ZINI YANG

Tel: 919-697-3124 | Email: zini.yang@duke.edu | Homepage: https://ziniyang.github.io/zini_yang/

RESEARCH INTEREST

My research interests are: (1) **Social Science for Trustworthy AI:** I aim to narrow the gap between large language models (LLMs) and human intelligence through computational social science (e.g., game theory, behavioral economics, and network science) to build more socially aware LLMs that interact reliably with humans; (2) **AI for Scienti & Social Good:** Analyzing how to better apply AI for scientific research and other high-impact domains(e.g., health), with a focus on the reliability and efficiency of information retrieval and data mining.

EDUCATION

Duke University Durham, United States
M.S. in Computer Science and Economics Aug 2024 – May 2026 (Expected)
Advisor: Dr. Emily Wenger
Major GPA: 3.96/4.00
Honors & Awards: Master's Scholarship Award, Dean's Research Award ($1000)

Shanghai University of Finance and Economics Shanghai, China
B.S. in Mathematical Economics Sep 2019 – Jul 2023
Advisor: Dr. Simin He
Major GPA: 3.73/4.00, Rank: 2/40
Honors & Awards: Graduate with Distinction, Academic Performance Scholarship (2019-2023)

RESEARCH EXPERIENCE

Research Area 1: Social Science for Trustworthy AI

LLMs Creativity Evaluation (Preparing for submission to ACL 2026) Durham, NC
Advisor: Dr. Emily Wenger, Dr. Richard So *Oct 2025 – Present*

- Developed a framework to quantify human creativity by modeling response boundaries as convex hulls in semantic embedding space and validated its robustness across tasks using bootstrapping.
- Proposed a human-grounded method to evaluate LLM creativity by modeling semantic boundaries of human responses and proposed the coverage rate metric to measure the proportion of human creative space reproducible by LLMs.
- Evaluated LLMs of varying sizes on three creativity task types (divergent thinking, convergent thinking, creative writing) using the coverage rate metric, finding consistent underperformance with coverage below 60% across all models.
- Identified task openness as a critical factor limiting LLM creativity through coverage rate analysis, with performance declining from $\sim$60% on structured convergent tasks to only $\sim$10% on open-ended creative writing.

Data Protection Mechnism Design Durham, NC
Advisor: Dr. Emily Wenger, Dr. Curtis R. Taylor *Feb 2025 – Present*

- Identified a critical asymmetry in data protection scenarios where defenders (e.g., artists, private information holders) possess less advanced technology than professional scrapers, and investigated how to exploit information advantages to deter attacks.
- Formulated the protection problem as a dynamic game where protectors choose defense levels periodically and attackers optimize attack timing.
- Developed a tractable Markov model with breach history as the state variable to analyze evolving deterrence dynamics.
- Derived optimal protection strategies that maximize long-term defense efficiency and provide design principles for protection systems.

Research Area 2: AI for Science & Social Goods

AI for Mathematical Conjecture Discovery and Counterexample Finding (ICML 2025 Workshop) Durham, NC
Advisors: Dr. Ethan Fang; Dr. Junwei Lu *Oct 2024 – Present*

- Conjecture discovery accepted at *ICML 2025 Workshop*: **Inequality Ranking and Inference System (IRIS)** — *Giving Mathematical Conjectures Numerical Value*.
- Designed four ranking metrics (sharpness, diversity, difficulty, and novelty) to evaluate conjecture quality, and re-engineered the `graffitiai` codebase into a modular pipeline enabling scalable conjecture scoring.
- Built an automated counterexample discovery system using GNNs for graph embedding and PPO-based RL agents for action sequence optimization, successfully refuting 95% of tested conjectures generated by IRIS.

- Built a self-play system with two agents (conjecture generator and counterexample searcher) that iteratively propose and test conjectures, demonstrating stable convergence and progressive improvement in conjecture quality through automated feedback loops.

LLM Based Vaccine Warning System Durham, NC

Duke NLP Group, Advisor: Dr. Bhuwan Dhingra *Jul 2025 – Present*

- Initiated an AI-driven framework for early warning detection of vaccine-related concerns by mining weak signals from large-scale open-source data.
- Built a scalable two-stage pipeline to extract vaccine-relevant content from Common Crawl (relevant content $<0.1\%$): fastText-based coarse filtering followed by Qwen model fine classification, reducing computational costs by 99% while maintaining high precision.
- Built a Qwen-based local API for vaccine-concern category (e.g., side effect, regulation concern) labeling; analyzed discrepancies between model and human annotations, crafted improved prompts, and boosted F1 from 43 % to 72.3 %.
- Developed an automated trend detection pipeline using UMAP for dimensionality reduction and HDBSCAN for clustering similar vaccine-related content, then generated interpretable trend reports with timelines through LLM-based summarization.

Research Area 3: Behavioral and Network Economics

Expectation Formation with Correlated Variables Shanghai, China

Research Intern, Advisor: Dr. Simin He *Jul 2022 – Oct 2023*

- Co-led the implementation of an oTree behavioral experiment following with two groups Baseline (A only) vs. Correlated (A+B).
- Conducted group-level significance analysis with OLS regressions and Mann–Whitney tests to compare experimental treatments.
- Built a fully reproducible pipeline (Bash/Makefile + R/renv) for end-to-end cleaning $\rightarrow$ estimation $\rightarrow$ figure/report, enabled one-click regeneration of results for review.

Public goods provision in a network formation game Shanghai, China

Research Intern, Advisor: Dr. Simin He *May 2022 – Nov 2022*

- Cleaned and merged oTree logs across groups, rebuilt bilateral link matrices, and computed network stats (density, degree, centrality).
- Ran group-level summaries and mixed-effects to compare *splitting* vs. *alternation* and link-cost treatments; delivered plots/tables reproducing the main findings.

INDUSTRY EXPERIENCE

Xiaohongshu (rednote) Shanghai, China

Data Analyst Intern *May 2023 – Aug 2023*

- Engineered a consumer-behavior model in SQL/Python to uncover consumption trends; produced analyses that supported 20+ online/offline acquisition campaigns and drove about 0.8M+ people participation.
- Collaborated with IT team to refine the recommendation algorithm by introducing new metrics and designing A/B tests to validate.
- Contributed to all stages of the "Find a Partner" sub-platform lifecycle, including user research, design, and integration; resulted in 0.4M+ user engagements (clicks and conversations) and a significant uplift in platform activity.

TEACHING EXPERIENCE

COMPSCI 653: Human-Centered Computing Duke University

Teaching Assistant *Fall 2025*

COMPSCI 230: Discrete Mathematics for Computer Science Duke University

Teaching Assistant *Spring 2025*

SKILLS

Statistical & Analytical: GLM (logistic/Poisson/NegBin), hypothesis testing, causal inference, A/B testing, time-series analysis, dimensionality reduction & clustering (PCA, k-means/HDBSCAN)
Programming: Python (NumPy, pandas, SciPy, Matplotlib, Seaborn, scikit-learn, PyTorch), SQL, R, SAS, Java
Machine Learning: Transformers & fine-tuning (LoRA/QLoRA), LSTM/RNN, XGBoost/Random Forest/Decision Tree, RAG, interpretable ML (SHAP/LIME)
Data Eng & Tools: ETL pipelines, feature engineering, Apache Spark, Hadoop, Databricks, Snowflake, BigQuery/Redshift, dbt, Airflow, MLflow, Docker/Kubernetes
Cloud/BI & DevOps: AWS/GCP, CI/CD (Git/GitHub), automated testing, Tableau/Power BI