Ryan (Yunxiang) Yan

ryan.yan.it@gmail.com | Google Scholar | raccoononion.github.io

RESEARCH INTEREST

- **Computational Creativity**: How to build autonomous AI agents for creative tasks such as novel generation and game development
- **Evaluation & Interpretability of AI**: How to evaluate and understand AI models through scalable, statistical approaches

EDUCATION

Georgia Institute of Technology 08/2023 – 12/2025
Master's, Computer Science GPA: 3.8/4.0

- Courses: NLP (4.0), Deep Learning for Text (4.0), Machine Learning (4.0), Knowledge-Based AI (4.0), Big Data Systems & Analytics (4.0), Computer Networks (4.0), Graduate Algorithms (4.0), *Modeling & Simulation, *Conversational AI

University of Notre Dame 08/2021 – 05/2022
Exchange Student GPA: 3.9/4.0

- Courses: Discrete Mathematics (4.0), Mathematical Statistics (4.0), Linear Regression (4.0), Algorithms (4.0), Investment Theory (4.0), ^Machine Learning (4.0), ^Neural Networks (4.0), ^Data Science (4.0)

Southern University of Science and Technology 09/2019 – 06/2023
Bachelor's, Financial Engineering Major GPA: 3.8/4.0

- Courses: Advanced Linear Algebra (90), Probability and Statistics (90), Financial Mathematics (90), Computer Architecture (96), Data Structures (90), Object-Oriented Programming (93), Databases (89), Big Data Science (91), Macroeconomics (99)

**In Progress; ^Cross-listed Graduate Course*

SELECTED PUBLICATIONS

[1] **Yunxiang Yan**, Tomohiro Sawada, and Kartik Goyal, "Cascaded Information Disclosure for Generalized Evaluation of Problem Solving Capabilities," *arXiv:2507.23776*. (**IJCNLP-AACL** 2025 Main)
[2] **Yunxiang Yan**, and Meng Jiang, "Constructing and Sampling Directed Graphs with Linearly Rescaled Degree Matrices," *arXiv:2507.23025*. (**KDD-UC** 2022)

RESEARCH EXPERIENCE

Contamination-Robust Math Word Problem Generation with Symbolic Sampling and RAG 08/2025 – Current
Research Lead, Advisor: Prof. Kartik Goyal, Larry P. Heck

- Proposed **MathRizz**, a dynamic evaluation framework that replaces static benchmarks (e.g., GSM8K) with symbolic expression sampling and Retrieval-Augmented Generation (RAG) to eliminate dataset contamination.
- Engineered an Actor-Critic pipeline to generate semantically coherent word problems and an LLM-as-a-Judge verifier (via GPT-5) that achieved 96% accuracy in detecting logical inconsistencies.
- Demonstrated that SOTA models suffer massive performance degradation on contamination-free data (e.g., GPT-4.1 accuracy dropped from 94.5% on GSM8K to 37.0%), proving that reported scores on static benchmarks are inflated.

Cascaded Information Disclosure for Generalized Evaluation of Problem Solving [1] 06/2025 – 08/2025
Research Intern, Advisor: Prof. Kartik Goyal

- Proposed a holistic evaluation framework ("Cascaded Information Disclosure") that assesses LLM problem-solving capabilities by revealing question information in stages, mitigating the overestimation bias inherent in standard static QA benchmarks.
- Demonstrated that standard benchmarks overestimate performance gaps by benchmarking diverse open and close source LLMs, showing that cascaded disclosure induces significantly higher-quality intermediate reasoning traces compared to direct questioning.

LLMFail: Probing LLM Failure Modes with Objectively Verifiable Synthetic Benchmarks 08/2024 – 12/2024
Research Intern, Advisor: Prof. Ling Liu

- Proposed a novel synthetic benchmarking framework to mitigate dataset contamination, utilizing procedurally generated tasks (e.g., string sorting, repeating decimal computation) to create an objectively verifiable and infinite problem space.
- Conducted a comprehensive analysis of 7 state-of-the-art LLMs (including Llama 3.1 and Gemma 2), revealing counter-intuitive failure modes in symbolic reasoning and instruction following despite high performance on static benchmarks.

Multivariate Time Series Forecasting with Transformers-based Graph Representation Learning 03/2023 – 07/2023
Undergrad Thesis & QR Intern Project at Ricequant, Advisor: Prof. Xinjie Wang

- Proposed a novel application of Time Series Attention Transformer (TSAT) to the Chinese stock market to predict alpha returns, capturing non-linear inter-stock correlations.
- Engineered a graph construction pipeline by implementing Super-Empirical Mode Decomposition (SEMD) to decompose daily returns into Intrinsic Mode Functions (IMFs) and constructing edge attribute with Pearson's correlation to model stock connectivity.
- Achieved an out-of-sample R^2 of **4.80%** using daily alpha returns alone, matching SOTA models requiring 94+ stock characteristics.

Constructing and Sampling Directed Graphs with Linearly Rescaled Degree Matrices [2] 09/2021 – 08/2022
REU (iSURE Program), Advisor: Prof. Meng Jiang

- Proposed a novel graph sampling framework for large-scale directed networks that linearly rescales the Joint Degree Matrix (JDM) and Degree Correlation Matrix (DCM) to accelerate complex network analysis.
- Developed a constructive sampling algorithm in SciPy, NumPy and NetworkX that generates representative subgraphs, providing a theoretical guarantee for the preservation of original in-degree and out-degree distributions.
- Validated the method on diverse benchmarks (including DBLP, CiteSeer, and synthetic scale-free networks), demonstrating that the sampled graphs retain key topological properties significantly better than random selection baselines.

WORK EXPERIENCE

Canaria Inc. New York, NY
Software Engineer Intern, AI/ML 06/2025 – 08/2025
- Designed distributed scraping pipelines on **AWS** across sources; multi-layer caching (**Redis** + **Mongo**) cut storage costs by **83%**
- Enhanced multi-model NLP serving system with a config-driven orchestrator (**FastAPI**), increasing GPU usage by **43%**
- Optimized semantic search quality on company databases, improving retrieval accuracy by **40%** on an internal benchmark

Georgia Institute of Technology Atlanta, GA
Teaching Assistant 05/2024 – Present
- NLP (Fall 2025), Internet Sys & Services (Spring 2025, Summer 2024), Big Data Systems (Fall 2024), supporting **400+** students
- In charge of developing and grading coding assignments and paper critique, conducting office hours and grading semester projects

Ricequant Shenzhen, China
Quantitative Researcher Intern, Equities 03/2023 – 07/2023
- Applied **Transformer**-based time-series modeling for alpha factor mining on Asian equities market, achieving **1.6 Sharpe ratio**

Southern University of Science and Technology Shenzhen, China
Teaching Assistant 02/2023 – 06/2023
- CS209 Computer Architecture, supporting **100+** students
- In charge of building JUnit test cases for coding assignments, holding tutorial sessions and grading semester projects

PROJECTS

GTLLMZoo: Interactive Visualization of Aggregated LLM Leaderboards Atlanta, GA
Research Intern 04/2024 – 06/2024
- Engineered GTLLMZoo, a **full-stack** leaderboard explorer, with backend scrapers that continuously ingest data from LiveBench and LMSYS Chatbot Arena; a frontend app (**Gradio**) deployed on Hugging Face Spaces
- Implemented filterable tables, structured model cards, and interactive visualizations, enabling real-time benchmarking

Kaggle – LLM Agent System for 20 Questions Remote
Team Lead 05/2024 – 08/2024
- **Gold** Medal Winner (10 / 832 teams; **Top 1%**) – Building LLM agents that compete in cooperative games [Kaggle Solution]
- Led agentic pipeline with **Phi-3,** improving guessing accuracy to **~90%** on 20 Questions via bisection + Chain-of-Thought
- Created a multi-tiered system of 2M common words using Wikipedia data and **GPT-4o mini** as scalable N-gram filters

Kaggle – Chatbot Arena Human Preferences Prediction Remote
Core Member 05/2024 – 08/2024
- **Silver** Medal Winner (21 / 1849 teams; **Top 1%**) – Predicting human preferences for LLM responses [Kaggle Solution]
- Constructed **LoRA** fine-tuning pipelines for **Gemma 2** (9B) and feature engineering improving accuracy by **23%**

Kaggle – AI Mathematical Olympiad - Progress Prize 1 Remote
Solo Competitor 04/2024 – 06/2024
- **Silver** Medal Winner (38 / 1161 teams; **Top 3%**) – Building AI models that can solve IMO level math problems [Kaggle Solution]
- Engineered a program-aided reasoning system by integrating **DeepSeekMath** with a custom Python REPL for iterative code execution, utilizing chain-of-thought prompting and self-consistency (majority voting) to minimize arithmetic hallucinations.

HONORS & AWARDS

Kaggle Competition Master (Top 0.3%) 08/2024 – Present
Graduate Teaching Assistantship (with full tuition waiver, total $88,000), Georgia Tech 08/2024 – 12/2025
International Exchange Program Scholarship (with full tuition waiver, total: $75,000), SUSTech 08/2021 – 05/2022
Award for Best Visualization, ASA DataFest 2022 Competition, University of Notre Dame 08/2022
Student Travel Award, SIGKDD 2022 08/2022
Award for Outstanding Volunteers, SUSTech 02/2021

SKILLS

Programming: Python, Java, C/C++, SQL, AWS, MongoDB, REST, Docker, Redis, Git, Bash, Linux
ML: PyTorch, vLLM, Distributed Training, PEFT, Knowledge Distillation, Supervised Finetuning, Reinforcement Learning
Language: English (Fluent, TOEFL 112, GRE 329); Mandarin (Native)