

AMIRHOSSEIN ROKNILAMOUKI

Phone: +1 (213) 354-3761 · roknilamouki.1@osu.edu · [\[LinkedIn\]](#) · [\[GitHub\]](#) · [\[Google Scholar\]](#)

RESEARCH PROFILE

Ph.D. candidate in Electrical and Computer Engineering at Ohio State University advancing **safe**, **Trustworthy**, and **scalable AI**. My current research includes **provably safe RL theory**, **distributed systems** for multi-objective **LLM alignment**, and **deep RL** methods for the **pre-training** and **fine-tuning** of intelligent agents. Before my Ph.D., my M.Sc. focused on autonomous control, **planning**, and **Model Predictive Control (MPC)** for safety-critical systems. My work has led to impactful research outcomes, including a first-author publication at **ICML 2025**.

EDUCATION

- **The Ohio State University** Columbus, OH
Ph.D. in Electrical and Computer Engineering Expected: May 2027
 - **Research Focus:** Ph.D. candidate in Electrical and Computer Engineering at The Ohio State University, focusing on *safe reinforcement learning (RL)*, *distributed LLM alignment*, and *pre-training and safe fine-tuning* of learning systems.
- **University of Tehran** Tehran, Iran
M.S. in Electrical and Computer Engineering [2022]
 - **Thesis:** Autonomous Surface Vessel Control
- **University of Tehran** Tehran, Iran
B.S. in Electrical and Computer Engineering [2019]

PUBLICATIONS

1. **Roknilamouki, A.**, Ghosh, A., Shi, M., Nourzad, F., Ekici, E., and Shroff, N. "Provably Efficient RL for Linear MDPs under Instantaneous Safety Constraints in Non-Convex Feature Spaces." *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, 2025.
2. **Roknilamouki, A.**, Dagefu, F. T., Shroff, N. B., et al. "Safe and Reliable Deep Reinforcement Learning for Covert Routing." *IEEE International Conference on Mobile Ad-Hoc and Smart Systems (MASS)*, 2025. (**Invited Paper**)
3. Nourzad, F., **Roknilamouki, A.**, Ekici, E., Liu, J., and Shroff, N. "FIRM: Federated In-client Regularized Multi-objective Alignment for Large Language Models." *Manuscript under review at a top-tier AI conference*, 2025.

RESEARCH & ENGINEERING EXPERIENCE

- **Graduate Research Assistant** Columbus, OH
The Ohio State University
 - Developed a novel **safe reinforcement learning algorithm** that corrects a foundational error in prior theory, enabling agents to *maintain provable safety guarantees even in highly complex, non-convex environments* (a core component of **Trustworthy AI**). Published at **ICML 2025**.
 - Architected a **communication-efficient distributed framework** for the **multi-objective alignment of generative AI models (LLMs)**, mitigating 'disagreement drift' through novel regularization to optimally balance conflicting human-defined objectives (e.g., helpfulness, harmlessness). **Provided** the first finite-time convergence guarantees for this setting.
 - Engineered a **deep reinforcement learning system for safe and reliable decision-making** in high-stakes communication networks, leveraging **deep RL pre-training and fine-tuning** to significantly improve robustness and end-to-end performance.
- **Graduate Research Assistant (M.Sc.)** Tehran
University of Tehran
 - Engineered an **autonomous control system** for surface vessels, integrating **Distributed Model Predictive Control (MPC)** and data-driven methods to adapt to real-world environmental variability while ensuring compliance with strict collision avoidance protocols.

SELECTED PROJECTS

- **Multi-Modal Foundation Models: Vision & Language** [Code Repository]
 - *Technologies: Python, PyTorch*
 - Implemented and trained a CNN-RNN network from scratch for the task of image captioning on the Flickr8k dataset, effectively combining visual and sequential modeling.
 - Implemented a SegNet encoder-decoder network for semantic segmentation on the CamVid dataset.
- **Embodied AI & Robotics: State Estimation** [Code Repository]
 - *Technologies: Python, Gazebo, Bayesian Inference*

- Developed a probabilistic state estimator by modeling robot kinematics and sensor noise using Bayesian methods to accurately represent real-world uncertainty.
- Engineered and deployed a particle filter from scratch that fused the motion and measurement models to perform robust, real-time localization of the robot within a simulated Gazebo environment.

- **LLM-Guided Reinforcement Learning for Exploration**

[[Code Repository](#)]

- *Technologies: Python, PyTorch*
- Adapted and extended the existing ELLM framework to investigate the performance of an Actor-Critic agent in an LLM-guided context.

HONORS AND AWARDS

- Ranked **1st** in the graduate class, Electrical Engineering, University of Tehran. [2019 – 2022]
- M.Sc. Admission from ECE Department, University of Tehran, as an exceptional talent student. [2019]
- Ranked **11th** among 129 Electrical Engineering undergraduate students, University of Tehran. [2018]
- Ranked **89th (Top 0.05%)** in the National University Entrance Exam for B.Sc studies among 181,846 participants. [2015]

SKILLS

Programming Languages:

Python, C++, C, R, MATLAB

ML Libraries & Frameworks:

PyTorch, TRL, Hugging Face, Scikit-Learn, NumPy, Pandas, OpenCV

Open Source Models:

Llama, GPT-2, TinyLlama

Machine Learning Areas: Safe RL, Deep RL, Generative AI (LLM) Alignment, Federated Learning, Pre-training & Fine-tuning

Developer Tools:

Git, GitHub, Jupyter Notebooks, Google Colab

Systems & Platforms:

Linux, SLURM, OSC Clusters, CUDA.

ACADEMIC ACTIVITIES & SERVICE

- **Oral Presenter, IEEE MASS** 2025
- **Poster Presenter, 8th US-Japan Digital Innovation Hub Workshop** 2024
- **Poster Presenter, AIMACCS Workshop** 2024
- **Reviewer, IEEE MILCOM, AISTATS**
- **Teaching Assistant**, University of Tehran (Selected Courses: *Machine Learning, Modern Control, Nonlinear Systems, Linear Algebra*)