Project Report – Initial Review

Project Title: QnA Bot – AI in Finance

Submitted by: [Your Name]

Mentor: [Mentor Name]

Submitted to: [Institute/Organization Name]

# I. Project Overview

* With the rapid rise of AI in the financial sector, there is a growing demand for systems that can provide accurate and instant information to users.
* A Question-and-Answer (QnA) Bot focused on finance leverages natural language processing (NLP) and AI to address user queries related to finance, investment, economics, and banking in real-time.
* This project aims to develop an AI-powered QnA bot that can respond accurately to user queries using domain-specific datasets and pretrained transformer models.

# II. Problem Statement

* The finance domain is complex, filled with technical jargon, and often difficult for non-experts to understand.
* Searching through documentation or financial portals for answers can be time-consuming and confusing.
* There is a need for a conversational AI that understands financial terminology, retrieves accurate answers, and is accessible to both finance professionals and the general public.

# III. Objectives

* To research and understand existing QnA systems and how they are adapted to finance.
* To identify and analyze datasets suitable for training a finance-specific QA model.
* To finalize the tools, technologies, and models for building the system.
* To explore the limitations and scope of AI in financial question answering.

# IV. Research Work Done

* Literature Review: QnA Systems – BERT, T5, RoBERTa, FinBERT.
* QnA Techniques: Extractive vs. Abstractive QA.
* Application in Finance: Banking chatbots, robo-advisors, financial support systems.

# V. Datasets Identified

* FinQA: Financial QA dataset with numerical reasoning.
* FiQA: QA from financial forums and blogs.
* Financial PhraseBank: Common financial expressions.
* Wikipedia and Investopedia: Used as supplementary sources.

# VI. Tools & Technologies Proposed

* Python, HuggingFace Transformers, FinBERT.
* Streamlit or Flask for web UI.
* Pandas, NumPy, Scikit-learn, NLTK.

# VII. Proposed Methodology

* Data Collection and Preprocessing.
* Model Selection and Fine-tuning.
* Context Retrieval using TF-IDF or DPR.
* Answer Extraction using transformer models.
* User Interface development with Streamlit.

# VIII. Use Cases

* Educational Tool: For students learning finance.
* Support Assistant: For professionals needing quick reference.
* FAQ Automation: Replacing static FAQs with dynamic answers.

# IX. Challenges & Risks

* Limited labeled financial QA datasets.
* Complex and ambiguous financial terminology.
* Performance and speed of large language models.

# X. Expected Outcomes (Phase 1)

* Research documentation on QA in finance.
* Dataset evaluation and preparation.
* Model selection report.
* Architecture for prototype.

# XI. Timeline (Initial Phase)

* Week 1: Literature Review.
* Week 2: Dataset collection and curation.
* Week 3: Model review and comparison.
* Week 4: Proposal submission and feedback.

# XII. References

* FinQA Dataset - https://github.com/czyssrs/FinQA
* FiQA Dataset - https://sites.google.com/view/fiqa/home
* FinBERT - https://arxiv.org/abs/2006.08097
* HuggingFace Transformers - https://huggingface.co/transformers/
* Investopedia - https://www.investopedia.com/

# XIII. Acknowledgement

* I would like to thank my mentor [Mentor Name] for guiding me during the initial research phase.
* Thanks to [Institute/Organization Name] for providing this opportunity to explore AI applications in finance.