FSE 570 Capstone Project Proposal
Earnings Call and Risk Intelligence Engine for Financial Decision Support

Shalin N. Bhavsar sbhavsa8@asu.edu
Ramanuja M. A. Krishna rkrish79@asu.edu
Freya H. Shah fshah14@asu.edu
Kruthika Suresh ksures21@asu.edu
Harihara Y. Veldanda hveldan1@asu.edu

Arizona State University

1. Problem Statement

Every public company in the United States hosts 10,000 earnings calls per annum, which generate transcripts ranging from 8,000 to 15,000 words. These transcripts include forward-looking statements, risk disclosures, and language patterns critically important for assessing earnings quality and downside risk. Yet, their unstructured and lengthy nature makes systematic analysis infeasible using manual review. The existing practice followed is manual reading, fundamental keyword search, or lagged market response. Manual analysis is a time-consuming process that can take hours or days for each earnings report. Furthermore, such a cycle does not scale, thereby creating early signals of suspected earnings manipulations or weakening fundamentals, and thus can only be identified in terms of their delayed market reactions and short-run mispricing associated with earnings announcement episodes. This project develops an Earnings Call and Risk Intelligence Engine that automatically infers sentiment, linguistic uncertainty, and incorporates signals from earnings calls with structured financial data, cutting analysis times from days to minutes. This problem qualifies as a data-driven engineering issue since it demands the development of generalized pipelines for ingest, preprocess, and integrate large volumes of heterogeneous unstructured text and structured financial time series data, followed by the application of well-developed machine learning models with tight temporal constraints.
The system must meet the following criteria:

- Scalable processing of over 10,000 earnings call transcripts every year.
- Extraction of sentiment, risk, and uncertainty signals both accurately and in an interpretable manner.
- The integration of unstructured text with structured financial time-series data.
- Predictive outputs to aid in timely post-earnings decisions.

Primary stakeholders include institutional investors, equity analysts, portfolio managers, and corporate risk teams. Improved interpretation of earnings disclosures might also, at a broader level, reduce the incidence of mispricing, improve capital allocation efficiency, and add to the market stability in periods of economic uncertainty.

2. Data Sources

Among the data sets used are many large, heterogeneous, and publicly available:

- **Earnings Call Transcripts:** There are 13,500 transcripts available from Seeking Alpha (seekingalpha.com/earnings) and SEC EDGAR (sec.gov/edgar) (2018-2023, S&P 500 companies). Each transcript is structured with metadata fields (ticker, date, quarter) and content fields (speaker roles, prepared remarks, Q&A sections, utterance text), with lengths ranging from 8,000 to 15,000 words. These transcripts provide the unstructured text needed for extracting sentiment, uncertainty, and risk signals critical to the prediction task.
- **Financial and Market Data**: Quarterly financial statement data reported on SEC EDGAR and formatted as tabular records with 28 attributes or features, such as revenue, EPS, net income, operating margins, leverage, and cash flows. Daily stock information from Yahoo Finance (finance.yahoo.com) contains OHLC information, volume, and adjusted close prices normalized for stock splits and dividends, and can be matched on stock and date.

Key challenges include cleaning the transcripts, segmenting speakers, performing temporal alignment between text and financial time-series data, and handling missing/noisy records. Other sourcing challenges - API rate limits, inconsistent structures of HTMLs across transcripts, and incomplete historical coverage for smaller-cap firms. All datasets are publicly available with no personally identifiable information.

3. Methodology

- **Prediction Target:** Binary classification of 3-day post-earnings abnormal returns, measured as market-adjusted returns, stock returns - S&P 500 returns. 3-day window is adequate to capture post-earnings drift (Bernard and Thomas, 1989) while minimizing unrelated noise.
- **Text Processing**: Transcript normalization, tokenization, lemmatization, and speaker-role identification (management vs. analysts).
- **Feature Extraction**: FinBERT generates sentiment scores and 768-dim embeddings (Financial PhraseBank F1: 0.97 vs. BERT 0.89; publicly available; 110M parameters for deployment). Loughran-McDonald dictionary quantifies linguistic uncertainty. Management-analyst sentiment divergence captures Q&A tone shifts.
- **Data Integration**: Early fusion concatenates mean-pooled FinBERT embeddings (768-dim) with normalized quarterly financials (28 features) forming 796-dim vectors per event, enabling sentiment-fundamental interaction learning. Temporal alignment prevents lookahead bias.
- **Predictive Modeling**: Logistic regression baseline. XGBoost primary model (gradient boosting for non-linear interactions, native missing value handling, SHAP interpretability). Hyperparameters via grid search (learning rate: [0.01, 0.1], depth: [4, 6, 8], trees: [50, 100, 150]). Random forest validates architecture-independence. SMOTE balances positive/negative classes.
- **Validation:** 70/30 chronological split with 5-fold time-series CV maintaining temporal order. Metrics: accuracy, F1, ROC-AUC. Baselines: sentiment-only, financial-only. Success: $\geq$3% ROC-AUC improvement (statistically significant via DeLong test, $p < 0.05$).

4. Project Management: Timeline and Risks

Week	Milestone	Monitoring Criteria
1	Finalize scope, datasets, and evaluation metrics	Scope approved, API access verified
2	Collect earnings calls and financial data	$\geq$90% data collected
3	Transcript cleaning and normalization	Parsing pipeline operational
4	Speaker segmentation and temporal alignment	$\geq$85% transcripts parsed
5	Baseline financial-only modeling	Baseline ROC-AUC $\geq$0.55 established
6	NLP feature extraction (FinBERT)	FinBERT embeddings for 100% of transcripts
7	NLP-enhanced model training	$\geq$3% ROC-AUC over baseline
8	Model validation and comparison	Cross-validation stability
9	Error analysis and refinement	Sensitivity metrics reviewed
10	Final report and presentation	All deliverables completed

Status Check-ins:

- **Status Update 1 (Week 4)**: Cleaned dataset with $\geq$85% parsing success, baseline model with documented ROC-AUC $\geq$0.55, data quality issues identified with resolution plan.
- **Status Update 2 (Week 8)**: Achieved integrated models with $\geq$3% ROC AUC improvement over the baseline, cross-validation of the models done, error patterns and model limitations documented.

Key Risks and Mitigation:

- **Data Quality Risk**: Monitored weekly via automated parsing logs tracking success rate by source. In case of less than 80% in any one source, the assigned data engineer shall, within 48 hours, apply source-specific parsing rules or exclude the source. Contingency: Pre-identified backup sources for transcripts-FactSet, Capital IQ-if primary sources fail.
- **Model Performance Risk**: To be monitored after every training based on the ROC-AUC delta from the baseline. If it is less than 2% at the end of Week 7, the lead will convene a model review to institute a preset sequence of fallbacks, in order: (1) feature pruning using SHAP values, (2) changing to LightGBM with pre-selected hyperparameters, and (3) ensemble baseline models if all else fails.
- **Resource Constraint Risk**: Monitored by processing time logs per batch. Should any batch take more than 8 hours per 1,000-transcript batch, the team shall immediately move to stratified sampling-reduce to 9,000 transcripts while maintaining sector representation-or switch from FinBERT to the lightweight version, DistilFinBERT-to ensure meeting of the Week 6 deadline.