Sarthak Singh

+1(240)3548400 | [Email](#) | [LinkedIn](#) | [Website](#)

EXPERIENCE

Influenced (Full Stack Developer) *Next.js, TypeScript, Tailwind CSS, AWS Amplify, Dynamo DB, LLM*
July 2024 – June 2025 *Remote, USA*

- Developed an AI-powered marketplace that connects brands with micro-influencers, enabling authentic social proof and massive reach. Implemented role-based access and AI-driven workflows for product listings, interest expressions, and collaboration management.

IBM (Software Development Engineer II) *Python, AWS, GO, Java, Docker, Kubernetes, PostgreSQL*
July 2021 – Jan 2024 *Kochi, India*

- Led the migration of data analytics pipelines from ElasticSearch to OpenSearch, reducing operational costs by up to $3M annually and improving search performance.
- Streamlined backend services for API management on IBM Cloud and AWS, optimizing traffic routing, security, and data processing, reducing inefficiencies and operational overhead.
- Spearheaded a security optimization project for managing sensitive data, resulting in a more cost-effective solution and earning the IBM Quarterly Cash Award for contributions.
- Enhanced the overall reliability and scalability of services, facilitating API integrations and improving customer-facing application performance.

Informatica (Software Engineer Intern) *Python, ETL, Oracle DB, Tableau, Java, AWS S3*
Jan 2021 – July 2021 *Bangalore, India*

- Contributed to customer support by resolving 100+ tickets, offering technical guidance and troubleshooting for issues related to the Informatica ETL tool.
- Provided direct technical support for integrating the Informatica ETL tool with databases such as PostgreSQL, MySQL, and AWS S3.

PROJECT & RESEARCH

AI Powered Solutions *LLM, RAG, Gradio, Streamlit, Groq, MCP, HF Pipelines, LangChain*
Oct 2025 – Dec 2025 [Link](#)

- RAG Voice Customer Support: An end-to-end voice-enabled customer support assistant combining Retrieval-Augmented Generation with real-time speech interaction.
- An AI-powered Indian Law Assistant leveraging RAG to ground responses in Indian legal statutes, utilizing MCP for seamless integration and processing of real-world legal data and case law analysis.
- GPT-OSS 120B CritiqueLoop: An automated critique → improve → repeat pipeline using Groq for ultra-fast inference.

Anomaly Detection in IoT Devices [Link](#)
Aug 2024 – Dec 2024

- Explored unsupervised learning algorithms for anomaly detection in IoT devices, eliminating the need for labeled data while achieving strong detection performance.

Research Article (Journal of Environmental Management) [Link](#)
July 2022 – Jan 2024

- Contributed to research on green procurement dynamics in the public sector, developing an assessment framework to help developing nations make data-driven decisions towards green environmental practices.

EDUCATION

University of Maryland College Park, MD
Master in Engineering, Software Engineering, Cumulative GPA: 3.6/4.0 *Jan 2024 – Dec 2025*
Teaching Assistant | *Introduction to Software Engineering ENPM611*

Vellore Institute of Technology Vellore, India
Bachelor of Technology in Computer Science and Engineering *June 2017 – May 2021*

TECHNICAL SKILLS

Programming: Python, GO, FastAPI, Java, SQL (PostgreSQL, MySQL), NoSQL, Redis, Kafka, Django, GraphQL
DevOps & Cloud: Docker, K8, Jenkins, CI/CD, Terraform, AWS(CloudWatch, Grafana, EC2, EKS, IAM, S3, Lambda)
Systems: Microservices, Distributed Systems, Scaling Systems, Object-Oriented Design, Unit Testing
Other: Git/GitHub, Windows, Linux, Data Structures & Algorithm, Communication, Innovation
AI: Cursor, NLP(transformer models), LLM(Claude,GPT), Deep Learning ML (Pytorch, SGD, scikit learn)