FINAL AI PROJECT REPORT

🔐 PrivateRAG AI

Your Private, Secure, Offline, Document-Grounded AI Assistant

Developed by: Dr. Subramani Suresh

Date : 2025

1. INTRODUCTION

Modern Artificial Intelligence systems predominantly rely on cloud-hosted Large Language Models (LLMs), requiring sensitive documents to be transmitted to third-party servers. This introduces critical risks related to **data privacy, regulatory compliance, hallucinations, auditability gaps, and loss of data sovereignty**—particularly in enterprise, research, legal, and regulated environments.

PrivateRAG AI was engineered to directly address these risks.

It is a **fully offline, document-grounded intelligence system** that enables users to interact with confidential documents **without any external connectivity**, while ensuring answers are **verifiable, auditable, and strictly evidence-based**.

This system is **not a general-purpose chatbot**.
It is a **controlled document intelligence platform** designed for **accuracy, safety, and trust**.

2 PROJECT OBJECTIVES

PrivateRAG AI is designed to:

- Enable **secure interaction with confidential documents**
- Provide **accurate, evidence-backed answers only**
- **Eliminate hallucinations entirely**
- Enforce **mandatory refusal** when evidence is insufficient
- Maintain **100% offline execution** and full data sovereignty
- Offer **two distinct interaction modes** for different user needs

Correctness and safety are always prioritized over answer coverage.

3. PROJECT SCOPE

✅ In Scope

- Local document ingestion and indexing
- Semantic retrieval using vector embeddings
- Retrieval-Augmented Generation (RAG)
- Dual interaction modes (Chatbot & Question)
- Guardrails, refusal logic, and confidence scoring
- Local storage of vectors, logs, and evaluation metrics

❌ Out of Scope

- Internet or cloud-based APIs
- Online LLM services
- Training or fine-tuning on user data
- Creative, speculative, or opinion-based responses
- Autonomous decision-making systems

4. SUPPORTED INPUTS

Supported Document Formats

- PDF (.pdf)
- Word (.docx)
- Text (.txt)
- Markdown (.md)
- CSV (.csv)

Upload Options

- Single document upload
- Multiple document upload
- Folder upload (recursive ingestion)

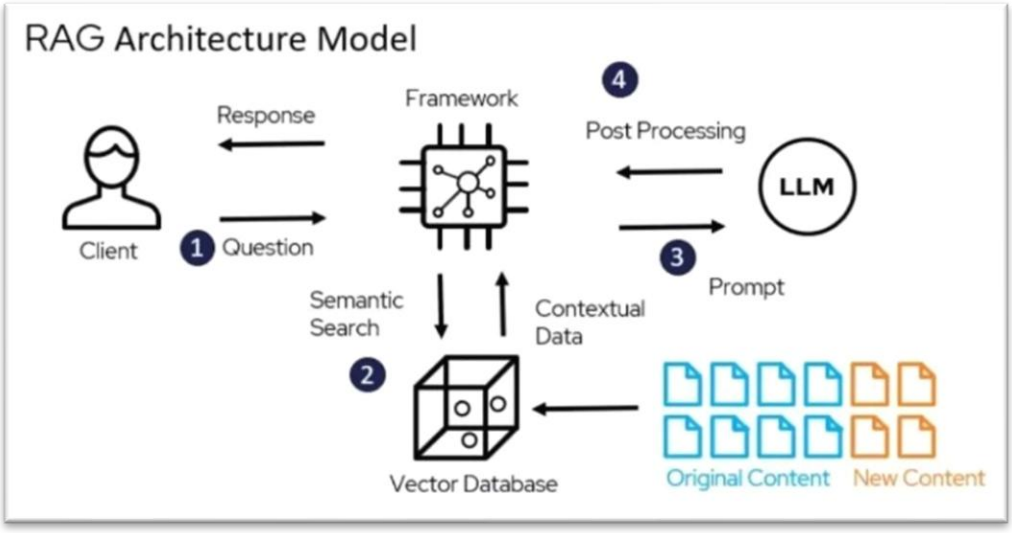
All documents are processed locally only.

5. SYSTEM ARCHITECTURE

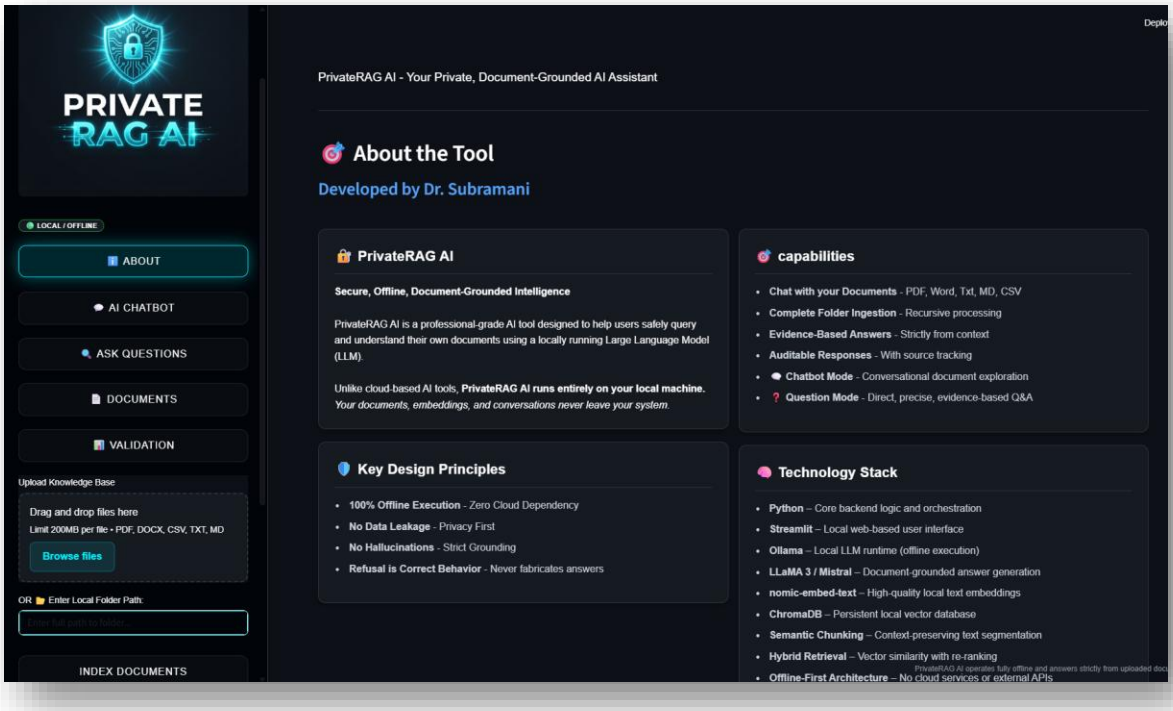
5.1 End-to-End Workflow

1. User uploads documents (single, multiple, or folder)
2. Documents are parsed and cleaned locally
3. Text is semantically chunked with contextual overlap

4. Chunks are converted into vector embeddings
5. Embeddings are stored in a local vector database
6. User interacts via the selected interface mode
7. Relevant document chunks are retrieved
8. Guardrails validate evidence sufficiency
9. Local LLM generates an answer or refusal
10. Confidence score and evidence sources are displayed



Landing Page



6. DUAL INTERACTION DESIGN

PrivateRAG AI supports **two interaction modes**, both powered by the **same backend pipeline** and **identical safety guarantees**.

🔁 Shared Characteristics (Both Modes)

Both interaction modes:

- Share the **same backend architecture**
- Use **identical retrieval logic**
- Apply the **same guardrails**
- Enforce **zero hallucination policy**
- Support:
    - Greeting handling
    - Context-aware follow-ups
    - Expandable evidence sections
    - Confidence score per answer
    - Clear refusal responses

💬 MODE 1: Chatbot Mode

Purpose: Conversational document exploration

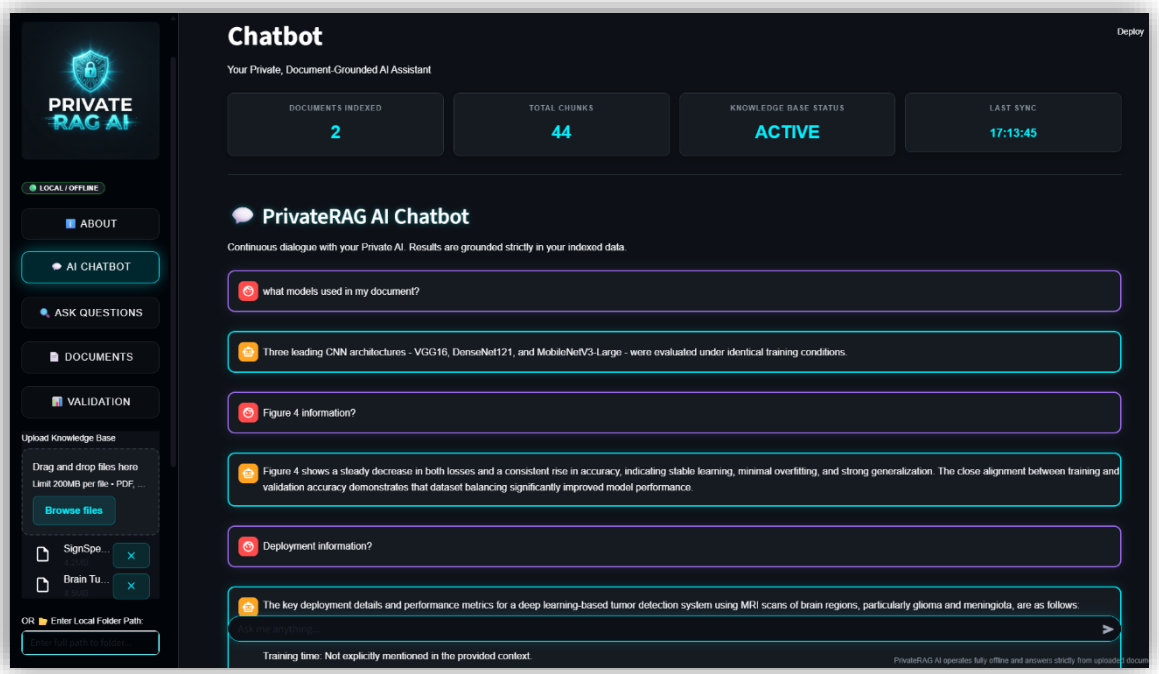
Features

- Natural conversational flow
- Multi-turn context retention
- Automatic greeting vs query detection
- Follow-up question support
- Evidence-backed answers only
- Expandable evidence panels
- Confidence score displayed with each response

Use Case

- Exploratory reading
- Research analysis
- Progressive understanding of documents

MODE 1: Chatbot Mode: Screenshot Result



❓ MODE 2: Question Mode

Purpose: Direct, precise, evidence-based Q&A

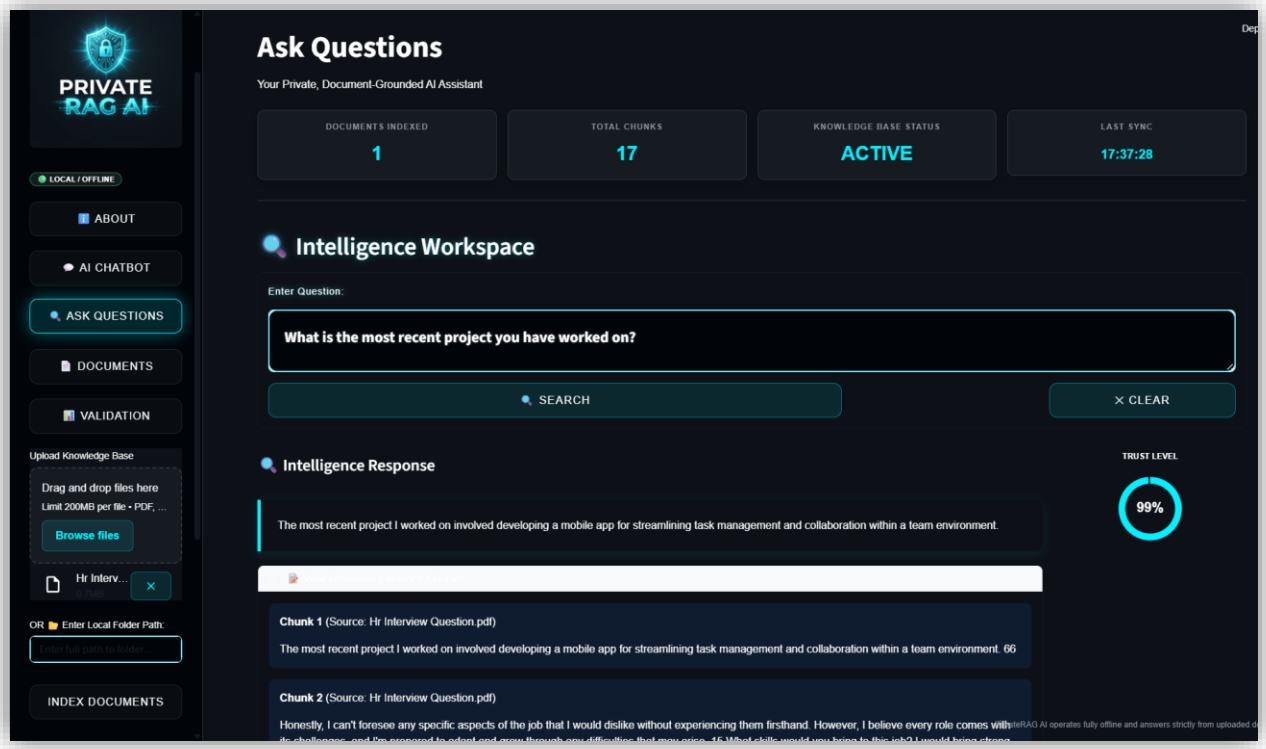
Strict Behavior

- Single-question input only
- No conversational noise
- Deterministic retrieval
- Strict evidence enforcement
- Binary output:
  - **Clear Answer** (with evidence), or
  - **Explicit Refusal**
- Source-focused responses
- Ability to **view retrieved evidence chunks**

Use Case

- Compliance checks
- Legal review
- Auditable Q&A
- Enterprise decision support

MODE 2: Question Mode Screenshot Result



7. CORE SYSTEM COMPONENTS

🖥️ User Interface

- Built with **Streamlit**
- Local web-based application
- Dark / cyberpunk UI theme
- Thinking / typing indicators
- Expandable evidence panels

📄 Document Ingestion Layer

- Fully offline loaders
- Metadata extraction
- Recursive folder ingestion

🧩 Text Chunking & Normalization

- Semantic chunking
- Context-preserving overlap
- Section-aware tagging

🧠 Embedding Engine

- Local embeddings via **Ollama**
- Model: `nomic-embed-text`
- Zero external calls

🗂️ Vector Database

- **ChromaDB**
- Persistent local storage
- High-performance similarity search

🔍 Retrieval Engine

- Deterministic hybrid retrieval
- Vector similarity search
- Evidence re-ranking
- Controlled retrieval parameters

🛡️ Guardrails & Validation

- Evidence sufficiency checks
- Mandatory refusal logic
- Zero hallucination tolerance

🤖 Answer Generation

- Local LLM execution
- Models: LLaMA 3 / Mistral
- Deterministic, document-grounded output

📊 Confidence & Evaluation Module

- Evidence-density-based confidence scoring
- Refusal accuracy tracking
- Evaluation metric logging

8. SECURITY & PRIVACY MODEL

- 100% offline execution
- No cloud connectivity
- No telemetry or tracking
- No external data sharing
- No training on user data
- Local-only storage (documents, vectors, logs)

Suitable for enterprise, legal, research, and regulated environments.

9. EVALUATION & PERFORMANCE SUMMARY

Latest Results

- Hallucinations: **0**
- Refusal Accuracy: **100%**
- Grounded Recall: **67%**
- Grounded Accuracy: **80%**
- Average Confidence: **47%**

Interpretation

- Strong safety guarantees
- Conservative, honest outputs
- Enterprise-aligned refusal behavior
- High trustworthiness

10. USE CASES

- Academic research & literature analysis
- Enterprise internal knowledge systems
- Legal & compliance document review
- Scientific & technical documentation
- Confidential organizational data exploration

11. LIMITATIONS

- Answers limited strictly to uploaded documents
- Performance depends on document quality
- Not suitable for creative or open-ended reasoning
- Requires sufficient corpus for high recall

12. CONCLUSION

PrivateRAG AI, developed by **Dr. Subramani**, is a **secure, offline, document-grounded intelligence system** featuring a **dual interaction design** that balances conversational flexibility with strict, auditable Q&A.

By eliminating hallucinations, enforcing refusal when evidence is missing, and maintaining complete data sovereignty, the system delivers **trustworthy, enterprise-ready AI intelligence**.