

Henry Vu

469-649-6967 | Vuducdung28@gmail.com | [linkedin.com/in/henry-vu27/](https://www.linkedin.com/in/henry-vu27/) | github.com/HenryVu27 | henryvu27.github.io

EDUCATION

University of Texas at Dallas

B.Sc. Computer Science - Intelligent Systems

- **GPA: 4.0/4.0.**

Richardson, TX

Sep. 2024 - Expected May 2026

University of Alberta

B.Sc. Computing Science with Honors

- **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.

Edmonton, AB

Sep. 2019 - May 2024

EXPERIENCE

Founding Machine Learning Engineer

eXRealityAI

Aug 2025

Dallas, TX

- Built voice-to-voice **RAG** system (Whisper STT → Qdrant hybrid retrieval + **LangGraph** agentic loop with Mistral-7B-q4 → Kokoro TTS) achieving sub-10s latency with local inference on Jetson Orin.
- Sole-developed and published GardenXR on the Meta Horizon Store, a Quest 3 mixed reality plant care assistant reaching **20,000+** users within 48 hours of launch.
- 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 multimodal workflows combining **YOLOv9** and **Wit.ai** across 2 XR applications.

Computer Vision Engineer

ThorMed Innovation

Feb 2025 – Present

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** to extract clinical images from ultrasound videos.

Research Assistant

Alberta Machine Intelligence Institute (Amii)

Apr. 2022 - May 2024

Edmonton, AB

- **Online Learning:** Implemented algorithms for online optimization problems using the online **primal-dual** framework. Improved competitive ratios beyond traditional worst-case analysis 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

ADHD Parenting Coach with Agents | *LangGraph, Gemini, Qdrant, FastAPI*

- Developed a parent-facing **ReAct** agent for evidence-based ADHD coaching.
- Designed a 4-tier **memory** architecture (rolling summaries, gated fact extraction, episodic memory, chat history) with **context** engineering (dynamic system prompt assembly, token-budgeted trimming, and profile injection).
- Built **LangGraph**-native input/output guardrails as graph nodes with structured **Gemini** classifiers for crisis detection, jailbreak prevention, and scope validation (medication/diagnosis boundaries). [REPO]

Clinical Gap Detection using Multi-Agent AI System | *LangGraph, Gemini, FAISS, FastAPI*

- Architected a **multi-agent** pipeline with 4 detection strategies (cross-reference, dropoff, symptom clustering, contradiction) coordinated via a supervisor pattern to identify clinical documentation gaps.
- Implemented hybrid retrieval combining **FAISS** vector with SQLite FTS5 keyword search, using **PubMedBERT** domain-specific embeddings, achieving **97%** detection coverage.
- Built two-phase orchestrator with intent classification and confidence-based validation with hallucination detection, reducing false positives through iterative refinement. [REPO]

TECHNICAL SKILLS

Languages: Python, Java, C++/C#, R, SQL, HTML/CSS, JavaScript, MATLAB.

Frameworks and Libraries: NumPy, Pandas, PyTorch, Qdrant, TensorRT-LLM, TensorFlow, HuggingFace.

Developer Tools: Git, GCP, Linux, Docker, MongoDB, Unity, Android Studio, R Studio.