Andy Tran

Dracut, MA | 978-995-9369 | andytranswe@gmail.com | linkedin.com/in/AndyTranSWE | github.com/AndyTranSWE

EDUCATION

University of Massachusetts Lowell — Lowell, MA

Bachelor of Science in Computer Science — GPA: 3.99/4.00 — December 2024

Master of Science in Computer Science — GPA: 3.93/4.00 — May 2025

TECHNICAL SKILLS

Languages: Python, C, C++, Javascript, SQL, Assembly (x86, MIPS)
Frameworks: PyTorch, Boost, Bootstrap, jQuery
Developer Tools: Git, GCC, Make, Valgrind, Google Colab, Jupyter Notebook, Weights & Biases, OpenSSH, Docker, Amazon Web Services
Libraries: NumPy, pandas, Matplotlib, Keras, PyTorch, scikit-learn, NLTK, HuggingFace, DeepSpeed, Accelerate, FlashAttention

PROJECTS

Master's Thesis | *PyTorch, HuggingFace, DeepSpeed, Accelerate, Weights & Biases*　　January 2025 – May 2025
- Designed and developed an intellectual property protection scheme for large language model training data
- Adapted fragile watermarking with custom datasets to detect changes in weights of 7-billion parameter models
- Implemented watermarking through fine-tuning, achieving 100% successful implantation
- Optimized GPU use with Accelerate configurations, DeepSpeed CPU offloading, and low-rank adaptation, reducing fine-tuning times against comparable published methods by 65%

Optimized ColonGPT | *PyTorch, DeepSpeed, HuggingFace, FlashAttention, Docker*　　January 2025 – April 2025
- Collaborated with a team to modify the multimodal language model ColonGPT to classify colonoscopy images with greater efficiency and accuracy than the public baseline
- Refactored preprocessing code and batching logic to increase modularity
- Optimized number of training samples and complexity of SigLIP visual encoder, reducing full training run times by approximately 80%
- Increased accuracy on the testing set from 83% to 88%

Song Classifier | *PyTorch, NumPy, pandas, scikit-learn, Matplotlib*　　September 2024 – December 2024
- Coordinated with a team to develop three machine learning models for automatic song labeling
- Implemented an encoder-only Transformer to classify songs into genres based on lyrics
- Identified and eliminated sources of data contamination between testing and training sets, facilitating valid evaluation
- Evaluated performance in multinomial classification with four target genres across 1,000,000 songs, achieving 65% accuracy
- Increased accuracy over baseline bidirectional recurrent neural network by 35%

Weight Predictor Model | *Python, Keras, scikit-learn, NumPy, pandas, Matplotlib*　　February 2024 – March 2024
- Coordinated team to create a machine learning model to estimate obesity levels according to environmental and social factors
- Collaborated over Google Colab to identify flawed model outputs, opportunities to remove outliers, and class imbalances
- Implemented an ensemble stacking model that estimated the probability of being overweight with an accuracy of 94.8%, using a dataset of over 2,000 individuals and 14 factors
- Cleaned dataset, improving accuracy by 10.4% using scikit-learn standardization functions

Adversarial Hangman | *C, Valgrind*　　January 2023 – May 2023
- Developed a terminal-based hangman game with custom classes that dynamically changes the target word based on player guesses
- Enabled game logic through creating new string, boolean, vector, and AVL tree structures
- Achieved properly functioning hangman behavior for 13,000 different English words
- Eliminated 100% of detected memory leaks using Valgrind and fixed program crashes