Henry Vu

469-649-6967 | henry@gmail.edu | linkedin.com/in/henry | github.com/henry | henry.com

Education

University of Texas at Dallas — Richardson, TX
M.Sc. Computer Science - Intelligent Systems — *Sep. 2024 - Expected May 2026*

University of Alberta — Edmonton, AB
B.Sc. Computing Science with Honors — *Sep. 2019 - May 2024*

- **GPA: 3.83/4.0**, graduated *Summa cum laude*. International Student Scholarship, Dean's list 2020 - 2024.
- **Coursework**: Deep Learning, RL, CV, NLP, Probability Theory, Optimization Theory, Algorithms, Databases.

Experience

Augmented Reality/Machine Learning Developer — Aug 2025
eXRealityAI — *Dallas, TX*

- Built voice-to-voice **RAG** system (Whisper STT → hybrid retrieval with BM25 + FAISS → Mistral-7B-q4 → Kokoro TTS) achieving sub-10s latency with local inference on NVIDIA Jetson Orin AGX.
- Engineered NLP pipeline with EmbeddingGemma-300m, BAAI cross-encoder reranking, query rewriting, and intent-aware retrieval reducing processing by **40-70%**.
- Integrated **Gemini 2.5** with structured JSON schema and context-aware prompt engineering, orchestrating stateful multimodal workflows combining **YOLOv9** and **Wit.ai** voice input across 2 XR applications.
- Mentored development teams on **LLM API** integration and prompt engineering for Meta Quest applications.

Computer Vision Engineer — Feb 2025 – Present
ThorMed Innovation — *Dallas, TX*

- Led domain-aligned transfer learning for NIH-funded bladder segmentation project. Pretrained **U-Net** and **SimSiam SSL** encoders on **9.2K** thyroid/breast ultrasound images, achieving **95.99%** Dice on downstream task.
- Enabled edge deployment via **4-bit PTQ** with **7x robustness improvement** over ImageNet initialization.
- Built automated data pipeline with **PyTorch** and **OpenCV**: extracted, preprocessed, and segmented **486 clinical images** from ultrasound videos with augmentation and quality validation frameworks.

Teaching Assistant — Jan 2025 – Dec 2025
University of Texas at Dallas — *Richardson, TX*

- Mentored **150+** students in Advanced Algorithms and Data Structures through technical and code reviews.

Research Assistant — Apr. 2022 - May 2024
Alberta Machine Intelligence Institute (Amii) — *Edmonton, AB*

- **Online Learning**: Implemented algorithms for online optimization problems using the online **primal-dual** framework. Improved competitive ratios beyond traditional worst-case analysis by using ML predictions.
- **Reinforcement Learning**: Conducted a comprehensive survey on adversarial, Markovian and restless multi-armed bandits. Benchmarked UCB, Exp3, Gittins Index, etc. in **Python** on real-world data. [REPO]

Selected Projects

Decode EEG using Multi-Modal Approach | *PyTorch, MATLAB, HuggingFace*

- Identified bad electrodes, filtered, and transformed EEG data using **ICA** and *Automagic* in **MATLAB**.
- **LLMs**: Implemented RoBERTa for word embedding, resulting in an increase of **274%**, **78%**, and **1.4%** in F_1-score compared to Gaussian, GloVe and BERT embeddings on a 10-label classification task.
- Developed a novel **EEG extraction** framework by combining a **convolution** and a **self-attention** module. Achieved consistent increases in F_1-score across all 4 embedding types. [PAPER][REPO]

Modeling Political Sarcasm in Online Discourse | *NLTK, spaCy, scikit-learn*

- Developed **15+** novel engineered features: linguistic, sentiment dynamics, context, sarcasm-specific, boosting traditional model F_1-scores by **2%** for political sarcasm detection using **100k+** Reddit comments.
- Fine-tuned **DistilRoBERTa** achieved superior F_1-score of **78.14%** and accuracy of **78.33%**. [PAPER][REPO]

Technical Skills

Languages: Python, Java, Linux, C++, R, SQL, HTML/CSS, JavaScript, MATLAB.
Frameworks and Libraries: NumPy, Pandas, PyTorch, vLLM, TensorRT-LLM, TensorFlow, HuggingFace.
Developer Tools: Git, Linux, Docker, MongoDB, Unity, Android Studio, R Studio.