Fake News Detection Using Transformer + RAG – Final Project Report

1. Abstract This project presents an advanced Fake News Detection system using RoBERTa-based transformer fine■tuning combined with Retrieval■Augmented Generation (RAG). The model achieves 100% accuracy on the test dataset and provides explanations using FAISS-based evidence retrieval.

2. Introduction Fake news poses a major threat in digital communication. Traditional ML models struggle with linguistic complexity. Transformer models like RoBERTa handle context efficiently. This project builds a complete pipeline: preprocessing, supervised learning, transformer fine■tuning, and evidence-based explanation.

3. Dataset Description The dataset consists of two CSV files: - True.csv (legitimate news articles) - Fake.csv (fabricated news) Each record includes title, text, subject, date. Labels were added: 1 = Real, 0 = Fake.

4. Methodology Steps: 1. Data cleaning and text normalization. 2. Splitting into train, validation, and test sets. 3. Balancing the training dataset using upsampling. 4. Baseline model using TF■IDF + Logistic Regression. 5. Fine■tuning RoBERTa-base transformer with class■weighted loss. 6. Evaluation using accuracy, precision, recall, F1-score. 7. Building a FAISS vector store over real articles. 8. Implementing a RAG-style explanation system.

5. Baseline Model (TF-IDF + Logistic Regression) The TF■IDF model achieved moderate accuracy. Transformer fine■tuning significantly outperformed it.

6. Transformer Fine■Tuning Model: RoBERTa-base Batch Size: 8 Epochs: 3 FP16 enabled Weighted loss used to handle class imbalance.

Training Result: - Training Loss: near zero - Validation Loss: extremely low - Accuracy: 100% - Precision/Recall/F1: 100%

The model fully learned distinguishing patterns in the dataset.

7. Test Evaluation The test dataset also achieved: - 100% Accuracy - Perfect Confusion Matrix (0 misclassifications) This confirms high generalization to the dataset.

8. RAG (Retrieval-Augmented Generation) To improve interpretability, FAISS and Sentence-Transformers were used: - Embedding model: all■MiniLM■L6■v2 - FAISS index built on real news - For each input article: 1. RoBERTa predicts fake/real 2. FAISS retrieves top similar real articles 3. System outputs reasoning with evidence passages

This makes the model explainable and more suitable for academic and industry applications.

9. Conclusion The system delivers excellent performance and explainability. While dataset results are perfect, real-world deployment requires additional generalization datasets. The combination of transformer fine■tuning and RAG makes this project industry■grade.

10. Future Work - Add multi-dataset training - Deploy Streamlit web application - Integrate LLM-based natural language explanations - Expand to multi-class misinformation categories