Linh Tran

ltran18@ur.rochester.edu | linkedin.com/in/linh-ph-tran/ | github.com/lnhtrn

EDUCATION

University of Rochester
MS in Computer Science, Artificial Intelligence Specialization *Anticipated 12/2025*

University of Rochester, GPA: 3.5
BS in Computer Science, Minor in Mathematics *12/2024*
- **Relevant Coursework**: Machine Learning, Natural Language Processing, Computer Vision

SKILLS

- **Languages & Tools:** Python, PyTorch, HuggingFace, Transformers, Pandas, Matplotlib, C++, Java
- **ML/AI:** NLP, Large Language Models (LLMs), RAG, Vector databases, Model deployment
- **Data:** Microsoft Azure, Data visualization, Unstructured data processing, MLOps workflows
- Cross-functional collaboration, Technical problem-solving, Clear communication, Adaptability

WORK EXPERIENCE

kLab, University of Rochester **Rochester, NY**
Deep Learning Research Assistant *01/2024 – Present*
- Utilized prompt engineering to conduct comprehensive analysis on racial bias in Vision-Language Assistants (VLAs), revealing racial biases in personality traits, skills, and occupational representation.
- Optimized neural network architectures and implemented bias mitigation techniques across 20+ computer vision models from HuggingFace using PyTorch Lightning.
- Improved model fairness and accuracy by 10-15% through strategic data augmentation and architecture modifications.
- Integrated CI/CD pipelines for automated testing and deployment of improved model versions.

Dye Lab, University of Rochester Medical Center **Rochester, NY**
Data Science Team Lead *09/2024 – Present*
- Lead multidisciplinary team bridging data science and social science students to analyze COVID data on health and social oppression.
- Leveraging NLP and prompt engineering techniques to build an LLM-based system for analyzing 10,000+ unstructured data responses on identity and oppression.

VinBigData **Remote**
Data Science Intern *05/2023 – 08/2023*
- Refined Large Language Model (LLM) question classifier using PyTorch and Transformers, achieving 87% accuracy on 100,000-question dataset.
- Enhanced model robustness by augmenting SQuAD 2.0 dataset with 50,000 negative sample pairs, improving DeBERTa chatbot accuracy by 4-10%.
- Developed and deployed a multilingual semantic analysis model using TorchServe, improving processing speed by 200%.

MBBank
Data Science Intern *06/2022 – 08/2022*
- Engineered data pipeline with Python and Selenium to build a 60,000-entry real estate database based on publicly available geographic data.
- Developed and optimized ensemble machine learning models (XGBoost, LightGBM, CatBoost) for real estate price prediction, achieving RMSE of 0.17.
- Improved website traffic by 120% through implementation of accurate price prediction models.

PROJECTS & ACHIEVEMENTS

DoubletDetection – Biomedical Data Science Hackathon, University of Rochester (link)
Team Leader, First Place, Undergraduate Division *08/2023*
- Led 3-person team through rapid development cycle, defining project scope under 48-hour deadline.
- Implemented and optimized models for detecting doublet cells in single-cell RNA sequencing data.
- Applied feature engineering, feature selection, and ensemble techniques to achieve highest accuracy in undergraduate division.