

ARPITA PATNAIK

Software Engineer | Backend & AI/ML Engineer | LLM Applications

Orlando, FL | Open to relocation (U.S.)

Email: patnaikarpita100@gmail.com | Portfolio: <https://arpita31.github.io/>

GitHub: <https://github.com/arpita31> | LinkedIn: <https://www.linkedin.com/in/arpitapatnaik10>

Professional Summary

Results-driven Software Engineer with 5+ years of experience in backend, distributed systems, and full-stack development. Experienced in building production systems using Python, FastAPI, and Django, with PostgreSQL and Redis for scalable, high-performance applications. Proven track record of reducing latency by 75% and designing resilient APIs. Experienced in integrating AI into production systems, including LLM-based applications, agentic workflows, and RAG pipelines. Passionate about building intelligent, scalable solutions with clean architecture.

Education

University of Florida, Gainesville

Aug 2023 – May 2025

Master of Science in Computer Science

GPA: 3.63/4

National Institute of Science and Technology, India

Aug 2012 – 2016

Bachelor of Technology in Information Technology

CGPA: 8.19/10

Experience

Mindtree Ltd. (Bengaluru, India) | Senior Software Engineer

Jul 2016 – 2021

- Reduced P90 latency by 75% (20s → 5s) by designing a distributed caching layer using Redis, optimizing query patterns, and implementing retry/backoff strategies.
- Architected real-time feature flagging system (Django + Redis) enabling dynamic rollout, A/B testing, and rollback, improving release reliability.
- Designed resilient backend systems with fault tolerance, retry mechanisms, and performance optimizations for high-traffic APIs.
- Built automated deployment pipelines (CI/CD) and integrated API-based update systems, reducing manual release overhead.
- Led test architecture improvements, increasing coverage to 75% and reducing production defects.
- Refactored multi-source data aggregation pipelines, improving data consistency and system maintainability.
- Developed scalable UI modules (WPF, WinForms) supporting enterprise-grade applications.
- Worked across end-to-end system architecture (API → caching → DB → client) ensuring scalability and performance under load.

Projects

Deal Hunter Ai - Multi-Agent Deal Intelligence System | Python, Llama 3.1, OpenAI, RAG, Modal.com, Gradio | Github Repo

- Architected production LLM-based AI system with 9 specialized agents orchestrating fine-tuned Llama 3.1-8B (QLoRA), RAG pipeline (400k embeddings), and ensemble ML for autonomous deal discovery and accurate price prediction.
- Deployed scalable GPU inference on Modal.com achieving sub-2s latency with real-time notifications, demonstrating end-to-end MLOps from model fine-tuning to production deployment and distributed systems design.

M3: Mood Movie Matcher | Python, PyTorch, Machine Learning, NLP, scikit-learn, Transformers | Github Repo

- Developed ML-powered movie recommendation system using RoBERTa sentiment analysis and Autoencoders on MovieLens dataset, enabling discovery of personalized best matches and hidden gem films through interactive command-line interface.

Safe Data Cloak and Lift The Veil | Python, SpaCy, scikit-learn, sentence transformers | Github Repo

- Designed scalable data privacy system utilizing NLP transformers for precise redaction of sensitive data including PII, dates, phone numbers, and domain-specific concepts.
- Built unredaction functionality using NLTK and Random Forest algorithms with enterprise API integration capabilities.

A.L.F.R.E.D. | Python, FastAPI, OpenAI GPT-4, React, Redis, PostgreSQL, Celery, Docker | Github Repo

- Architected event-driven LLM-based decision engine combining behavioral signal processing, GPT-4 inference, and rule evaluators with confidence scoring, cooldown logic, and feedback loops for real-time interventions.
- Built scalable full-stack system with FastAPI async APIs, PostgreSQL, Redis (cache + queue), and Celery pipelines; implemented financial simulations and on-device ML for low-latency, personalized insights.

Technical Skills

Languages: Python, C#, JavaScript, TypeScript, SQL, HTML, CSS, React

Backend: FastAPI, Django, Django REST Framework (DRF), Flask, REST APIs

AI and ML: LLMs, Agentic AI, RAG, Transformers, PyTorch, TensorFlow, NLP, Fine-tuning, OpenAI API, Pandas

Systems Design / Architecture: Distributed Systems, Microservices, Event-Driven Architecture, Caching (Redis), Fault Tolerance

Databases: PostgreSQL, Redis

Cloud and DevOps: AWS, Docker, CI/CD, Jenkins, Linux, Git

Testing and Quality: Pytest, XUnit, TDD, Postman

Certifications and Coursework

Certifications: AWS Certified Cloud Practitioner (April 2025), Meta Full Stack Developer

Coursework: Machine Learning, Analysis of Algorithms, Distributed Operating Systems, Computer Networks, Data Engineering, Computer Network Security

Leadership and Extracurricular

- Received Achievement Award Scholarship by University of Florida for academic excellence and software engineering achievements.
- Teaching Assistant for Analog Electronic Circuits, guiding 50+ undergraduates in circuit design, analysis, and technical problem-solving.
- Directed coordination and logistics for 50+ large-scale events at the O'Connell Center, some hosting 10,000+ attendees, ensuring flawless execution, operational efficiency, and exceptional participant experiences through strategic planning and team leadership.