Wenhao He

wenhaohe8@gmail.com GitHub LinkedIn **+1 347-824-4797**

Summary

Detail-oriented **MSc** graduate in AI from **University at Buffalo**, specializing in **data analytics** and **machine learning**. Proficient in statistical analysis, data wrangling, and predictive modeling using **Python, SQL, Hadoop, and Spark**. Experienced in deriving actionable insights from large datasets and presenting findings using advanced visualization tools like **Tableau and Matplotlib**.

Education

- **University at Buffalo, The State University of New York** *Aug 2022 - May 2024*
  Master of Science in Engineering Science focus on Artificial Intelligence
  Coursework: Machine Learning, Deep Learning, Computer Vision, Reinforcement Learning, Data Intensive Computing, Pattern Recognition

- **University at Buffalo, The State University of New York** *Aug 2018 - May 2022*
  Bachelor of Science in Computer Science
  Coursework: Data Structures, Computer Networking, Web Application, Algorithms and Complexity, Systems Programming, Theory of Computation, Computational Linguistics, Malware Analysis

Experiences

CAN International Corp: Software Engineer Internship *Apr 2024 - Present*
New York, NY · Remote

- Technology used: **React Native, JavaScript, Expo, Microsoft Azure, Firebase**
- Developed a full-stack CAN member app with job searching and mentor-mentee matching functionalities
- Implemented secure data storage and authentication using Azure and Firebase, enabling real-time data fetching
- Collaborated with a remote team to enhance app features and improve user experience
- Conducted extensive testing and debugging to meet all functional and performance requirements

Projects

Audio Cloning: Project Leader & Software Engineer *Aug 2023 - Feb 2024*
Developed SV2TTS model to replicate a judge's voice from the Brown v. Board case

- Technology used: **Python, PyTorch, Matplotlib, Librosa**
- Implemented a **three-stage pipeline deep learning SV2TTS** to clone voices unseen during the training
- Applied UMAP for efficient dimensionality reduction, preserving key characteristics of voice data
- Conducted experiments with different text lengths and both seen and unseen texts to enhance audio generation
- Achieved 90% voice replication accuracy through auditory comparison, enhancing digital record authenticity

Costumer Maximum Open Credit Prediction System: Software Engineer *Aug 2023 - Jan 2024*
ML model to estimate customer credit limits using comprehensive financial data

- Technology used: **Python, Numpy, Pandas, Scikit-learn, Hadoop**
- Designed an intuitive GUI for easy interaction with the prediction system
- Preprocessed data from bank customer dataset, implementing cleaning and transformation techniques
- Conducted EDA with Hadoop on large datasets, gaining insights into key variables for credit predictions
- Developed and compared machine learning models to estimate credit limits, optimizing performance

Skills

Cloud/Databases: MongoDB, MySQL, AWS
Frameworks: Tensor-Flow, Scikit-learn, Hadoop, React/React Native, Flask, Node.JS
Developer Tools: GitHub, Git, Docker, Jupyter Notebook, PyCharm, Visual Studio Code