Kaushik Chaturvedula

+1 (224) 678-1562 | **kaushikchaturvedula@gmail.com** | **Portfolio** | **LinkedIn** | **GitHub** | **Leetcode**

SUMMARY

Software Engineer with 1+ years of professional experience specializing in full-stack development, scalable distributed systems, and AI-powered solutions, with a strong focus on building user-centric applications. Proficient in C++, Python, Java, C#, and JavaScript (TypeScript, Node.js, React, Next.js), with expertise in frameworks like PyTorch, FastAPI, and Spring Boot. Experienced in designing and deploying cloud-native applications on AWS, GCP, and Azure using Docker, ensuring scalability and maintainability. Skilled in SQL/NoSQL databases and RESTful/GraphQL API integrations. Strong background in data engineering, backend optimization, and AI/ML technologies, including deep learning, NLP, RAG, and generative AI. Passionate about developing secure, high-performance backend architectures, ensuring system reliability, and leveraging AI to enhance user safety and engagement in large-scale interactive platforms.

EDUCATION

Purdue University, Indiana – Master of Science in Computer Science (Jan 2024 - May 2025 | GPA: 4.0/4.0)
National Institute of Technology Warangal, India – Bachelor of Technology in Mechanical Engineering (Aug 2018 - May 2022)
Ranked **9227 out of over 1.2 million students** (top 0.8%) in JEE Mains 2018, India's national engineering entrance exam.

EXPERIENCE

Medical Informatics Engineering (IN, U.S.) – Software Development Intern (Jan 2025 - Present)
● Working on building a semantic query engine for medical and EHR documents using RAG, NLP, LLMs and transformers like DistilBERT for query intent classification, Opensearch, Redis, and various optimizations to enhance response speed and accuracy.

Purdue University – Teaching Assistant and Research Assistant (Sept 2024 - Jan 2025)
● Assisted in delivering course material, grading assignments, and providing student support for Programming Language Design. Contributed to high-performance cosmological simulations research using Python on datasets >300TB in HPC environments.

Wibmo (Bengaluru, India) – Associate Software Engineer (Jul 2022 - Apr 2023)
● Developed and optimized backend services for a Risk-based Authentication Engine in a leading FinTech firm, enhancing fraud detection and prevention, including identifying money laundering and BIN attacks, ensuring secure transaction processing.
● Architected scalable microservices for high-traffic apps, improving reliability, performance, and optimizing transaction processing time.
● Developed asynchronous, event-driven systems with RabbitMQ and Kafka, achieving a 20% reduction in latency.

Freecharge (Mumbai, India) – Full Stack Developer Intern (May 2021 - Jul 2021)
● Optimized microservices, led memory caching initiatives, and improved load balancing to enhance efficiency and system reliability.

Ravgins (India) – Front-end Developer Intern (Jun 2020 - Aug 2020)
● Built web/mobile applications from scratch using various front-end frameworks (like Angular) and tools, enhancing user engagement.

PROJECTS

■ **Semantic Query Engine**: A context-aware retrieval system that enhances search relevance using AI, RAG, NLP, LLMs, embedding models, Opensearch, FastAPI REST APIs, and Redis caching for efficient natural language query answering at near real-time speed.
■ **LexiPhylax:** A high-accuracy hate-speech classification engine built using PyTorch, Hugging Face Transformers, and LLMs, leveraging intelligent embeddings and contextual awareness for precise toxic language detection/classification. Currently developing a 400M parameter BERT-like transformer from scratch using PyTorch for enhanced classification accuracy and fit.
■ **SwiftNet**: A high-performance C++ networking library utilizing io_uring, kqueue, and coroutines for low-latency, scalable server applications. Features request pipelining and virtual thread offloading for efficient CPU utilization and high-throughput handling.

TECHNICAL SKILLS

● **AI and Machine Learning**: Natural Language Processing, Computer Vision, Deep Learning, PyTorch, AI tools, RAG
● **Full-Stack Development**: Spring Boot, Node.js, Express.js, Django, FastAPI, React, Angular, Next.js, Cypress, Testing Frameworks
● **Software and Data Management**: Web Services, API Design, Distributed Systems, Real-time Data Processing, System Design
● **Databases and Operating Systems**: MariaDB, MySQL, PostgreSQL, MongoDB, Couchbase, MacOS, Linux, Windows
● **High Performance Computing (HPC)**: Parallel Programming, Multi-threading, Virtual Threads, Coroutines, MPI, OpenMP
● **Programming Languages**: C, C++, Java, Python, C#, JavaScript, SQL, N1QL, PowerShell, Systems and Socket Programming
● **Technologies and Platforms**: Git, Docker, Redis, RabbitMQ, Apache Kafka, Postman, JMeter, AWS, Azure, GCP, JIRA, Heroku