

TANISHA PRITHA

Nagpur, India | [E-mail](#) | [LinkedIn](#) | [GitHub](#) | [LeetCode](#) | 7368825023

PROFESSIONAL SUMMARY

Full-Stack AI Developer with a strong backend focus, experienced in building efficient, scalable systems for RAG pipelines and LLM-powered applications. Skilled in FastAPI, asynchronous Python, and PostgreSQL-based retrieval to deliver reliable, low-latency AI workflows. Experienced in prompt design, agent orchestration, and end-to-end system integration, with emphasis on performance, correctness, and production readiness.

TECHNICAL SKILLS

AI / GenAI: LLMs, RAG, Embeddings, Agentic Workflows, Prompt Engineering, Transformers, LoRA, Evaluation

Backend: Python, C++, FastAPI, REST APIs, Async I/O, JWT, RBAC, System Design

Frontend: React.js, Next.js, TypeScript

Data & Infra: PostgreSQL, PgVector, Redis, FAISS, Docker, AWS EC2, Linux

Document AI: OCR (PaddleOCR), PDF Parsing

Strengths: Performance Optimization, Security-First Engineering, Observability

EXPERIENCE

AI Engineer Intern — AptlyHired (Startup) | Nov 2025 – Dec 2025 | Remote, New Delhi

- Developed **multi-format document ingestion** (PDF, DOCX, TXT, PNG, JPG) with OCR using PaddleOCR.
- Built an end-to-end **RAG pipeline** with chunking, embeddings, metadata storage, and contextual retrieval.
- Implemented **Upload → Store → Chat** workflow enabling users to query their own documents.
- Improved retrieval accuracy through optimized chunking, noise filtering, and prompt design.
- Integrated the entire system into a functional web UI used by internal testers.

PROJECTS

Multi-Agent Company Compliance Engine | FastAPI, Python 3.11, PostgreSQL, pgvector, RAG, LLMs | [GitHub](#)

- Built a **multi-agent AI system** using Planner-Reasoner-Verifier agents to perform forensic audits of long-form documents.
- Designed an **agent consensus workflow** to finalize decisions only after cross-verification, reducing hallucinations in high-stakes outputs.
- Implemented a **production-grade RAG pipeline** with PostgreSQL + pgvector for semantic evidence retrieval with traceable citations.
- Engineered **layout-aware PDF processing** (PyMuPDF) to extract structured content and spatial coordinates for source-level verification.
- Optimized **LLM cost and latency** through semantic chunking and prompt control.
- Added **observability and audit integrity** via live execution traces and immutable **SHA-256 snapshots**.

MeetMe — Meeting Intelligence (Chrome Extension) | Next.js, Deepgram STT, LLMs, Embeddings | [GitHub](#)

- Built an **AI-powered Chrome extension** that performs real-time speech-to-text on live meetings using Deepgram.
- Implemented a **rolling live summary** that updates continuously during meetings and videos.
- Extracted key signals to detect **relevance, conflicts, and red flags** in real time.
- Integrated **semantic matching** against user-uploaded, chunked, embedded documents to surface contextual insights.
- Developed a **client-side Next.js frontend** optimized for low-latency in-meeting usage.

EDUCATION

B.Tech in Mechanical Engineering | Government College of Engineering, Nagpur | Graduating: June, 2027

CERTIFICATIONS

Introduction to Machine Learning: Art of the Possible (AWS) | HackerRank SQL (Intermediate) | Data Modeling for Amazon DocumentDB (AWS)