# Ayman Mahfuz

US Citizen | ayman.afeef@gmail.com| (512)-705-8897 | LinkedIn: aymanmahfuz | Website: Ayman Mahfuz

## 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*

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, MATLAB, C++, Flask, Django, Pandas, NumPy, Scikit-learn, Ruby, ARM64, Postgresql, CUDA

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

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

## EXPERIENCE

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

*Software Engineer Research Assistant – Center of Media Engagement, Moody College of Communications*

- Engineered large-scale robust **Python** pipelines for scraping, preprocessing, and uploading **50M+** news articles and **70M+** comments to **BigQuery**, employing **APIs**, sitemaps, **HTML parsing**, **Pandas**, and **NumPy**. Developed dynamic **dashboards** using **SQL**, **Matplotlib**, and **Looker Studio** to track data collection progress and fill gaps programmatically.
- Led **machine learning** initiatives, **fine-tuning** a **DistilBERT** model (Hugging Face) to classify news headlines and comments with 99% accuracy and high **precision, recall,** & **F1 score**. Conducted advanced research on clickbait trends and personal stories in comments, leveraging **NLP, CUDA** and extensive **data analysis** to derive insights for upcoming **publications** on misinformation.

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

- Led a 3-member team in developing **advanced ML models** for **abdominal organ segmentation**, leveraging **MONAI** frameworks and **TransUNet** to enhance pancreas segmentation accuracy from MRI scans. Implemented and **fine-tuned TransUNet** and **ResNet-50** with **ViT** models using **PyTorch**, achieving significant improvements in Dice Score metrics.
- Engineered robust **Python** data preprocessing **pipelines** using **Nibabel, Pydicom, NumPy**, and **H5py** for efficient handling of large-scale **3D MRI datasets**. Conducted comprehensive data analysis and visualization using **Scikit-learn, Statsmodels**, and **Matplotlib** to assess model robustness and drive insights into contextual impacts on performance.

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

- Led team in designing and implementing comprehensive **Python** script to assess **MedAgents'** diagnostic consistency using **Autogen** library and **GPT-4 API**. Engineered robust data pipelines with **Pandas** and **NumPy** for preprocessing, cleaning, and introducing variations to medical questions, managing data in **JSON** format.
- Developed models to evaluate contextual impacts on MedAgents' diagnostic consistency. Conducted **advanced data analysis** using **Scikit-learn, Statsmodels**, and **Matplotlib**, employing techniques such as **ANOVA, Chi-square tests**, **Cohen's Kappa**, and **logistic regression** to derive actionable insights and enhance **AI reliability** in medical question answering.

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

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

- Developed NLP-driven chatbot to improve online engagement, contributing to CHI 2024 conference paper. Conducted text analytics and Python scripting for data analysis, producing key linguistic insights and visualizations.

**Lockheed Martin** *Jun 2022 – Oct 2022*

*Software Engineer Intern*

- Optimized **CRM** workflows & refined **Configuration Database** through **JavaScript** & **RPA** integration

## PROJECTS

**Inkwell: YouTube for Books**

- Engineered a full-stack book-sharing platform using **React**, **Django**, and **PostgreSQL**, featuring a comprehensive **RESTful API** with **50+ endpoints**, **JWT authentication**, **real-time analytics** & **custom rich text editing.** Integrated **AWS** for scalable cloud storage & **CI/CD pipelines** using **GitHub Actions** for **automated testing** and **deployment.** Optimized performance through **scalable database schemas** & **efficient data loading** techniques

**Leetcode Matchmaker**

- Developed a web application that finds and 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.