Amit Kiran Rege

amit.rege@colorado.edu · github.com/amitrege · amitrege.github.io · +1 (303) 960-4240

PhD Candidate, Computer Science · CU Boulder · Advisor: Prof. Claire Monteleoni

Research in theoretical machine learning, with a focus on AI safety and control: applying RL theory to LLM post-training, foundational multi-agent bandits, and the mathematics of interpretability. Published at AISTATS, L4DC (oral), IEEE CDC, and Physical Review Applied; 6 papers currently under submission.

RESEARCH

Reinforcement Learning Theory · Multi-Agent Bandits · Data Attribution for Interactive Learning · AI Interpretability & Safety · Generative Models

SELECTED PUBLICATIONS

Multi-Agent Lipschitz Bandits	AISTATS 2026
Flickering Multi-Armed Bandits	L4DC 2026 **[Oral]**
A Unified Framework for Locality in Scalable MARL	L4DC 2026
Incentivized Lipschitz Bandits	IEEE CDC 2025
Where Did Your Model Learn That? — Label-free Influence for SSL	arXiv 2024
Hamiltonian Learning using Machine Learning	Physical Review Applied 2024

+ 6 papers under submission: RL attribution theory, oracle separations for RL, interpretability auditability, KL-regularized agent monitorability

OPEN SOURCE PROJECTS

tinygrad-classroom	Interactive autograd: scalar neuron → backprop, built for teaching	`amitrege/tinygrad-classroom`
NeuroSteer	Stepwise activation steering library for Llama-style decoder LLMs	`amitrege/NeuroSteer`
tracemap	Training-data influence with query heatmaps for image classifiers	`amitrege/tracemap`
rl-attr-code	Exact RL data attribution distinguishing replay vs. interventional effects	`amitrege/rl-attr-code`

EXPERIENCE

Investment Associate — Deming Center Venture Fund, CU Boulder *Apr 2025 – Present*

Source and evaluate early-stage deep-tech companies for CU Boulder's student venture fund

Led investment in Great Sky (AI chip startup); closed at fund's highest valuation ever ($50M)

Student Engineer — DARPA Subterranean Challenge, Team MARBLE *Summer 2019*

Built object detection & localization pipeline for robots navigating lightless underground mines

Placed **4th of 11 teams** worldwide against NASA JPL, CMU, and UPenn

Research Intern — École Normale Supérieure (advised by Prof. Cezara Dragoi) *Feb – Jul 2018*

Formally verified round-based synchronicity of Raft and Viewstamp consensus protocols

EDUCATION

PhD, Computer Science · University of Colorado Boulder	2020–2026 (expected)
MS, Computer Science · University of Colorado Boulder	GPA: 4.0/4.0 · 2020
BE (Hons.), Computer Science · BITS Pilani–Goa	2018

SKILLS & TOOLS

Python · PyTorch · JAX · NumPy · C++ | RL Theory · Bandits · AI Interpretability · Generative Models · Formal Verification · LLM Fine-tuning

TALKS & INDUSTRY ENGAGEMENT

LLMs from Scratch — Spencer Fane LLP (Industry Talk)	Nov 2025
Frontiers of Generative AI — Ensemble Innovation Ventures	Jan 2026
Theoretical Machine Learning — KOA 94.1 FM iHeart Radio	Jan 2025

Guest Lecturer: CSCI 7000 (Adv. ML), CSCI 4622 (ML), CSCI 5802 (Data Science)