

SHIVAM PANDEY

+91 7457996994 | spdev912@gmail.com | linkedin.com/in/spdev912 | github.com/spdev18 | shivam-pandey.tech

WORK EXPERIENCE

TechOTD Solutions Private Limited — Software Developer

Aug 2024 – Present

- Owned end-to-end delivery of three production AI products — LegalDrishti (Indian legal research & document evaluation), Lexmate (Taiwan legal multi-agent assistant), and VoxaAI (Android transcription app) — across Next.js, FastAPI, LangGraph, and native Android (Java) stacks.
- Designed hybrid RAG retrieval combining Milvus/Pinecone vector search with PostgreSQL full-text and boolean filtering, fused via Reciprocal Rank Fusion, with NLP query understanding, legal entity extraction, and synonym expansion tailored to jurisdiction-specific corpora.
- Built LangGraph multi-agent orchestration routing across Claude, GPT, and Llama 3 backends for statute search, case-law retrieval, and drafting; shipped a clause-by-clause Document Evaluation engine producing compliance scores and human-review workflows against Indian-law templates.
- Integrated Deepgram Nova-3 for multilingual real-time speech-to-text with speaker diarization and Gemini for AI summarization, with Firebase Auth, Firestore sync, Google Play Billing, and iText7 PDF export on a Material Design 3 Android client.
- Containerized and shipped platforms on AWS EC2 via Docker Compose (Nginx, PostgreSQL, Milvus, Ollama) with GitHub Actions CI/CD enabling zero-touch deployments on every push to main.

PROJECTS

LegalDrishti

- AI-powered Indian legal research and document-evaluation platform built on Next.js + FastAPI, ingesting judgments, statutes, circulars, and form templates via OCR pipelines (Docling + PyMuPDF) with role-based access control.
- Engineered hybrid retrieval fusing Milvus vector search (768-dim embeddings) with PostgreSQL full-text search via Reciprocal Rank Fusion, layered with NLP query understanding and synonym expansion against a self-hosted Ollama LLM backend.
- Productionized on AWS EC2 with a 9-container Docker Compose deployment, Nginx reverse proxy, and GitHub Actions CI/CD pipeline.

Lexmate

- Taiwan legal AI assistant using LangGraph multi-agent orchestration with specialized graphs for statute search, case-law retrieval, and document generation across Claude 3, GPT, and Llama 3 backends.
- Implemented Pinecone semantic search (1536-dim embeddings), memory/reflection agents for personalized legal writing, Deepgram transcription, and LangSmith observability with PDF/DOCX export in a Yarn monorepo.

VoxaAI

- Android (Java) app integrating Deepgram Nova-3 for multilingual real-time transcription with speaker diarization and Google Gemini for AI summarization with a follow-up Q&A chat interface.
- Implemented Firebase Auth, Firestore for cross-device persistence, Google Play Billing for premium subscriptions, and PDF export via iText7 with a Material Design 3 animated recording UI.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, Java, SQL

Frameworks & Runtimes: Next.js, React, Node.js, FastAPI, LangGraph, LangChain, Tailwind CSS, SQLAlchemy, Pydantic, AndroidX

AI / LLM: Claude, GPT, Gemini, Llama 3, Deepgram Nova-3, Ollama, LangSmith, RAG Pipelines, Multi-Agent Systems

Databases: PostgreSQL, Milvus, Pinecone, Firebase Firestore, Supabase

DevOps & Cloud: Docker, Docker Compose, Nginx, AWS EC2, GitHub Actions, Linux (Ubuntu)

EDUCATION

B.Tech in Computer Science Engineering

Graduated July 2024

NIET, Greater Noida

Relevant Coursework: Python, OOP (Java), Data Structures & Algorithms, DBMS, Operating Systems, Computer Networks