

Muhammed Ustaomeroglu

mustaome@andrew.cmu.edu [Website](#) [Google Scholar](#) [LinkedIn](#)

Education

Carnegie Mellon University

- *Ph.D. in Electrical and Computer Engineering, GPA: 4.0/4.0*

*Pittsburgh, PA
08/2023–05/2028*

Bilkent University

- *B.Sc. in Physics, CGPA: 3.99/4.0, Ranking: 1st/51*
- *B.Sc. in Electrical and Electronics Engineering, CGPA: 3.98/4.0, Ranking: 2nd/256*

*Ankara, Türkiye
09/2018–01/2023*

Experience

Research Intern (Machine Learning)

NVIDIA

- Investigated memorization in language and vision-language models and developed methods to mitigate sensitive-data retention while preserving task performance.

*Herndon, VA
05/2026–08/2026*

Graduate Research Assistant (Machine Learning)

Carnegie Mellon University

- Develop theoretical and empirical frameworks for understanding large language model (LLM) computation, spanning mental state, planning, attention mechanisms, generalization, and alignment.
- Proposed *Representational Effective Theory*, which coarse-grains LLM hidden-state trajectories into learned macrostates for interpretation, early behavioral prediction, and causal intervention. *Related:* [[webpage](#)]
- Study implicit LLM planning from an information-theoretic perspective, quantifying horizon, branching, and history dependence across tasks. *Related:* [[paper](#)]
- Develop methods for shaping LLM behavior, including mitigating emergent misalignment. *Related:* [[paper](#)]
- Analyze LLM generalization through hidden-state geometry, fine-tuning data structure, pretraining distributions, and training channels, with applications to emergent and subliminal misalignment. *Related:* [[paper 1](#)] [[paper 2](#)]
- Perform theoretical analyses of Transformer architectures, including self-attention variants, convergence/generalization properties, and long-range interaction modeling. Propose and empirically validate new model architectures inspired by theoretical insights. *Related:* [[paper](#)]

*Pittsburgh, PA
08/2023–Present*

Undergraduate Research (Theoretical Physics & Electrical Engineering)

Türkiye & Italy 2020–2022

- **University of Ferrara:** Developed a phase-space and Wigner-function formulation for composite quantum systems.
- **Nurol Makina:** Developed a computer-vision system for detecting road anomalies, with applications to military convoy safety.
- **Middle East Technical University:** Studied relativistic tidal Love numbers of spinning and nonspinning black holes using differential geometry.
- **TUBITAK:** Studied homodyne-based quantum random-number generation and information-theoretic measures of randomness.
- **Bilkent University / NANOTAM:** Conducted research in nonlinear photonic simulation and transmission-line systems.

Selected Publications

- **Towards *Effective Theory* of LLMs: A Representation Learning Approach**
M. Ustaomeroglu, G. Qu
Under Review [[Paper](#)] [[Webpage](#)][[Code](#)]
ICML 2026 Workshop on Foundations of Deep Generative Models (Spotlight)
Mechanistic Interpretability Workshop at ICML 2026
- **Language Model Planning from an Information Theoretic Perspective**
M. Ustaomeroglu*, B. Askin*, G. Joshi, C. Joe-Wong, G. Qu
International Conference on Learning Representations (ICLR), 2026 [[Paper](#)]
- **BLOCK-EM: Preventing Emergent Misalignment via Latent Blocking**
M. Ustaomeroglu, G. Qu

International Conference on Machine Learning (ICML), 2026 [Paper] [Code]

- **Transformer-Based Scalable Multi-Agent Reinforcement Learning for Networked Systems with Long-Range Interactions**
V. Sinha, M. Ustaomeroglu, G. Qu
European Control Conference (ECC), 2026 (Best Student Paper) [Paper]
- **Emergent and Subliminal Misalignment Through the Lens of Data-Mediated Transfer**
B. Askin*, M. Ustaomeroglu*, A. Nayak*, G. Joshi, G. Qu, C. Joe-Wong
Under review [Paper]
ICML 2026 Workshop on Foundations of Deep Generative Models
- **A Theoretical Study of (Hyper) Self-Attention through the Lens of Interactions**
M. Ustaomeroglu, G. Qu
International Conference on Machine Learning (ICML), 2025 [Paper]

Selected Awards

- **CyLab Presidential Fellowship**, Carnegie Mellon University, 2027.
- **Best Student Paper Award**, European Control Conference (ECC), 2026. Awarded to 1 paper.
- **Leo Finzi Memorial Fellowship**, Carnegie Mellon University, 2025–2026.
- **Carnegie Institute of Technology Dean’s Fellowship**, Carnegie Mellon University, 2023–2024.
- Bilkent University EEE Department Awards: Senior/Voluntary Projects, Entrepreneurship, Social Awareness, and Academic Excellence Awards.
- **Bilkent University Comprehensive Scholarship**, full tuition waiver and stipend, 2017–2023.
- TUBITAK Scholarship: undergraduate stipend, 2019–2023.
- **Nationwide University Entrance Examination**, ranked 506th among 2.5 million students in Türkiye, 2017.

Skills

Selected CMU Coursework in Machine Learning: Theoretical and Empirical Foundations of Modern Machine Learning; Advanced Machine Learning; Advanced Introduction to Machine Learning; Fundamentals of MDPs and Reinforcement Learning; Intermediate Statistics.

Mathematical Coursework: My training in theoretical physics and electrical engineering included coursework covering linear algebra, information theory, differential geometry, differential equations, complex analysis, advanced calculus, probability theory, geometric algebra, group theory, signals and systems, statistical mechanics, and stochastic models.

Programming and Computing: Experience with Python and machine-learning tools including PyTorch, NumPy, TensorBoard, and SLURM-based high-performance computing workflows.

Service and Other Experience

- **Conference Reviewer:** ICML, ICLR, NeurIPS.
- **Teaching Assistant:** [Introduction to Machine Learning for Engineers \(18-661\)](#), Carnegie Mellon University, Spring 2024 and Fall 2026 semesters.
- **Student Leadership:** Board Member, Carnegie Mellon University Turkish Student Society.
- Hosted international ambassadors from more than 50 countries during the 2021 Malazgirt Commemoration in Türkiye.