Masih Eskandar

masih.eskandar@gmail.com | meskandars.github.io | GitHub | Google Scholar | LinkedIn

Summary

Ph.D. candidate in Continual Learning and Robust ML at Northeastern University. Experience developing LLM-based dialogue agents, activation engineering, continual learning, and adversarial defenses. Passionate about safe and scalable ML systems for healthcare and language models.

Education

- **Northeastern University** | **Advisor: Prof. Jennifer Dy** — *09/2022 - 09/2026*
  Ph.D. in Electrical Engineering (In Progress) — Boston, MA
  M.S. in Electrical Engineering - Computer Vision, Machine Learning and Algorithms (12/2024)
  - **Courses**: Machine Learning & Pattern Recognition - Big Data Sparsity and Control - Advanced Computer Vision - Advances in Deep Learning - Verifiable Machine Learning - Advanced Machine Learning - Statistical Inference
  - Current GPA: 3.88/4.00
- **Sharif University of Tech.** | **Advisor: Prof. M.H. Rohban** — *09/2018 - 06/2022*
  B.Sc. in Computer Engineering — Tehran, Iran
  - **Courses:** Linear Algebra - Probability and Statistics - Advanced Information Retrieval - Natural Language Processing (NLP) - Signal Processing
  - GPA: 3.96/4.00 (top 5% of class)

Experience

- **Northeastern Univeristy** | **Machine Learning Lab @ SPIRAL** — *09/2022 - Curr.*
  Research Assistant — Boston, MA
  - Developed STAR, a regularization method using weight perturbations to reduce catastrophic forgetting (ICLR 2025)
  - Proposed ADAPT, an adversarially robust prompt-tuning method for Vision Transformers (TMLR 2025)
  - Implemented deep learning methodologies for dermatology, including image generation using stable diffusion, multi-modal LLMs, and feature matching
  - Developing Transformer-based models for tissue-specific and patient-specific splice site predictions of RNA-seq data
  - Developing theoretically verifiable continual learning algorithms for safety-critical applications
  - Developed the lab website!
- **Technical University of Munich** | **CAMP** — *06/2021 - 11/2021*
  Research Intern — Munich, Germany
  - Developed a novel method for explaining model predictions for individual samples or classes using various input augmentations in conjunction with information bottleneck methods
- **Sharif University of Tech.** | **Robust/Interpretable ML lab** — *06/2020 - 06/2021*
  Research Assistant — Tehran, Iran
  - Proposed an efficient single-step adversarial attack generation method for performing adversarial training while avoiding overfitting (ZeroGrad, ISWA 2023)

Publications

- **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
- **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, Verilog, SQL, Assembly (MIPS)
- **Frameworks:** PyTorch, Tensorflow, JAX
- **Tools:** Numpy, Pandas, Huggingface, Git, Docker

Open-source Contributions

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