Nicholas Crispino

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

Research Interests: Foundation Model • Agentic AI • LLM Alignment & Safety

Education

University of California, Santa Cruz **Santa Cruz, CA**
PhD in Computer Science *09/25–05/29 (expected)*

- **Advisor:** Chenguang Wang

Washington University in St. Louis **St. Louis, MO**
PhD in Computer Science *08/24–08/25*

- **Advisor:** Chenguang Wang

Bachelor of Science *09/20–12/23*

- **Majors:** Computer Science + Economics (primary), Statistics (double)
- **GPA:** 4.00/4.00

Preprints

[1] 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.

Publications

[1] 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. VMDT: Decoding the Trustworthiness of Video Foundation Models. In Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025).

[2] S Kolasani, M Saplin, **N Crispino**, K Montgomery, J Quincy Davis, M Zaharia, C Wang, C Wang. LLM CHESS: Benchmarking Reasoning and Instruction-Following in LLMs through Chess. In Workshop on Foundations of Reasoning in Language Models (FoRLM @ NeurIPS 2025)

[3] V. Siu*, **N. Crispino***, D. Park, N. W. Henry, Z. Wang, Y. Liu, D. Song, C. Wang. SteeringSafety: A Systematic Safety Evaluation Framework of Representation Steering in LLMs. In Workshop on Socially Responsible and Trustworthy Foundation Models (ResponsibleFM @ NeurIPS 2025)

[4] V. Siu, N. W. Henry, **N. Crispino**, Y. Liu, D. Song, C. Wang. RepIt: Steering Language Models with Concept-Specific Refusal Vectors. In Workshop on Socially Responsible and Trustworthy Foundation Models (ResponsibleFM @ NeurIPS 2025)

[5] 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).

[6] 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).

[7] **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).

(*) denotes equal contribution.

Projects

Lead 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/massgen/MassGen.

Awards

- Dean's Select Fellowship (2024)
- Cox Family Fellowship (2024)
- Undergraduate Engineering Valedictorian (2024)
- Ernest D. Weiss Junior Award for Academic Excellence – Computer Science and Engineering (2023)
- Brian Blank Award in Mathematics (2023)
- Antoinette Frances Dames Award for Productive Scholarship in Engineering (2022)

Teaching

Teaching Assistant	**St. Louis, MO**
Natural Language Processing – Co-Head TA	*09/23– 12/23*
Analysis of Algorithms	*02/23– 05/23*
Introduction to Computer Science	*01/21– 12/23*

Technical Skills

Languages: Proficient in Python (Transformers, PyTorch, NumPy, scikit-learn, pandas, Matplotlib). Familiarity with Linux, Java, R, Matlab, Stata, C, C++, CUDA, SQL.