

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

*Herndon, VA
05/2026–08/2026*

- Investigate memorization mechanisms in large language and vision-language models and develop methods to control how models store, retrieve, and generalize training information, including suppressing the learning of private or sensitive data while preserving task performance.

Graduate Research Assistant (Machine Learning)

Carnegie Mellon University

*Pittsburgh, PA
08/2023–05/2028*

- 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:* [\[paper\]](#)

Visiting Researcher (Quantum Information Theory)

University of Ferrara

*Ferrara, Italy
07/2022–10/2022*

- Proposed a new phase space structure and Wigner function formulation for composite systems. Regardless of the system's degrees of freedom, the function defines states in a 2D phase space, which has applications in systems with high degrees of freedom.

Undergraduate Researcher (Computer Vision)

Nurol Makina

*Ankara, Türkiye
10/2021–05/2022*

- Designed an AI system that detects anomalies on the road, prospective application of which was mine detection for military convoys.

Undergraduate Researcher (General Relativity Theory)

Middle East Technical University

*Ankara, Türkiye
09/2021–04/2022*

- Used *differential geometry* to work on the relativistic theory of tidal Love Numbers for spinning and nonspinning black holes.

Undergraduate Research Intern (Quantum Random Number Generation)

TUBITAK

*İstanbul, Türkiye
06/2021–07/2021*

- Generated quantum random numbers via the Homodyne detection method and used *information theory* to develop new methods for measuring the randomness level of the quantum random numbers.

Undergraduate Researcher (Photonic Simulation)

Bilkent University

*Ankara, Türkiye
02/2021–03/2021*

- Used probabilistic differential equations such as *master equation* to make some small changes on an exiting detailed code to simulate physics of ultrafast nonlinear lasers.

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]

Skills

Machine Learning: Have a grasp of the field thanks to courses I took (Advanced Deep 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) at Carnegie Mellon University.

Mathematical Maturity: Thanks to my background in theoretical physics and engineering I learned several applied math subjects: linear algebra, information theory, differential geometry, differential equations, complex calculus, advanced calculus, probability theory, geometric algebra, group theory, signals&systems, statistical mechanics, stochastic models.

Programming: Libraries relevant to machine learning (e.g. pytorch, numpy, tensorboard etc.) and SLURM.

Selected Awards and Other Experience

- Reviewer for ICML, ICLR, NeurIPS.
- Teaching Assistant for [Introduction to ML for Engineers \(18-661\)](#) in Spring 2024.
- Leo Finzi Memorial Fellowship for the 2025-2026 academic year.
- Carnegie Institute of Technology Dean's Fellowship for 2023-2024
- Bilkent University Department of Electrical and Electronics Engineering Senior/Voluntary Projects, Entrepreneurship, Social Awareness, and Academic Excellence Awards.
- One semester tuition scholarship at the National University of Singapore.
- Bilkent University Comprehensive Scholarship: full tuition waiver & stipend during the B.Sc. Program
- TUBITAK Scholarship: stipend during the B.Sc Program.
- TURING Scholarship: stipend during the B.Sc Program.
- Ranked 506th/2.5 million in nationwide university entrance exam in Türkiye.
- Ministry of National Education Scholarship: stipend during secondary and high school education.
- Board member of Carnegie Mellon University Turkish Student Society.
- Conducted a responsible research and innovation (RRI) inquiry on hydroelectric power plants, investigated Teknopar firm via RRI guidelines.
- Hosted and guided ambassadors from more than fifty different countries in a Turkish national celebration event.