

WILL AN

+1(424) 475-1233 ◇ Atlanta, GA
anxiaodong0329@gmail.com ◇ [linkedin](#) ◇ [GitHub](#)

RESEARCH INTERESTS

LLM post-training (SFT, RLVR), evaluation and LLM-as-a-judge, agentic systems and tool use, AI for science.

EDUCATION

Ph.D. in Physics, Georgia Institute of Technology	Expected Dec 2027
M.S. in Computational Science and Engineering, Georgia Institute of Technology	Expected Dec 2027
B.S. in Physics, The Hong Kong Polytechnic University	May 2022

PUBLICATIONS

Xiaodong An, Guangze Luo, Flavio H. Fenton (2026). [WEB-PDE-LLM: A Multi-Agent Framework for Automated Real-Time Web-Based GPU PDE Solving](#). *Under review, NeurIPS 2026 Workshop on Sim2Science*

Xiaodong An^{*}, Afrouz Delshad^{*}, Elizabeth M. Cherry, Flavio H. Fenton (2026). [Do LLM Judges Evaluate What We Ask Them to Measure? A Controlled Benchmark of Complexity and Chaos in Scientific Time Series](#). *Under review, NeurIPS 2026 JUDGE Workshop*. ^{*}Equal contribution.

Yansong Wen, Chengyun Fan, Xiaodong An, et al. (2026). [An Agentic Framework for Auditable Public Health Policy Communication](#). *Under review, Nature Health*.

Xingyu Shen, Xiaodong An, et al. (2026) [Can a Teenager Fool an AI? Evaluating Low-Cost Cosmetic Attacks on Age Estimation Systems](#) *arXiv preprint*, arXiv:2602.19539

Xiaodong An, et al. (2025). [Quantifying The Complex Spatiotemporal Chaos of Cardiac Fibrillation in Ionic Models Across Parameter Regimes](#). *arXiv preprint*, arXiv:2508.14303

EXPERIENCE

Machine Learning Engineer	May 2025 – Aug 2025
Scam AI	Berkeley, CA

- Led a team of 5 on the attack side to automate **adversarial data generation** with **diffusion** and **GAN**-based face-swap models, producing **10,000+** attack samples for detector training and evaluation.
- Tuned hyperparameters and **expert-routing** configurations for the defense team’s **mixture-of-experts (MoE)** detector in **PyTorch**, raising accuracy on the generated attack sets from **83%** to **90%+**.
- Contributed to the design of a **real-time face-swap detection** pipeline for live video streams, covering **frame-level inference**, **temporal consistency** checks, and **model quantization** for on-device latency.

PROJECTS

RLVR Post-Training for PDE Solver Code Synthesis LoRA SFT on Qwen2.5-Coder-7B over 8K rejection-sampled, execution-verified NumPy/PyTorch solvers, then **GRPO** with a **verifiable reward** (L_2 error). **FSDP** + **TRL** rollouts on A100.

SciAgent Bench: A Benchmark for Agentic AI in Scientific Discovery (On-going, Scale AI × Georgia Tech) [Link](#). Initiated a joint **benchmark and harness** for **agentic** research workflows: procedurally generated tasks with **programmatic verifiers**, expert rubrics on real lab data, and open-ended tasks under fixed **tool-call budgets**, scored by containerized **rollouts**, **process and outcome verifiers**, and **LLM-as-judge**.

SKILLS

LLM	HuggingFace Transformers, TRL, PEFT/LoRA, GRPO/DPO, vLLM
Systems	HPC, WebGL, Docker, FastAPI, Git
Languages	Python, MATLAB, JavaScript, HTML/CSS