

An Cao

📞 +1-4162549939 📩 an.cao.link@gmail.com 💬 ancaoai 💡 CA2528357431 🌐 An Cao 🌐 MyWebsite

EDUCATION

University of Toronto

Master of Science in Applied Computing (AI concentration) - GPA: 3.93/4.0

Sep 2024 – Dec 2025

Toronto, ON, Canada

- Recipient of Vector Scholarship in Artificial Intelligence for top 100 AI master's students in Ontario

Huazhong University of Science and Technology

Bachelor of Engineering in Software Engineering (AI track) - GPA: 3.96/4.0

Sep 2020 – Jun 2024

Wuhan, Hubei, China

- Excellent Undergraduate Graduates, Merit Students, Outstanding Undergraduates in Academic Performance

TECHNICAL SKILLS

- **Language & Frameworks:** Python, Pytorch, HuggingFace, DeepSpeed, Langchain, LangGraph, OpenAI, Vertex AI
- **ML Techniques:** Multi-Modal AI, Vision-Language Model, Audio-Language Model, Image Segmentation, Parallel Training, Model Finetuning, Model Evaluation, Model Deployment, Multi-Agent System, AI Agent, RAG, Embedding Model, Vector Database, Pydantic, Explainable AI, Data Augmentation, Computer Vision, NLP
- **Models:** LLaMa, LLaVa, Qwen, Transformer, BERT, LoRA, SAM, Unet, CLIP, Diffusion, ViT, Audio Flamingo
- **Tools & Infra:** GCP, Azure, AWS, SQL, MongoDB, PostgreSQL, Docker, FastAPI, Flask, Django, MLflow, DVC

EXPERIENCE

BMO

Toronto, ON, Canada

Machine Learning Engineer

Feb 2026 – Present

Focus: *FinTech AI Solution Research and Implementation*

- Translated advanced models into production-ready solutions that deliver tangible value in enterprise environments.
- To be continued

Modiface (L'Oréal AI Lab)

Toronto, ON, Canada

Machine Learning Intern

May 2025 – Dec 2025

Project: *Digital Dermatologist: Foundational Explainable Vision-Language Model for Skin Health*

- Employed Vertex AI for data augmentation, expanding dataset by 15 times and imputing absent text modality.
- Applied semi-supervised learning to synthesize missing segmentation masks in partially annotated skin datasets.
- Fused SAM and LLaMa for multimodal outputs of skin issue segmentation masks and visually-grounded text insights.
- Boosted diagnostic accuracy and coverage by 31% over SOTA via parallel quantized LoRA finetuning on 4 A100s.
- Achieved a 23% increase in IoU for skin concern segmentation, outperforming existing specialized models.
- Utilized contrastive learning to finetune an embedding model, enabling chat-based product recommendations.

Vector Institute

Toronto, ON, Canada

AI Technical Specialist

June 2025 – Jan 2026

Focus: *System Architecture and Technical Support for Multimodal AI and Agentic AI Systems*

- Integrated Qwen multi-modal LLM into audio-text RAG system for real-time voice-based grounded health insights.
- Led a RAG system optimized for live tabular stock data, matching human performance with 85% retrieval recall.
- Designed a multi-agent system for Anti-Money Laundering, attaining 76% info inconsistency detection accuracy.

Machine Learning Associate

Sep 2024 – May 2025

Project: *DiligenceGPT: AI for Due Diligence*

- Applied LLMs to extract finance data from unstructured documents and real-time online info with 89% coverage.
- Built a RAG agent on vector database of minute-level live data, delivering business analysis with traceable datapoints.
- Designed a LLM+ML quantitative company evaluator with 97% consistency, exceeding human expert performance.

Project: *Audience Builder: Conversational Database Agent for Synthetic Society*

- Created a conversational RAG agent that suggests relevant database values, accelerating complex query formulation.
- Solved retrieval bottlenecks by query LLM decomposition and embedding model finetuning, improving recall by 49%.

PUBLICATIONS

Diff-STAR (First Author, Published by IMAVIS, SCI Q1)

A. Cao and G. Shen, "Diff-STAR: Exploring student-teacher adaptive reconstruction through diffusion-based generation for image harmonization," *Image Vis. Comput.*, vol. 151, p. 105254, Nov. 2024, doi: 10.1016/j.imavis.2024.105254.

LisaCLIP (First Author, Accepted by IJCNN as ORAL)

A. Cao, Y. Zhou, and G. Shen, "LisaCLIP: Locally Incremental Semantics Adaptation towards Zero-shot Text-driven Image Synthesis," in 2023 International Joint Conference on Neural Networks (IJCNN), Jun. 2023, pp. 1–10. doi: 10.1109/IJCNN54540.2023.10191516.