

ZEESHAN SHAIKH

📍 Pune, Maharashtra, India ✉ zinczee97@gmail.com ☎ 9890895611 📺 in/zeezinc 🌱 zeezinc.github.io

SUMMARY

AI-focused backend engineer with around 4 years of experience building production-grade LLM, NLP, and document intelligence systems, delivering scalable AI workflows and data-driven platforms using Python and cloud technologies.

EXPERIENCE

Gen AI Data Enthusiast

Hexalytics

June 2024 – Present, Texas, USA

- Developed **Generative AI chatbots and analytics solutions** by integrating **large language models (LLMs)** with **vector-based RAG pipelines** and multi-step AI workflows to support enterprise data needs, handling **huge number of document queries per month**.
- Built and delivered **AI-powered platforms** for **education, analytics, and real estate**, reducing manual effort by **50–70%** and improving system performance by **up to 3×** through caching and backend optimization.
- Designed and maintained **Python-based backend systems using FastAPI** to automate **document ingestion, search, summarization, and report generation** across **10k+ unstructured documents**.
- Implemented **end-to-end document AI pipelines** using **Python, LlamaParse, Whisper, OCR tools, metadata-aware chunking, and embeddings**, enabling efficient processing of **PDFs, Word, PPT, CSV, URLs, audio, and video**.
- Applied **LLMs for natural language interfaces**, including **chatbots and Natural Language to SQL insight generation**, reducing analyst dependency by **~50%** and improving data-driven decision-making.
- Deployed **cloud-native solutions on AWS** using **Docker and CI/CD pipelines**, supporting **production-scale deployments** with high availability and scalability.

Software Developer

HftSolution

October 2021 – February 2024, Gurugram, India

- Designed and developed **enterprise-grade backend systems** using **Java and Spring Boot**, delivering scalable solutions across **healthcare, facility management, and payroll domains**.
- Led development of **core functional modules** (prescriptions, vitals tracking, payroll processing), improving operational efficiency and reducing manual errors by **up to 60–70%** through validations and optimized workflows.
- Migrated legacy **monolithic applications to Spring Boot microservices**, improving system scalability, maintainability, and deployment flexibility.
- Built secure **RESTful APIs** with role-based access controls and data validation, ensuring **data integrity and compliance** for sensitive healthcare and financial information.
- Implemented optimized **database queries and reporting workflows**, improving data access and analysis performance by **~30%**.
- Improved system stability and quality by enforcing **rigorous testing, debugging, and code reviews**, reducing user-reported issues by **up to 80%**.

Software Developer Intern

Transition Computing

December 2019 – June 2020, Pune, India

- Assisted in migrating a **legacy backend application** from **.NET Framework to .NET Core**, improving **application maintainability, scalability, and alignment with modern backend development practices**.
- Implemented and optimized **database operations** using **Entity Framework Core (ORM)**, reducing **manual SQL scripting** and improving **development efficiency** across multiple **data access layers**.
- Used **GitHub** for **version control**, following best practices for **commits, branching strategies, pull requests, and collaborative code management**.
- Tested and validated **RESTful APIs** using **Postman** and **Swagger UI**, verifying **request/response workflows** across **multiple endpoints** to ensure **data integrity and API reliability**.
- Participated in **code reviews**, analyzing submitted changes and assessing the impact on **application functionality, performance, and stability**.

PROJECTS

AI-Powered RFP & Document Intelligence Suite

- Built an **end-to-end AI platform** for processing and analyzing **RFP / PDF / Word / PPT / CSV / URL / audio / video documents**.
- Engineered advanced **RAG/CAG pipelines, multi-model prompting, and structured extraction workflows**.
- Implemented **file ingestion and parsing** using **LlamaParse, OCR, PyMuPDF, Whisper (audio), and custom chunking logic**.
- Automated **TOC extraction, header-section mapping, metadata tagging, and python-docx report generation**.
- Reduced **manual Request for Proposal review time by 60%** and improved **extraction accuracy by ~40%**.
- Technologies:** Python, LangChain, LlamaParse, ChromaDB, Groq, Vertex AI, FastAPI, Docker.

Smart Teaching System

- An **intuitive generative AI system** used by **multiple Middle-East educational institutes**.
- Built **AI tools** for **lesson planning, assignments, MCQs, summaries, vocabulary generation, email writing, chapter insights, text-to-speech, and speech-to-text**.
- Integrated **fine-tuned LLMs** for **personalized academic content generation**.
- Reduced **teachers' manual workload by 70%** and improved **student experience** with **auto-generated learning materials**.
- Technologies:** OpenAI, Fine-Tuning, FastAPI, React, Python, Prompt Engineering.

Erdi Document Chatbot

- **Interactive conversational chatbot** allowing users to ask questions over **PDFs, Word, PPT, CSV, audio, video, and URLs**.
- Built **ingestion pipelines** including **OCR, file converters, markdown extraction, URL scrapers, and metadata-aware chunking**.
- Implemented **multi-document memory, follow-up reasoning, and collection versioning**.
- Enabled **dynamic context retention** with **high-accuracy vector retrieval**.
- **Technologies:** LangChain, Groq, ChromaDB, FastAPI, React, LlamaParse, Whisper, Python.

Student Information Bot

- **AI agent** that converts **natural language** into **SQL queries** and generates **data-driven insights**.
- Built **NL → SQL → execution pipelines** with an **auto-correction layer** to fix **SQL syntax errors**.
- Provided **advanced analytics** including **gender analytics, enrollment trends, and school performance insights**.
- Reduced **data analyst dependency by 50%** through **automated query generation and insight delivery**.
- **Technologies:** Claude, FastAPI, Snowflake, Python, SQL.

Property Finder Dashboard

- **High-performance backend** powering **real-estate search** for a **Middle-East platform**.
- Built **dynamic filter pipelines** and **caching layers**, improving **query performance by 3×**.
- Integrated **AWS services, asynchronous schedulers, cron jobs, and Redshift analytics** for scalable data processing.
- Reduced **Redshift load by 40%** through **aggressive caching strategies**.
- **Technologies:** Python, FastAPI, AWS (EC2, ECS, Secrets Manager, Redshift), Redis, Boto3, Cron, Docker.

Third Eye – Traffic Surveillance System

- Leveraged **advanced machine learning algorithms** and **image recognition techniques** to analyze **vehicle movements** and detect **traffic violations** in **images and videos**.
- Extracted critical information from **vehicle number plates** using **computer vision**, enabling detection of **signal violations, white line violations, triple riding, and helmet non-compliance**.
- Analyzed **10,000+ images and videos** to enhance **traffic management and enforcement**, contributing to **road safety improvements**.
- **Technologies:** Computer Vision, Machine Learning, Python, OpenCV, YOLO.

EDUCATION

Master’s in Science(Computer Science)

Savitribai Phule Pune University · Pune, India · 2020 · 8.25

Bachelor’s in Science(Computer Science)

Savitribai Phule Pune University · Pune, India · 2018

Indian School Certificate(ISC)

The Bishop's School(CAMP) · Pune, India · 2015

Indian Certificate Of Secondary Education (ICSE)

The Bishop's School (CAMP) · Pune, India · 2013 · 87%

SKILLS

Programming & Backend: Python, FastAPI, Flask, Django

AI & Machine Learning: NLP, RAG Pipelines(chunking, metadata, scoring), LangChain, LlamaIndex, Groq, Vertex AI, Whisper, Transformers (HuggingFace), Embeddings (sentence-transformers, OpenAI embeddings),PyTorch, TensorFlow

Document & Data Processing: Document Parsing & Classification, Table Extraction, Metadata Extraction, File Conversion Pipelines, Basic Computer Vision (YOLO), PyPDF2, pdfplumber, PyMuPDF, Unstructured.io, Tesseract OCR

Automation & Pipelines: Multi-step AI Pipelines, LLM Workflow Automation, Task Orchestration, Python Scripting

Data Engineering & Analysis: Pandas, NumPy

APIs & Integrations: REST API Development, OAuth/JWT, JSON-based services, Third-party AI API Integrations

Databases: PostgreSQL, MySQL, Redis, Snowflake, Vector DBs (ChromaDB, Qdrant)

Testing: PyTest, Postman, Swagger UI

DevOps & Deployment: Docker, GitHub Actions, Jenkins, AWS (EC2, ECS, Secrets Manager, Redshift), CI/CD

Version Control: Git, GitHub, Bitbucket

Architecture: RAG Architecture, Event-Driven Workflows, Document Automation Pipelines, Template-Based Content Generation, Metadata Search Systems

Soft Skills: Problem-solving, Critical Thinking, Communication, Collaboration