Masih Eskandar

masih.eskandar@gmail.com | +1 781 366 6801 | Boston, MA, USA
Website | GitHub | Google Scholar | /in/meskandars/

Summary

Machine learning researcher with 5+ years of experience in continual learning, adversarial robustness, and scalable AI. Published in top-tier venues (ICLR, TMLR). Developed foundation models for real-world applications such as skin cancer diagnosis and genomics. Passionate about trustworthy AI for critical domains such as genomics, healthcare, and safety-critical systems.

Education

Northeastern University

- Ph.D. in Electrical Engineering (In Progress) — 2022 - 2027 (Expected)
- M.S. in Electrical Engineering — 2022 - 2024

Sharif University of Technology

- B.S. in Computer Engineering — 2018 - 2022

Professional Experience

Northeastern Univerisity | Machine Learning Lab @ SPIRAL — *2022 - Curr.*
Doctoral Candidate

- Developed STAR, a regularization method using weight perturbations for continual learning methods, resulting in up to 15% improvements in accuracy (ICLR 2025)
- Proposed ADAPT, an adversarially robust prompt-tuning method for Vision Transformers, increasing robustness by up to 40% compared to existing methods (TMLR 2025)
- Implemented deep learning methodologies for dermatology, including image generation using stable diffusion, multi-modal LLMs, and feature matching to enhance the skin cancer diagnostic pipeline
- Developing interpretable Transformer-based models for splice site prediction using RNA sequencing data
- Developing theoretically verifiable continual learning algorithms for safety-critical applications (e.g. healthcare)

Technical University of Munich | CAMP — *2021*
Research Intern

- Developed a novel method for explaining the predictions of machine learning models using information theory

Sharif University of Tech. | Robust/Interpretable ML lab — *2020 - 2022*
Undergraduate Research Assistant

- Proposed an efficient adversarial training method for deep neural networks, solving key overfitting issues and leading to up to 50% increase in robust accuracy (ISwA 2023)

Publications

- **Grounding Multimodal Large Language Models with Quantitative Skin Attributes: A Retrieval Study** — **Under Review (2025)**
  M. Torop, **M. Eskandar** , N. Kurtansky, J. Liu, J. Weber, O. Camps, V. Rotemberg, J. Dy, K. Kose
- **CerCE: Towards Verifiable Continual Learning** — **Under Review (2025)**
  M. Eskandar, F. Tohidian, A. Kashiri, M. Everette, J. Dy
- **DISCO: Disentangled Communication Steering for Large Language Models** — **Under Review (2025)**
  M. Torop, A. Masoomi, **M. Eskandar**, J. Dy
- **STAR: Stability-Inducing Weight Perturbation for Continual Learning** — **ICLR 2025**
  M. Eskandar, T. Imtiaz, D. Hill, Z. Wang, J. Dy
- **SAIF: Sparse Adversarial and Interpretable Attack Framework** — **TMLR 2025**
  T. Imtiaz, M. Kohler, J. Miller, Z. Wang, **M. Eskandar**, M. Sznaier, O. Camps, J. Dy
- **ADAPT to Robustify Prompt Tuning Vision Transformers** — **TMLR 2025**
  M. Eskandar, T. Imtiaz, Z. Wang, J. Dy
- **ZeroGrad: Costless conscious remedies for catastrophic overfitting in the FGSM adversarial training** — **ISwA 2023**
  Z. Golgooni, M. Saberi*, **M. Eskandar***, M.H. Rohban

Skills

- **Programming Languages:** Python, C++/C, R, Java, Golang
- **Frameworks:** PyTorch, Tensorflow, JAX
- **Tools:** Numpy, Pandas, Huggingface, Git, Docker, Auto-LiRPA

Open-source Contributions

- **Mammoth:** Integrated STAR into the Mammoth continual learning library, enabling reproducible benchmarking and broader accessibility