader' from the PyPDF2 library.
2. The text content from each page is concatenated to form a single string representing the entire document.
3. The get_text_chunks() function is called. This function takes the text string as input and splits it into smaller chunks. It uses 'RecursiveCharacterTextSplitter' from the LangChain Library to split the text into chunks of size 5000 with an overlapping of 1000.
4. The get_vector_store() function is called. In the context of Natural Language Processing, the Vector store is a data structure which stores the vector representation of textual data. This function takes the text chunk as input. It generates vector embeddings for each text chunk using the Google Generative AI embeddings..These embeddings are stored in a vector store created using the FAISS Library.
5. The get_conversational_chain() function is called. This function creates a conversational chain using the LangChain library. It takes the vector store and a prompt template as input.
6. First, it initializes the gemini-1.5-flash model which is our large language model for generating responses.
7. It creates a memory component 'ConversationBufferMemory' to store the conversation history.

Step 3: Handling User Queries

1. Now as the PDFs are processed, the user enters a question and clicks on get response.
2. The user_input function is called. It takes the user's question as input. This function interacts with the conversational chain stored in the streamlit session state to retrieve a response.
3. The vector representation of the query is created and is compared with the vector representations present in the vector store.
4. The most similar vectors are then retrieved based on the similarity.
5. The conversation history is updated and displayed to the user.

Step 4: Displaying the Response

1. First, the code checks if there is any conversation history stored in streamlit session state. If it is there, it iterates through each message in the history.
2. For each message, it checks if the message index ('i') is even or odd. If it is even, it is the user's message and if it is odd, it is the bot's message.
3. If it is the response, it checks if the response contains table formatting by looking for '|' and '---' characters.
4. If yes, then it parses the response into a dataframe, eliminates dirty values and displays it as a table.
5. Then, it extracts the data from the table into a data.csv file.
6. This data is string data. It contains words like million, billion, M, B, %, $, etc. We convert those characters into empty string and then the entire value into float and plot the graph.
7. But, if there is no table in the response, it is displayed as a plain text.

Step 5: Generating FAQs

1. We have predefined categories and questions.
2. The response is retrieved the same way as user queries are answers.

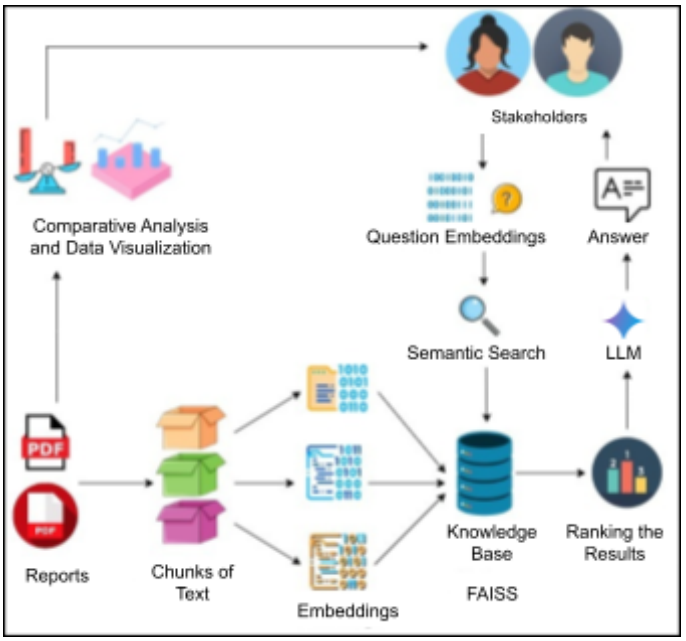
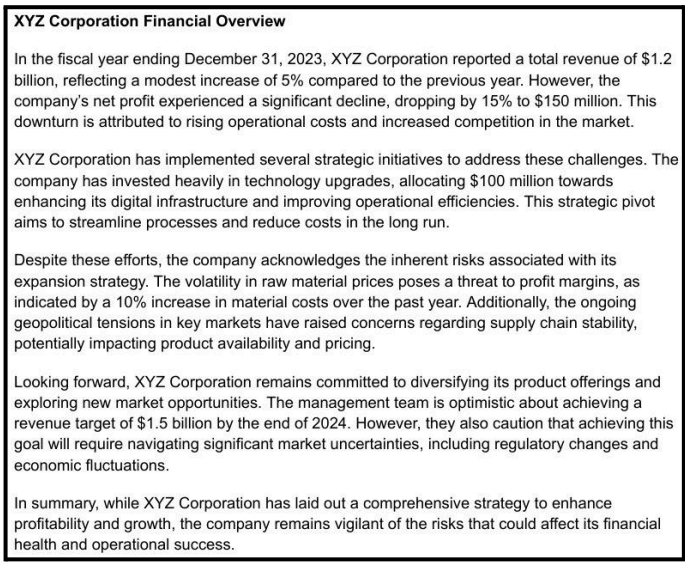


Fig. 1 System Architecture based

VI. Results

The input taken here is a sample document which contains some information about the company XYZ Corporation.



XYZ Corporation Financial Overview

In the fiscal year ending December 31, 2023, XYZ Corporation reported a total revenue of $1.2 billion, reflecting a modest increase of 5% compared to the previous year. However, the company's net profit experienced a significant decline, dropping by 15% to $150 million. This downturn is attributed to rising operational costs and increased competition in the market.

XYZ Corporation has implemented several strategic initiatives to address these challenges. The company has invested heavily in technology upgrades, allocating $100 million towards enhancing its digital infrastructure and improving operational efficiencies. This strategic pivot aims to streamline processes and reduce costs in the long run.

Despite these efforts, the company acknowledges the inherent risks associated with its expansion strategy. The volatility in raw material prices poses a threat to profit margins, as indicated by a 10% increase in material costs over the past year. Additionally, the ongoing geopolitical tensions in key markets have raised concerns regarding supply chain stability, potentially impacting product availability and pricing.

Looking forward, XYZ Corporation remains committed to diversifying its product offerings and exploring new market opportunities. The management team is optimistic about achieving a revenue target of $1.5 billion by the end of 2024. However, they also caution that achieving this goal will require navigating significant market uncertainties, including regulatory changes and economic fluctuations.

In summary, while XYZ Corporation has laid out a comprehensive strategy to enhance profitability and growth, the company remains vigilant of the risks that could affect its financial health and operational success.

Fig. 2 Input Document - Financial Overview of XYZ Corporation

The proposed solution will yield the following outcomes:

A. Statements Depicting Risk

The statements depicting risk are the excerpts or phrases that highlight the potential risks to the concerned company. These risks can be regarding the operations, finances, reputation, etc. The challenges depicting risk can be regulatory changes, supply chain disruptions, economic downturns, cybersecurity threats, or competitive pressures.

The Gemini-1.5-flash model identifies them using the following:

1) Semantic Understanding

Using natural language understanding (NLU), Gemini identifies sentences and phrases that convey uncertainty, challenges, or potential threats.

2) Keyword Detection

It scans for risk-related terms such as "challenge," "uncertainty," "adverse impact," "vulnerability," and "risk."

3) Contextual Analysis

By analyzing the surrounding context, Gemini differentiates between routine business updates and genuine risk indicators.

4) Training Data

The model is fine-tuned on large datasets containing risk statements from similar financial reports, enabling it to recognize patterns.

Fig. 3 showcases the statements depicting risks that were extracted by the system from the financial overview which was given as an input.

Statements Depicting Risk:

- "The company's net profit experienced a significant decline, dropping by 15% to $150 million."
- "This downturn is attributed to rising operational costs and increased competition in the market."
- "The volatility in raw material prices poses a threat to profit margins, as indicated by a 10% increase in material costs over the past year."
- "Additionally, the ongoing geopolitical tensions in key markets have raised concerns regarding supply chain stability, potentially impacting product availability and pricing."
- "Despite these efforts, the company acknowledges the inherent risks associated with its expansion strategy."
- "However, they also caution that achieving this goal will require navigating significant market uncertainties, including regulatory changes and economic fluctuations."

Fig. 3 Statements Depicting Risk for XYZ Corporation

B. Negative Impact Statements

The statements that directly mention the adverse effects faced by the company or the effects that the company may face in the future. This could be due to some internal or external factors. The examples can include losses from lawsuits, revenue declines due to market conditions, or negative feedback from stakeholders.

The Gemini-1.5-flash model identifies them using the following:

1) Sentiment Analysis

Gemini uses sentiment analysis to classify sections of text as negative, neutral, or positive, flagging phrases that have a distinctly negative tone.

2) Phrase Extraction

It looks for phrases indicating setbacks, losses, or failures, such as "declined revenue," "adverse effects," "unexpected losses," or "regulatory penalties."

3) Section Focus

Specific sections, like "Management Discussion and Analysis" or "Operational Challenges," are prioritized for deeper analysis.

Fig. 4 showcases the negative impact statements that were extracted by the system from the financial overview which was given as an input.

Fig. 4 Negative Impact Statements for XYZ Corporation

Fig. 5 Quantifiable Data Classified into Positive and Negative Trends for XYZ Corporation

C. Quantifiable Data Classified Into Positive And Negative Trends

This involves analyzing numerical data provided in the reports, such as financial statements, to determine trends. Positive trends include growth in revenue, increased profit margins, or higher customer acquisition rates, while negative trends include declining sales, rising debts, or shrinking market share.

The Gemini-1.5-flash model identifies them using the following:

1) *Data Extraction*

Gemini identifies and extracts numerical data from financial tables, charts, and statements, including income statements, balance sheets, and cash flow reports.

2) *Trend Detection*

Using historical comparisons (e.g., year-over-year or quarter-over-quarter changes), it determines whether trends are positive or negative.

3) *Sentiment and Impact Analysis*

Phrases accompanying numerical data, such as "strong growth" or "significant decline," are analyzed to classify trends effectively.

4) *Rule-Based Models*

Gemini applies predefined financial rules (e.g., a decline in net income over consecutive periods is flagged as negative) to supplement its ML-based approach.

Fig. 5 showcases the quantifiable data classified into positive and negative trends that were extracted by the system from the financial overview which was given as an input.

D. An Analysis of All The Extracted Data - Risk Analysis

This step involves synthesizing all identified risks, impacts, and trends into a coherent narrative to assess the overall risk profile of the company. It includes understanding the likelihood of risks materializing, their potential impact, and the company's preparedness to address them.

The Gemini-1.5-flash model gives the assessment summary using the following:

1) *Correlation of Insights*

Gemini synthesizes extracted data, connecting risk statements, negative impacts, and quantifiable trends to create a comprehensive picture of the company's risk profile.

2) *Model Outputs*

It uses probabilistic models to evaluate the likelihood and severity of risks.

3) *Risk Taxonomy*

Gemini maps extracted data to predefined risk categories (e.g., operational, financial, or compliance risks) based on its understanding of similar documents.

Fig. 6 showcases the risk assessment summary which was generated by the system from the financial overview which was given as an input and the previous results.

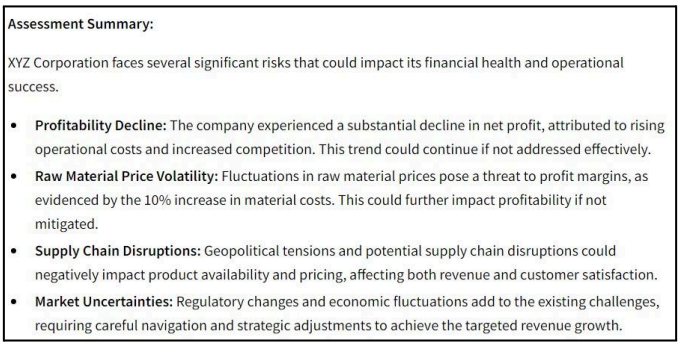


Fig. 6 Risk Assessment for XYZ Corporation

E. Risk Mitigation Suggestions

Based on the risk analysis, this involves proposing actionable steps the company can take to reduce exposure to identified risks. Suggestions can be operational, strategic, or technological and may include diversifying supply chains, enhancing cybersecurity measures, or adopting new technologies.

The Gemini-1.5-flash model gives the suggestions using the following:

1) Generative Capabilities

Leveraging its generative abilities, Gemini provides actionable recommendations based on best practices and historical data.

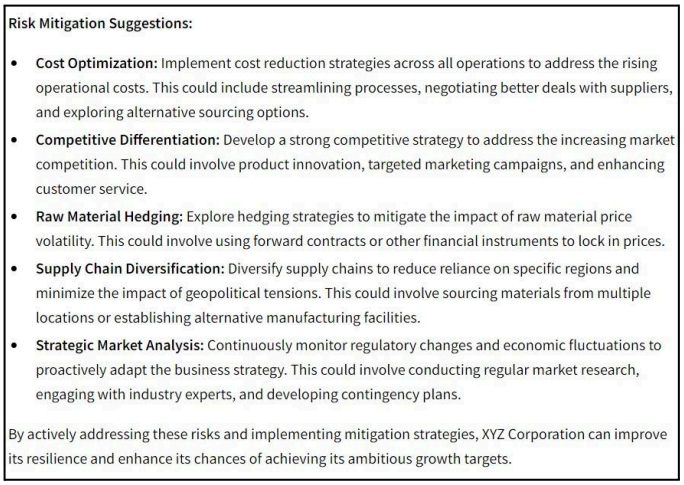
2) Trend-Based Recommendations

For example, if a downward trend in revenue is identified, Gemini might suggest cost-cutting measures or diversifying revenue streams.

3) Mitigation Mapping

The model correlates risks to solutions provided in the company report or external sources, enabling it to suggest precise mitigation strategies.

Fig. 7 showcases the risk mitigation suggestions which were generated by the system from the financial overview which was given as an input.



Risk Mitigation Suggestions:

- **Cost Optimization:** Implement cost reduction strategies across all operations to address the rising operational costs. This could include streamlining processes, negotiating better deals with suppliers, and exploring alternative sourcing options.
- **Competitive Differentiation:** Develop a strong competitive strategy to address the increasing market competition. This could involve product innovation, targeted marketing campaigns, and enhancing customer service.
- **Raw Material Hedging:** Explore hedging strategies to mitigate the impact of raw material price volatility. This could involve using forward contracts or other financial instruments to lock in prices.
- **Supply Chain Diversification:** Diversify supply chains to reduce reliance on specific regions and minimize the impact of geopolitical tensions. This could involve sourcing materials from multiple locations or establishing alternative manufacturing facilities.
- **Strategic Market Analysis:** Continuously monitor regulatory changes and economic fluctuations to proactively adapt the business strategy. This could involve conducting regular market research, engaging with industry experts, and developing contingency plans.

By actively addressing these risks and implementing mitigation strategies, XYZ Corporation can improve its resilience and enhance its chances of achieving its ambitious growth targets.

Fig. 7 Risk Mitigation for XYZ Corporation

VII. Technologies Used

A. Google Gemini API

Gemini is a family of highly capable multimodal models developed at Google [14]. The system is powered by the Gemini-1.5-flash model.

B. FAISS (Facebook AI Similarity Search)

Faiss is a library for ANNS. The core library is a collection of source files written in standard C++ without dependencies. Faiss is used in many configurations. Hundreds of vector search applications rely on it, both within Meta. [15] and externally. The proposed system deals with a large amount of textual data. Hence, FAISS is always preferred for scalability as it avoids a drop in performance. The purpose of this utilization is to perform similarity search on the vectorized representations of the documents and for a quick and easy retrieval.

C. Streamlit

Streamlit is an open-source Python framework for data scientists and AI/ML engineers to deliver dynamic data apps with only a few lines of code. [16] With this framework, you can easily build interactive visualization plots, models, and dashboards without having to worry about the underlying web framework or deployment infrastructure used in the backend [17]. Even though a Streamlit app is easy to build and deploy, it is not scalable. Hence, for large scale applications, frameworks like Flask, a Python-based framework would be preferred.

D. Langchain

Langchain is a custom Large Language Model tailored for organizations [18]. LLMs have been rapidly adopted due to their capabilities in a accept a text string (prompt) and output a text string [19]. range of tasks, including essay composition, code writing, explanation, and debugging [19]. They

VIII. Potential Biases and Inaccuracies in LLM Outputs

LLMs, while powerful, may exhibit biases due to training data limitations. These biases can lead to inaccurate risk assessments, especially in subjective or nuanced financial contexts. Hallucinations, where the model generates incorrect information, pose a challenge. For risk analysis, such inaccuracies can misrepresent a company's financial health, leading to flawed strategic decisions. Mitigating these issues requires integrating LLMs with external verification mechanisms and structured financial data.

Biases in LLMs often stem from the data they are trained on, which may include historical financial documents reflecting systemic biases. For instance, if past reports underrepresented risks in certain industries, the model may downplay potential threats. Additionally, LLMs might overemphasize commonly reported risks while overlooking emerging financial threats due to a lack of sufficient training data.

Another challenge is the tendency of LLMs to struggle with real-time market updates. Unlike traditional financial models that integrate live data feeds, LLMs depend on their training corpus, which may become outdated. This can lead to risk assessments that do not fully account for recent economic disruptions, regulatory changes, or unexpected financial crises, making human oversight essential for ensuring accuracy.

IX. Future Scope

The current implementation of the risk assessment system has considerable utility at accelerating evaluation of the financial documents and risk reports' compilation. However, there are multiple areas where the system can be extended and improved to better meet the needs of organizations and enhance its capabilities:

A. Customizable Report Formats

Another important feature which should be anticipated in the future is the possibility of an implementation to define the layout and structure of the risk reports produced. It has become a norm for different companies to have preferred style of reporting, including sections preferred in the report, preferred kind of diagrams or tables, or industry-specific language. It was proposed that, through a reporting feature that can be customized, the user would be able to prepare the report according to internal rules or other requirements.

B. Linking of documents to more than one document type

The present system of analyzing and interpreting the PDF-file format of financial reports may be expanded in subsequent versions for other formats, such as Excel tables, Word documents or even access to database information. This would enhance flexibility of the system in handling accounting information from various data feeds from different sources, and would also ensure compatibility to various documentation processes by companies.

X. Conclusion

This research outlines a simple way of automating financial risk assessment through Gemini 1.5 Flash coupled with enhanced user interface. The system is extremely effective to scan financial documents, flag risks and offer actionable intelligence. Automating risk analysis therefore enables organizations to take quicker and better decisions. Additional additions such as report features, or industry-specific models will enhance the system's flexibility making the system relevant for companies as they grapple with sophisticated risk levels.

References

[1] Environmental and Public Health, Journal of. "Retracted: Prediction and Analysis of Corporate Financial Risk Assessment Using Logistic Regression Algorithm in Multiple Uncertainty Environment." *Journal of Environmental and Public Health 2023 (2023): n. pag.*

[2] Judijanto, Loso, Sitti Hartati Hairuddin, Subhan Subhan and Baren Sipayung. "Analysis of the Effect of Risk Management and Compliance Practices on Financial Performance and Corporate Reputation in the Financial Industry in Indonesia." *The Es Accounting And Finance (2024): n. pag.*

[3] Balaji, Saradha & Shreshta, Lolakpuri & Sujatha, K.. (2024). A Study on Risk Management in Corporate Business. Involvement International Journal of Business. 1. 197-209. 10.62569/iijb.v1i3.26.

[4] Sun, Weihua. (2024). Research on Corporate Financial Risk Management and Countermeasures. International Journal of Global Economics and Management. 3. 59-65. 10.62051/IJGEM.v3n1.07.

[5] Aven, T. (2015). Risk assessment and risk management: Review of recent advances on their foundation. European Journal of Operational Research, 253(1), 1–13. https://doi.org/10.1016/j.ejor.2015.12.023

[6] Cao Y, Chen Z, Pei Q, Dimino F, Ausiello L, Kumar P, Subbalakshmi KP, Ndiaye PM. RiskLabs: Predicting Financial Risk Using Large Language Model Based on Multi-Sources Data. arXiv preprint arXiv:2404.07452. 2024 Apr 11.

[7] Teixeira, Ana & Marar, Vaishali & Yazdanpanah, Hamed & Pezente, Aline & Ghassemi, Mohammad. (2023). Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting. 340-348. 10.1145/3604237.3626902.

[8] Yu G, Wang X, Li Q, Zhao Y. Fusing LLMs and KGs for Formal Causal Reasoning behind Financial Risk Contagion. arXiv preprint arXiv:2407.17190. 2024 Jul 24.

[9] Yang, Tianyi & Li, Ang & Xu, Jiahao & Su, Guangze & Wang, Jiufan. (2024). Deep Learning Model-Driven Financial Risk Prediction and Analysis. 10.20944/preprints202406.2069.v1

[10] Vivien E. Jancenelle Susan Storrud-barnes Rajshekhar Javalgi , (2017)," Corporate Entrepreneurship and Market Performance: A Content Analysis of Earnings Conference Calls ", Management Research Review, Vol. 40 Iss 3 pp. -

[11] Kimbrough, M.D. (2005), "The effect of conference calls on analyst and market underreaction to earnings announcements", The Accounting Review, Vol. 80 No. 1, pp. 189-219.

[12] P. K. Aithal, D. A. U. and G. M., "Analyzing Tone of the Annual Report - An Indian Context," 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), THIRUVANANTHAPURAM, India, 2022, pp. 536-542, doi: 10.1109/SPICES52834.2022.9774247. keywords: {Instruments;Signal processing algorithms;Companies;Signal processing;SPICE;Stock markets;Information technology;Portfolio Management;Tone;Initial Public Offerings;Parallel Algorithms;Message Passing Interface}

[13] Tocchini, F., & Cafagna, G. (2022, March 9). The ABCs of ESG reporting: What are ESG and sustainability reports, why are they important, and what do CFOs need to know. https://www.wolterskluwer.com/en/expert-insights/the-abcs-of-esg-reporting#:~:text=ESG%20reporting%20is%20the%20disclosure,organizations%20to%20do%20the%20same.

[14] G. Team *et al.*, "Gemini: a family of highly capable multimodal models," *arXiv.org*, Dec. 19, 2023. https://arxiv.org/abs/2312.11805

[15] M. Douze *et al.*, "The Faiss library," *arXiv.org*, Jan. 16, 2024. https://arxiv.org/abs/2401.08281

[16] "Streamlit Docs." https://docs.streamlit.io/

[17] Sreeram a, Adith & Sai, Jithendra. (2023). An Effective Query System Using LLMs and LangChain. International Journal of Engineering and Technical Research. 12.

[18] K. Pandya and M. Holia, "Automating Customer Service using LangChain: Building custom open-source GPT Chatbot for organizations," *arXiv.org*, Oct. 09, 2023. https://arxiv.org/abs/2310.05421

[19] Topsakal, Oguzhan & Akinci, T. Cetin. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. International Conference on Applied Engineering and Natural Sciences. 1. 1050-1056. 10.59287/icaens.1127.

Paper 2 :

Metadata Extraction from Legal Contracts Using Large Language Models

Mrs. Sujata Khedkar[#1], Tasmiya Sarfaraz Khan[#2], Srushti Satish Sambare[#3],

Purtee Santosh Mahajan[#4], Ketaki Dhananjay Nalawade[#5]

[#]*Computer Engineering Department, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai - 400071, India*

[1]sujata.khedkar@ves.ac.in

[2]2021.tasmiya.khan@ves.ac.in

[3]2021.srushti.sambare@ves.ac.in

[4]2021.purtee.mahajan@ves.ac.in

[5]2021.ketaki.nalawade@ves.ac.in

***Abstract*— Contracts are formal agreements with legal force that govern interactions between people or entities. Different sections and clauses of the contracts are defined regarding obligations, rights, and responsibilities of each party involved. Due to the complexity as well as the volume of legal contracts, extracting key metadata is highly important in ensuring efficiency for legal professionals. This paper addresses the use of Large Language Models to automate the extraction of core metadata from legal contracts, including clauses such as liability, insurance, and jurisdiction, and organizing it into a structured, tabular format for easier analysis. We make use of the OpenAI Azure LLM to produce a full-stack solution, able to parse and structure this metadata into a format that is usable. The study addresses problems related to large prompt sizes and various optimization strategies that could be applied to such systems in order to attain better efficiency and accuracy. In addition, we include real-time FAQ generation from the legal contracts. The goal behind this research therefore is to verify the potential that LLMs may bring to streamline legal workflows while ideally addressing technical and operational challenges found in large-scale metadata extraction tasks.**

Keywords— Legal Contracts, LLM, Clauses, FAQ, Metadata.

I. Introduction

A contract is a legally binding agreement that specifies the terms and conditions of the parties involved and is signed by at least two of them. Since contracts are usually written in text, there is a lot of room for NLP applications in the field of legal papers. However, contract language is repetitive with high inter-sentence similarities and sentence matches, in contrast to most natural language corpora that are typically used in NLP research. [1]

In the legal domain, collecting information from contracts poses three major hurdles for existing methods. The primary difficulty is the scarcity of data needed to train or fine-tune algorithms for high accuracy. Another difficulty is the large size of many contracts, which frequently exceeds the processing capacity of current transformer topologies. Transformer-based models have a limited sequence length that they can accommodate. Contracts that exceed the limit may need to be broken into smaller portions, complicating the analyzing process. The third challenge is that contracts contain a mix of long and short entities, such as full clauses like non-complete and audit rights, and short ones like names and dates. [2]

Metadata extraction is the process of identifying and extracting specific data points from unstructured documents, and is becoming increasingly important for legal

professionals. This is particularly true for clauses related to liability, insurance, indemnity, termination, and jurisdiction, which are crucial for understanding the legal implications of contracts. Manually reviewing large, complex contracts for specific clauses can be time-consuming and error-prone. By extracting information about liability, insurance, indemnity, termination, and jurisdiction, etc legal professionals can better assess the potential risks associated with a contract.

To overcome these challenges, in this paper we proposed to extract metadata from an unstructured format to a structured format using LLM. This metadata will be formatted into a table. We proposed four tables: a quarterly report, which includes various clauses and important entities; a termination table, which includes all the termination clauses; an indemnification table, which includes all the indemnification clauses; and a basic table, which includes all important dates from the contracts and party names.

II. Related work

This paper proposes an unsupervised two-step approach for extracting knowledge from long legal contracts. It overcomes the token limitation in the basic LLMs and the lack of training data. As stated in the abstract, authors propose a query-based summarization model that extracts relevant sentences, thus reducing the text size without losing any core information. The summary is fed into GPT-3.5 to generate accurate metadata extraction. This method does not rely on the necessity of training domain-specific models unlike supervised models but depends on the pre-trained capabilities of LLMs which it uses to extract short and lengthy contract entities without fine-tuning. The performance exhibited was better and brought about state-of-the-art results in zero-resource settings, especially when contrasted with fine-tuned supervised models on domain-specific data. [2] This paper explores the application of Large Language Models (LLMs) in law by designing generative AI to optimize contract management. As already mentioned in the abstract, it uses Retrieval-Augmented Generation (RAG) to improve the analyzing ability and generate legal clauses of LLMs. The system uses semantic embeddings for accurate document retrieval and generation of contract clauses that are contextually relevant. It also employs Azure OpenAI models using alertness of prompting optimization to solve issues on contract complexity; it makes use of retrieval as well as LLMs to reduce the time lawyers take in reading and analyzing a contract, thereby making it possible to draft new contracts fast and accurately. As part of future work, the paper mentions the implementation of Optical Character Recognition (OCR) technology. [3] This will enable the system to handle non-digital legal documents, expanding its application breadth while boosting the depth of the analysis process. It will also enable the system to extract text from scanned papers or PDFs. Optical Character Recognition, or simply OCR, is certainly one of the most important technologies by means of which printed or hand-written text can be scanned and converted to digital versions, saving literally hundreds of hours of time and labor when doing data entry and document management. As highlighted in the paper [4], OCR automates the process of extracting text from various sources, such as images and scanned documents, and converts it into searchable and editable digital content. This technology plays a vital role in industries that handle large volumes of data, enabling organizations to streamline operations, reduce errors, and improve overall productivity.

The paper [7] provides a methodology for automating the reporting of contracts and obligation extraction with LLMs, focusing particularly on filling in the DORA compliance templates and on extracting obligations from the contract. This introduction discusses how the emergence of LLMs like GPT-3 and GPT-4 creates opportunities for new avenues for automation with respect to automating complex tasks like information extraction from legal texts. These models were used in the experiments of this paper as a means of extracting specific data from legal contracts to test their accuracy in compliance and contractual obligation tasks. The system was shown to be highly accurate in filling compliance templates with up to 97.71%, but less so consistent in the extraction of obligations in legal contracts, with up to 70.56% accuracy in some cases. These results showcase the capabilities of LLMs for the purpose of changing the face of legal document management by saving manpower and increasing accuracy. Large language models (LLMs) have found increasing use in the legal sphere, demonstrating how these technologies may be tailored to industry-specific requirements. Prominent models comprise LegalBERT[8], CaseLaw-BERT[9], and FinBERT[15] which have been refined from generic models to tackle assignments including anticipating legal judgments, reviewing contracts, and other domains necessitating an intricate comprehension of legal terminology and concepts. These models demonstrate at least two aspects: the potential for LLMs to increase efficiencies in legal workflows and the role of domain-specific training in reaching almost precise relevance with respect to professional applications

III. Proposed solution

A. System Architecture

The full-stack solution built around OpenAI's Azure LLM is designed to automate the extraction of key metadata from legal contracts. The system architecture comprises several key components:

1) Input (Contract Documents):

Users can upload contract documents in various formats, such as PDF or DOCX. To ensure compatibility and efficient processing, the system extracts the text from these contracts. If the contract is in a non-digital format (e.g., a scanned image), Optical Character Recognition (OCR) is used to extract the text. Tools such as Pytesseract are used to ensure high accuracy when converting image-based contracts into text format.

2) LLM for Clause Extraction:

- Once the text is extracted, it is passed through the OpenAI Azure LLM for clause extraction. The LLM is specifically prompted to identify key clauses, such as liability, insurance, and jurisdiction.
- The system uses semantic embeddings and context-aware language models to recognize specific patterns and extract relevant metadata from the contract clauses, which are often complex and filled with legal jargon.
- The LLM processes the contract and returns the extracted metadata in a structured output format.

3) Structured Output in Tabular Format:

The extracted metadata is organized into a tabular format for easy review and further analysis. Each row in the table corresponds to a metadata type (e.g., Liability, Insurance, Jurisdiction) with columns containing the extracted information.

Our proposed solution extracts four tables:

Basic Table: This table includes columns such as Agreement Type, Client, Service Provider and Dates. Dates include:

- Document date: Date when the document was created.
- Effective date: Date when the agreement becomes effective
- End date: Date when the agreement ends.

Termination Table: This table includes:

- Termination clauses: The specific sections or provisions in the contract that outline the conditions under which the agreement can be terminated. This includes reasons for termination such as breach of contract, failure to meet obligations, or termination for convenience. The clauses also specify any notice periods, penalties, or other conditions for ending the contract.
- Clause number
- Who can terminate the agreement: Specifies which party or parties (e.g., Client, Service Provider, or both) have the right to terminate the agreement under the conditions set forth in the termination clause. Some contracts allow only one party to terminate, while others grant termination rights to both parties under certain circumstances.

Report: This table includes Client name:

- Liability: This term refers to the legal responsibility for damages or breaches under the contract. It can be capped at a certain amount or percentage of the contract's value (e.g., Work Order value). If capped, the liability is limited to the specified value mentioned in the "LIMITATION OF LIABILITY" clause. This clause protects one party by restricting their financial obligations in the case of legal claims.
- Uncapped liability: Certain liabilities may be uncapped, meaning they have no limit on the amount of compensation required. Common uncapped liabilities include claims related to death, personal injury, fraud, or breach of law. If liabilities are capped, this section will list the specific types of liabilities that remain uncapped.
- Warranty: Warranties ensure the quality or performance of products or services. If exceptions are present in the warranty (e.g., "except for certain conditions"), they should be noted. Otherwise, it is marked as "Standard" or "Not available" if the warranty section is absent.
- Indemnity: A legal obligation where one party agrees to compensate the other for losses, damages, or liabilities arising from specified actions or circumstances. If exceptions are present (e.g., indemnity does not apply in certain situations), they should be noted. Otherwise, it is marked as "Standard" or "Not available" if the indemnity section is absent.
- Services/Damage: Refers to the obligations regarding services provided under the contract and any potential damages. If the damages are ongoing, this should be noted. A cap may be specified, which limits the amount of compensation for damages, either as a fixed amount or percentage.
- Jurisdiction/Law: The geographic location (country and city) whose laws govern the contract and where disputes arising under the contract will be resolved.
- Termination for convenience:A clause allowing one party, typically the service provider or client, to terminate the contract without cause. This provides flexibility to exit the agreement at any time, with advance notice as per the contract terms.
- Insurance: Refers to the types and amounts of insurance required to cover potential liabilities or risks in the agreement. This clause outlines the minimum coverage necessary for the parties involved.

The output is exported to csv format, which can be easily integrated into existing contract management systems or used for reporting purposes. The structure of the table includes fields like:

- Clause Type: The specific clause extracted (e.g., Liability).
- Extracted Text: The actual content of the clause from the contract.
- Summary: A brief summary generated by the LLM to give overall content

B. Optimizing Prompts Sizes

Large Language Models (LLMs) are intended to be effectively adapted to specific activities through prompt optimization. An appropriate prompt, whether textual or visual, improves the model's output to better fit the user's task [13]. One of the primary challenges of using Large Language Models (LLMs) like OpenAI's Azure model is handling large input text sizes, as legal contracts can be lengthy and exceed token limitations. To overcome this, the following strategies are employed:

1) Text Chunking:

The contract text is divided into smaller chunks

based on logical sections such as clauses, headings, or paragraph breaks. The LLM processes each chunk independently and extracts metadata for each section. This ensures that even large contracts can be processed within the token limit of the LLM without losing context or accuracy.

2) Summarization:

For particularly large sections, a summarization step is added. The system uses the LLM to generate summaries of each section before extracting metadata. This reduces the overall token count while preserving the key information needed for metadata extraction. This step is useful when contracts contain extensive descriptions or legal explanations that may not be directly relevant to the metadata extraction process but are still necessary to retain for legal clarity.

3) Hierarchical Querying

Instead of sending the entire contract at once, the system performs hierarchical querying, where initial queries extract broad metadata (e.g., major sections like "Liability" or "Insurance"). Subsequent, more specific queries are then used to refine the extraction, targeting specific details within each clause. This two-step querying process minimizes the amount of text the model has to process in one go, ensuring efficient and focused metadata extraction.

C. *Real-time FAQ Generation*

One of the extra features of the system is the ability to generate real-time FAQs from the text of a contract. This is particularly useful for legal professionals who might be needed to provide some quick response to the fundamental questions related to a contract. The process of generating FAQs involves:

1) Natural Language Processing (NLP) for Query Understanding:

When a user asks a question related to the contract (e.g., "What are the liability limits?"), the system uses NLP to interpret the question and match it to the relevant sections of the contract. The OpenAI Azure LLM then searches the contract text to locate the clauses most likely to answer the user's query.

2) Answer Generation:

Once the relevant text is identified, the LLM generates a natural-language answer to the user's question. This answer includes the direct clause as well as a concise summary or explanation for better comprehension.

For example, if asked about payment obligations, the system extracts the payment terms and provides a summarized version, highlighting key details like deadlines and penalties.

3) Answer Generation:

Once the relevant text is identified, the LLM generates a natural-language answer to the user's question. This answer includes the direct clause as well as a concise summary or explanation for better comprehension.

For example, if asked about payment obligations, the system extracts the payment terms and provides a summarized version, highlighting key details like deadlines and penalties.

4) Continuous Learning:

The FAQ system is designed to learn from previous interactions. Over time, it builds a knowledge base of frequently asked questions and their corresponding answers, improving its efficiency and accuracy with repeated use. It also integrates user feedback to refine the generated answers, ensuring that the system evolves based on the needs of its users.

By integrating real-time FAQ generation, the system reduces the time legal professionals spend searching for specific information within contracts, allowing for quicker decision-making and better client support.

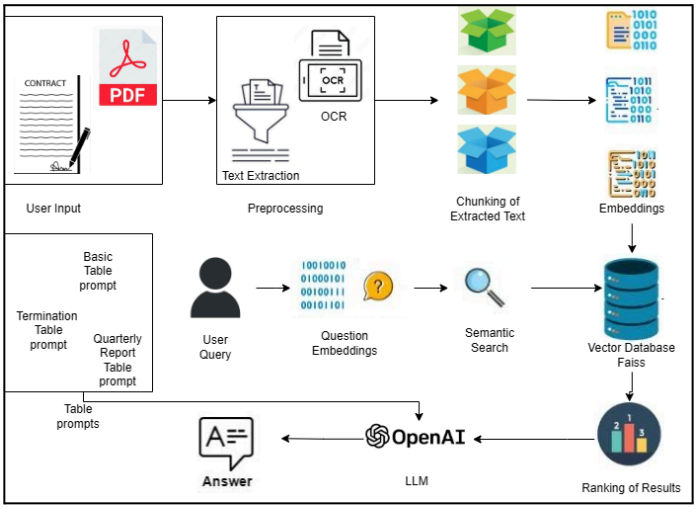


Fig 1. System Architecture

IV. Experiments and Results

Sample_MSA.pdf

MetaData

	Agreement Type	Client	Service Provider	Document Date	Effective Date
0	----------	------	----------	----------	----------
1	Master Services Agreement	Not specified	Celonis	May 2021	The earlier of (A) your i

Fig. 2 Basic Table

Fig. 3 Termination Table

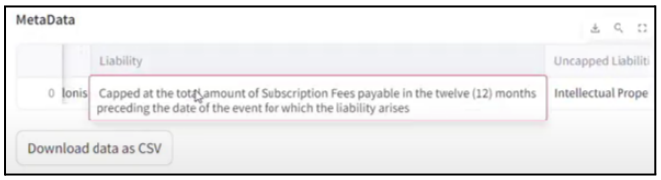


Fig. 4 Liability clause of Report Table

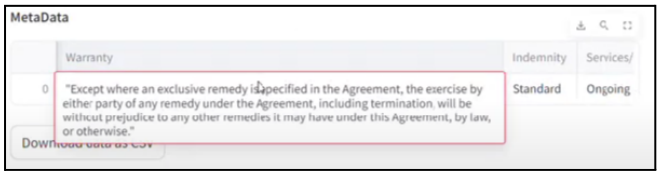


Fig. 5 Warranty clause of Report Table

Fig. 6 Report Table

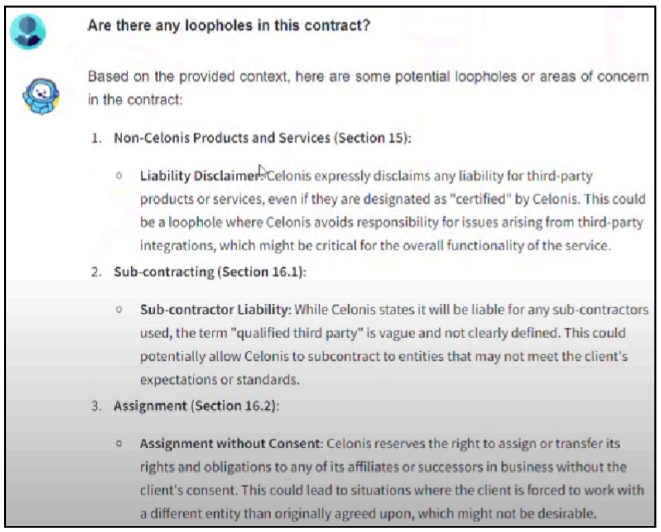


Fig. 6 Chatbot Q&A

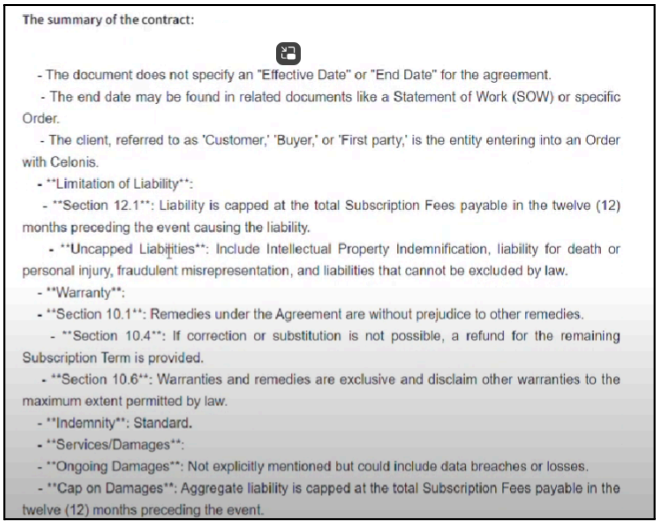


Fig. 7 Summary of Quarterly Report

V. Limitations

A. *Scalability Issues with Large Contracts*

- The proposed system relies on Large Language Models (LLMs) such as OpenAI's Azure LLM, which have inherent token limitations. Legal contracts often exceed the maximum context length that these models can process in a single query, requiring text chunking and hierarchical querying. This can lead to loss of contextual understanding between related clauses spread across different sections of a contract.
- Processing multiple large contracts simultaneously could significantly increase computational load, leading to higher latency and processing costs when deployed in a real-world legal setting with a high volume of contracts.

B. *Handling Diverse Legal Documents*

- Contracts vary significantly in structure, language, jurisdiction, and legal terminology. While the proposed system performs well with common legal clauses (e.g., liability, insurance, jurisdiction), it may struggle with domain-specific contracts such as those in healthcare, real estate, or mergers and acquisitions, where unique clauses exist.
- The system primarily focuses on English-language contracts, and while future work mentions multilingual support, legal translation and jurisdictional variations in phrasing and clause structuring remain a major challenge.

VI. Future Work

To address the current limitations and further improve the system, several potential enhancements are proposed:

A. *Overcoming Token Limitations:*

There are many strategies that can be pursued to overcome token limitations:

- Text Chunking with Overlap: Implement smarter chunking mechanisms with overlapping content to divide the text into chunks, which may help preserve continuity in clauses. This can ensure that even

clauses spanning over multiple sections are captured appropriately.

- Hierarchical summarization: In case one has a very huge contract, that input can be summarized in a hierarchical procedure wherein a summary is made first, and then the summary will be processed toward metadata extraction. It will reduce the amount of the input text while conserving the most important information for the end user. We aim to prototype and test hierarchical summarization within the next quarter, focusing initially on reducing text size while maintaining critical information.
- Memory-Augmented Models: Using memory-augmented LLMs or longer context windows, for instance in GPT-4, which extends token capability, the model can process huge blocks of text while having contextual information from previous sections.

B. Multi-lingual Contract Support:

Most contracts are multilingual. International business has a significant share of contracts in multiple languages. It would be very precious to expand the system to accommodate such multilingual contracts. There would be scope for using multilingual models like mBERT or XLM-R. More importantly, the model could be fine-tuned for legal documents in different languages- French, Spanish, German, or Arabic. In the next six months, we plan to introduce support for European languages (French, Spanish, German), followed by Arabic and Asian languages in the next year.This would enable the system to operate with contracts spread across various legal jurisdictions and less dependent on translation services.

VII. Conclusion

This work demonstrates the capacity of LLMs, such as the OpenAI Azure LLM, to automate the extraction of essential metadata from legal contracts. The system quickly identifies and extracts all the pertinent clauses concerning liability, insurance, jurisdiction, and terms of payment, then structures them in tabular form for easy analysis and review. We have proposed four tables which provides with effective analysis of legal contracts. In addition, real-time generation of FAQs based on the content of a contract has more functionality integrated into the system: legal professionals can now query particular details of a contract and receive an accurate response instantly. It has reduced manual efforts in analyzing contracts, with LLMs instead hastening workflows, minimizing human error, and improving the efficiency associated with legal processes. The result is the ability of LLM-based systems in changing legal professionals' interfaces with complex contracts, especially in metastuffs like data extraction and compliance checking.

References

[1] Dan Simonson, Daniel Broderick, and Jonathan Herr. 2019. The extent of repetition in contract language. In Proceedings of the Natural Legal Language Processing Workshop 2019, pages 21–30, Minneapolis, Minnesota. Association for Computational Linguistics.

[2] Zin, May Myo, et al. "Information Extraction from Lengthy Legal Contracts: Leveraging Query-Based Summarization and GPT-3.5." *Legal Knowledge and Information Systems*. IOS Press, 2023. 177-186.

[3] Mongoli, Alessio. *The Use of LLMs in the Legal Field: Optimizing Contract Management with Generative Artificial Intelligence.* Diss. Politecnico di Torino, 2024.

[4] A. D et al., "Image Text Detection and Documentation Using OCR," *2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC)*, Coimbatore, India, 2024, pp.410-414,doi:10.1109/ICSSEECC61126.2024.10649443.

[5] Sisodia, Prakhar, and Syed Wajahat Abbas Rizvi. "Optical character recognition development using Python." *Journal of Informatics Electrical and Electronics Engineering (JIEEE)* 4.3 (2023): 1-13.

[6] Mukherjee, Sumita, et al. "OCR Using Python and Its Application." *Journal of Advanced Zoology* 44 (2023).

[7] Geerligs, Cornelis. *Information Extraction from Contracts Using Large Language Models*. Diss. 2024.

[8] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion An droutsopoulos. “LEGAL-BERT: The muppets straight out of law school”. In: arXiv preprint arXiv:2010.02559 (2020). DOI: https://doi.org/10.48550/arXiv.2010.02559.

[9] Lucia Zheng, Neel Guha, Brandon R. Anderson, Peter Henderson, and Daniel E. Ho. When DoesPretraining Help? Assessing Self-Supervised Learning for Law and the CaseHOLDDataset. 2021. DOI: https://doi.org/10.48550/arXiv.2104.08671.arXiv:2104.08671[cs.CL]

[10] McCamus, John. *The Law of Contracts, 3/e*. University of Toronto Press, 2020.

[11] Fosbrook, Deborah, and Adrian C. Laing. *The AZ of contract clauses*. Bloomsbury Publishing, 2022.

[12] Wang, Brydon T. "Prompts and large language models: A new tool for drafting, reviewing and interpreting contracts?." *Law, Technology and Humans* 6.2 (2024): 88-106.

[13] Sabbatella, Antonio, et al. "Prompt Optimization in Large Language Models." *Mathematics* 12.6 (2024): 929.

[14] Frieda Josi, Christian Wartena, and Ulrich Heid. 2019. Detecting Paraphrases of Standard Clause Titles in Insurance Contracts. In RELATIONS - Workshop on meaning relations between phrases and sentences, Gothenburg, Sweden. Association for Computational Linguistics.

[15] Dogu Araci. “Finbert: Financial sentiment analysis with pre-trained language models”. In: arXiv preprint arXiv:1908.10063 (2019). DOI: https://doi.org/10.48550/arXiv.1908. 10063.

[16] TomBrownetal.“Language Models are few-shot learners”. In: Advances in neural information processing systems 33 (2020), pages 1877–1901. DOI: 10.48550/arXiv.2005.14165.

[17] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in Neural Information Processing Systems* 33 (2020): 9459-9474.

[18] Bulles, John, Hennie Bouwmeester, and Anouschka Ausems. "A best practice for the analysis of legal documents." *On the Move to Meaningful Internet Systems: OTM 2019 Workshops: Confederated International Workshops: EI2N, FBM, ICSP, Meta4eS and SIAnA 2019, Rhodes, Greece, October 21–25, 2019, Revised Selected Papers*. Springer International Publishing, 2020.

[19] Wei-Lin Chiang et al. “Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference”. In: arXiv preprint arXiv:2403.04132 (2024). DOI: https://doi.org/10. 48550/arXiv.2403.04132.

[20] Chiang, Wei-Lin, et al. "Chatbot arena: An open platform for evaluating llms by human preference." *arXiv preprint arXiv:2403.04132* (2024)

Paper 3 :

Annual Public Report Analysis ChatBot Using LLM

Mrs. Sujata Khedkar[#1], Purtee Santosh Mahajan[#2], Tasmiya Sarfaraz Khan[#3],

Srushti Satish Sambare[#4], Ketaki Dhananjay Nalawade[#5]

[#]*Computer Engineering Department, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai - 400071, India*

[1]sujata.khedkar@ves.ac.in

[2]2021.purtee.mahajan@ves.ac.in

[3]2021.tasmiya.khan@ves.ac.in

[4]2021.srushti.sambare@ves.ac.in

[5]2021.ketaki.nalawade@ves.ac.in

***Abstract*— In today's data-driven world, interpretation of public annual reports might become a very time-consuming process for the stakeholders. This paper introduces the concept called an Annual Public Report Analysis Chatbot using LLM, through which automatically going through reports and generating intuitive answers through AI will be possible. The chatbot is designed to help stakeholders extract key insights from annual public reports, making the information easier to understand and more actionable.Utilizing AI capabilities, this system does not merely provide correct responses to questions it receives from the user but also provides comparative analysis and data visualization options to further help the decision-making process.The intent of this project is to change the way that stakeholders interact with complex reports making the process associated with analyzing those reports streamlined and empowering such decision-makers to go data-driven. The chatbot processes public reports by extracting and organizing the information into a structured knowledge base, enabling efficient retrieval of relevant insights. It helps users uncover key information from annual reports, making the content clearer and more actionable. By leveraging AI, the system provides accurate answers to user queries and offers comparative analysis and data visualization, supporting informed decision-making. One of its standout features is text-to-speech functionality, which delivers responses audibly, enhancing accessibility and inclusivity.**

Keywords— Large Language Models (LLMs), Annual Public Report, AI Chatbot, Data Visualization, Text-to-Speech, Corporate Data Analysis, Semantic Search, Stakeholder Engagement.

I. Introduction

The annual report is the major source of information to investors. Investors will read the annual report to analyze the performance of the company.The annual report is composed of the director's report, auditors report, financial statements, and notes to the accounts. Financial statements are in quantitative form while the director's report, notes to the accounts and auditors report are in qualitative form. Investors, in general, will read the annual report to make the investments. Investors will invest in the company if the company is performing well [10], then there will be positive narration in the annual report else there will be negative narration [1].With the rise of the Internet and the information age, an increasing amount of data is being uncovered and consumed by individuals. As a key vehicle for conveying corporate information, financial statements have grown progressively longer. This expansion is driven by the need to meet the diverse demands of information users, while also adhering to the standards of accounting information quality. Longer reports mean more information. While the increased information has both benefits and problems for the users. [2] . Also predicting bankruptcy is important before investing because bankruptcy status is a clear indicator of fiscal health [6].

However, manually analyzing these extensive documents can be time-consuming and prone to human error.Due to the sheer volume and complexity of these documents, they are highly labor-intensive and susceptible to human error. Recent advances in Natural Language Processing (NLP), particularly in Large Language Models (LLMs), have significantly enhanced the performance of text analysis systems by machines. Our chatbot leverages the capabilities of such models in efficiently and intelligently extracting insights from annual reports. The rapid advancement of Large Language Models (LLMs) has revolutionized the extraction and interpretation of financial data, particularly in quantifying market sentiment derived from sources such as corporate disclosures, financial news, and social media. These insights play a crucial role in influencing market movements and guiding investment decisions. LLMs have also demonstrated significant potential in the domain of Financial Time Series Analysis [8], where they contribute to forecasting trends, detecting anomalies, and classifying financial data[9]. Despite ongoing debate regarding their efficacy in this area, LLMs excel in capturing complex temporal dependencies within financial datasets through their advanced deep learning architectures. One of their most notable capabilities is reasoning, allowing them to not only analyze data but also simulate cognitive processes similar to human decision-making.[7] This reasoning extends into Financial Planning, where LLMs assist in generating investment strategies and supporting decisions by synthesizing vast amounts of data. Furthermore, their application in Agent-based Modeling enhances the simulation of financial ecosystems, enabling the study of market behaviors and economic activities through dynamic interactions between agents and their environments[3].

This paper aims to design an annual public report chatbot that uses large language models to analyze unstructured corporate data. Users can upload or search for reports in this system, which are processed in real time to produce answers, comparative analyses, and visual data insights. This will really aid the decision-making process by ensuring a very efficient and user-friendly understanding of corporate reports.

II. Related work

The researchers discovered that the language used in annual reports is a crucial indicator of a company's financial well-being and performance [1]. They found that a positive tone is associated with profitable companies and has an impact on investor behavior. Investors are more likely to invest in companies with a positive tone and avoid those with a negative tone. The tone was measured using the Loughran-McDonald sentiment word lists, providing a structured method for evaluating qualitative financial disclosures. The study also identified a connection between tone and stock price movements, indicating that the sentiment expressed in these reports influences market perceptions and investment decisions, particularly in initial public offerings. While tone is important, the study was focused on Indian companies and suggests that incorporating quantitative data and broader contexts could improve the analysis.

A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges [3]: This paper delves into the application of Large Language Models (LLMs) in examining annual reports, with a focus on various important aspects. LLMs like BioFinBERT are utilized for gauging sentiment to forecast stock price movements based on the emotions conveyed in regulatory filings and legal documents. They also shine in extracting information, enabling stakeholders to swiftly condense crucial data for well-informed decision-making. Additionally, LLMs assist in flagging anomalies by pinpointing inconsistencies within financial statements, ensuring precision in reporting. Their capabilities extend to numerical analysis, allowing for the comprehension of intricate financial data through multi-step calculations. However, the analysis encounters challenges such as high computational expenses, the potential for misinterpreting financial terminology, susceptibility to adversarial attacks, and the presence of excessive information that might obscure crucial insights.

Automatic Analysis of Annual Financial Reports: A Case Study [4]: The analysis of annual reports involved the use of various methodologies and focused on examining 10-K reports from Ford Motor Company, General Motors Company, Google (Alphabet Inc.), and Yahoo! Inc. in the automotive and IT industries over a decade (2005-2014). Researchers specifically extracted non-financial sections from the reports, particularly Part I and Items 7 and 7A from Part II, to capture management opinions on past and future performance. They conducted linguistic analysis by considering document length, sentiment using a sentiment dictionary, the prevalence of trust and doubt keywords, and various discursive features such as personal versus impersonal pronoun ratios. Differential content analysis involved using TF-IDF weighting to identify characteristic terms for each year, while correlation analysis examined connections between linguistic characteristics and financial performance. Nevertheless, the research encountered constraints, such as a limited sample size that might not accurately reflect wider market patterns, a restricted time period that might fail to capture notable changes in reporting methods, difficulties in extracting data due to inconsistent formats, and the subjective nature of linguistic analysis. Furthermore, the conclusions might not be applicable to other types of reports or regulatory landscapes beyond the U.S., indicating the necessity for additional research using larger and more diverse samples to affirm and build upon these findings.

Narrative analysis of annual reports: A study of communication efficiency [5] : The researchers analyzed the content of annual reports, specifically focusing on the Management Discussion and Analysis (MD&A) sections. They used the Flesch Reading Ease formula to assess the readability of the texts. This method allowed them to measure the language complexity in the reports and track changes in readability over the years under review, especially during the global recession from 2008 to 2012. The objective was to evaluate how external factors, such as economic downturns, affected the clarity and openness of corporate communications.

InvestLM [12] is especially adept at analyzing annual reports because it was trained on texts that are primarily financial in nature and can come up with contextually relevant answers. Given its training on a finely curated financial dataset, InvestLM understands complex financial language and concepts found within annual reports; thus, it may obtain essential information from financial statements, management discussions, or risk factors. Its performance has been rated just as excellent in expert evaluations as with models like GPT-3.5 and GPT-4, making it truly useful in investment decision-making. InvestLM really shines in setting up risks and opportunities in the context of a company's financial health and management expectations. Automation of large volumes of text saves much time for financial professionals to do more by lending itself well beyond annual reports to tasks such as sentiment analysis and document classification.

III. Input Sources

The system deals with two main input sources, explained as follows:

A. Direct Report Upload

Our chatbot facilitates users directly to upload the annual public reports in PDF format. The system supports the upload of multiple files, enabling detailed comparative analysis across different reports. The reports include detailed financial statements, directors' reports, the auditors' reports, and other

detailed information on a firm's performance during the fiscal year. Stakeholders can quickly analyze specific reports for insights without having to search through them manually through the direct uploads.

B. Company Name and Year Search

In addition to direct uploads, users can search for annual public reports by entering the company's name and the desired year. The system fetches all the reports available on the internet using API ,allowing users to download them and then upload the report PDF to the chatbot. The feature aims at making data collection easy, thus providing ease in accessing historical reports and comparing across various periods.

IV. Proposed solution

The system would have detailed and user-friendly analysis of public annual reports with the possibility of diverse data entry means. The user could upload PDF files with annual public reports or apply the search functionality of the chatbot to find [11] and download relevant reports regarding the specified company and year. Once the reports are uploaded, the text content will be extracted and broken down into more manageable pieces and interpreted in vector form-through embeddings-and stored efficiently inside a vector database, for example, Faiss. As soon as a user poses a question, the chatbot processes the query by converting it into embeddings; the app then conducts a semantic search of the stored data to decide what's the most relevant piece of text regarding the question asked. The chatbot then ranks the information from the selected source according to relevance and contextual importance.

Using LLM, the chatbot produces context-sensitive precise answers and directly answers questions of users regarding fact type, number, or type of reasoning questions. To provide better understanding in complex data, the chatbot provides data visualization in graphical trends and comparisons for those questions that contain numerical information. Furthermore, it also provides for comparative analysis when there are multiple reports uploaded, showing the comparison of differences and similarities that will give the user an overall view of the company's performance. There is also a FAQ available to its chatbot categorized into strategic goals, financial performance, operational highlights, risk mitigation, and initiatives toward sustainability; from there, one can easily find an answer to the most commonly asked question.

The inclusion of a text-to-speech function further enhances user accessibility of the chatbot, and the responses generated can now be received in listening format. All this puts into action sophisticated AI techniques that will perform multi-dimensional analysis on annual public reports, thus empowering stakeholders with a tool that does more than facilitate easy interpretation of data but also makes and executes decisions based on data-driven inputs with clarity and precision.

V. Methodology

Step 1: Uploading Files

The users can directly upload annual public report PDFs (single or multiple) into the chatbot.Additionally the user can search for a specific company's annual report by entering the company's name and desired year. The chatbot fetches relevant reports from the internet, which the user can download and then upload to the chatbot for analysis. The user uploads PDF files using Streamlit's file upload component. After selecting the file, the user clicks on the 'Process PDF' button to initiate processing.

Step 2: Processing the Files

To initiate the handling of a PDF file, the process_pdf() function is called. It makes use of the PdfReader class of the PyPDF2 library for the purpose of text extraction on each page of the PDF. The pages' extracted text is joined together to create one complete string that summarizes the contents of the document. The get_text_chunks() function is called here, which takes the created string and transforms it into smaller parts using LangChain's RecursiveCharacterTextSplitter. It uses RecusiveCharacterTextSplitter and splits the text in parts of 5000 characters with a 1000 character overlap between parts. Then, the get_vector_store() function is called, in which these text snippets are encoded as vector embeddings using Google Generative AI embeddings. The produced embeddings are stored in a vector database, built with the help of Weaviate Library for fast semantic searches.

Step 3: Building the Conversational Chain

To establish a conversational chain, the method get_conversational_chain() is invoked. This method employs the LangChain framework for building the conversational structure. We load the Gemini-pro model from the library which is called ChatGoogleGenerativeAI and assign it as the language model (LLM) intended for the generation of responses. There is a memory component in the form of ConversationBufferMemory which stores the conversation history that enables the chatbot to remember the flow of the dialogue. The dialog chain is set in such a way that the already created vector store is used as a retriever, making it possible to search for any relevant text chunk in accordance to the user's queries. The LLM is connected to the retriever and the prompt template by the ConversationalRetrievalChain.from_llm() method, thus forming a chain that makes use of retrieved information to generate a response.

Step 4: Handling User Queries

After processing and indexing the PDF document, the user can type his or her inquiry into the input area. The user's request is responded to by a conversational chain that is maintained in Streamlit's session state and is provided with the question as input. The request is then vectorized and the resulting vector is compared to all the existing vectors in the vectorized database in order to locate the most relevant text chunks. The system retrieves the most analogous vectors on the basis of their meanings and then invokes the LLM to assemble an answer. The history of dialogue is freshed with the user's query and the system's reply and is shown to the user.

Step 5: Generating FAQs

In order to create FAQs, certain Questions and categories are premade. The geography of the responses to the FAQ section is created in a manner similar to that of the user queries. The conversational outlay fetches and answers the questions using information stored in the vector database

Step 6: Displaying the Response

The application inspects the conversation logs that are maintained in Streamlit's session state and iteratively scans through the messages one by one. The formatting positions the user messages and AI messages alternating one below the other for the clear presentation of the dialogue. Further, these tabular results are known because of certain characters being present along with the AI's response (for instance, | and ---). This tabulated information is extracted, sanitized, and presented in a neat form. If there is no tabulated data, then the result is given to the end user as normal text. Additionally, a text-to-speech feature is available for users who prefer listening to the responses, ensuring a user-friendly experience.

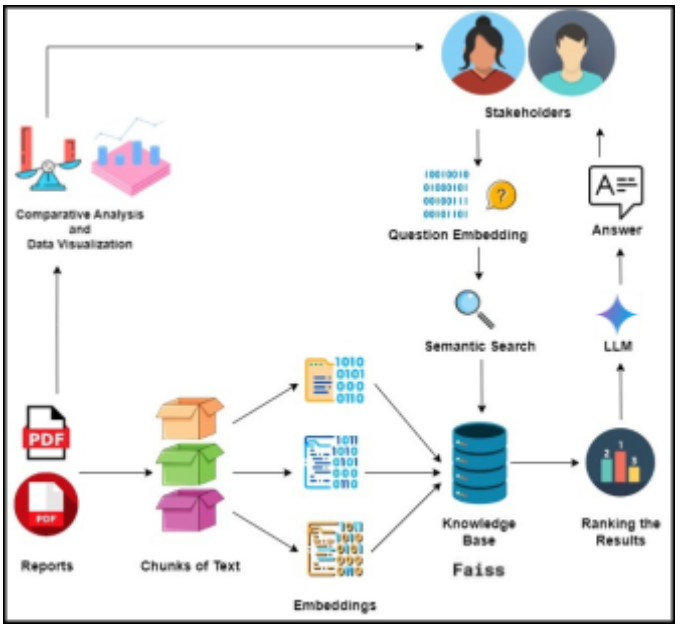


Fig. 8 System Architecture based

VI. Experiments and Results

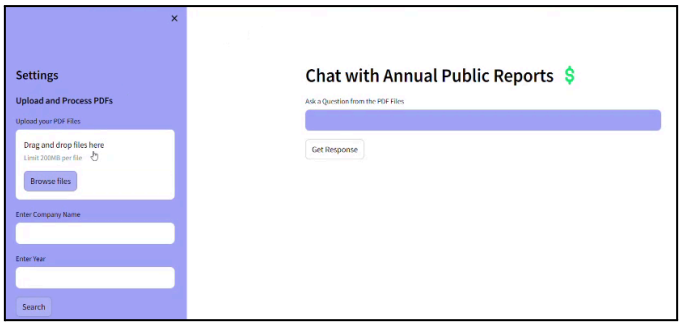


Fig 1. ChatBot Home Page

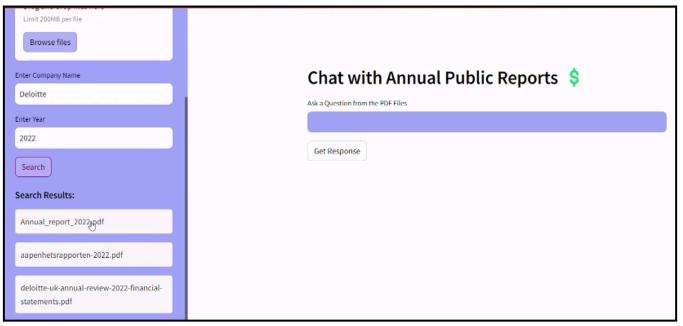


Fig 2. Results after entering the name of any company and particular year (Deloitte and 2022 in this case)

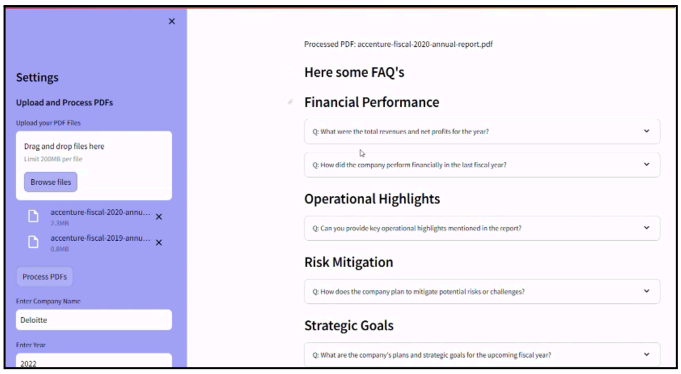


Fig 3. FAQs generated category wise with answers taken from the 2 uploaded PDFs

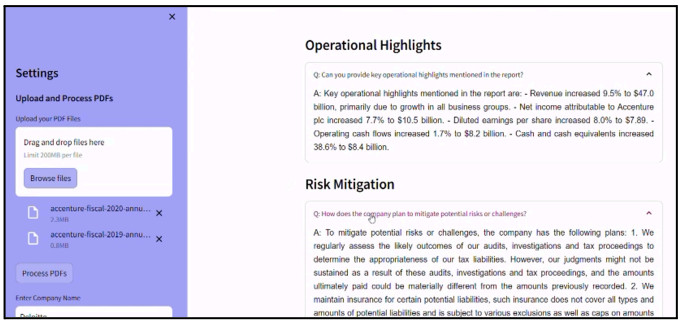


Fig. 4 Few answers retrieved from the two input documents in FAQ section

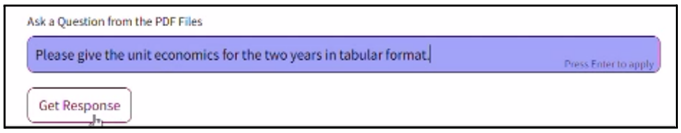
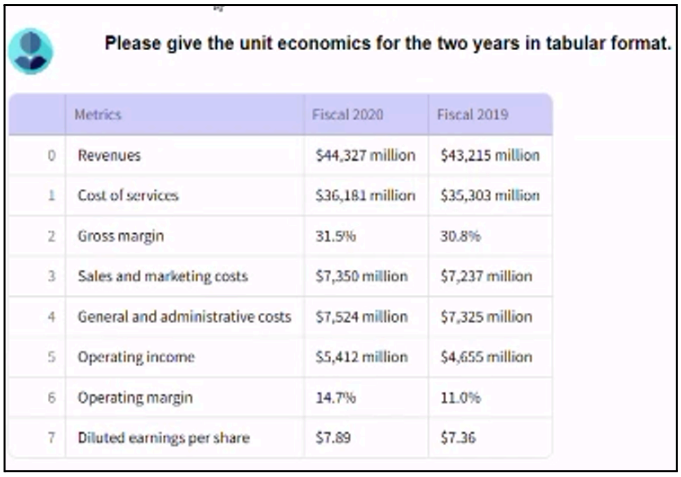


Fig. 5. Question entered by the user



	Metrics	Fiscal 2020	Fiscal 2019
0	Revenues	$44,327 million	$43,215 million
1	Cost of services	$36,181 million	$35,303 million
2	Gross margin	31.5%	30.8%
3	Sales and marketing costs	$7,350 million	$7,237 million
4	General and administrative costs	$7,524 million	$7,325 million
5	Operating income	$5,412 million	$4,655 million
6	Operating margin	14.7%	11.0%
7	Diluted earnings per share	$7.89	$7.36

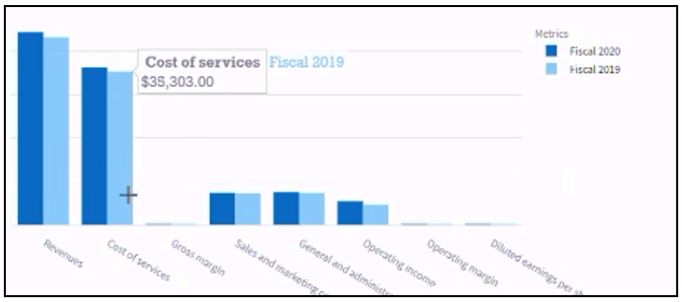


Fig 6. Answer to the question given by the user

VII. Technologies Used

A. *Google Gemini API*

Gemini is a family of highly capable multimodal models developed at Google [13]. The system is powered by the Gemini-1.5-flash model.

B. *FAISS (Facebook AI Similarity Search)*

Faiss is a library for ANNS. The core library is a collection of source files written in standard C++ without dependencies. Faiss is used in many configurations. Hundreds of vector search applications rely on it, both within Meta [14] and externally. The proposed system deals with a large amount of textual data. Hence, FAISS is always preferred for scalability as it avoids a drop in performance. The purpose of this utilization is to perform similarity search on the vectorized representations of the documents and for a quick and easy retrieval.

C. *Streamlit*

Streamlit is an open source Python framework that allows data scientists and AI/ML engineers to write dynamic data applications in a few lines of code [15]. In this framework, you can build interactive visualization plots, models, and dashboards without worrying about the underlying web framework or the deployment infrastructure used in the backend [16]. Though it is quite easy to build and deploy a Streamlit app, it is not scalable. So for large scale applications, frameworks like Flask, which is a Python-based framework, would be preferred.

D. *Langchain*

Langchain is an in-house built Large Language Model for any organization [17]. LLMs are being used very rapidly as they can effectively execute a vast range of tasks, some of which include essay composition, code writing, explanation, and debugging [18]. They take in a text string (prompt) and return a text string [18].

VIII. FUTURE WORK

The current implementation of the risk assessment system has considerable utility at accelerating evaluation of the financial documents and risk reports' compilation. However, there are multiple areas where the system can be extended and improved to better meet the needs of organizations and enhance its capabilities:

- Customizable Report Formats: Another important feature which should be anticipated in the future is the possibility of an implementation to define the layout and structure of the risk reports produced. It has become a norm for different companies to have preferred style of reporting, including sections preferred in the report, preferred kind of diagrams or tables, or industry-specific language. It was proposed that, through a reporting feature that can be customized, the user would be able to prepare the report according to internal rules or other requirements.
- Linking of documents to more than one document type: The present system of analyzing and interpreting the PDF-file format of financial reports may be expanded in subsequent versions for other formats, such as Excel tables, Word documents or even access to database information. This would enhance flexibility of the system in handling accounting information from various data feeds from different sources, and would also ensure compatibility to various documentation processes by companies.

IX. CONCLUSION

This research outlines a simple way of automating financial risk assessment through Gemini 1.5 Flash coupled with enhanced user interface. The system is extremely effective to scan financial documents, flag risks and offer actionable intelligence. Automating risk analysis therefore enables organizations to take quicker and better decisions. Additional additions such as report features, or industry-specific models will enhance the system's flexibility making the system relevant for companies as they grapple with sophisticated risk levels.

REFERENCES

[1] P. K. Aithal, D. A. U. and G. M., "Analyzing Tone of the Annual Report - An Indian Context," 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), THIRUVANANTHAPURAM, India, 2022, pp. 536-542, doi: 10.1109/SPICES52834.2022.9774247.
keywords: {Instruments;Signal processing algorithms;Companies;Signal processing;SPICE;Stock markets;Information technology;Portfolio Management;Tone;Initial Public Offerings;Parallel Algorithms;Message Passing Interface},

[2] Yuyan, Guo. "An analysis of the growing length of the annual reports." The Frontiers of Society, Science and Technology 5, no. 2 (2023).

[3] Nie, Yuqi, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, and Stefan Zohren. "A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges." arXiv preprint arXiv:2406.11903 (2024).

[4] Smailović, Jasmina & Žnidaršič, Martin & Valentinčič, Aljoša & Loncarski, Igor & Pahor, Marko & Martins, Pedro & Pollak, Senja. (2018). Automatic Analysis of Annual Financial Reports: A Case Study. Computación y Sistemas. 21. 10.13053/cys-21-4-2863.

[5] Srinivasan, Padmini, and Ana Cristina Marques. "Narrative analysis of annual reports: A study of communication efficiency." (2017).

[6] Mai, Feng, Shaonan Tian, Chihoon Lee, and Ling Ma. "Deep learning models for bankruptcy prediction using textual disclosures." European journal of operational research 274, no. 2 (2019): 743-758.

[7] Hadi, Muhammad Usman, Qasem Al Tashi, Abbas Shah, Rizwan Qureshi, Amgad Muneer, Muhammad Irfan, Anas Zafar et al. "Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects." Authorea Preprints (2024).

[8] Yu, Xinli, Zheng Chen, Yuan Ling, Shujing Dong, Zongyi Liu, and Yanbin Lu. "Temporal Data Meets LLM--Explainable Financial Time Series Forecasting." arXiv preprint arXiv:2306.11025 (2023).

[9] Lee, Jean, Nicholas Stevens, Soyeon Caren Han, and Minseok Song. "A survey of large language models in finance (finllms)." arXiv preprint arXiv:2402.02315 (2024).

[10] Olayinka, Aminu Abdulrahim. "Financial statement analysis as a tool for investment decisions and assessment of companies' performance." International Journal of Financial, Accounting, and Management 4, no. 1 (2022): 49-66.

[11] https://newsapi.org/docs

[12] Yang, Yi, Yixuan Tang, and Kar Yan Tam. "Investlm: A large language model for investment using financial domain instruction tuning." arXiv preprint arXiv:2309.13064 (2023).

[13] G. Team *et al.*, "Gemini: a family of highly capable multimodal models," *arXiv.org*, Dec. 19, 2023. https://arxiv.org/abs/2312.11805

[14] M. Douze et al., "The Faiss library," arXiv.org, Jan. 16, 2024. https://arxiv.org/abs/2401.08281

[15] "Streamlit Docs." https://docs.streamlit.io/

[16] Sreeram a, Adith & Sai, Jithendra. (2023). An Effective Query System Using LLMs and LangChain. International Journal of Engineering and Technical Research. 12.

[17] K. Pandya and M. Holia, "Automating Customer Service using LangChain: Building custom open-source GPT Chatbot for organizations," *arXiv.org*, Oct. 09, 2023. https://arxiv.org/abs/2310.05421

[18] Topsakal, Oguzhan & Akinci, T. Cetin. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. International Conference on Applied Engineering and Natural Sciences. 1. 1050-1056. 10.59287/icaens.1127.

Paper 4 :

AI driven Investment Insights Using ESG Prediction Models

Mrs. Sujata Khedkar[#1],Ketaki Dhananjay Nalawade[#2], Tasmiya Sarfaraz Khan[#3],

Purtee Santosh Mahajan[#4], Srushti Satish Sambare[#5]

[#]*Computer Engineering Department,Vivekanand Education Society's Institute of Technology, Chembur, Mumbai - 400071, India*

[1]sujata.khedkar@ves.ac.in

[2]2021.ketaki.nalawade@ves.ac.in

[3]2021.tasmiya.khan@ves.ac.in

[4]2021.purtee.mahajan@ves.ac.in

[5]2021.srushti.sambare@ves.ac.in

***Abstract*— In order to develop a tried-and-true investment strategy that takes advantage of the correlation between environmental, social, and governance (ESG) factors and financial success, this study looks at the statistical influence of ESG concerns on economic investment. As mandated reporting requirements are implemented and investors take sustainability into account when making investment choices, there is an increasing demand for transparent and reliable ESG ratings. The goal of this paper is to examine several approaches that may be applied to forecast the ESG ratings of businesses.**

Keywords— ESG factors , financial performance , ESG ratings.

I . Introduction

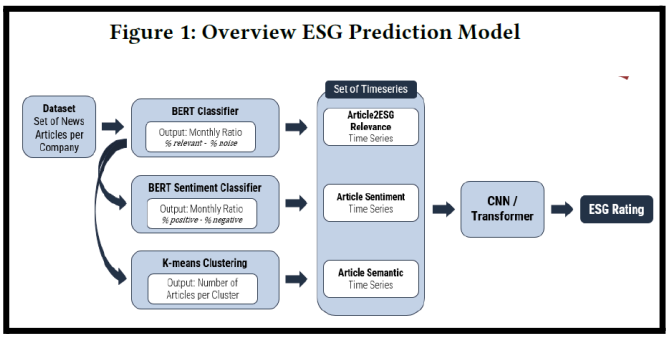
Environmental, social, and governance, or ESG, compliance is a relatively new field of study that has grown in prominence recently due to the importance of these issues in the worldwide discourse[3]Businesses, investors, governments, and society at large all care about ESG and related topics, and they are increasingly prioritizing ESG compliance when entering into agreements[1].Consequently, as national policies and investor interest in ESG investments take corporate ESG factors into account, the size of the investments is also growing quickly, and each company's expectations and interest in ESG management are rising sharply. However, there is a dearth of data-driven analysis and discussion of ESG trends[5].

II . Motivation

Predicting ESG ratings entirely automatically using natural language processing (NLP) algorithms, without the need for human judgement, could save costs for businesses and, most importantly, be accessible to small and medium-sized businesses. Additionally, automatic methods ought to guarantee that the ratings are clear and perhaps reconstructable by any stakeholder[1].ESG standards are used by most socially conscious investors to evaluate investments.Investors frequently use the term "environmental, social, and governance" (ESG) to evaluate corporate policies and forecast future financial performance[4].

III. Literature review

[1] **Model inputs:**
Text Classification: News articles were categorized as either ESG-relevant or irrelevant using BERT-based algorithms. A weak-supervision strategy was utilized to classify a portion of data using semantic similarity between articles and ESG category definitions.
Sentiment Analysis: SieBERT was used to forecast the sentiment of ESG-related publications, resulting in a sentiment time series for each firm. The sentiment ratios for good and negative news were calculated monthly.
Semantic analysis involved grouping the articles by content using DistilBERT embeddings and a k-means clustering technique. Six clusters were found, with each depicted as a time series that tracked dominating subjects for each organization.



ESG Rating Prediction Models : Several deep learning models were used to forecast ESG scores based on the generated time series. Four models were assessed: basic CNN (convolutional neural network) , Deep CNN , CNN with single or many Transformer layers.

The models saw the task as either a classification or regression issue, with the input being a 9x12 matrix comprising nine categories of monthly time series data for each organization. The appendix included model hyperparameters, optimisers, and settings, as well as early termination conditions to prevent overfitting.

[2] Four testable hypotheses or notions are formulated to address the research question: (1) HPS1: Firms with lower

ESG risk ratings tend to have lower returns on investment; (2) HPS2: The risk associated with stocks tends to increase in tandem with ESG risk level; (3) HPS3: Firms with lower ESG Risk achieve higher returns than those with higher ESG Risk; and (4) HPS4: Invest stocks with low ESG risk and short stocks with high ESG risk to achieve extraordinary returns.

The following are the steps to test Hypotheses HPS1 and HPS2:
✓ Load two .csv datasets (S&P500, ESG data).
✓ Calculate the Expected Returns E[r] and the total risk for all stocks.
✓ Merge the data frame with the E[r], total risk, and ESG data. The merged data frame is shown in Table 2.
✓ Graph the Expected Returns and ESG Risk.
✓ Graph Expected Returns and Total Risk.
✓ Graph Total Risk and ESG Risk.
✓ Investigate correlations across all variables.
✓ Remove all rows with missing ESG Risk observations.
✓ Find the Correlation between E[r] and ESG Risk using: ‘‘np.corrcoef(df_master[‘Expected Return’], df_master[‘ESG Risk Score’])’’

The process of estimating the returns of an ESG portfolio involves the following steps:
✓ Import pandas and load the ESG.csv file.
✓ Handling of missing observations.
✓ Sort the ESG Firms into quintile buckets (‘Q1,’ ‘Q2,’ ‘Q3,’ ‘Q4,’ ‘Q5’) using the qcut method available in the Pandas library.
‘‘df_esg[‘Quintile ESG Rank’] = df_esg[‘ESG Risk Score’].transform(lambda x: pd.qcut(x, 5, labels = quintile_rank_labels))’’
✓ Merge S&P 500 Returns and ESG risk data frames.
‘‘df_returns_esg = df_returns.merge(df_esg, left_on =‘Ticker’, right_on =‘Symbol’)’’
✓ We find the average return for every stock belonging to a particular portfolio based on its ESG rank. Estimate the return on equally weighted ESG portfolios by calculating the average return, grouped by date and ESG Rank
‘‘quintile_returns = df_returns_esg.groupby([‘Date’, ‘Quintile ESG Rank’])[‘Returns’].mean()’’.
✓ We drop the missing observations in the quintile_ returns data frame. The daily quintile-returns data frame displays the average returns of ESG-risk firms, which are sorted into buckets by quintile
Calculate the average E[r] across quintile ESG Portfolios using ‘‘quintile_returns.mean()’’.

[3] **Data Collection:** The study makes use of the Sustainalytics ESG Risk Rating emphasis Database, which contains ESG ratings for 5,012 organizations across 11 sectors, with an emphasis on financial institutions. **Textual and Empirical Analysis:**Topic Modelling (LDA) is a machine learning approach used to extract latent themes from analysts' textual comments on ESG ratings. The model found 13 key topics related to ESG performance across industries. **Sentiment Analysis:** Natural Language Processing (NLP) was used to identify positive or negative attitudes in analyst remarks and relate them to particular ESG problems.**Sector Focus:** The research looked at sector-specific ESG matters, with a particular emphasis on governance difficulties for financial institutions and environmental concerns for sectors such as energy and utilities. **Comparison of Best and Worst Performers:** Financial firms were classified as "Best" or "Worst" based on ESG ratings, and their relevant themes and attitudes were examined.

[4] **Data Collection :** ESG ratings for India's top 500 firms were gathered from the MSCI (Morgan Stanley Capital International) and Refinitiv platforms. Financial measures such as ROA (Return on Assets) and ROE (Return on Equity) were also acquired using the Trendlyne platform. **Machine learning models include:** Several machine learning approaches, including Regression, were used to investigate the link between ESG disclosure and company performance. Random Forest, Support Vector Regression, K-Nearest Neighbour, and Neural Networks were the models utilized in the study.**Measuring performance :**

The study focused on two financial measures (ROA and ROE) to assess business performance and how ESG ratings affected these ratios. **Analysis:**The study employed statistical approaches to investigate how ESG disclosures affected financial results and sustainable investments, with the goal of determining the impact of ESG transparency on decision-making and corporate performance.

[5] **Data collection:** Between May 2006 and December 2021, 16 media outlets provided ESG-related news stories. Following filtering, 7,049 articles were examined. **Period Division:** The data was separated into three periods. ESG was first introduced between 2006 and 2018. 2019-2020: ESG disclosure regulations become more widespread and are adopted by large institutions. 2021: Increased ESG implementation following COVID-19. **Topic Modeling:** The Latent Dirichlet Allocation (LDA) technique was used to determine major ESG topics and trends for each era.

[6] **Data collection:** It makes use of the 850,000 ESG-related stories in the Dow Jones News dataset.**Text parsing:** Stanford CoreNLP is used to process the text in order to extract entities (such as individuals or organizations) and the connections between them.Creating a **Knowledge Graph:**Relationship and Entity Extraction:

Verbs and nouns are recognised to create triples (subject-predicate-object).Refinement involves the use of semantic tools such as WordNet and sentence transformers to integrate related items and relationships.**Evaluation:** The method's accuracy in identifying legitimate relationships in news items was 85%.**ESG Analysis:** The knowledge graph that resulted (7.2 million statements, 4 million entities) showed patterns in the discourse surrounding ESG, including increasing attention to corporate governance, gender identity, and climate change over time.

[11] **Framework:** ESGReveal comprises three components. The ESG Metadata Module establishes ESG standards, criteria, and indications for data extraction. The Report Preprocessing Module structures ESG reports by removing text and table data and prepares them for analysis. The LLM Agent Module retrieves and extracts particular ESG data based on metadata and structured reports.ESG reports are analyzed to form a knowledge base. LLMs then retrieve and extract pertinent data using prompts generated by the ESG metadata. **Application:** ESGReveal was tested on ESG reports from 166 Hong Kong Stock Exchange businesses, and several LLMs (including GPT-4) were evaluated for data extraction accuracy. **Results:** GPT-4 was the most accurate, with 76.9% for data extraction and 83.7% for disclosure analysis. The study also discovered inconsistencies in company ESG disclosures, notably in environmental and social reports.

[12] Data gathered from 348 Indonesian non-financial enterprises (2021-2022).The variables include idiosyncratic risk, financial report quality (earnings management), ESG disclosure (GRI standards), risk disclosure (COSO ERM), and audit quality.**Analysis:** Structural Equation Modeling (SEM) was used to analyze correlations and the moderating influence of audit quality. **Findings:** ESG disclosure and audit quality decreased idiosyncratic risk, although financial report quality and risk disclosure had no significant effect. The influence of the other factors on risk was not enhanced by audit quality.

[13] Data collection makes use of Refinitiv's ESG ratings and basic data. The Heterogeneous Ensemble Model enhances prediction accuracy by combining machine learning methods such as XGBoost, CatBoost, and feedforward neural networks. Feature Selection extracts pertinent financial indicators from the core information. The model is trained using past data and tested using data that hasn't been seen before.By offering a more unbiased, data-driven method, it seeks to address the shortcomings of conventional ESG ratings.Model performance is evaluated using a variety of metrics.

[15] **Data collection:** phrases classified as irrelevant, quasi-related, or pertinent to ESG subjects are gathered from Corporate Sustainability Reports (CSRs).**Model training:** Using this dataset, a Transformer-based model (such as BERT) is refined to categorize phrases according to their applicability to ESG considerations.**Transfer Learning:** To illustrate the model's capacity for generalization, the trained model is applied to earnings call transcripts in order to identify ESG conversations without the need for additional training.**Evaluation:** BERT obtained the highest score (78.3%) when the model's performance was assessed using F1-scores.

[16] **Data Collection:** Three thousand headlines from The New York Times, Reuters, and The Independent were collected.**Annotation:** Participants categorized headlines and determined the importance of ESG factors; inter-annotator agreement was evaluated for consistency.**Sentiment Analysis:** A variety of ML and DL techniques were used to categorize headlines as neither positive nor negative.**Training the Model:** Random Under-Sampling was employed to correct class imbalance in an annotated dataset.
Based on weighted sentiment ratings, the ESG-Miner tool computed scores for the three ESG domains.**Evaluation:** The accuracy of the tool was assessed by contrasting its performance with manual annotations.**Validity Considerations:** Subjective annotations, sample size, and other possible threats to validity were examined.

IV. References

[1] Tanja Aue, Adam Jatowt, Michael Färber Predicting Companies' ESG Ratings from News Articles Using Multivariate Time Series Analysis arXiv:2212.11765 [q-fin.GN] https://doi.org/10.48550/arXiv.2212.11765

[2] K. R. Teja and C. -M. Liu, "ESG Investing: A Statistically Valid Approach to Data-Driven Decision Making and the Impact of ESG Factors on Stock Returns and Risk," in IEEE Access, vol. 12, pp. 69434-69444, 2024, doi: 10.1109/ACCESS.2024.3401873.

[3] Marco Mandas, Oumaima Lahmar, Luca Piras, Riccardo De Lisa,ESG in the financial industry: What matters for rating analysts?,Research in International Business and Finance, Volume 66,2023,102045,ISSN 0275-5319, https://doi.org/10.1016/j.ribaf.2023.102045.

[4] E. Twinamatsiko and D. Kumar, "Incorporating ESG in Decision Making for Responsible and Sustainable Investments using Machine Learning," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1328-1334, doi: 10.1109/ICEARS53579.2022.9752343.

[5] H. Seo, D. H. Jo and Z. Pan, "Big Data Analysis of 'ESG' News Using Topic Modeling," 2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science (BCD), Danang, Vietnam, 2022, pp. 183-187, doi: 10.1109/BCD54882.2022.9900604.

[6] S. Angioni, S. Consoli, D. Dessì, F. Osborne, D. Reforgiato Recupero and A. Salatino, "Exploring Environmental, Social, and Governance (ESG) Discourse in News: An AI-Powered Investigation Through Knowledge Graph Analysis," in IEEE Access, vol. 12, pp. 77269-77283, 2024, doi: 10.1109/ACCESS.2024.3407188.

[7] Jaeyoung Lee, Misuk Kim, ESG information extraction with cross-sectoral and multi-source adaptation based on domain-tuned language models, Expert Systems with Applications, Volume 221,2023,119726,ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.119726.

[8] Shen, H. Assessment of financial risk pre-alarm mechanism based on financial ecosystem using BPNN and genetic algorithm. Soft Comput 27, 19265–19279 (2023). https://doi.org/10.1007/s00500-023-09317-z

[9] T. R. Teor, I. A. Ilyina and V. V. Kulibanova, "The Influence of ESG-concept on the Reputation of High-technology Enterprises," 2022 Communication Strategies in Digital Society Seminar (ComSDS), Saint Petersburg, Russian Federation, 2022, pp. 184-189, doi: 10.1109/ComSDS55328.2022.9769074.

[10] Gunnar Friede, Timo Busch & Alexander Bassen (2015) ESG and financial performance: aggregated evidence from more than 2000 empirical studies, Journal of Sustainable Finance & Investment, 5:4, 210-233, DOI: 10.1080/20430795.2015.1118917

[11] Yi Zou, Mengying Shi, Zhongjie Chen, Zhu Deng, ZongXiong Lei, Zihan Zeng, Shiming Yang, HongXiang Tong, Lei Xiao, Wenwen Zhou ESGReveal: An LLM-based approach for extracting structured data from ESG reports https://arxiv.org/abs/2312.17264 [cs.CL]

[12] Arfiansyah, Z., Murwaningsari, E., & Mayangsari, S. (2024). Financial Report Quality, ESG Disclosure, Risk Disclosure, Audit Quality, and Idiosyncratic Risk: Evidence from Indonesian Companies. Journal of Risk and Auditing, 6(1), 4839-4858.

[13] Krappel, M., Bogun, M., & Borth, D. (2021). Predicting ESG Ratings Using a Heterogeneous Ensemble Model. In KDD-MLF '21, August 14–18, 2021, Singapore.

[14] Sokolov, Alik & Mostovoy, Jonathan & Ding, Jack & Seco, Luis. (2021). Building Machine Learning Systems for Automated ESG Scoring. The Journal of Impact and ESG Investing. 1. 39-50. 10.3905/jesg.2021.1.010.

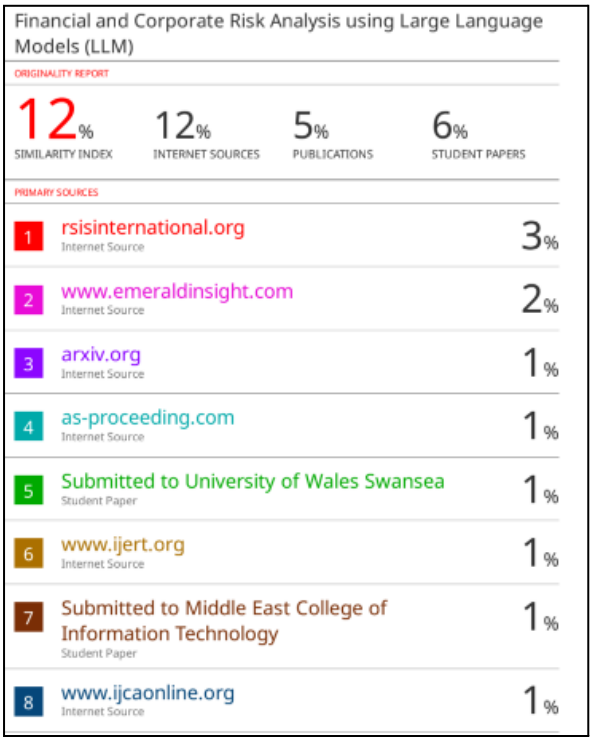
[15] Raman, Natraj & Bang, Grace & Nourbakhsh, Armineh. (2020). Mapping ESG Trends by Distant Supervision of Neural Language Models. Machine Learning and Knowledge Extraction. 2. 453-468. 10.3390/make2040025.

[16] Jannik Fischbach, Max Adam, Victor Dzhagatspanyan, Daniel Mendez, Julian Frattini, Oleksandr Kosenkov, Parisa Elahidoost Automatic (2024).ESG Assessment of Companies by Mining and Evaluating Media Coverage Data: NLP Approach and Tool https://arxiv.org/abs/2212.06540 [cs.IR]

b. Certificate of publication

c. Plagiarism report

Paper 1 : Financial and Corporate Risk Analysis using Large Language Models.



Financial and Corporate Risk Analysis using Large Language Models (LLM)

ORIGINALITY REPORT

12%	12%	5%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	rsisinternational.org Internet Source	3%
2	www.emeraldinsight.com Internet Source	2%
3	arxiv.org Internet Source	1%
4	as-proceeding.com Internet Source	1%
5	Submitted to University of Wales Swansea Student Paper	1%
6	www.ijert.org Internet Source	1%
7	Submitted to Middle East College of Information Technology Student Paper	1%
8	www.ijcaonline.org Internet Source	1%

Paper 2 : MetaData Extraction from Legal Contracts using Large Language Models

Metadata Extraction from Legal Contracts using LLM

ORIGINALITY REPORT

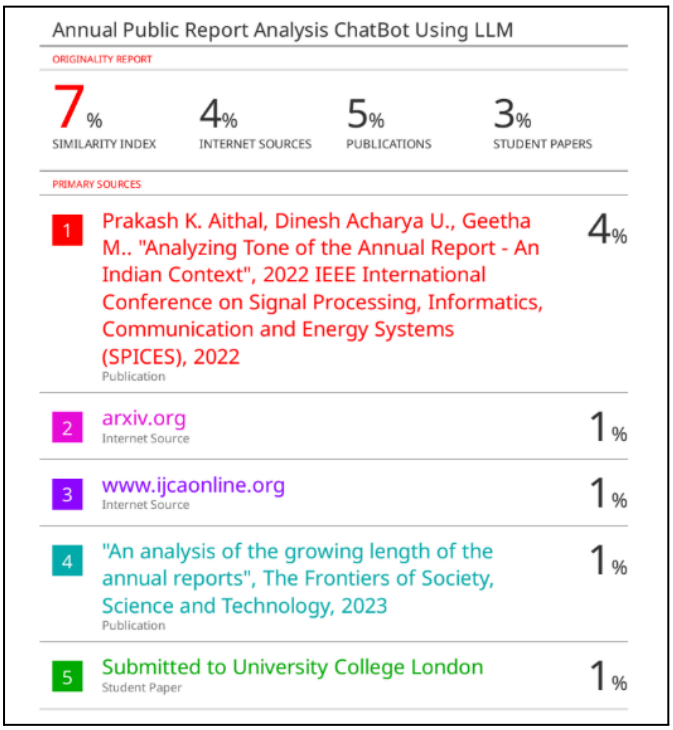
5%	2%	4%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Legal Knowledge and Information Systems", IOS Press, 2023 Publication	3%
2	www.gabormelli.com Internet Source	1%
3	www.ijcaonline.org Internet Source	1%
4	Submitted to HCUC Student Paper	1%

Exclude quotes On
Exclude bibliography On
Exclude matches < 1%

Paper 3 : Annual Public Report Analysis Chatbot using LLM



Annual Public Report Analysis ChatBot Using LLM

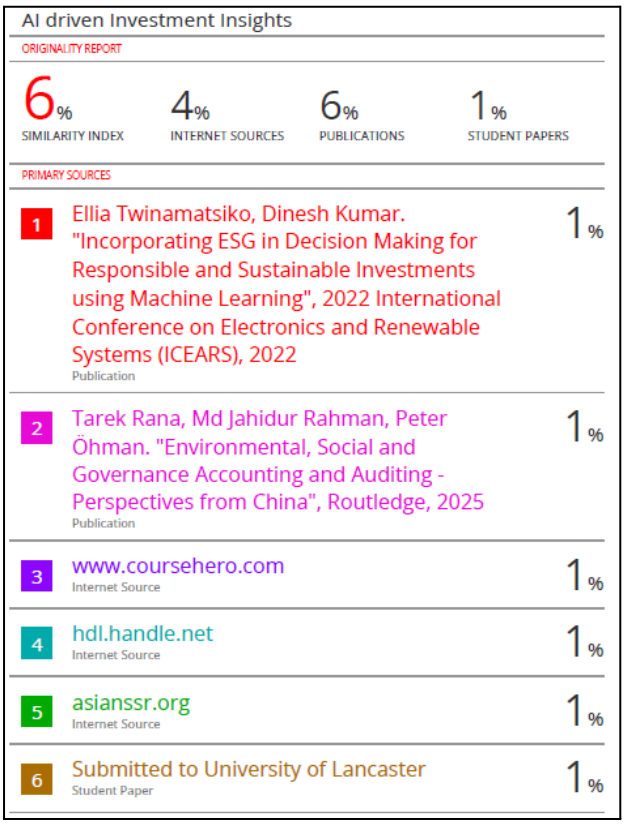
ORIGINALITY REPORT

7%	4%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Prakash K. Aithal, Dinesh Acharya U., Geetha M.. "Analyzing Tone of the Annual Report - An Indian Context", 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), 2022 Publication	4%
2	arxiv.org Internet Source	1%
3	www.ijcaonline.org Internet Source	1%
4	"An analysis of the growing length of the annual reports", The Frontiers of Society, Science and Technology, 2023 Publication	1%
5	Submitted to University College London Student Paper	1%

Paper 4 : AI driven Investment Insights Using ESG Prediction Models



AI driven Investment Insights

ORIGINALITY REPORT

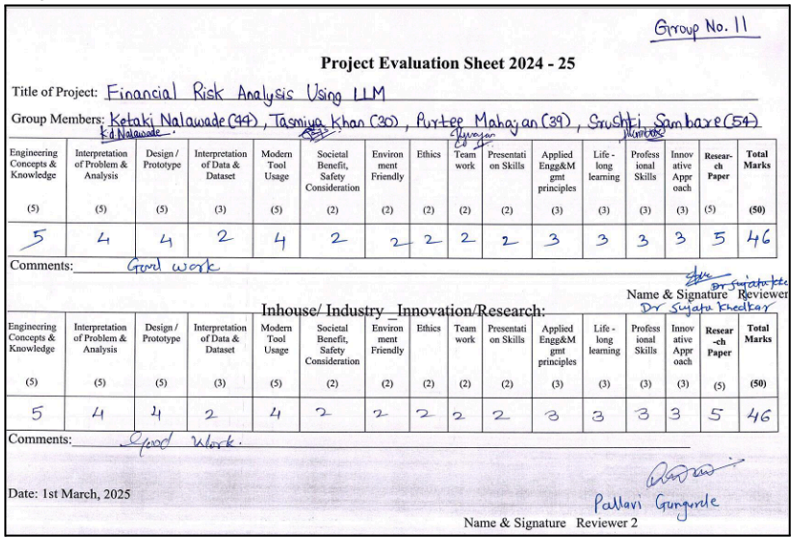
6%	4%	6%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ellia Twinamatsiko, Dinesh Kumar. "Incorporating ESG in Decision Making for Responsible and Sustainable Investments using Machine Learning", 2022 International Conference on Electronics and Renewable Systems (ICEARS), 2022 Publication	1%
2	Tarek Rana, Md Jahidur Rahman, Peter Öhman. "Environmental, Social and Governance Accounting and Auditing - Perspectives from China", Routledge, 2025 Publication	1%
3	www.coursehero.com Internet Source	1%
4	hdl.handle.net Internet Source	1%
5	asianssr.org Internet Source	1%
6	Submitted to University of Lancaster Student Paper	1%

d. Project review Sheet

Project Review Sheet 1 :



Group No. 11

Project Evaluation Sheet 2024 - 25

Title of Project: Financial Risk Analysis Using LLM

Group Members: Ketaki Nalawade (94), Tasmiya Khan (30), Purtej Mahajan (39), Srushti Sambare (54)

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation on Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
5	4	4	2	4	2	2	2	2	2	3	3	3	3	5	46

Comments: Good work

Name & Signature Reviewer: Dr Sujata Khedkar

Inhouse/ Industry _Innovation/Research:

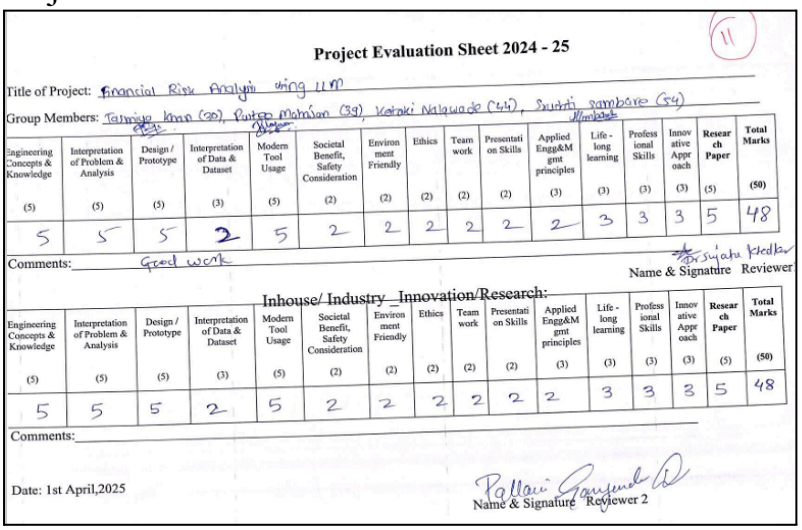
Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation on Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
5	4	4	2	4	2	2	2	2	2	3	3	3	3	5	46

Comments: Good Work.

Date: 1st March, 2025

Name & Signature Reviewer 2: Pallavi Gungarde

Project review sheet 2 :



11

Project Evaluation Sheet 2024 - 25

Title of Project: Financial Risk Analysis using LLM

Group Members: Tasmiya Khan (30), Purtej Mahajan (39), Ketaki Nalawade (41), Srushti Sambare (54)

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation on Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
5	5	5	2	5	2	2	2	2	2	2	3	3	3	5	48

Comments: Good work

Name & Signature Reviewer: Dr Sujata Khedkar

Inhouse/ Industry _Innovation/Research:

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation on Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
5	5	5	2	5	2	2	2	2	2	2	3	3	3	5	48

Comments:

Date: 1st April,2025

Name & Signature Reviewer 2: Pallavi Gungarde

2. Award certificate for project competition

a. Pradarshini 25 Winner

Vivekanand Education Society's Institute Of Technology, Mumbai

An Autonomous Institute affiliated to University of Mumbai, approved by AICTE

In association with VESIT - IQAC & VESIT - IIC

Department Of Computer Engineering

CERTIFICATE OF EXCELLENCE

This to certify that Tasmiya Khan of D17A has won 1st prize at Department Technology Day, **Pradarshini '25** conducted on 21st March 2025.

DR. MRS. NUPUR GIRI, HOD, COMPUTER DEPARTMENT

DR. MRS. GRESHA BHATIA, DHOD, COMPUTER DEPARTMENT

PROJECT COORDINATOR, COMPUTER DEPARTMENT

Vivekanand Education Society's Institute Of Technology, Mumbai

An Autonomous Institute affiliated to University of Mumbai, approved by AICTE

In association with VESIT - IQAC & VESIT - IIC

Department Of Computer Engineering

CERTIFICATE OF EXCELLENCE

This to certify that Srushti Sambare of D17A has won 1st prize at Department Technology Day, **Pradarshini '25** conducted on 21st March 2025.

DR. MRS. NUPUR GIRI, HOD, COMPUTER DEPARTMENT

DR. MRS. GRESHA BHATIA, DHOD, COMPUTER DEPARTMENT

PROJECT COORDINATOR, COMPUTER DEPARTMENT


Vivekanand Education Society's Institute Of Technology, Mumbai
An Autonomous Institute affiliated to University of Mumbai, approved by AICTE
In association with VESIT - IQAC & VESIT - IIC
Department Of Computer Engineering

CERTIFICATE OF EXCELLENCE

This to certify that Purtee Mahajan
of DITA has won 1st prize at Department Technology Day, **Pradarshini ' 25** conducted on 21st March 2025.

DR. MRS. NUPUR GIRI, HOD, COMPUTER DEPARTMENT
DR. MRS. GRESHA BHATIA, DHOD, COMPUTER DEPARTMENT
PROJECT COORDINATOR, COMPUTER DEPARTMENT


Vivekanand Education Society's Institute Of Technology, Mumbai
An Autonomous Institute affiliated to University of Mumbai, approved by AICTE
In association with VESIT - IQAC & VESIT - IIC
Department Of Computer Engineering

CERTIFICATE OF EXCELLENCE

This to certify that Ketaki Nalawade
of DITA has won 1st prize at Department Technology Day, **Pradarshini ' 25** conducted on 21st March 2025.

DR. MRS. NUPUR GIRI, HOD, COMPUTER DEPARTMENT
DR. MRS. GRESHA BHATIA, DHOD, COMPUTER DEPARTMENT
PROJECT COORDINATOR, COMPUTER DEPARTMENT

b. Habitia (18th October 2024) :


VIVEKANAND EDUCATION SOCIETY
H.A.B.I.T FOUNDATION

CERTIFICATE
OF ACHIEVEMENT
This certificate is presented to

Srushti Sambare

For Winning 'SECOND PRIZE' in HABITia's 'Pitch Perfect' competition organised by HABIT Foundation Incubation Center , VES Trust held on 18th October 2024

Mr. Vijay Talreja, Director, HABIT FOUNDATION VES TRUST
Dr.Jyoti Chandwani, Nodal Officer, HABIT Foundation, VES Trust, Director E-Cell VESIM


VIVEKANAND EDUCATION SOCIETY
H.A.B.I.T FOUNDATION

CERTIFICATE
OF ACHIEVEMENT
This certificate is presented to

Tasmiya Khan

For Winning 'SECOND PRIZE' in HABITia's 'Pitch Perfect' competition organised by HABIT Foundation Incubation Center , VES Trust held on 18th October 2024

Mr. Vijay Talreja, Director, HABIT FOUNDATION VES TRUST
Dr.Jyoti Chandwani, Nodal Officer, HABIT Foundation, VES Trust, Director E-Cell VESIM


VIVEKANAND EDUCATION SOCIETY
H.A.B.I.T FOUNDATION

CERTIFICATE
OF ACHIEVEMENT
This certificate is presented to

Ketaki Nalawade

For Winning 'SECOND PRIZE' in HABITia's 'Pitch Perfect' competition organised by HABIT Foundation Incubation Center , VES Trust held on 18th October 2024

Mr. Vijay Talreja, Director, HABIT FOUNDATION VES TRUST
Dr.Jyoti Chandwani, Nodal Officer, HABIT Foundation, VES Trust, Director E-Cell VESIM


VIVEKANAND EDUCATION SOCIETY
H.A.B.I.T FOUNDATION

CERTIFICATE
OF ACHIEVEMENT
This certificate is presented to

Purtee Mahajan

For Winning 'SECOND PRIZE' in HABITia's 'Pitch Perfect' competition organised by HABIT Foundation Incubation Center , VES Trust held on 18th October 2024

Mr. Vijay Talreja, Director, HABIT FOUNDATION VES TRUST
Dr.Jyoti Chandwani, Nodal Officer, HABIT Foundation, VES Trust, Director E-Cell VESIM