Atharva Chundurwar

Tempe, AZ, USA | atharvachundurwar841@gmail.com | +1(480)743-6844 | LinkedIn | GitHub

EDUCATION

Master of Science in Computer Science   **Aug 2024 - May 2026**
Arizona State University, United States   **(CGPA-3.89/4.0)**
Relevant Courses: Distributed Database Systems, Natural Language Processing, Data Mining, Machine Learning.

Bachelor of Technology in Information Technology with Specialization in Data Science   **May 2020 - May 2024**
Symbiosis Institute of Technology, Pune, India   **(CGPA-7.89/10)**
Relevant Courses - Data Structures and Algorithms, Operating System, Data Science, Database Management, Machine Learning.

PROFESSIONAL EXPERIENCE

miniOrange | Business Analyst Intern   **Jan 2024 - May 2024**
- Engineered **automated data pipelines using SQL** for real-time analytics, reducing manual effort by 60% and built dynamic Tableau dashboards to track system performance metrics across 5+ production deployments.
- Conducted market research on **customer behavior**, analyzing **1,000+ data points** to drive **strategic planning**, optimize product lifecycle, and improve product marketing initiatives.
- Partnered with the backend engineering team to integrate secure authentication protocols **(SAML, OAuth, LDAP)** into client systems, debugging auth flows and ensuring cross-platform compatibility in Windows and Linux environments.
- Formulated **User Success Stories** and gathered customer feedback to guide product improvements and introduced client incentive to increase customer satisfaction ratings by over 20%.
- **Revamped operational strategy**, clarified specific requirements for USA and Europe customers by engaging with over 10 clients resulting in **5+ project scope changes** saving over 2 weeks' time.

ACADEMIC PROJECTS

Manual Analysis of Reasoning Chains and Auto Evaluator using NLP, Python, Transformers, Deep Learning
- Evaluated open-source and closed-source LLMs (**Llama-3-8B-Instruct and Gemini-Pro 1.5 002**) on SOTA logical reasoning tasks using LogiQA and ARCT datasets.
- Developed a robust error taxonomy to categorize reasoning chain failures into **5 broad and 9 subcategories**.
- Built a custom **GPT-based** auto evaluator that achieved **81.98% accuracy** in classifying reasoning errors into core categories.
- Leveraged the use of technologies: Python, Hugging Face, Google Colab, Llama-3-8BInstruct and Gemini-Pro 1.5-00.2 Models.

LLM Powered Assessment Automation using GPT-4o and Langchain
- Developed an **AI-powered MCQ generator** using **GPT-4o**, automating personalized **quiz creation and auto-grading**, saving faculty significant time by auto-generating unique questions and evaluating answers per student.
- Designed multiple **LangChain-based pipelines**, including a **quiz generation chain** and an **evaluation chain**, and integrated them using **Sequential Chaining** for structured processing.
- Built a **Flask API** for a user-friendly interface, allowing seamless quiz generation and refinement.
- Automated **data processing and export**, enabling quizzes to be saved in **CSV format** for easy review and distribution.

AI-Med, Medical Chatbot using RAG and Vector Database
- Architected a **Retrieval-Augmented Generation (RAG)** pipeline to semantically retrieve medical knowledge from the Gale Encyclopedia using chunked embeddings and Pinecone vector DB improving query response relevance and latency.
- Implemented cosine similarity-based semantic indexing for low-latency, high-throughput query retrieval.
- Integrated **LangChain** with a locally hosted Meta **LLaMA 2** model and deployed a production-ready Flask application that fused vector-based document retrieval with dynamic LLM-based generation.
- Focused on optimizing vector storage layout and retrieval efficiency to simulate high-throughput, low-latency access patterns, echoing distributed storage behavior in cloud-scale systems.

Quantitative Analysis of Dysrhythmia using DL and ML
- Implemented deep learning models and machine learning algorithms to predict dysrhythmia with **85% accuracy.**
- Recommended model was adopted at Symbiosis Hospital, **reducing diagnostic time by 20%,** thereby optimizing patient care.
- Presented, published and authored a research paper at International IEEE Conference demonstrating the model's capabilities to over **100 attendees**.
- Employed technologies: Python, Deep Learning Models: VGG16-CNN and Modified CNN, GoogLeNet-CNN, ML Models: Random Forest, Decision Tree, SVM.

TECHNICAL SKILLS

- **Programming Languages:** Python, C, C++, SQL, R, PostgreSQL, HTML, CSS, Shell/Bash, Linux.
- **Technologies**: LangChain, Llama-Index, Ollama, Vector Databases (Pinecone, FAISS, Chroma DB), Hugging Face, AWS, PySpark, MongoDB, MySQL, Tableau, Power BI, CI/CD, Git, Hadoop, MS Office, Flask, Streamlit, GCP, Excel, DBT, MLflow, DVC.
- **Skills:** Machine Learning, Gen-AI, Data Science, Natural Language Processing, Deep Learning, Transformers, Data Analysis, Business Analysis, Data Analytics, Business Analytics, Advanced Mathematics, Data Engineering, Data Mining, Data Visualization, Agile, Database Management, MLOps, DevOps.
- **Libraries and Frameworks:** TensorFlow, PyTorch, BitsandBytes, Keras, Transformers, Hugging Face, LangChain, Llama-Index, OpenAI, Google-GenerativeAI, BeautifulSoup.