

ATHAR SAYED

PROFESSIONAL SUMMARY

AI Software Engineer and M.Tech Artificial Intelligence candidate with hands-on experience in developing intelligent applications and scalable software systems. Experienced in building AI-driven solutions and contributing to enterprise-grade projects through industry and research initiatives.

PROJECTS

- Context-Grounded LLM Assistant for Recruiter-Facing AI Search** ([Live](#)) 2025
- Built a **production-grade LLM assistant** enabling natural-language search across **projects, skills, experience, and publications** in a live portfolio environment.
 - Designed a **context-grounded inference pipeline** with dynamic content injection, enforcing **strict context boundaries** to minimize hallucinations.
 - Implemented **prompt constraints, deterministic decoding (temperature = 0.3), and response validation** for consistent and reliable outputs.
 - Deployed a **serverless GenAI backend** using **LLaMA-3.1-8B-Instruct**, optimized for low-latency inference.
- IntelliTube – AI-Powered YouTube Insight Engine** ([Code](#)) 2025
- Developed a privacy-first, end-to-end AI pipeline for video insight extraction, running **100% locally** without third-party APIs.
 - Built an audio transcription pipeline using Faster Whisper, achieving **~95% transcription accuracy** across diverse accents and content types.
 - Implemented **NLP-based sentiment analysis** on **100+ scraped comments per video**, achieving **~91% classification accuracy** with multilingual translation support.

WORK EXPERIENCE

Diebold Nixdorf, Mumbai, India

Generative AI Intern

February 2026 - Present

- Contributing to **enterprise-scale** Generative AI solutions by enhancing an internal LLM-powered chatbot through prompt engineering, response evaluation, and output optimization to improve accuracy and reliability.
- Optimizing the **Retrieval-Augmented Generation (RAG)** pipeline by refactoring preprocessing modules, cleaning legacy code, and restructuring functions to improve maintainability, modularity, and pipeline efficiency.
- Performing prompt experimentation and LLM output evaluation to analyze model behavior, improve contextual responses, and reduce hallucinations.
- Supporting LLM **workflow optimization** by improving context handling and response consistency across the chatbot system.

FortytwoLabs, Pune, India

Software Engineer

June 2023 - July 2024

- Managed build & release processes for 6+ production C++ libraries, ensuring stable **cross-platform deployment** (Linux, Windows, macOS).
- Implemented **cryptographic features (policy signature modules)** to improve system security and reduce processing time
- Performed code testing on **143 unit tests** using **Google Test (GTest)** and **JUnit**, achieving **95% code coverage** and reducing bugs by **50%**.
- Conducted **API and performance testing** using **Postman** and **JMeter** to identify bottlenecks and improve system performance by **25%**.
- Documented 100+** system error codes, helping reduce troubleshooting time by **30%**.

FortytwoLabs, Pune, India

Technical Intern

Jan 2023 - May 2023

- Gained hands-on exposure to DevOps practices, including **CI/CD pipelines**, Docker containers, and Linux-based **deployment workflows**.
- Assisted in configuring Nginx and Linux-based reverse proxy setups** to optimize application routing & improve deployment reliability
- Configured Apache ActiveMQ within Docker containers on CentOS**, enabling reliable message brokering & client-server communication.

CERTIFICATIONS

- [Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate - Oracle](#) 2025
- [Oracle Cloud Infrastructure 2025 Certified Generative AI Professional - Oracle](#) 2025

PUBLICATION

- [IOT Based Crowd Detection and Stampede Avoidance using Predictive Analysis - IEEE Xplore · Apr 4, 2023](#) 2023

SKILLS

- Programming & Analytics:** Python | Data Structures & Algorithms | Object-Oriented Programming
- Tools & Libraries:** TensorFlow | PyTorch | Keras | Streamlit | Gradio | Fast-API | Flask | Apache Spark | Kafka
- Cloud & LLM/Agentic AI Frameworks:** LangChain | RAG Pipelines | Agentic Workflows | Prompt Engineering | Model Context Protocol (MCP) | LLaMA | Oracle Cloud AI | GCP | Microsoft Azure
- DevOps & MLOps:** Docker | Kubernetes | Jenkins | Version Control | Git | Prometheus | Grafana | Kubeflow | CI/CD
- Databases:** MongoDB | SQL | MySQL
- Frontend & Full-Stack :** JavaScript | Node.js | REST APIs | Client-Server Architecture | API Integration

ACADEMIC DETAILS

- Mukesh Patel School of Technology, Management & Engineering, NMIMS** 2024 - 2026
- M. Tech. – Artificial Intelligence (CGPA:9.10/10)
- Symbiosis Institute of Engineering and Technology, Symbiosis International University** 2019 - 2023
- B.Tech. – Electronics & Telecommunication Engineering
 - Honors in Mechatronics & Automation (CGPA: 7.11/10)