TEXAS A&M UNIVERSITY
CORPUS CHRISTI

Thesis Defense:	Abiodun Adedeji
Title:	FinaGent: Financial LLM Benchmarking Pipeline
Committee Chair:	Dr. Wenlu Wang
Date and time:	Wednesday, July 8, 2026 3:30 pm - 5:00 pm
Place:	CI - 109

Abstract:

This paper introduces FinaGent, a benchmarking pipeline for large language models (LLMs) in the financial domain, specifically aimed at assessing and enhancing their reasoning abilities on SEC EDGAR filing data. The project focuses on four major publicly traded companies: JPMorgan Chase (JPM), Apple (AAPL), Pfizer (PFE), and ExxonMobil (XOM). It comprises five key steps: gathering financial data, building training datasets, fine-tuning with QLoRA, developing an agentic pipeline, and performing multi-metric evaluation.

The training dataset, containing 6,901 domain-specific examples, was assembled from FinQA (6,251 numerical reasoning questions), FinanceBench (150 open-ended questions), and 500 SEC EDGAR passage summaries. Qwen2.5-14B-Instruct was fine-tuned using QLoRA (4-bit NF4, LoRA rank-16, with 68.8 million trainable parameters), resulting in FinaGent-LLM. This model achieved a final training loss of 0.2188 over 1,227 steps on an NVIDIA A100- SXM4-40GB GPU.

A LangGraph-based agentic pipeline, called FinaGent-Agent, was developed with three nodes: Retriever, Calculator, and Reasoner. It was iterated through four versions (V1–V4). FinaGent-LLM achieved a 4.5% exact match rate on the FinQA benchmark, compared to 4.0% for the base model (+0.5%). On a custom 41-question benchmark, FinaGent-Agent V4 scored 2.83 out of 5 as an LLM-Judge, versus 3.54 for the base model, indicating that domain-specific RAG introduces noise on general benchmarks. Four failure modes of the agent were identified and addressed across different pipeline versions.