Roya Daneshi

Linkedin | Github | 970-581-5445 | Email | Personal website

PhD researcher in Artificial Intelligence with a focus on Human–AI interaction and Explainable AI, seeking an industry research internship to apply academic findings to real-world problems.

Education

2025 – Present Ph.D. in Computer Science (Artificial Intelligence), **Colorado State University**
roya.daneshi@colostate.edu

2020 – 2025 Bachelor's Degree in Computer Engineering, **Sharif University of Technology** (GPA: 18.85/20, 3.86/4)
roya.daneshi@sharif.edu

Research Interests

Human–AI Mental Model Reconciliation
Human-Aware Planning
Explainable AI and Trust in Human-AI Systems
Artificial Intelligence

Publications

Implications of AI Involvement for Trust in Expert Advisory Workflows Under Epistemic Dependence.
Under review at *ACM FAccT 2026*.

Research Experience

Dr. Sarath Sreedharan's HAPI Laboratory August 2025 - Present
Research Focus: Artificial Intelligence, Mental Model Reconciliation, Explainable AI, and Human-Aware Planning

- Framed mental model reconciliation as a search problem over model modifications.
- Formalized trust and explanation trade-offs in human–AI decision-making workflows.
- Evaluating trust in AI systems and human experts assisted by AI.
- Studying human preferences among explainable AI, explicable AI, and optimal AI systems.
- Developing an agentic AI system for the Countdown game and introducing a new benchmark for the task.

Dr. Hossein Sameti's SLPL Laboratory August 2024 – March 2025
Research Focus: Investigating the Theoretical Feasibility of Artificial General Intelligence (AGI)

- Identified and formalized fundamental requirements for AGI-based systems.
- Collected and categorized perspectives both supporting and opposing AGI achievability.
- Analyzed theoretical pathways toward AGI and assessed its potential achievability.
- Conducted a comparative analysis of existing theories and evidence to evaluate feasibility.

Dr. Mohammad Hossein Rohban's RILM Laboratory March 2024 – August 2024
Research Focus: Out-of-distribution detection and anomaly detection in machine learning models

- Trained robust deep learning models based on ResNet architectures for detecting out-of-distribution (OOD) inputs.
- Generated synthetic OOD data using multiple techniques, including image cropping, patch shuffling (puzzle), noise injection, and diffusion models.
- Conducted experiments on diverse benchmark datasets, including ImageNet, CIFAR-10, CIFAR-100, MNIST, Fashion-MNIST, and MVTec AD.
- Evaluated model performance using standard metrics such as AUC, precision, and recall.

Skills

Languages	**Persian** (Native), **English** (C1, TOEFL Score: 95 – R23, L28, S22, W22), **Arabic** (Intermediate)
Technical Skills	Python, PyTorch, NumPy, Scikit-learn, PDDL, Linux, Git, Docker, LaTeX Java, C, R, Go, JavaScript, React, CSS, HTML, Selenium, Android, Swift, Verilog
Soft Skills	Teamwork, Time management, Strong willingness to learn, Effective communication, Decision-making under pressure, Strong presentation skills, Active listening, Dependable and responsive
Interests	Music, Biking, Kayaking

Work Authorization

Valid F-1 visa with U.S. work authorization.

Teaching Assistant

Discrete Structures and their Applications	(January 2025 – Present)
Introduction to Artificial Intelligence	(August 2025 – December 2025)
Signals and Systems	(September 2024 – February 2025)
Linear Algebra	(September 2024 – February 2025)
Software Testing	(June 2024 – September 2024)
Web Programming	(October 2023 – February 2024)
ENG Probability and Statistics	(October 2022 – February 2023)
Volunteer Programming Course Assistant for new university students	(October 2022 – February 2023)

Honors and Awards

Completed the **CITI Program** Group 2: SOCIAL/HUMANISTIC/BEHAVIORAL RESEARCH and obtained certification	2025
Completed the **Scrum Foundations** seminar as a volunteer and obtained certification	2023
Ranked 78th among more than 155,000 participants in the national university entrance exam	2020