

# PANSHUL JINDAL

panshuljindal@gmail.com | +1(551)-2410325 | www.linkedin.com/in/panshuljindal | www.github.com/panshuljindal

Software Engineer with **1.5+ years** of hands-on experience building scalable **AI systems** and deploying **LLMs to production**. Proven track record in backend infrastructure, multi-agent architectures, retrieval-augmented generation (RAG) and LLM fine-tuning.

## EDUCATION

### University of Illinois, Urbana Champaign

Jan 2024 - May 2025

Masters in Computer Science

CGPA: 4.0/4.0

Relevant Coursework: Machine Learning for Software Engineering, Artificial Intelligence, User-Centered ML

### Vellore Institute of Technology, India

Jul 2019 – Apr 2023

Bachelor of Technology – Computer Science Engineering

CGPA: 9.18/10.00

## SKILLS

**Languages** – Python, Go (Golang), Java, C++, SQL, Shell, TypeScript

**LLMs & GenAI** – OpenAI API, Anthropic Claude, Hugging Face Transformers, LangChain, LangGraph, Multi-Agent Systems, Retrieval-Augmented Generation (RAG), Prompt Engineering, LLM Fine-tuning (SFT, PEFT), Foundation Models (LLaMA, Mistral)

**Machine Learning** – PyTorch, OpenCV, Scikit-learn, Sentence Transformers, Model Context Protocol

**Backend** – FastAPI, Flask, Django, Node.js, Express, Redis, gRPC, REST APIs, Postman, Datadog, Retool

**MLOps and Cloud** – AWS (Bedrock, EC2, Lambda, SageMaker, ECR), GCP, Azure, Firebase, Docker, GitHub Actions, MLflow, Weights & Biases, CI/CD

**Databases** – Pinecone, FAISS, ChromaDB, PostgreSQL, MySQL, MongoDB, DynamoDB, Neo4j, Elasticsearch

## EXPERIENCE

### Software Engineer - Visiostack (OpenAI, Supervised Fine Tuning (SFT), Python)

Jun 2025 – Present

- Built a **retrieval-augmented generation (RAG)** pipeline leveraging OpenAI embeddings and ChromaDB vector store to convert natural language queries into structured JSON outputs aligned with a given schema.
- Fine-tuned** a domain-specific **AI Agent** via supervised fine-tuning on synthetic data (generated using **OpenAI's batch API**), boosting query accuracy and response consistency for the enterprise search application.
- Applied advanced prompt engineering (few-shot examples and chain-of-thought prompting) to enhance the model's ability to generalize query handling across multiple use cases, increasing integration success rates.

### Research Assistant - RailTEC, UIUC (Azure, Python, PySpark)

Jun 2024 – May 2025

- Developed a LLM-powered query interface for railway diagnostics to convert **NL queries to SQL** for data mining and reporting.
- Designed a **monitoring framework** for inspection systems, combining PySpark pipelines with **LLM-based analysis** of anomaly reports to improve fault detection and explainability.
- Implemented **modular backend microservices** on Azure to ingest and analyze terabytes of sensor and image data, supporting real-time infrastructure health monitoring with high availability.

### Software Engineer - Kwikpic AI Solutions (AWS, NodeJS, MongoDB, Python)

Jan 2023 – Dec 2023

- Built an **LLM-driven personalization engine** for multi-channel user notifications, dynamically tailoring content to each user. This personalization resulted in higher click-through rates and improved user retention.
- Developed scalable REST APIs to power user authentication, photo delivery, serving over **10 million users** across platforms.
- Enhanced a facial recognition engine**, achieving a 98.9% accuracy while reducing image preprocessing latency by over 30%.
- Achieved a 40% reduction in AWS costs by implementing auto-scaling and moving static workloads to serverless Lambda functions.

### Software Developer Intern - Student Welfare Office (MongoDB, Python)

Mar 2022 – Sept 2022

- Built backend APIs for VITNav, a smart campus app with NLP-based event categorization and garnering over **50K+ downloads**.
- Designed NoSQL schemas with AI-driven event tagging using **spaCy** and optimized for real-time notification delivery.
- Led deployment of backend on campus Linux servers using PM2 and Nginx, ensuring **stability under load and zero downtime**.

## PROJECTS

### TaskPilot: Agentic Workflow Engine for Autonomous LLM-Orchestrated Tasks

(FASTAPI, Docker, OpenAI API, Redis)

- Architected a modular multi-agent system where autonomous AI agents plan, execute, and validate tasks using tool integrations.
- Built Redis-backed task queues with FastAPI and a real-time memory store, enabling scalable automation with persistent context.
- Developed a framework supporting streaming outputs and error recovery, simulating end-to-end DevOps workflows through AI agents.

### AutoTest: AI-Powered Code Coverage Booster

(OpenAI API, DeepSeekCoder, CI/CD, Pytest)

- Engineered a tool that leverages a **code-specialized LLM** (DeepSeekCoder) to analyze codebases and synthesize unit tests.
- Integrated with CI/CD to raise coverage from ~40% to **100%** by identifying untested paths and **generating relevant test cases**.
- Built an **auto-refinement loop** where failing tests trigger iterative adjustments (using execution logs) to improve subsequent outputs.

## LEADERSHIP & COMMUNITY

### Technical Lead - Apple Developers Group, VIT

- Led AI-focused workshops and mentorship programs for **200+ students**, providing hands-on training in machine learning applications.
- Spearheaded the creation of internal tooling and documentation standards to streamline project collaboration for new contributors.