Ruiyu Wang

rwan388@emory.edu | Atlanta, GA, USA | 470-905-8972

Research Interests

LLM Agent: Developing language-model agents for planning, tool use, and feedback-driven learning, with a focus on efficient, robust, and well-evaluated behavior.
AI for Health and Science: Leveraging foundation models and structured reasoning for scientific discovery and applied tasks such as clinical prediction and biomedical literature mining.
Multimodal & RL Methods: Using multimodal modeling and reinforcement learning to enhance agent capabilities and scientific workflows (e.g., perception, planning, and decision support).

Education

Emory University Atlanta, GA
(*Current*) Cumulative GPA: 3.918/4.000 08/2022 - 05/2025 (*Expected*)
Bachelor of Science in Computer Science
Bachelor of Science in Applied Mathematics and Statistics

Research Experience

Emory Graph Mining Lab, Department of Computer Science, Emory University Atlanta, GA
Undergraduate Research Assistant
PI: Dr. Carl J. Yang

Enhanced Atrial Fibrillation Prediction with Pre-training and Transfer Learning 08/2024 – 02/2025

- Developed hypergraph-transformer pretraining pipelines on a large stroke cohort (n=7,780) to learn transferable patient visit embeddings.
- Transferred compact 32-D embeddings to an embolic stroke of an undetermined source(ESUS) cohort (n=510), integrated with baseline clinical features, enabling improved atrial fibrillation (AF) risk prediction with lightweight classifiers, with 5–15% AUROC gains over from-scratch baselines
- Co-authored the project (3rd author); Manuscript ready for submission.

MedAssist: Knowledge Graph and LLM-based Retrieval agent 01/2025 – 04/2025

- Built a retrieval-augmented pipeline that unifies external medical knowledge with internal Electronic Health Record (EHR) tables via an LLM agent, document retriever, and SQL/tool executors
- Converted medical literature into structured knowledge graphs; designed evaluation metrics (pair variance, direction variance, type variance) to trade off coverage vs. stability across runs.
- Co-authored the project (3rd author); accepted by the WWW 2025 Demo Track.

ACERAG: Retrieval-Augmented Generation with Large Model Prompting 03/2025 – 05/2025

- ACERAG proposes a self-play RAG framework where a single LLM alternates as Decomposer (multi-hop query breakdown) and Solver (evidence integration) to improve reasoning over retrieved context.
- Ran experiments across multiple QA benchmarks, implementing and prompting baseline LLMs for fair comparison and standardized evaluation.
- Co-authored the project (4rd author); accepted by the NeurIPS 2025.

BioMedJImpact: LLM-Based Impact Analysis for Biomedical Journals 2024 – 2025

- Constructed **BioMedJImpact**, a large-scale dataset of 1.74M PMC articles from 2,744 biomedical journals with citation, collaboration, and LLM-derived AI engagement features for journal impact modeling.
- Designed a three-stage Gemma-12B LLM pipeline over titles/abstracts to detect AI-related work and assign AI subfields, validated via human evaluation.
- Analyzed how collaboration intensity and AI engagement relate to citation impact and quartile rankings.
- Led the project end-to-end and authored the manuscript, submitted to PAKDD 2026 (under review).

Knowledge Graph–Aided Clinical Prediction via SFT LLM 05/2025 – Present

- Developing a KG-guided clinical prediction framework using MIMIC-III, integrating structured reasoning chains into LLaMA-8B fine-tuning for interpretable disease detection.

– Achieved AUC/AUPR performance comparable to traditional ML baselines on a 1,000-patient cohort, while providing transparent, clinically aligned reasoning.
– Demonstrated strong data efficiency (maintaining performance with ˜400 patients) and improved cross-dataset generalization.
– Led the project end-to-end and authored the study, preparing for an AMIA 2026 submission.

Department of Mathematics, Emory University Atlanta, GA
Undergraduate Research Assistant
Supervisor: Dr. Yuanzhe Xi

Reinforcement Learning for Multigrid Tuning (HYPRE) 09/2025 – Present

– Formulated multigrid tuning in HYPRE as an RL problem across mixed-type BoomerAMG hyperparameters (strength thresholds, smoothing parameters, coarsening and relaxation types).
– Implemented a policy-based tuner that adapts multi-parameter configurations across solves, removing the need for Bayesian GP-based pre-tuning.
– Benchmarking contextual bandits vs. full episodic RL against GPTune / GPTuneBand to evaluate sample efficiency and stability.

Publications & Presentations

Xu, R., Shi, W., **Wang, R.**, Zhou, J., and Yang, C. (2025).
"MedAssist: LLM-Empowered Medical Assistant for Assisting the Scrutinization and Comprehension of Electronic Health Records."
WWW '25: Companion Proceedings of the ACM on Web Conference, pp. 2931–2934.

Xu, R., Zhuang, Y., Dong, Z., **Wang, R.**, Yu, Y., Ho, J. C., Zhang, L., Wang, H., Shi, W., and Yang, C. (2025).
"AceRAG: Advancing Reasoning-Intensive Retrieval-Augmented Generation via LLM Self-Play."
NeurIPS 2025 (spotlight).

Wang, R., Xie, Y., Hu, X., Yang, J.C., and Lu, J. (2026).
"BioMedJImpact: A Comprehensive Dataset and LLM Pipeline for AI Engagement and Scientific Impact Analysis of Biomedical Journals."
Proceedings of the 30th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2026). (Submitted)

Industry Experience

JINGDONG (JD), China May 2024 – June 2024
AI Intern
– Fine-tuned LoRA models to generate high-fidelity images of JD's mascot, improving brand-aligned output quality.
– Documented AI workflows for text-to-image, video generation, and object recognition for a non-AI team.

Tongcheng Holdings Limited, China June 2024 – August 2024
AI Engineer Intern
– Refined the speaker diarization system, reduced the EER from 0.87% to 0.37%.
– Fine-tuned the model backbone and QMF, boosting internal recall from 79% to 96%.
– Contributed to DeepTrip, an LLM-based travel assistant, designing an evaluation framework using structured LLM prompts to assess factual accuracy, constraint adherence, and response clarity.

Leadership & Service Experience

Emory University Mathematics Association (EUMMA) Fall 2025 – Present
Vice President Atlanta, GA
– Co-organized academic programming including Honor Research Panel, Graduate School Panel, and Major/Minor Panel; coordinated faculty and student panelists and managed event logistics.
– Provided competition advising for **MCM**, **DataFest**, and **Kaggle** teams, including team formation, modeling strategy, evaluation methods, and report writing.
– Led and organized preparation workshops and mock judging sessions to strengthen competitive performance.

Honors & Awards

Honorable Mention, The Mathematical Contest in Modeling Spring 2025
Dean's List, Emory College of Arts and Sciences Fall 2023 – Fall 2024
Best Team and Insight Award, DataFest 2025, Emory College of Arts and Sciences Spring 2025

Professional Competencies

- Programming Languages: Python, Java, C, ARM Assembly, HTML/CSS
- Core Competencies: Machine Learning, Deep Learning, Data Analysis, Full-Stack Web Development, LLM fine-tuning, LLM agents, Retrieval-Augmented Generation
- Models & Algorithms: Transformer, GNN, CNN, LSTM, SVM, KNN, Random-Forest
- Databases & SQL: MySQL, PostgreSQL, MongoDB, Redis
- Developer Tools & Frameworks: Git, Docker, PyTorch, Visual Studio, PyCharm

Relevant Coursework

- Computer Science: Data Structures and Algorithms, Computer Architecture, Systems Programming, Database Systems, Analysis of Algorithms
- Machine Learning: Machine Learning, Deep Learning, Natural Language Processing, Data Mining, Probabilistic Machine Learning
- Mathematics & Statistics: Linear Algebra, Mathematical Statistics, Numerical Analysis, Mathematics of Data Science