# PaperSafe: Intelligent Research Paper Summarizer and Verifier using Groq LLaMA

## Theme:

AI for Education / Research and Innovation

## Problem Statement:

Researchers, students, and reviewers face difficulty understanding, summarizing, and verifying research papers due to their increasing volume and technical complexity. Traditional search tools (Google Scholar, Scopus, etc.) only retrieve documents — they don’t explain content, verify facts, or compare models. Manual reading and interpretation consume several hours per paper, leading to delays in research, duplication of effort, and incomplete comprehension.

## Proposed Solution:

PaperSafe is an AI-driven PDF Question–Answer and Verification platform that allows users to:  
- Upload research papers / PDFs  
- Ask natural-language questions about methodology, results, or conclusions  
- Receive factual, grounded answers directly derived from the text  
- Get confidence scores (Grounding %) showing how well the AI answer is supported by the source document  
  
The system combines:  
1. Document Retrieval & Chunking using LangChain + FAISS  
2. Contextual Understanding using Groq LLaMA 3.3–70B (via Groq API)  
3. Embeddings-based Search using SentenceTransformer (all-MiniLM-L6-v2)  
4. Explainable AI layer showing retrieved evidence  
5. Session History & Export (CSV report for reviewers)

## Technical Architecture:

Frontend: Streamlit (Python-based interactive UI)  
Backend: Groq API (LLM inference), LangChain, FAISS  
Embedding Model: all-MiniLM-L6-v2 (Sentence Transformers)  
Hosting: Streamlit Cloud / GitHub Actions  
Environment: Python 3.10, Virtual Env, dotenv (secured API keys)

## Workflow Overview:

1. User uploads a research paper PDF.  
2. PDF is split into semantic text chunks using RecursiveCharacterTextSplitter.  
3. Each chunk is embedded and stored in a FAISS vector database.  
4. When a query is entered, top relevant chunks are retrieved.  
5. Groq LLaMA generates a factual answer, constrained to the retrieved context.  
6. A grounding analysis highlights what percentage of the answer tokens appear in the source chunks.  
7. Users can view or download the full Q&A history for audit and report generation.

## Key Features:

* Natural Language Q&A from uploaded PDFs
* Cross-paper model comparison (future update)
* Auto-grounding and fact-verification
* Explainable chunk display (retrieval transparency)
* Downloadable Q&A audit report (CSV)
* Secure API usage via .env or Streamlit Secrets

## Expected Impact:

- Reduces paper review time by up to 70% for students, professors, and researchers.  
- Improves academic transparency — each answer is traceable to its source text.  
- Assists in plagiarism detection and fact verification in literature reviews.  
- Enables inclusive learning for non-native researchers or beginners.

## Team Composition:

|  |  |  |
| --- | --- | --- |
| Name | Role | Skills |
| Arul Selvam P | Team Lead | AI, NLP, LangChain, Python |
| Member 2 | Frontend Developer | Streamlit, UX Design |
| Member 3 | Data Engineer | FAISS, APIs |
| Member 4 | Research Analyst | Paper Curation, Testing |

## Future Enhancements:

- Multi-document comparison (for literature review automation)  
- Voice-based question input and narration output  
- Integration with Scopus / IEEE APIs  
- Summarization + Similarity-based clustering  
- Explainable AI Dashboard for reviewers

## Live Demo & Repository:

🔗 Live Demo: https:// papersafe-sih2025-7kzdfnr5hp4jcp9yn8nmuc.streamlit.app/  
🔗 GitHub Repository: https://github.com/arulselvamp/papersafe-sih2025

## Tagline:

“From Paper to Insight — Instantly. Making Research Review Smarter, Faster, and Transparent.”