ABSTRACT

Financial documents such as regulatory filings, earnings call transcripts, financial newswire articles, and analytical narratives form the backbone of modern decision-making in capital markets. These documents contain dense numerical, textual, and semantic information that must be accurately interpreted, structured, and reasoned over by analysts, regulators, and automated systems alike. Despite significant progress in natural language processing, financial language understanding remains challenging due to domain-specific terminology, large and evolving label spaces, extreme document lengths, numerical grounding requirements, and the need for precise reasoning. Traditional discriminative models, while effective in constrained settings, often struggle to scale to extreme label regimes, generalize to unseen concepts, or maintain factual and numerical reliability.

This thesis advances a unifying perspective: *financial language understanding can be effectively framed as a structured generation problem, guided by explicit task instructions and semantic supervision*. Under this paradigm, large pre-trained language models are adapted in a parameter-efficient manner to generate semantically meaningful outputs such as label descriptions, salient summary points, relevant facts, and executable reasoning steps, instead of selecting from fixed label inventories. This shift enables improved scalability, interpretability, robustness to rare and unseen labels, and stronger generalization across diverse financial text sources.

The first part of this thesis focuses on *Extreme Financial Numeral Labeling (XFNL)*, where numerals in SEC filings must be tagged with the correct XBRL concepts from a taxonomy containing thousands of labels. We propose FLAN-FinXC, a parameter-efficient instruction-tuning framework built on FLAN-T5 that treats XBRL tagging as a generative task. By leveraging rich XBRL tag documentations as

semantic supervision and generating these descriptions directly, the proposed approach substantially outperforms state-of-the-art extreme classification baselines on the FiNER and FNXL datasets, while exhibiting strong zero-shot capability for unseen tags.

The second part addresses *bullet-point summarization of earnings call transcripts*, which are long, conversational financial documents requiring extreme compression while preserving factual and numerical fidelity. We propose FLAN-FinBPS, a scalable two-stage framework that combines unsupervised question-based contextual extraction with parameter-efficient instruction-tuned generative summarization. This design eliminates the need for costly supervised extractive training while significantly improving factual consistency, numerical precision, and overall summary quality. Extensive experiments on the ECTSum benchmark demonstrate substantial gains over existing state-of-the-art summarization methods.

The third part of the thesis investigates *financial numerical reasoning*, a task that requires identifying relevant facts from long financial contexts and performing multi-step numerical computations. We propose FINDER, a generative retriever–generator framework that replaces static retrieval with instruction-tuned fact extraction and employs Program-of-Thought reasoning for executable numerical inference. A dynamic in-context example selection strategy further improves robustness and generalization. The proposed method establishes new state-of-the-art results on FinQA and ConvFinQA.

The final part generalizes the structured generation paradigm to *multi-label text classification of financial and non-financial documents*. We introduce a domain-agnostic, parameter-efficient generative framework that generates label descriptions instead of predicting label indices, enabling effective semantic alignment between texts and labels. In the financial setting, the approach is evaluated on large-scale finance newswire and enterprise datasets involving market, risk, and event categorization. A dual-objective loss combining token-level supervision with representation-level semantic similarity further enhances performance, yielding consistent improvements over strong baselines, particularly on rare labels and in zero-shot settings.

Overall, this thesis demonstrates that instruction-guided, parameter-efficient generative modeling provides a powerful and unified foundation for financial language understanding. By explicitly incorporating task instructions, semantic grounding, and structured generation, the proposed methods advance the state of the art across multiple challenging financial NLP tasks and offer a scalable pathway toward robust, generalizable, and interpretable financial AI systems.

Keywords: Financial NLP, Instruction Tuning, Generative Models, Extreme Classification, XBRL Tagging, Bullet-Point Summarization, Earnings Call Transcripts, Financial Numerical Reasoning, Program-of-Thoughts, Multi-label Text Classification, Finance Newswire Analysis, Parameter-Efficient Fine-Tuning, LoRA, Large Language Models