Nicholas Crispino

✉ ncrispino0@gmail.com   🌐 ncrispino.github.io/

Summary

Second-year PhD student in computer science focused on LLM agents and interpretability. Excited about scaling multi-agent approaches to improve downstream task performance and expanding the real-world use cases of agentic workflows, as well as exploring critical areas of LLM safety.

Education

University of California, Santa Cruz
Santa Cruz, CA

PhD in Computer Science   *September 2025–*
- **Advisor:** Chenguang Wang

Washington University in St. Louis
St. Louis, MO

PhD in Computer Science   *August 2024–August 2025*
- **Advisor:** Chenguang Wang, director of the WashU NLP Group
- **Awards:** Dean's Select Fellowship (2024), Cox Family Fellowship (2024)

Bachelor of Science   *September 2020–December 2023*
- **Majors:** Computer Science + Economics (primary), Statistics (double)
- **GPA:** 4.00/4.00
- **Awards:** Undergraduate Engineering Valedictorian, Ernest D. Weiss Junior Award for Academic Excellence – Computer Science and Engineering, Brian Blank Award in Mathematics, Antoinette Frances Dames Award for Productive Scholarship in Engineering

Preprints

V. Siu[*], **N. Crispino**[*], D. Park, N. W. Henry, Z. Wang, Y. Liu, D. Song, C. Wang. *SteeringControl: Holistic Evaluation of Alignment Steering in LLMs.* In arXiv preprint 2509.13450.

V. Siu, N. W. Henry, **N. Crispino**, Y. Liu, D. Song, C. Wang. *RepIt: Representing Isolated Targets to Steer Language Models.* In arXiv preprint 2509.13281.

L Phan, A Gatti, Z Han, N Li, W Zhang, **N Crispino**, C Wang, D Li, J Shen, K Montgomery, H Szlyk, T Wang, S Yoe, A Wang, D Hendrycks, many others. Humanity's Last Exam. In arXiv preprint 2501.14249.

Under Review

S Kolasani, M Saplin, **N Crispino**, K Montgomery, J Quincy Davis, M Zaharia, C Wang, C Wang. (2025). LLM CHESS: Benchmarking Reasoning and Instruction-Following in LLMs through Chess.

Publications

Y Potter*, Z Wang*, **N Crispino***, A Xiong*, K Montgomery*, F Pinto, E Chang, Y Chen, C Christodoulopoulos, M Ziyadi, R Gupta, C Wang, B Li, D Song. (2025). VMDT: Decoding the Trustworthiness of Video Foundation Models. In Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025).

V Siu, **N Crispino**, Z Yu, S Pan, Z Wang, Y Liu, D Song, C Wang. COSMIC: Generalized Refusal Identification in LLM Activations. In Findings of the Association for Computational Linguistics (ACL 2025).

J Tu, Z Ni, **N Crispino**, Z Yu, M Bendersky, B Gunel, R Jia, X Liu, L Lyu, D Song, C Wang. MLAN: Language-Based Instruction Tuning Improves Zero-Shot Generalization of Multimodal Large Language Models. In Proceedings of the 3rd Workshop on Towards Knowledgeable Foundation Models (KnowFM @ ACL 2025).

N Crispino, K Montgomery, F Zeng, D Song, and C Wang. (2024). Agent Instructs Large Language Models to be General Zero-Shot Reasoners. In International Conference on Machine Learning (ICML 2024).

Projects

Core Contributor – MassGen

- Contributing to development and strategic direction of the open-source multi-agent LLM scaling package MassGen, designing and implementing new features to enhance capabilities.
- GitHub: https://github.com/Leezekun/MassGen.

Teaching

Teaching Assistant   **St. Louis, MO**

Natural Language Processing – Co-Head TA   *September 2023 – December 2023*
Analysis of Algorithms   *February 2023 – May 2023*
Introduction to Computer Science – Section Lead   *January 2021 – December 2023*