SAURABH CHAUHAN

schau57@uis.edu | (217) 862-4640 | Springfield, IL | [LinkedIn](#) | [GitHub](#)

EDUCATION

Master of Science in Computer Science | University of Illinois Springfield |

Expected May 2026 | GPA: 3.8/4.0

Technical Coursework: Deep Learning, Neural Networks, Machine Learning, Data Mining, Cloud Computing, Algorithms & Computation

Bachelor of Engineering in Computer Engineering | Pune University, India

Graduated May 2021

TECHNICAL SKILLS

AI/ML: TensorFlow, Keras, PyTorch, scikit-learn, Hugging Face Transformers, spaCy, NLTK, OpenAI API, LangChain, RAG, Fine-tuning

Languages: Python (asyncio, type hints), Java (Spring Boot, JVM optimization), SQL, JavaScript

Backend: Django, Flask, FastAPI, REST APIs, gRPC, WebSocket, Microservices, Apache API SIX

Data: PostgreSQL, SQL Server, Redis, FAISS, ChromaDB, Pandas, NumPy

Cloud/DevOps: AWS (EC2, S3, Lambda), GCP (Compute Engine, Cloud Storage), Docker, CI/CD (Jenkins, GitLab CI, GitHub Actions), Prometheus, Grafana

PROFESSIONAL EXPERIENCE

Product Dossier Solutions Pvt Ltd | Pune, India

Software Engineer (AI-ML) | Jun 2023 – Jul 2024

- Architected RAG-based document Q&A chatbot serving 10K+ users with 95% accuracy by implementing LangChain, Mistral-7B, and FAISS vector database, reducing response time by 120x from 6s to 50ms
- Secured 11 AI microservices handling 110K+ daily requests by engineering Apache API SIX gateway with JWT authentication, rate limiting, and load balancing in Spring Boot, reducing security vulnerabilities by 40%
- Accelerated software deployment by 67% from 45 to 15 minutes by building Docker-based automation pipeline with Apache Freemarker, supporting 10+ daily test deployments and 5-6 monthly production releases
- Reduced incident response time by 30% by implementing Prometheus and Grafana dashboards tracking latency and error rates across production ML services
- Led cross-functional team of 6 engineers conducting 15+ code reviews weekly and participating in architecture discussions with C-level executives, establishing Python best practices (OOP, logging standards) across the organization

Dasha Krit Technology Pvt Ltd | Pune, India

Software Engineer | Apr 2021 – Apr 2023

- Built Django-based multi-tenant SAAS platform from v0 to v1.1 with 8-10 modular features by architecting custom authentication, RBAC (6 user roles), and FSM workflows, supporting estimated 50+ concurrent enterprise users
- Enabled real-time stock market visualization for 100+ active traders by engineering WebSocket data pipeline with SQLAlchemy and integrating Apache Superset dashboards
- Delivered 20+ features across 2-year product lifecycle by leading sprint planning, API design, and database architecture as technical lead in 2-person development team
- Achieved 100% test coverage of v1.1 features by developing data-driven test suite with 100+ automated test cases using Robot Framework and Playwright

University of Illinois Springfield | Springfield, IL

Website Development Intern | Apr 2025 – Present

- Improved mobile accessibility for 5,000+ prospective students by building WCAG 2.1-AA compliant web components using Drupal, Twig, and Bootstrap for College of Health, Science & Technology sites
- Optimized faculty directory search performance by 95% to <150ms by implementing REST API with caching and vanilla JavaScript, serving 50+ profiles with sub-second filtering
- Prototyping OpenAI-powered admissions chatbot handling 200+ FAQ queries with contextual answers and real-time site content search, featuring persistent chat history and token-limited API management

AI/ML PROJECTS

ASL to Text Classification | GitHub
- Improved ASL video classification accuracy by 32% from 62% to 82% by fine-tuning VideoMAE transformer on 239-class dataset and introducing Universal Temporal Sub-sampling technique for GPU-constrained hardware
- Reduced training time to 2 hours on H100 GPU by optimizing pipeline with AdamW optimizer, cosine decay scheduling, and early stopping on 8GB memory constraints

Invoice IQ - Automated Invoice Processing | GitHub
- Achieved 95% accuracy in extracting 8 key invoice entities by developing end-to-end NER pipeline with GPU-accelerated OCR and custom spaCy transformer model trained on 200+ invoices
- Built production-ready Flask REST API with entity validation and structured JSON output, demonstrating full-stack ML deployment capabilities

Neurofinity - AI Mind Map Generator | GitHub | 4th Place, UIS Hackathon 2025
- Converted 30+ minute audio recordings into structured mind maps by building Streamlit app with Whisper API transcription, Hugging Face Transformers NLP pipeline, and Graphviz visualization, reducing note-taking time by 80%

CERTIFICATIONS

AWS Machine Learning Essentials (LinkedIn Learning, Nov 2025) | AWS Essentials (LinkedIn Learning, Jun 2024)