Ayman Mahfuz

US Citizen | ayman.afeef@gmail.com| (512)-705-8897 | LinkedIn: aymanmahfuz | GitHub: AymanMahfuz27

EDUCATION

The University of Texas at Austin, Austin, TX *May 2027*

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

SKILLS

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

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

Certifications: Data Scientist: Machine Learning – Codecademy, Software Design Principles – Codecademy

EXPERIENCE

The University of Texas at Austin, Austin, TX *Aug 2023 – Pres*

Software Engineering Research Assistant – Department of Communication

- Engineered robust **Python** pipelines for scraping, preprocessing, and uploading data to BigQuery. Employed various scraping techniques (**APIs, sitemaps, HTML parsing**), leveraged **Pandas** and **Numpy** heavily, and ensured data integrity with automated cron jobs for daily updates. Processed over **50 million news articles** and **70 million comments** from 12+ domains, maintaining a comprehensive database dating back to 1851.
- Developed dynamic dashboards and visualizations using **SQL, Pandas, Matplotlib**, and **Looker Studio** to track data collection progress, identify and programmatically fill data gaps, enhancing the database's reliability and accessibility for research purposes.
- Led **machine learning** initiatives, creating and evaluating models to classify news headlines and comments. Fine-tuned a **DistilBERT** model using Hugging Face, achieving high accuracy and performance metrics (**precision, recall, F1 score**) through comprehensive evaluations.
- Conducted advanced research projects: analyzed clickbait trends over a decade and investigated the prevalence of personal stories in comments. Leveraged machine learning algorithms and extensive data analysis to derive actionable insights, contributing to upcoming research publications that aim to understand and combat the spread of misinformation online.

Machine Learning Research Assistant – Dell Medical School *Aug 2023 – Pres*

- Led a team of 3 members to revolutionize abdominal organ segmentation by developing advanced machine learning models for medical imaging. Built on existing MONAI frameworks to segment the pancreas from MRI scans, enhancing accuracy and performance using the latest vision transformer technologies and the TransUNet library.
- Engineered robust data preprocessing pipelines: Designed and implemented Python scripts for preprocessing 3D MRI images, including clipping, normalization, and extraction of 2D slices, using libraries such as Nibabel, Pydicom, NumPy, and H5py. Ensured efficient data handling and storage for large-scale medical datasets, facilitating downstream analysis and model training.
- Developed and fine-tuned machine learning models for pancreas segmentation using advanced architectures like TransUNet and ResNet-50 with ViT. Implemented model training, validation, and evaluation processes using PyTorch, achieving significant improvements over existing benchmarks. Utilized metrics such as Dice Score to assess model performance.
- Conducted comprehensive data analysis and visualization: Employed statistical and machine learning techniques to analyze the impact of contextual changes on model performance. Utilized tools such as Scikit-learn, Statsmodels, and Matplotlib to perform descriptive statistics, correlation analyses, and visualize segmentation results, driving insights into model behavior and robustness.

Machine Learning Research Assistant – School of Information *Feb 2024 – Pres*

- Led a team in the systematic design and implementation of a comprehensive Python script for assessing the consistency of MedAgents in diagnosing medical conditions. Utilized the Autogen library and GPT-3.5 API, incorporating features such as exponential backoff for API rate limiting, offline script execution, and organized data collection. Ensured robust performance through well-structured code, modular design, and extensive documentation.
- Engineered robust data pipelines using Pandas and NumPy for data preprocessing and cleaning. Developed methods to introduce variations to medical questions, ensuring clean and structured datasets for analysis. Managed data storage and processing in JSON format, facilitating efficient and accurate data handling.
- Developed models to evaluate the impact of contextual changes on the diagnostic consistency of MedAgents. Assessed the influence of variations such as patient age or symptom modifications on model responses, identifying patterns and key

factors affecting diagnostic outcomes. Utilized advanced machine learning techniques to analyze and interpret results, contributing to the project's goal of enhancing the reliability of AI in medical question answering.

- Conducted advanced data analysis and visualization using a variety of statistical and machine learning techniques. Employed tools such as Scikit-learn, Statsmodels, and Matplotlib for analyses including descriptive statistics, correlation matrices, ANOVA, Chi-square tests, Cohen's Kappa, and logistic regression. Created comprehensive dashboards and visualizations to track progress, identify patterns, and derive actionable insights, significantly contributing to the project's overall success.

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

Software Engineering Research Assistant: "Towards Designing a Question-Answering Chatbot for Online News"

- Developed an NLP-driven chatbot, improving online engagement & analyzing human-chatbot dynamics, contributing to a CHI 2024 conference paper
- Conducted advanced text analytics & Python scripting for data analysis, producing key linguistic insights & visualizations

Lockheed Martin, Remote *Jun 2022 – Oct 2022*

Software Engineer Intern

- Optimized CRM workflows & refined the Configuration Database, enhancing operational efficiency & data integrity through JavaScript & RPA integration

AT&T, Remote *Jun 2021 – Aug 2021*

Summer Extern

- Engaged with industry leaders & collaborated on tech-focused strategies, enhancing AI & technology insight, & contributing a unique international perspective towards shaping future tech innovation.

PROJECTS

Inkwell: YouTube for Books *Team Lead*

- Developed Inkwell, a dynamic book-sharing platform, leading full-stack development with React, PostgresQL, & Django, enhancing user experience & preparing for future growth & monetization

AI Dermatologist *Team Lead*

- Leading the AI Dermatologist project to innovate skincare recommendations with a single facial scan using vision transformers, public Kaggle datasets, and content-based machine learning filtering

Leetcode Matchmaker *Independent Project*

- Developed a web application that finds and displays LeetCode problems solved similarly to a given problem to aid in interview prep, using Machine learning, React for the frontend, & Flask for the backend.