

CHAO CAO

caochao0306@gmail.com | [linkedin.com/in/chao-cao-167037246](https://www.linkedin.com/in/chao-cao-167037246)

SUMMARY

Ph.D. candidate in Computer Science (Dec. 2026) with 3+ years building machine learning models end to end, from design and training to evaluation and deployment. Specialized in **reinforcement learning** and **Transformer-based foundation models**, LLM systems, and large-scale data pipelines using Python, SQL, and AWS; 990+ citations (NeurIPS, AAAI).

EDUCATION

University of Texas at Arlington, <i>Ph.D. in Computer Science and Engineering</i>	Dec. 2026 (Expected)
University of Maryland at College Park, <i>Master in Data Science and Analytics</i>	Dec. 2022
Nanjing University of Posts and Telecommunications, <i>Bachelor in Computer Science and Technology</i>	Jul. 2021

WORK EXPERIENCE

Full-Stack Software Development Engineer Intern GyriQAI, Inc.	May. 2025 – Aug. 2026
<i>Python, SQL, AWS Lambda, Docker, asyncio, Playwright, Flask, Gemini/GPT-4o</i>	<i>Athens, GA (remote)</i>

- Engineered **20+ fault-tolerant data pipelines** (asyncio, Playwright, checkpoint/resume) to retrieve and combine data from varied sources, producing a 40K+-school dataset, 366K contacts, and **1M+ structured Q&A samples**.
- Built an LLM-powered automation suite across 5 platforms with multi-agent orchestration over GPT-4o + Gemini (**160+ pytest tests**); red-teamed the production chatbot via a 232-cell matrix, surfacing 3 vulnerabilities.
- Shipped data products full stack on **AWS Lambda** and Docker, including a Flask + SQLite dashboard monitoring **65K posts**, a serverless Leaflet map, and a bulk-email platform with RFC 8058 unsubscribe.

Graduate Research Assistant University of Texas at Arlington, advised by Dr. Dajiang Zhu	Jan. 2023 – Present
<i>Deep Learning, Computer Vision, Medical Imaging, Brain Network Modeling</i>	<i>Arlington, TX</i>

- Designed and validated **deep learning models** to represent and match brain folding patterns and networks, yielding two first-author **oral** papers at IEEE ISBI 2026 and a first-author paper at ISBI 2024.
- Co-developed **Transformer-based** and LLM-based multimodal models for medical image representation learning, published at **NeurIPS 2026**, AAAI 2026, and MICCAI 2024–2025.

Research Assistant University of North Carolina at Chapel Hill, BRAIN Lab	May. 2024 – Aug. 2024
<i>PyTorch, Deep Learning, Computer Vision, Cortical Surface Modeling</i>	<i>Chapel Hill, NC</i>

- Led **Deep-GyratNet**, the first Spherical U-Net PyTorch framework for GyratNet extraction; designed combined loss functions that reached **96.40 F1** and cut per-subject runtime from 237s to 14s (**17× faster**).

RESEARCH & PROJECT EXPERIENCE

Large Vision-Language Model Post-Training for Multimodal Math Reasoning	Jan. 2026 – May. 2026
<i>Qwen2.5-VL-32B, PyTorch, Megatron-LM, verl, vLLM, CPT/SFT/RL</i>	<i>Arlington, TX</i>

- Designed a scalable **CPT → SFT → RL** pipeline for a 32B vision-language foundation model with Megatron-LM, training on **20M image-text pairs** and 100B text tokens, then full SFT on 770K reasoning samples.
- Built a **GSPO reinforcement learning** stage with verl and vLLM on 80K curated samples with rule-based rewards (98.5% verification accuracy), lifting image-based math accuracy by up to **+20%** and text-only by **+12%**.

Multi-Agent Clinical Decision Support System	Feb. 2026 – May. 2026
<i>LangGraph, FastAPI, GraphRAG, PostgreSQL, Docker</i>	<i>Arlington, TX</i>

- Built a LangGraph 5-agent clinical decision support system with GraphRAG retrieval, PHI de-identification, ICD-10 coding, and drug interaction checks, deployed as Dockerized **FastAPI** services at **94% diagnostic accuracy**.

Radiology NLI Benchmarking with Large Language Models	Nov. 2023 – Jun. 2024
<i>LLM Evaluation, Medical NLP, Benchmarking</i>	<i>Arlington, TX</i>

- Built **RadQNLI**, a **9,838-pair** radiology NLI benchmark, and evaluated GPT-4/ChatGPT zero-shot, few-shot, and CoT prompting against open-source LLMs and fine-tuned BERT baselines.

PUBLICATIONS

Research Impact: 990+ citations, h-index 11, i10-index 12 | Google Scholar

Selected: First-author IEEE ISBI 2026 (**2 Oral**), ISBI 2024, and IEEE Transactions on Big Data 2025 (co-first); co-author at NeurIPS 2026, AAAI 2026, and MICCAI 2024–2025.

TECHNICAL SKILLS

ML & Modeling: PyTorch, Reinforcement Learning, Transformers, LLMs, Megatron-LM, verl, vLLM, LangGraph

Languages & Data: Python, SQL (PostgreSQL, SQLite), JavaScript, asyncio, Playwright

Cloud & Tools: AWS Lambda, Docker, FastAPI, Flask, pytest, Linux, Git