AISHWARYA NAIR

anair@umass.edu | 6692077462 | San Jose, CA | aishwaryanair.me | LinkedIn | Github

SUMMARY

Software Engineer with experience building end to end systems, APIs, and automation tools using Python, FastAPI, and cloud platforms like AWS and OpenShift. Developed scalable services integrating LLMs and vector search, and contributed to internal tooling for data and ML workflows. Skilled in system design, CI/CD, and cross-functional collaboration for delivering production-ready software.

SKILLS

- **Programming:** Python, SQL, JavaScript, AJAX, C++, Shell scripting, Typescript
- **Machine Learning & AI:** LangChain, LlamaIndex, OpenAI Fine-Tuning API, Scikit-learn, ML pipelines, NLP models (BERT, MPNet), ChromaDB, MLFlow
- **Frameworks & Libraries:** TensorFlow, PyTorch, Keras, JAX, OpenCV, Flask, Django, FastAPI
- **Cloud & Infrastructure:** AWS, Azure, Kubernetes, Docker, OpenShift
- **Big Data & Orchestration:** Databricks, PySpark, Apache Spark
- **API Development & Optimization:** REST API, Pytest, MagicMock

WORK EXPERIENCE

Data Scientist - Co-Op   *May 2024 - January 2025*
Boehringer Ingelheim (Leading global biopharma company)

- Optimized Databricks job orchestration to **reduce pipeline latency by 60 seconds;** integrated results into Streamlit dashboards for near-real-time stakeholder visibility.
- Designed and deployed GPT-4-based RAG pipelines using LangChain, LlamaIndex, and ChromaDB; integrated OCR processing and TypeScript frontend, **reducing manual workload by over 4,000 FTE hours monthly.**
- Automated data transfer between OpenShift PVC and AWS S3 using Kubernetes, Docker, FastAPI, MLFlow and Jenkins; replaced manual workflows with scalable backend services, **saving 2,000+ FTE hours annually.**
- Identified and resolved performance bottlenecks in LLM API deployments via stress testing of GPT-4o and Claude models; benchmarked latency across Azure and Anthropic endpoints to inform optimization strategy.
- Built backend service integrating GPT-4 classification for document triage, **enabling 50% reduction in manual review time and delivering $30K+ in cost savings.**
- Implemented ML-based anomaly detection system for protocol deviations using Spark, XGBoost, and logistic regression; handled large, imbalanced datasets and **achieved 85% balanced accuracy, saving 3,000+ FTE hours monthly.**

Artificial Intelligence Engineer   *September 2021 – August 2023*
Heystack Inc. (Early-stage start up)

- Built and deployed multi-class emotion classification models using BERT and MPNet; **achieved 70% accuracy** across 23 emotion labels and delivered to PepsiCo's food science team.
- Fine-tuned transformer models to **improve text classification accuracy by 20% for VFCorp**, enabling structured analysis of large unlabelled datasets.
- Developed an end-to-end inference system using Flask, Docker, SQL, and JavaScript on AWS EC2, **handling 2,000+ model queries per request via scalable RESTful APIs.**
- Implemented clause extraction pipeline with morphological parsing and dependency trees to **improve downstream transformer model accuracy by 5%.**
- Used OpenAI's Fine-Tuning API to optimize classification performance, **achieving a 2% model accuracy gain** in production workflows.

PROJECTS

Forest fire prediction using LSTM models — Student team leader   *January 2020 - May 2021*
- Awarded **$15,000 Microsoft AI for Earth grant**; presented project outcomes at INCET 2021.
- Deployed LSTM-based forest fire forecasting model on Azure using TensorFlow, achieving 85% accuracy for early risk detection.
- Led a team of 6 to build a Django-based web application visualizing real-time fire predictions with geospatial mapping and automated data preprocessing.

ASSIST — Team leader (Github URL)   *August 2019 - May 2020*
- Selected as a **regional finalist** at the eYantra Ideas Competition hosted by IIT Bombay.
- Developed a computer vision model using TensorFlow and MobileNet to detect braille, ramps, and other accessibility features in architecture, **achieving 83% object detection accuracy.**
- Developed a voice-enabled interface using PHP and Bootstrap to deliver accessible, screen-free navigation support.

LCP — Team leader (Github URL)   *September 2019 - April 2020*
- **Super Team awardee of Project Deep Blue** competition organized by Mastek & Majesco.
- Predicted Mumbai's 2030 water demand using Python and Scikit-learn, applying data mining and ML algorithms to achieve **86% accuracy** and support future planning decisions.

EDUCATION

Masters of Science in Computer Science   *Sept 2023 – May 2025*
University of Massachusetts Amherst
- **GPA: 3.9**
- Coursework: Reinforcement Learning, Machine Learning, Applied Statistics, Distributed Operating Systems.