MIKA OKAMOTO

mikahokamoto@gmail.com • mika-okamoto.github.io • mokamoto • mika-okamoto •

EDUCATION

Georgia Institute of Technology, *B.S. in Computer Science* • GPA: 4.00 Aug. 2023 – May 2026

EXPERIENCE

Member of Technical Staff • Decagon Jun. 2026 – Present

- Enhanced an opaque simulation framework for CX agents through LLM-based diagnosis of chat logs against a data-driven failure taxonomy, iterating with real users to create truly useful automations that accelerated debugging by 3x across 100+ deals.
- Integrated customer telemetry into Decagon's observability platform, unblocking agent hill-climbing over customer-indexed success metrics, which uncovered improvement opportunities that closed multi-million dollar contracts.
- Consolidated alerting into a standard framework that eliminates coverage gaps and reliably detects incidents.
- Built a system to enhance agents by translating customer test specs into formal simulations, identifying misalignment between programmed and anticipated agent behavior, and iteratively co-optimizing agents and simulations.

Co-founder and Lead Researcher • Trace AI Labs Jan. 2026 – Present

- Created PACT, a multi-metric evaluation framework to assess whether enterprise AI assistants follow compliance rules in 48 realistic scenarios across regulated domains like healthcare, revealing no existing model is reliably safe.
- Analyzed the psychology of why AI agents violate legal constraints under user pressure tactics, showing alignment between LLM behavior and prevailing human behavioral theories like deterrence and legitimacy [AIES '26].

Research Assistant • Georgia Tech Human-Centered AI Lab Aug. 2024 – May 2026

- Instrumented an explainable system that profiles LLMs' strengths and weaknesses and selects the best model for tasks under cost constraints [MLSys YPS '25], and generalized it to routing complex agent workflows [HCXAI '26].
- Traced where social reasoning in LLMs originates in pretraining corpora using training-data attribution [COLM '26].
- Developed counterfactual explanation methods that translate complex agentic-workflow behavior and errors into clear, actionable insights for developers through a visual, interactive interface. [HCXAI '26].
- Led FLaME, the first holistic benchmark of LLMs over diverse financial analysis and reasoning tasks [ACL '25].

Research Assistant • Georgia Tech Research Institute Jan. 2024 – Dec 2025

- Built an agentic sandbox with a chat interface for creating, executing, and debugging scientific experiments, infusing context from retrieved research papers and code to empower researchers to rapidly iterate on their ideas.
- Proposed novel graph algorithms for decomposing scientific claims into atomic parts for automatic verification.
- Architected MethodBench, a framework for evaluating LLM competence in determining the solution to open-ended, but previously solved, scientific problems [ARR Preprint; DARPA Contract HR001125C0302].

Software Engineer Intern • Two Sigma Jun. 2025 – Aug. 2025

- Designed metrics to provide traders actionable feedback on their stock ideas and market timing, surfacing suboptimal trading patterns and informing more profitable quantitative models distilled from human trading intuition.
- Built an AWS-native ETL pipeline using Lambda and Step Functions to process metrics in Postgres, incorporating observability, testing, and documentation to ensure reliable and rapid insight delivery to quantitative researchers.

ACTIVITIES

- **Reviewing/Program Committee Engagements**: ACL Rolling Review ERC Member (2026), ACM/AAAI Conference on AI Ethics and Society PC Member (2026), Eleuther AI Summer of Open AI Research Application Reviewer (2026)
- **Awards/Fellowships**: Supervised Program for Alignment Research (2026); GT Fintech Research Fellow (2025)
- **Research Presentations**: AAAI/ACM Conference on AI Ethics and Society (2026), ACM CHI Workshop on Human Centered Explainable AI(2026), MLSys Young Professionals Symposium (2025), COLM Workshop on Agent Behavior (2026), COLM Workshop on Scientific Understanding of Foundation Models (2026)

SKILLS

- **Programming**: Python, R, Java, SQL, Bash, C++, HTML, JavaScript, TypeScript, Go
- **Technologies**: Git, Github, Docker, AWS, Hugging Face, ReactJS, Flask, REST APIs, Postgres, Coding Agents
- **Data and ML**: PyTorch, TensorFlow, Scikit, Pandas, NumPy, Jupyter, MLFlow, Matplotlib, Seaborn, R-Studio
- **Natural Language**: LLM and Agent Evaluation, LLM-as-judge, Fine-tuning, Agent Orchestration, Explainable AI

PUBLICATIONS † denotes equal contribution

[1] **M. Okamoto**, A. K. Erol. "PACT: Can Enterprise AI Assistants Be Trusted Under Pressure?" *Under review*, 2026.

[2] **M. Okamoto**, A. K. Erol, K. Erol. "Why Do AI Agents Break Rules? How Framing, Context, and Social Signals Shape Compliance." *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, 2026. Also at the COLM Workshop on Agent Behavior.

[3] G. Matlin, C. Chakraborty, S. Eom, **M. Okamoto**, R. Castilla, L. Jaburi, A. Deng, T. Min, L. Quirke, S. Biderman, M. Riedl. "Where Does Social Reasoning Come From? Capability Provenance in Language Models." *Conference on Language Modeling (COLM)*, 2026.

[4] **M. Okamoto**, G. Sarti. "Encoded Early, Used Late: Where Transformers Begin to Act on an Inferred Partner's Expertise." *COLM Workshop on Scientific Understanding of Foundation Models*, 2026.

[5] **M. Okamoto**, A. Erol, M. Riedl. "Explainable Model Routing for Agentic Workflows." *Workshop on Human-Centered Explainable AI (HCXAI) at CHI*, 2026.

[6] B. Johnson, **M. Okamoto**, J. C. Kerce, F. Fekri. "Method Bench: Evaluating Path-to-Feasibility Reasoning in the Sciences." *Under review*, 2026.

[7] M. Singh†, G. Kim†, **M. Okamoto**†, A. Ammavajjala, A. Alabdulkarim, G. Mansi, M. Riedl. "Counterfactual Explanations for Agentic Workflows." *Workshop on Human-Centered Explainable AI (HCXAI) at CHI*, 2026.

[8] S. Xiong, O. Gungordu, B. Johnson, **M. Okamoto**, J. C. Kerce, F. Fekri. "DeepVerify: Evidence-Based Expert-Level Scientific Claim Verification." *Preprint*, 2026.

[9] G. Matlin†, **M. Okamoto**†, H. Pardawala†, Y. Yang, S. Chava. "FLaME: Holistic Finance Language Model Evaluation." *Findings of the Association for Computational Linguistics (ACL)*, 2025.

[10] **M. Okamoto**, A. Erol, G. Matlin. "Trust by Design: Skill Profiles for Transparent, Cost-Aware LLM Routing." *MLSys Young Professionals Symposium*, 2025.

[11] A. Iyer†, **M. Okamoto**†, A. Lorenzo, B. Johnson, M. Kim, C. Kerce, M. Gonzalez. "Enhancing Military Family Readiness and Resilience Programs Through LLM-Generated Synthetic Data and Natural Language Interfaces." *Military Health System Research Symposium (MHSRS)*, 2025.

[12] B. B. Johnson†, T. M. Sanches†, **M. Okamoto**†, K. M. Nguyen, C. A. Ortez, O. Polesskaya, A. A. Palmer. "RATTACA: Genetic Predictions in Heterogeneous Stock Rats Offer a New Tool for Genetic Correlation and Experimental Design." *bioRxiv*, 2023.

[13] O. Polesskaya, E. Boussaty, R. Cheng, O. A. Lamonte, T. Y. Zhou, E. Du, T. M. Sanches, K. M. Nguyen, **M. Okamoto**, A. A. Palmer, R. Friedman. "Genome-Wide Association Study of Age-Related Hearing Loss in CFW Mice Identifies Multiple Genes and Loci, Including Prkag2." *Journal of the Association for Research in Otolaryngology*, 2025.