# Reviewer Recommender System — Project Report

## 1. Overview

The Reviewer Recommender System is an AI-driven application designed to automatically suggest potential research paper reviewers based on the semantic similarity between a submitted paper and a collection of existing academic works.   
The system leverages state-of-the-art NLP models to identify the most contextually relevant experts from a large dataset of research papers.

## 2. Objective

The goal of this project is to automate and optimize the reviewer selection process in academic publishing by understanding the thematic alignment between submitted papers and previously published research.

## 3. System Architecture

The pipeline consists of five major components:  
1. PDF Text Extraction— Converts uploaded research papers into plain text using Apache Tika.  
2. Preprocessing — Cleans, normalizes, and structures the extracted text.  
3. Embedding Generation — Uses SentenceTransformer models (e.g., all-mpnet-base-v2) to represent papers as numerical vectors in semantic space.  
4. FAISS Indexing— Efficiently stores and retrieves embeddings using cosine similarity for top-k candidate retrieval.  
5. CrossEncoder Reranking — Uses a transformer-based CrossEncoder (e.g., MiniLM) to re-evaluate semantic similarity between the query and retrieved candidates for final ranking.

## 4. Models Used

• SentenceTransformer (`all-mpnet-base-v2`) — Generates high-quality embeddings that capture sentence meaning and context.   
• CrossEncoder (`cross-encoder/ms-marco-MiniLM-L-6-v2`) — Reads query and candidate together for fine-grained semantic comparison.   
• FAISS (`IndexFlatIP`) — Enables scalable similarity search across thousands of vectors in milliseconds.

## 5. Implementation Details

- Language: Python 3.12   
- Frameworks: PyTorch, SentenceTransformers, FAISS, Streamlit   
- Hardware: Supports CUDA GPUs (NVIDIA RTX 4090 optimized)   
- Directory structure: Modular design with `src/` for backend logic, `ui/` for Streamlit app, and `data/` for input papers and processed text.

## 6. Streamlit User Interface

The interactive Streamlit UI provides an intuitive way to:  
- Upload PDF papers.  
- Extract and preview text.  
- Select models for embedding and reranking.  
- View ranked reviewers with similarity percentages and snippets.  
- Customize retrieval parameters (`top\_k`, `rerank\_k`).

## 7. Workflow Summary

1. Upload a PDF or input custom text.  
2. Extract and preprocess textual content.  
3. Encode the text into embeddings.  
4. Retrieve top-k semantically similar papers via FAISS.  
5. Rerank using CrossEncoder for precision.  
6. Aggregate scores per author and display ranked reviewers.

## 11. Conclusion

The Reviewer Recommender System demonstrates how transformer-based NLP models and semantic search can be used to intelligently match researchers with relevant expertise. The architecture is modular, GPU-optimized, and expandable, making it suitable for academic institutions, publishers, or AI-driven research tools.