Danying Xu

Education

New York University *New York, NY*
Master of Science in Computer Engineering, Grade: 3.89/4.0 Sep. 2023 - May. 2025

Southeast University *Nanjing, China*
Bachelor of Engineering in Artificial Intelligence Sep. 2019 - Jun. 2023

Technical Skills

Programming Python (Scikit-learn/Pytorch/Tensorflow/OpenAI), C++/C, Linux, Java, SQL, Cypher, LaTeX
Tools Azure, Hugging Face, Heroku, MySQL, PostgreSQL, Git, Github, Apache Spark/Hadoop, Hive, Dockers, K8s, Neo4j
Methodologies Machine Learning, Deep Learning, NLP, CV, LLM, Database Management, Dockers, Statistics, Data Analysis

Professional Experiences

Machine Learning Engineer *New York, United States*
Global AI May. 2024 - Aug. 2024

- Deployed a news chatbot based on LangChain on **Heroku** using **Git** for continuous integration.
- Used **Bucketeer** for **AWS S3-compatible** storage to efficient handle data input, reducing storage costs by 30%.
- Created a storage optimization method with auto updates for faster web responses, cutting reprocessing time by 95%.

Machine Learning Engineer *New York, United States*
Global AI Jan. 2024 - Apr. 2024

- Established a **Postgres** database on 300k+ GDELT news data, boosting time efficiency by 20%.
- Automated database upgrading with parallel processing which enhanced overall efficiency by 15%.
- Analyzed and visualized descriptive statistics on 1 million MSCI US Index stocks to identify trends and anomalies.
- Developed a **LSTM** model with a 0.4 MSE score, enhancing stock forecasts by 19.7% over the baseline.

Software Development Engineer in Test *Nanjing, China*
Huawei Nanjing Research & Development Center Aug. 2022 - Sep. 2022

- Conducted **Gray Box Testing** by examining 143 static path graphs with thousands of functions in **C/C++**.
- Performed **White Box Testing** using FUZZ test technology on 872 code files in **Linux**, expected to improve product performance by 30%.

Projects

AI-Generated Text Detection *New York, NY*
New York University Apr. 2024 - May. 2024

- Used Apache Spark to handle and analyze 50k+ human-written and AI-generated sentences.
- Researched state-of-the-art text detection models and trained 3 baseline models with average accuracy of 86%.
- Finetuned the **BERT** model as a light language model with the accuracy of 93.2%.
- Utilized the **LlaMa2** model on Google Colab and **Hugging Face** with prediction accuracy of 65.4%.
- Deployed the **ChatGPT3.5 model** through **Azure OpenAI** and **Azure Notebook** with accuracy of 80.1%.

Text Gender Bias Rewriter *Nanjing, China*
Southeast University Dec. 2022 - Jun. 2023

- Designed an NLP framework to reduce data gender bias via pattern transform, neural translation and data aggregation.
- Implemented **Seq2Seq** and **Seq2Seq attention** models (character/word level) on 148k+ Chinese sentences on **Pytorch**.
- Devised the Word-Embedding Association Test to Chinese evaluated on **CBOW** model, reducing gender bias by 45.4%.
- Performed Coreference Resolution downstream task using **wwm-RoBERTa** model, maintaining consistent performance around 92% after reducing gender bias.
- Conducted Sentimental Analysis on **TextCNN** model with consistent performance of 80% after reducing gender bias.

Deep Learning-based Explanatory Brain Science *Nanjing, China*
Southeast University Nov. 2020 - May. 2022

- Extracted 1.2 million images from 1297 videos of trained monkeys playing Pac-Man game using **Python (OpenCV)**.
- Designed a **ConvRNN model with AlexNet and LSTM** on **TensorFlow**, predicting player moves with 84.6% accuracy.
- Performed **Class Activation Map (CAM) heatmaps** for activation layer visualizations.
- Modified the **Grad-CAM heatmap** for each layer to interpret the brain's decision-making mechanism with visualizations.

Awards and Honors

May 2022 **Finalist** Interdisciplinary Contest in Modeling (ICM) by COMAP *United States*
Nov 2021 **Outstanding Award** 17th "Challenge Cup" Chinese National Competition *China*