Ayman Mahfuz

ayman.afeef@gmail.com| (512)-705-8897 | http://www.aymanmahfuzportfolio.com/

EDUCATION

The University of Texas at Austin, Austin, TX *August 2023 - May 2027*

Double B.S. in Computer Science & Mathematics, Minor in Business, Concentration in Machine Learning & Artificial Intelligence

Courses: Data Structures, Computer Architecture, Computer Systems, Discrete Math, Linear Algebra, Statistics & Probability

SKILLS

Programming & Libraries: Python, Java, C, JavaScript, HTML/CSS, SQL, PHP, Node.js, React.js, C++, Flask, Django, Pandas, NumPy, Scikit-learn, Ruby, ARM64, Postgresql, CUDA

Tools: IntelliJ, VSCode, Eclipse, Google Cloud Platform, Jupyter Notebooks, Git, AWS

EXPERIENCE

The University of Texas at Austin - Center of Media Engagement *Aug 2023 – Pres*

Software Engineer Research Assistant

- Engineered large-scale robust Python pipelines for scraping, preprocessing, & uploading 50M+ news articles & 70M+ comments to BigQuery, employing APIs, sitemaps, HTML parsing, Pandas, & NumPy. Developed dynamic dashboards using SQL, Matplotlib, & Looker Studio to track data collection progress & fill gaps programmatically
- Led machine learning initiatives, fine-tuning a DistilBERT model (Hugging Face) to classify news headlines & comments with 99% accuracy & high precision, recall, & F1 score. Conducted advanced research on clickbait trends & personal stories in comments, leveraging NLP, CUDA & extensive data analysis to derive insights for upcoming publications on misinformation.
- Designed and deployed a research platform with React/Tailwind, Flask, and Firebase, featuring 3 interactive games, real-time analytics tracking 15+ metrics, MTurk integration, and 99.9% uptime serving 1,000+ participants.

The University of Texas at Austin – Dell Medical School *Aug 2023 – Pres*

Machine Learning Research Assistant

- Led a 3-member team in developing advanced ML models for abdominal organ segmentation, significantly improving pancreas segmentation accuracy using MedSAM 2, MONAI, TransUNet, & ResNet-50 with ViT models in PyTorch.
- Engineered Python pipelines for preprocessing large 3D MRI datasets & conducted comprehensive data analysis using Scikit-learn, Statsmodels, & Matplotlib to assess model robustness & performance

The University of Texas at Austin – School of Information *Feb 2024 – Pres*

Machine Learning Research Assistant

- Conducting research on diagnostic reasoning in multiagent LLM systems for medical queries, assessing consistency and accuracy in responses with statistical analyses. Developed Python scripts using Autogen and GPT-4 API to test if multiagent LLMs reason reliably across varied and misleading contexts.

The University of Maryland, College Park *Jun 2022 – Jan 2024*

Software Engineer Research Intern: "Towards Designing a Question-Answering Chatbot for Online News"

- Developed NLP-driven chatbot with Python & NLTK, co-authored CHI 2024 conference paper which had linguistic insights

Lockheed Martin *Jun 2022 – Oct 2022*

Software Engineer Intern

- Optimized CRM workflows & refined Configuration Database through JavaScript & RPA integration. Enhanced data accuracy by 25% & streamlined internal processes, resulting in 30% improved operational efficiency

PROJECTS

Inkwell: YouTube for Books

- Engineered a full-stack book-sharing platform using React, Django, & PostgreSQL, featuring a comprehensive RESTful API with 50+ endpoints, JWT authentication, real-time analytics, custom rich text editing, AWS integration, intelligent search functionality, & an advanced multi-step upload process with draft saving, while implementing scalable database schemas & efficient data loading techniques to optimize performance for complex user-book interactions

Leetcode Matchmaker

- Developed a web application that finds & displays LeetCode problems solved similarly to a given problem using cosine-similarity on problem vectors, leveraging Machine learning techniques, utlized React for the frontend, & Flask for the backend

Pintos Operating System (C)

- Implemented core OS components in C for Pintos, including a priority scheduler, user programs, virtual memory, and a filesystem; achieved 100% test coverage for synchronization primitives, system calls, and memory management.