Xinyu Guo Ph.D. Student

Homepage Google Scholar
+1 520-501-4268 xinyuguo@arizona.edu
Mountain View, CA github/xinnyuguo

EDUCATION

University of Arizona, Tucson, AZ — 2024 - Present
Ph.D. in Computer Science, Advisor: Prof. Mihai Surdeanu, GPA: 3.86/4.0
Research Interests: LLM Reasoning, Reinforcement Learning, Reasoning Mechanisms

Shandong University, Shandong, China — 2019 - 2023
B.E. in Computer Science, Graduated with Honors, GPA: 3.8/4.0
Core Modules: Artificial Intelligence (95), Discrete Mathematics (92)

RESEARCH OVERVIEW

Large language models increasingly reason through knowledge-intensive and open-ended tasks. My research focuses on **LLM-based reasoning** in these settings, along two directions:

- **RL Credit Assignment for Reasoning Supervision**: credit assignment is key to RL for reasoning, which needs fine-grained credit beyond RLVR's outcome-only rewards. I tackle it from two angles: (1) *efficient intermediate rewards*, self-derived from the model's own correct reasoning without annotation or an LLM judge (CogRE [1]); and (2) *from rewards to advantages*, disentangling the varying query-specific rubrics in critic learning to estimate per-rubric values and advantages (DVE [2]).
- **Reasoning Mechanisms**: looking inside LLMs to understand how concepts and their logical structure support reasoning, by aligning sparse autoencoder features with concepts (AlignSAE [5]) and taxonomies (TaxoSAE [4]).

PUBLICATIONS

RL Credit Assignment for Reasoning Supervision

[1] **Xinyu Guo**, Zhengliang Shi, Minglai Yang, Mahdi Rahimi, Mihai Surdeanu. The Answer Lies Within: Self-Derived Rewards Enable Explainable Relation Extraction.
EMNLP, Main, 2026

- Identified two failure modes of LLMs on realistic one-shot relation extraction: attending to irrelevant tokens, and predicting relations at a granularity mismatched with human annotation schemas
- Proposed Hit@Dict, a self-derived intermediate RL reward that credits relation-bearing phrases in the reasoning trace, using a dictionary mined from the model's own correct explanations; no LLM judge needed
- Improved F1 by **+73.7%** (relative) over accuracy-only RL on one-shot NYT29 and human-rated explanation quality by **+54%**, with faster and more stable training

[2] **Xinyu Guo**, Zhengliang Shi, Minglai Yang, Chen Xu, Lingyong Yan. Disentangled Value Estimation: Rethinking Scalar Critics in Rubric-Based RL.
Under Review, 2026

- Identified a bottleneck in the critic model for rubric-based RL: the critic sees only the query and response prefix, yet is trained on a scalar reward that collapses many query-specific rubrics, so it must implicitly recover the reward specification before it can estimate values

- Proposed Disentangled Value Estimation: a shared critic conditioned on one rubric at a time estimates per-rubric values and token-level advantages, aggregated with the rubric weights
- Outperformed PPO and GRPO in **13 of 15** model–benchmark settings across instruction following, chat, and science QA (up to **+8.0 points** over PPO), with lower value error and advantages concentrated on rubric-relevant tokens

[3] Minglai Yang, **Xinyu Guo**, Utkarsh Tyagi, Mian Zhang, Razvan Dumitru, Sunjie Hou, Yunzhong He, Daniel Yue Zhang, Ying Liu. Rubric Dropout: A Simple Way to Mitigate Reward Hacking in Rubric-as-Reward RL.
Under Review, 2026

- Showed that rubric-as-reward RL reward-hacks out of distribution: the training judge's score keeps rising while a stronger gold judge's score peaks and falls by **22 points** on ResearchQA
- Proposed Rubric Dropout, which randomly drops rubric criteria at every step with a mask shared across each GRPO rollout group, at no extra judge cost
- Raised the out-of-distribution gold score at every matched checkpoint (**+6–7 points** on ResearchQA, **+1–2 points** on HealthBench-Hard) and reduced both hacking measures, with no in-domain cost

Reasoning Mechanisms

[4] **Xinyu Guo**, Minglai Yang, Clayton T. Morrison, Eduardo Blanco, Mihai Surdeanu. TaxoSAE: Rethinking Sparse Autoencoders through Taxonomic Logic.
Under Review, 2026

- Showed that existing SAEs learn a flat feature space, so concept readout lacks taxonomic consistency and control lacks the granularity to scope an edit
- Proposed TaxoSAE, which mirrors a taxonomy with stacked SAEs, one per level, and a taxonomic alignment loss that activates a concept whenever one of its hyponyms is active; only terminal concepts need supervision
- Improved hypernym readout of unseen abstract concepts by **+31.3 points** and swap success by **+25.5 points**; suppressing a concept removes only its own answers (**0–2%** vs. **91–100%** on others)

[5] Minglai Yang, **Xinyu Guo**, Zhengliang Shi, Jinhe Bi, Steven Bethard, Mihai Surdeanu, Liangming Pan. AlignSAE: Concept-Aligned Sparse Autoencoders.
TMLR, 2026

- Showed that unsupervised SAEs fragment human-defined concepts across many entangled features, making feature-level interventions unreliable
- Proposed a "pre-train, then post-train" curriculum that binds each ontology concept to a dedicated latent slot while keeping a free feature bank for reconstruction
- Enabled reliable concept swaps (**94.8%** success on 1-hop QA) and **9–∞×** better disentanglement than standard SAEs; used the aligned slots to probe multi-hop reasoning and grokking-like dynamics

Other Publications

[6] Minglai Yang, **Xinyu Guo**, Renxiong Wang, Daniel Yue Zhang, Yunzhong He. From Copying to Composition: How Language Models Develop In-Context Learning.
Under Review, 2026

[7] Minglai Yang, Xinyan Velocity Yu, Pengyuan Li, **Xinyu Guo**, Zhenting Qi, et al. Dr. DocBench: A Comprehensive Benchmark for Expert-Level and Difficult Document Parsing.
NeurIPS Workshop (LCFM), 2026

[8] Yunze Xiao, Jienan Chen, **Xinyu Guo**, Minghao Liu, Zihan Wang, Weihao Xuan. RecuritOpsBench: When Agents Fail to Ask for the Information They Need.
Under Review, 2026

[9] Zhengliang Shi, . . . , **Xinyu Guo**, . . . , Zhaochun Ren. Deep Research: A Systematic Survey.
Under Review, 2025

[10] Zhaoze Sun, Qiyan Deng, Chengliang Chai, Kaisen Jin, **Xinyu Guo**, Han Han, Ye Yuan, Guoren Wang, Lei Cao. QUEST: Query Optimization in Unstructured Document Analysis.
VLDB, 2025

INTERNSHIPS

Abaka AI, Mountain View, CA — May - Oct. 2026
Research Scientist Intern

- **Agentic RL Training**: built an internal agentic RL training framework supporting MCP-based tool use and RL training; post-trained Qwen3-8B on **2K** synthesized tasks, raising Pass@1 on τ^2-bench Retail from **31.6%** to **60.2%** (RetailServe blog).
- **RL environments & task synthesis**: built a knowledge-grounded pipeline for scaling RL environment construction and task synthesis, where deep research and knowledge graphs supply the domain knowledge that guides both code agents in building environment software and the synthesis of training tasks.
- **Video QA pipeline**: built a large-scale audio-visual QA generation pipeline for multiple-choice and open-ended questions across **13** video domains, with multi-stage annotation, leakage and difficulty probing, and model-based verification.
- **Image QA pipeline**: constructed challenging image QA data requiring multi-step visual reasoning over real-world images, with hard distractors, adversarial validation, and no overlap with public test sets.
- **Document parsing**: curated data and built the multi-model evaluation harness for Dr. DocBench (NeurIPS 2026 LCFM Workshop).

RESEARCH EXPERIENCE

CLU Lab, University of Arizona, Tucson, AZ — Mar. 2025 - Present
Graduate Research Assistant
Advisors: Prof. Mihai Surdeanu, Prof. Clayton T. Morrison, Prof. Eduardo Blanco

- Research topics: NLP, LLM reasoning mechanisms.
- Publications: CogRE [1] (EMNLP), AlignSAE [5] (TMLR), TaxoSAE [4] (under review).

CAES Research Center, Shandong University, China Sep. 2023 - Sep. 2024
Undergraduate Research Assistant
Advisor: Prof. Zhongxing Yu

- **C++ type conversion bugs**: built a dataset of **226** real-world type conversion bugs from **10** major C++ projects (e.g., V8, PyTorch, OpenCV, MySQL), summarized root-cause and fix patterns, and designed a tool to automatically detect and repair them.
- **Code clone detection & repair** (undergraduate thesis): applied SemanticCodeBERT, which extends GraphCodeBERT with semantic flow graphs and five structure-aware pre-training tasks; outperformed GraphCodeBERT on BigCloneBench clone detection (F1) and Transformer baselines on Bug-Fixes code repair (BLEU).

Oxford Winter Exchange Program, Oxford University, UK Feb. 2022
Advisor: Prof. Jeremy Pitt

- Topic: Human-centered AI.

GRANTS & SERVICES

Co-PI, ABC SEED Grant, Arizona Astrobiology Center
Reviewer, ACL Rolling Review (ARR) 2026, NeurIPS Workshop (LCFM) 2026

AWARDS & HONORS

Galileo Circle Scholarship, University of Arizona 2026
Science Graduate College Fellowship, University of Arizona 2024, 2025
Outstanding Undergraduate Thesis Award, Shandong University 2023
Outstanding Graduate Student, Shandong University 2023
Academic Scholarship, Shandong University 2022
18th Certified Software Professional Examination, Top 4.57%, CCF 2021

SKILLS

Programming Languages: Python, C/C++, MATLAB, Go, Java, HTML
Libraries & Frameworks: PyTorch, NumPy, vLLM, VeRL, Hugging Face Transformers, SQLite, MySQL, Spring Boot, JavaScript, Qt

TEACHING

University of Arizona, Tucson, AZ 2024, 2025
Teaching Assistant
Courses: CSC337 Web Programming, CSC460 Database, CSC477/577 Computer Vision

Shandong University, Shandong, China 2023
Teaching Assistant
Course: Software Engineering