Logan Cross

Curriculum Vitae

Professional Profile

Dedicated researcher with a strong foundation in computer science, machine learning, and cognitive neuroscience. Proven track record in interdisciplinary research, aiming to drive advancements in human-centered AI.

Research Roles

2022 – Present **Stanford University** *Postdoctoral Scholar, Department of Computer Science*

- Hypothetical Minds: Scaffolding Theory of Mind (ToM) for Multi-Agent Tasks with LLMs
  - Built LLM-based agent for complex multi-agent tasks that outperforms LLM and deep RL baselines
  - ToM module infers other agent's latent states with a natural language approximation of Bayesian inference
  - Compared human experimental data to Hypothetical Minds, diagnosing similar performance and reasoning
- Advancing LLM-based Adaptive Teaching for Human and LLM Students
  - Developed a novel framework for adaptive teaching using LLMs
  - Applied pedagogical principles to enhance knowledge distillation in small LLMs
- Selective Context Augmentation Pipeline (SCAP) for Social Reasoning Evaluations
  - Augments social NLP benchmarks w/ synthetic contextual dialogues to improve LLM evaluations
  - Validated w/ LLM/human study showing reduction in ambiguity and improved inter-annotator agreement

locross@stanford.edu • *locross93.github.io*
logan-cross-55861979

May - Oct 2021 **Google DeepMind** *Research Scientist Intern*

- Advisor: Jane Wang
- Prototyped meta reinforcement learning models on a challenging meta-learning benchmark
- Developed a battery of cognitive tests of agents' abilities in a learning curriculum

2015 – 2022 **California Institute of Technology** *Graduate Research Assistant*

- Advisor: John P. O'Doherty
- Compared representations in deep reinforcement learning algorithms to neural representations in the human brain for dynamic, naturalistic tasks
- Led extensive research projects to study the neurobiological mechanisms of decision-making
- Leveraged machine learning to create cutting-edge fMRI data analysis tools
- Developed computational models of human learning and decision-making

Education

2015 - 2022 **California Institute of Technology**

PhD. in Computation and Neural Systems

Thesis: *The Neural Mechanisms of Value Construction*

Advisor: John P. O'Doherty

2011-2015 **University of Southern California**

B.S. in Neuroscience

Honors Thesis: *Classifying the Grateful Mind: Pattern Classification to Reveal Circuits for Value Judgment and Perspective Taking*

Advisor: Antonio Damasio

Publications

2024 **Cross, L.**, Xiang, V., Bhatia, A., & Yamins, D., & Haber, N.(2024). Hypothetical Minds: Scaffolding Theory of Mind for Multi-Agent Tasks with Large Language Models. Selected for oral presentation at the Workshop on Open-World Agents (OWA-2024), NeurIPS 2024. *arXiv* preprint arXiv:2407.07086.

2024 Sun, F-Y., S I, H., Yi, A., Zhou, Y., Zook, A., Tremblay, J., **Cross, L.**, Wu, J., & Haber, N. (2024). FactorSim: Generative Simulation via Factorized Representation. In *Advances in Neural Information Processing Systems*.

2024 **Cross, L.**, Xiang, V., Haber, N., & Yamins, D. (2024). Animate Agent World Modeling Benchmark. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

2023 Xiang, V., **Cross, L.**, Fränken, J. P., & Haber, N. (2023). From Centralized to Self-Supervised: Pursuing Realistic Multi-Agent Reinforcement Learning. *arXiv* preprint arXiv:2312.08662.

2023 Pool, E. R., Pauli, W. M., **Cross, L.**, & O'Doherty, J. P. (2023). Neural substrates of parallel devaluation-sensitive and devaluation-insensitive Pavlovian learning in humans. *Nature Communications*, 14(1), 8057.

2023 Iigaya, K., Yi, S., Wahle, I. A., Tanwisuth, S., **Cross, L.**, & O'Doherty, J. P. (2023). Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nature Communications*, 14(1), 127.

2021 **Cross, L.**, Cockburn, J., Yue, Y., & O'Doherty, J.P. (2021). Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron*, 109(4), 724-738.

2017 Suzuki, S., **Cross, L.**, & O'Doherty, J. P. (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nature Neuroscience*, 20(12), 1780.

Conference Presentations

2024 **Cross, L.**, Xiang, V., Bhatia, A., Yamins, D., & Haber, N. Hypothetical Minds: Scaffolding Theory of Mind for Multi-Agent Tasks with Large Language Models. *Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada.

2024 Sun, F-Y., S I, H., Yi, A., Zhou, Y., Zook, A., Tremblay, J., **Cross, L.**, Wu, J., & Haber, N. FactorSim: Generative Simulation via Factorized Representation. *Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada.

2024 **Cross, L.**, Xiang, V., Haber, N., & Yamins, D. Animate Agent World Modeling Benchmark. *CogSci 2024*, Rotterdam, Netherlands.

2023 Xiang, V., **Cross, L.**, Fränken, J. P., & Haber, N. From Centralized to Self-Supervised: Pursuing Realistic Multi-Agent Reinforcement Learning. *NeurIPS workshop on Agent Learning in Open-Endedness (ALOE)*, New Orleans, Louisiana.

2023 Xiang, V., **Cross, L.**, & Haber, N. Flexible Social Dynamics Emerge Through Model-Based Intrinsic Motivation. *International Workshop on Intrinsically Motivated Open-ended Learning (IMOL)*, Paris, France.

2023 Binder, F., **Cross, L.**, Friedman, Y., Hawkins, R., Yamins, D., and Fan, J. Advancing Cognitive Science and AI with Cognitive-AI Benchmarking. Organized workshop at *CogSci 2023*, Sydney, Australia.

Advisory Roles

2024 – Present **Mic&Pose** *Technical Advisor*

- Advising AI-powered communication coaching startup during pre-seed fundraising
- Providing technical guidance on real-time posture feedback and speech analytics systems
- Supporting development of multimodal AI models for communication skill assessment

2024 – Present **Read the Room (RTR)** *Technical Advisor*

- Advising nonprofit platform for real-time sentiment analysis and community feedback
- Guiding development of privacy-preserving data aggregation systems
- Supporting implementation of transparent and ethical data collection practices
- Contributing to platform architecture for scalable anonymous opinion gathering

Technical Skills

Programming	Python, C++, MATLAB, R, git, software engineering, Unity3D, Docker, SQL
Machine Learning	PyTorch, Tensorflow, JAX, sklearn, Ray, RLlib, Wandb, Tensorboard, reinforcement learning, LLMs, GPT API, computer vision
Mathematics and Statistics	Calculus, linear algebra, Bayesian modeling, probability
Leadership	Extensive experience mentoring, teaching, and working in multidisciplinary teams

Honors and Awards

2022-Present	Wu Tsai Neurosciences Institute Interdisciplinary Postdoctoral Scholar Award
2016-2022	NIH/NIDA Diversity Supplement fellowship
2019	ICML Conference Student Travel Award

2019 Diversity & Inclusion Travel Grant, ICML
2014 McNair Scholar

Teaching

2017-2019 Teaching Assistant – CNS 251: Human Brain Mapping: Theory and Practice
2018 Teaching Assistant – CNS 102: Brains, Minds, and Society

Organizations

2019-Present Black in AI
2015-2022 Black Scientists and Engineers of Caltech

Web

Github https://github.com/locross93/
Website https://locross93.github.io/
LinkedIn https://www.linkedin.com/in/logan-cross-55861979/