SINDHUJA CHADUVULA
Associate Applied ML Specialist

Toronto, Canada | +1 437-799-7401 | schaduvu@uwaterloo.ca | https://www.linkedin.com/in/sindhuja-chaduvula/

TECHNICAL SKILLS

Languages and Scripting: Python, C/C++, R, SQL, Javascript, Bash/Shell.
ML/AI Frameworks: Langchain, Langsmith, Langgraph, Phidata, CrewAI, LlamaIndex, OpenAI, OLLama, Hugging Face, Pytorch, Tensorflow, PySpark, Pandas, Numpy, OpenCV, NLTK, MLflow, Weights & Biases.
Generative AI & NLP: RAG, Finetuning LLMss, Prompt Engineering, Computer Vision, Deep Learning, Time Series Forecasting.
Cloud & DevOps: AES(S3), Azure, GCP, Docker, Kubernetes, Github Actions, Linux.
Soft Skills: Communication Skills, Time Management, Leadership, Adaptability, Teamwork, Decision Making, Problem Solving

EDUCATION

University of Waterloo   **Jan 2023 – Apr 2024**
Master of Engineering in Electrical and Computer Engineering
GPA – **3.90/4.00**
Key Focus: Data Modelling & Analysis, Image Processing, Text Analytics (NLP), Data Analysis & Management

EXPERIENCE

Associate Applied ML Specialist, Vector Institute (Toronto, Canada)   **Sept 2025 – Present**
- **Generated 125 synthetic text scenes and 125 MCQs** in the hiring and legal domain using GPT-4o Mini to train a bias detection model.
- **Researched AI explainability methods** (intrinsic, post-hoc, concept-based, mechanistic, and prototype-based) as part of an AI Safety project to improve model interpretability.
- Conducted **DPO training** with hint sampling and repeated sampling for **1,000 samples** of the Skywork reward preference dataset using the Qwen2-7B model.

Machine Learning Associate, Vector Institute (Toronto, Canada)   **May 2025 – Sept 2025**
- Developed a LangChain-based RAG pipeline to parse and semantically index **1,000-page** *Schedule of Benefits* documents using LlamaParse and Qdrant, **reducing manual look-up time for physicians by over 80%.**
- Optimized document retrieval by engineering hybrid strategies (OpenAIEmbeddings + BM25) with billing code filters, achieving **Recall@5 of 0.9** and **BERTScore of 0.8**, significantly boosting clinical response accuracy.
- Independently built the end-to-end retrieval and generation stack using Azure OpenAI models, **improving response precision and enabling scalable billing code assistance for clinical users.**
- Presented weekly results to **C-level stakeholders** to align delivery with strategic goals in a resource-constrained environment

Data Scientist, Genellipse (Mississauga, Canada)   **Aug 2024 – Dec 2024**
- Led a 4-member AI team to automate SEC 10-Q filings by fine-tuning LLMs (Gemma 2B, LLaMA 3, Mistral 7B, OpenAI) on **600 historical filings** across 75 public companies, **reducing manual workload by 60%.**
- Built scalable ETL pipelines using the EDGAR API to extract filings, process structured/unstructured data, and prepared LLM-ready datasets in AWS S3, **accelerating document readiness by 70%.**
- **Containerized** Mistral-based submodels using **Docker** for distinct SEC filing sections, and integrated **automated JSON structure validation** to ensure **reliable downstream parsing** and **deployment**.
- Delivered a functional PoC for SEC 10-Q generation with **40% template-level fill accuracy**, aligned with ethical compliance protocols and highlighting model adaptation gaps for roadmap planning.

AI Engineer, AIFocal (Markham, Canada)   **Sept 2023 – Oct 2023**
- Built a PDF-aware QA chatbot using Gemma 2B and LangChain RAG, integrating Milvus for retrieval to **enable accurate summarization and long-form reasoning over user-uploaded content.**
- Designed a **prompt routing system** using keyword + semantic similarity for dynamic query handling.
- Achieved **85% accuracy on synthetic QA benchmarks**, validating system performance for extracting answers from complex PDFs.
- Contributed to prompt engineering, retrieval tuning, and vector DB design as part of a 3-member Agile team that **delivered a working demo within 4 weeks.**

Assistant System Engineer, Tata Consultancy Services (Hyderabad, India)   **Oct 2021 – Dec 2022**
- Built **churn prediction models** (**Random Forest, XGBoost**) with **87% accuracy** on 5 years of telecom data; visualized key features using **Matplotlib** and **Seaborn.**
- Engineered **preprocessing pipelines** (capping, flooring, outlier handling) to support churn analysis; collaborated in Agile cycles to deliver actionable insights for retention.

TECHNICAL PROJECTS

Multi-PDF Conversational Assistant
Tech: Langchain, FAISS, Open AI, Streamlit, Information Retrieval
- Built an **AI-powered chatbot** to parse and answer questions from multiple PDFs using **vector indexing** and **semantic search**.
- Integrated **FAISS** with **LangChain** for efficient retrieval and deployed the solution with a **Streamlit UI**.