

Khawaja Abaid Ullah

khawajaabaid.github.io | github.com/KhawajaAbaid | linkedin.com/in/khawajaabaid | khawaja.abaid@gmail.com | 585-710-6778
PhD researcher at RIT's KAS Lab studying secure and adaptive learning systems, with current work on antidistillation defenses against model extraction and reinforcement learning for multi-agent autonomous systems. Broader interests include LLM-assisted reasoning and optimization.

Education

Ph.D. in Electrical and Computer Engineering

Jun 2026 - Present

Rochester Institute of Technology

Rochester, NY

- Graduate Research Assistant fellowship with financial support.

M.S. in Artificial Intelligence

Aug 2024 - May 2026

Rochester Institute of Technology

Rochester, NY

- GPA: 4.0/4.0; fully funded by the U.S. Department of State Fulbright Scholarship.

B.S. in Computer Science

2017 - 2021

University of Narowal

Narowal, Pakistan

- Graduated first in the Computer Science class of 2021.

Research Experience

Graduate Research Assistant - Khojasteh Autonomous Systems Laboratory

May 2026 - Present

Rochester Institute of Technology

Rochester, NY

- Develop inference-time defenses against distillation-based model extraction, designing reproducible PyTorch benchmarks that quantify the trade-off between legitimate model utility and extraction efficacy.
- Research reinforcement learning and multi-agent autonomous systems, with emphasis on coordination, resource allocation, and diagnosing learning and reward-design failure modes; translate results into manuscripts and open-source research code.

Selected Publications

K. A. Ullah, M. J. Khojasteh. “*ADS-C: Antidistillation Sampling for Classification*.” arXiv:2607.15467, 2026. [Paper](#) | [Code](#)

- Proposed a margin-aware inference-time defense that preserves every served top-1 prediction while reducing distilled-student accuracy by 29.6 pp on CIFAR-10, 17.4 pp on CIFAR-100, and 13.3 pp on Tiny-ImageNet.

K. A. Ullah, M. J. Khojasteh. “*Reversing the Incentive to Distill: A Review of Test-Time Antidistillation Defenses for Classification*.” Manuscript under review, 2026.

- Built a unified matched-cost benchmark and a method-agnostic top-1-preservation wrapper; directed perturbations drove distilled students 34.5 pp (CIFAR-10) and 47.7 pp (CIFAR-100) below the hard-label baseline.

H. Mahmud, **K. A. Ullah**, M. J. Khojasteh, J. Heard, P. David. “*How Do Large Language Models Judge Social Attraction? Evidence from Theory-Grounded Persona Ratings Across Multiple LLMs and Humans*.” arXiv:2608.09717, 2026. [Paper](#) | [Code](#)

- Evaluated social-judgment reliability and human-LLM alignment across 34 LLMs and 198 human participants, including controlled analyses of gender presentation.

Industry Experience

Software Development Engineer, AI/ML Intern

May 2025 - Aug 2025

Amazon Web Services (AWS)

East Palo Alto, CA

- Engineered core Apache Spark testing modules for a next-generation internal benchmarking platform, closing feature-parity gaps with its legacy predecessor across targeted benchmarking workflows.
- Exceeded the original project scope by independently identifying and implementing additional benchmarking capabilities, earning strong recognition from team leadership.

Open-Source Experience

Open-Source Developer

May 2022 - Aug 2024

Independent

Remote

Teras - [GitHub](#)

- Created a Keras 3-based tabular deep-learning library supporting JAX, PyTorch, and TensorFlow. Implemented research architectures for classification, regression, data imputation, and synthetic-data generation.

Keras

- Contributed Beta and Binomial random-sampling functionality to the multi-backend Keras 3 API across JAX, PyTorch, and TensorFlow.

micrograd_c - [GitHub](#)

- Implemented Andrej Karpathy’s micrograd from scratch in C. micrograd is scalar autograd engine and neural-network library with a PyTorch-like API.

Additional Technical Experience

Junior Python Developer (Remote)

Feb 2023 - May 2023

Comundo

Copenhagen, Denmark

- Built semi-supervised web-scraping and data-curation pipelines for energy-consumption data; cleaned and aggregated outputs for CO2 estimation and downstream ML forecasting.

Selected Research Projects

Fontus - Multi-Agent RL for Aerial UAV Recharging

Dec 2025

- Built a custom PettingZoo Parallel multi-agent environment with vectorized NumPy dynamics, energy-aware transitions, and proximity-triggered charging; trained parameter-shared IPPO/PPO policies with PyTorch and Ray RLlib.
- Derived a no-charge baseline and diagnosed a reward-hacking “lazy agent” local optimum: the trained policy learned inactivity rather than refueling, motivating off-policy/replay-based exploration.

codriver - Resource-Efficient Contextual Code Generation

Dec 2024

- Built a custom 260M-parameter Gemma code-generation prototype and applied staged knowledge distillation from Gemma 2 2B and CodeGemma 2B over text and code datasets for efficient local inference.

Technical Skills

Languages: Python, C, C++

ML / AI: JAX, Flax, PyTorch, TensorFlow, Keras, NumPy

Systems / Robotics: Apache Spark, ROS 1, ROS 2, Arduino

Methods: Reinforcement Learning, Multi-Agent RL, Knowledge Distillation, Model Extraction, ML Security, LLM Training

Teaching, Service & Honors

Teaching Assistant - Principles of Robotics

Aug 2026 - Present

Rochester Institute of Technology

Rochester, NY

- Develop lecture and laboratory material for undergraduate and graduate robotics course; labs use ROS 1, ROS 2, and Arduino.

Reviewer, IEEE Signal Processing Letters

Jun 2026 - Present

Honors: Fulbright Scholar (U.S. Department of State); Kaggle Expert.