Ali Keramati

2212 Verano Rd S, Irvine, CA

+1 (951) 258-0875 a.kera@uci.edu alikera.github.io alikera ali-keramati Google Scholar

Education

University of California Irvine **Sep. 2024 – Jun. 2029**

Ph.D in **Education Data Science** (GPA: 4/4) *Irvine, CA, USA*

- **Relevant Courses:** *Neural Networks and Deep Learning, Deep Generative Models, Neural Architecture of Language, Language Models in Cognitive Science, Learning Analytics Practicum, Qualitative Research Methods*

University of Tehran **Sep. 2019 – Jul. 2024**

B.Sc. in **Computer Engineering** (GPA: 18.51/20) + Minor in **Psychology and Education** (GPA: 19.21/20) *Tehran, Iran*

- **Major Degree Courses:** *Discrete Mathematics, Data Structures and Algorithms, Artificial Intelligence, Machines and Language Theory, Engineering Probability and Statistics, Internet Engineering, Foundations of IT, Large Language Models*
- **Minor Degree Courses:** *Educational Psychology, Instructional Design, Curriculum Development, Intro to Psychological Tests, Teaching Methods, Qualitative and Quantitative Research Methods, Philosophical Schools and Educational Theories*

Research Interests

- **GenAI**: LLMs/GANs/Diffusion Models/Flows/VAEs
- Multi-Agent and **Agentic AI**: Tool-use/RAG/Memory
- **LLM-as-Judge**: Automated Evaluation/Benchmarking
- **Fine-Tuning** & Alignment: LoRA/Adapters/RLxF
- **HCI**: Human-Centered AI/Human-in-the-Loop AI
- Robust, Safe, Explainable, Trustworthy, Responsible AI
- Personalized Recommender Systems & Conversational-AI
- Cognitive User Modeling & Human Behavior Analysis

Work Experience

Research Scientist Intern at Harvard University **Jun. 2025 - Sep. 2025**

Learning Media Lab, PI: Prof. Y. Xu *Cambridge, MA*

- Developed Curio 2, a LangGraph-based science chatbot that utilizes graph-driven dialogue and state management for enhanced Transparency and Explainability **(XAI)**, while promoting **Trustworthy AI** through Child-Safe design, robust dialogue management, and safeguards against harmful content. System alignment was achieved via a self-correction pipeline grounded in **Constitutional AI** principles, specifically adapted to K–12 educational standards.
- Compared two voice interaction architectures, **realtime (speech-to-speech)** and **chained**, to evaluate trade-offs between latency, transparency, pedagogical control, and age-appropriate conversational learning experiences in educational voice agents.

Volunteer Research Scientist **Oct. 2025 - Present**

Oumi: Open Universal Machine Intelligence *Remote*

- Improving **LLM judges** for better reasoning and reduced inconsistencies; *TrustJudge*, a framework for more reliable evaluations; Analyzing LLM benchmark failures and proposed solutions.

AI/ML Engineer **Sep. 2022 - Jul. 2024**

Tadvin Farayand Co. *Tehran, Iran*

- Analyzed structured ERP and unstructured industrial data; designed AI-driven workflows to enable autonomous collaboration across ERP modules for inventory analysis, document processing, and predictive maintenance.

Selected Publications

[1] Keramati A. and Warschauer M. **"Application of Multi-Agent Systems for Essay Scoring"**, NeurIPS 2025

[2] Keramati A., Jie C., Warschauer M. and Shi Y. **"Simulating Students' Java Programming Errors with Large Language Models"**, The International Conference on Learning Analytics & Knowledge (LAK 2026) (Under Review)

[3] Tajik E., Shahrokhian B., Borchers C., Keramati A., Simon S., Pal S. and Sankaranarayanan S. **"Disagreement as Data: Reasoning Trace Analytics in Multi-Agent systems"**, The International Conference on Learning Analytics & Knowledge (LAK 2026) (Under Review)

[4] Keramati A. and Warschauer M. **"The Order Matters: Sequential Fine-Tuning of LLaMA for Coherent Automated Essay Scoring"**, ICLR 2026 (Under Review)

[5] Keramati A. and Warschauer M. **"MA6S: Multi-Agent Supervisor System for Essay Scoring and Feedback Generation with Teacher in the Loop "**, CSCL 2026 (Under Review)

[6] Keramati A., Zhou S., Mehrotra S., and Warschauer M. **"MADRAG as Judge: A Multi-Agent Debate and RAG System for Rubric-Based Evaluation"**, (Pre-print)

[7] Fallah A., Keramati A., Nazari M., and Mirfazeli F. **"Automating Theory of Mind Assessment with a LLaMA-3-Powered Chatbot: Enhancing Faux Pas Detection in Autism"**, *14th International Conference on Computer and Knowledge Engineering (ICCKE 2024)*

[8] Shahhoseini N., Taghiyareh F., and Keramati A. **"iTAG: Easy, Rapid, Automatic Intelligent Tagging for Educational Contents"**, *2025 29th International Computer Conference, Computer Society of Iran (CSICC)*

[9] Keramati A., Keramati M. R., and Arefian M. H. **Students' Reflection on the Effect of Collaborative Learning on Learning Environment and Academic Achievement in Online Reflective Platforms**, *Reflective Practice*

Honors & Awards

NeurIPS 2025 — AI Education Workshop Program **Dec. 2025**
Accepted Oral Presentation *San Diego, CA*

Scholarship Recipient, LearnLab Summer School at Carnegie Mellon University **Jul. 2025**
Awarded funding to attend the program on Computer Science Education Research track at CMU *Pittsburgh, PA*

Graduate Ambassador Candidate – TPC Mentor Fellowship at *UC Irvine* **Jun. 2025**
Chosen to oversee mentor–mentee coordination and selection, enhancing communication and program engagement *Irvine, CA*

Top 10 — B.Sc. Computer Engineering (2019 cohort) at *University of Tehran* **Jul. 2024**
Ranked within the top 10 of the graduating cohort based on cumulative GPA *Tehran, Iran*

Top 0.1%, Konkour — National University Entrance Exam **Aug. 2019**
Ranked among hundreds of thousands of examinees; admitted to B.Sc. in CE at Iran's top-ranked university *Iran*

Selected Research Projects

University of California, Irvine

- *PapyrusAI:* Designed a GPT-based personalized generative writing coach leveraging an **agentic AI** pipeline capable of autonomous decision-making for tool invocation, web search, and dynamic **RAG** utilization. Integrated fine-tuned **BERT** for labeling argumentative essays and LLaMA-3.1-8B with **LoRA** + 4-bit **quantization** for essay evaluation.
- Built and deployed LLM-powered collaborative classmates with controllable dialogue behaviors in live classroom environment; conducted post-interaction surveys to evaluate students' perceptions of engagement and learning support.
- Analyzed multi-year quiz data to assess the impact of AI usage on student outcomes. Modeled **student behavior** over time and **benchmarked** LLMs on the same assessments to evaluate their proficiency in answering quiz questions.
- As **TA**, developed an **automated grading** and **feedback** system for weekly assignments and co-built an **AI detection system** using LLMs to identify AI-generated submissions based on writing patterns, timing, and key concepts.

LearnLab Summer School (2025) at Carnegie Mellon University

- Developed a multi-agent **student simulator + verifier** framework using LLMs to generate Java programs with realistic logical errors. The simulator used **Input-Output**, **Chain-of-Thought**, and **Self-Refine** prompting to produce buggy code. Implemented a **RLAIF** loop where the verifier agent provided structured feedback on error quality, which was then used to refine the simulator. Evaluated error diversity and fidelity using **ASTs** and **ZSS** tree edit distance, benchmarking against 74,000+ real-world CodeWorkout submissions.

University of Tehran

- Developed a multi-agent chatbot for open learner modeling, using a multi-stage conversational flow to dynamically **profile** and update learner characteristics based on their in-chat interactions. Integrated a **RAG** system that uses a learner's profile as a query to retrieve the top 3 relevant profiles, aligning personalized activity recommendations.
- Automated generation of interactive learning modules from **IEEE LOM metadata** and integrated Avida-ED simulations; applied genetic algorithms to personalize activity parameters and difficulty levels.
- Implemented Theory of Mind assessment for individuals with Autism by developing a **LLaMA-3 chatbot** to administer the Faux Pas Recognition Test. The system uses a multi-phase pipeline that features interactive rehabilitation dialogues with adaptive hints and an automated scoring phase using an **LLM-as-judge** to generate clinician-ready reports.

Technical Skills

Programming Languages: C/C++/C#, Python, Java, JavaScript, R, HTML/CSS, Verilog, Assembly, Stata, SQL
Frameworks & Libraries: LangChain, AutoGen, ReAct, PyTorch, TensorFlow, React, Spring
Technologies & Tools: CUDA, Git, Docker, Kubernetes, Maven, Makefile, Tomcat
Databases: MySQL, Postgres, MongoDB, Neo4j, Redis, ElasticSearch
Operating Systems: Linux, Windows, macOS
Qualitative Analysis Tools: ATLAS.ti, Dedoose

Certifications

Generative AI with Large Language Models **Jul. 2022**
DeepLearning.AI — Coursera *Online*

Fundamentals of Reinforcement Learning **Aug. 2023**
University of Alberta — Coursera *Online*

Introduction to EdTech **Sep. 2023**
EDHEC Business School — Coursera *Online*

LangChain: Chat with Your Data **Jul. 2024**
DeepLearning.AI *Online*

Microsoft Certified: Azure AI Fundamentals **Oct. 2025**
Microsoft Learn *Online*