

# JINMING NIAN

jnian@scu.edu ◊ (530)854-1122 ◊ Santa Clara, CA ◊ [Google Scholar](#) ◊ [Website](#)

## SUMMARY

---

Experienced research assistant and PhD candidate specializing in ML and AI, with a strong background in retrieval, NLP, retrieval augmented generation, and AI reasoning. Skilled in scientific programming with Python/PyTorch, with additional experience in software development and distributed systems. Actively publishing research on retrieval and LLM-related topics.

## EDUCATION

---

### Ph.D. in Computer Science and Engineering

*September 2023 - Present*

Santa Clara University

*Advisor:* [Prof. Yi Fang](#)

### M.S. in Computer Science and Engineering

*September 2021 - June 2023*

Santa Clara University

*Relevant Courses:* Machine Learning, Deep Learning, Information Retrieval, Natural Language Processing, Reinforcement Learning

### B.S. in Physics, Minor in Music

*September 2017 - June 2021*

University of California, Davis

## WORK EXPERIENCE

---

### Research Assistant, Santa Clara University

*April 2023 - Present*

- Evaluated the reranking effectiveness of LLMs based on their likelihood of generating correct answers for a given query and diverse document contexts, producing a synthetic dataset with quality comparable to human labels for training dense retrievers
- Developed methods using LLMs to generate controlled hallucinations from news articles, creating highly relevant questions that ask for information absent in the document, mimicking RAG scenarios where the retrieved document is unfit for the question
- Analyzed the usefulness of LLM internal attention matrices as features for training a document reranking model
- Developed web-based tools to facilitate human annotation, results aggregation, and inter-annotator agreement calculations

## PAPERS & PRE-PRINTS

---

- **W-RAG: Weakly Supervised Dense Retrieval in RAG for Open-domain Question Answering** — [Link](#)  
[Jinming Nian](#), Zhiyuan Peng, Qifan Wang, Yi Fang
- **ScopeQA: A Framework for Generating Out-of-Scope Questions for RAG** — [Link](#)  
Zhiyuan Peng, [Jinming Nian](#), Alexandre Evfimievski, Yi Fang
- **Evaluating Social Biases in LLM Reasoning** — [Link](#)  
Xuyang Wu, [Jinming Nian](#), Zhiqiang Tao, Yi Fang

## PROJECTS

---

### R1-Zero for Information Retrieval

*Feb 2025 - Present*

- Gathered training data: filtered hard queries where both lexical and dense retrieval methods fail; collected irrelevant documents that are consistently retrieved by strong retrievers as hard negative documents
- Redesigned reward modeling based on GRPO for the re-ranking task to encourage self-discovered CoT
- Leveraged Huggingface's Open-R1 as the base framework to train, test, and evaluate on the challenging BRIGHT dataset

### Atari Agent (ELEN 552, Reinforcement Learning)

*November 2023 - December 2023*

- Trained an AI agent to play Pong using Reinforcement Learning methods including a Deep Q-learning Network, Agent memory for experience replay, fixed Q-target, and the epsilon-greedy strategy
- Achieved 89% win-rate against a programmed player with an average winning score of 21:15