





Sara Ghazanfari

 github.com/SaraGhazanfari
 [saraghazanfari.github.io](https://github.com/saraghazanfari)
 [linkedin.com/in/sara-ghazanfari](https://www.linkedin.com/in/sara-ghazanfari)
 sg7457@nyu.edu

EDUCATION

Ph.D. Candidate	Jan. 2023 - Jan. 2027
Electrical and Computer Engineering Department, New York University (NYU)	GPA: 3.85/4.0
<i>Advisor:</i> Siddharth Garg, Farshad Khorrami	
<i>Research Area:</i> Deep Learning, Adversarial Machine Learning, Computer Vision	
<i>Courses:</i> Probability & Stochastic (A), Deep Learning (A), Algorithmic Machine Learning & Data Science (A ⁻), Linear Systems (A), Machine Learning for Cybersecurity (A).	
Master of Science	Sep. 2018 - July 2021
Computer Engineering Department, Sharif University of Technology (SUT)	GPA: 4.0/4.0
Bachelor of Science	Sep. 2013 - Feb. 2018
Computer Engineering Department, Sharif University of Technology (SUT)	GPA: 3.6/4.0

PUBLICATIONS

Towards Unified Benchmark and Models for Multi-Modal Perceptual Metrics	Sent to CVPR 2025
EMMA: Efficient Visual Alignment in Multimodal LLMs	Sent to ICLR 2025
LipSim: A Provably Robust Perceptual Similarity Metric	ICLR 2024
R-LPIPS: An Adversarially Robust Perceptual Similarity Metric	ICML Workshop 2023

SKILLS

Programming Languages: Python, C, Java, R.
Machine Learning Tools: Scikit-Learn, Spacy, Pytorch.
Deep Learning: Vision Foundation Models, Multimodal Large Language Models (MLLM), Large Multimodal Models (LMM), Multimodal Perception Systems, Vision-Language Models (VLMs).
Database: PostgreSQL, Redis, Elasticsearch.
Software and Tools: Microsoft Office, MATLAB, Jupyter Notebook, PyCharm, Git/GitHub, Unix Shell, Kafka.
Web frameworks: Django, Flask.

TEACHING EXPERIENCES

Head TA, Machine Learning	Feb. 2021 - Aug. 2021
TA, Modern Information Retrieval	Feb. 2020 - Aug. 2020
TA, Machine Learning	Feb. 2019 - Aug. 2019

RESEARCH INTERESTS

Multimodal Perception, Large Multimodal Models (LMMs), Multimodal LLMs

PROJECTS

Towards Unified Benchmark and Models for Multi-Modal Perceptual Metrics

In this our work, we propose UniSim-Bench, the first benchmark to track the progress of perceptual similarity metrics across uni- and multimodal tasks. We identify the limitations of current specialized perceptual metrics in generalizing to unseen datasets and perceptual tasks. We propose UniSim, a set of multi-task perceptual models which are a first step towards general-purpose perceptual metrics. Together, UniSim-Bench and UniSim lay the groundwork for understanding the challenges of learning automated metrics that broadly mimic human perceptual similarity, beyond narrow, task-specific applications.

EMMA: Efficient Visual Alignment in Multimodal LLMs

In this work, we propose EMMA (Efficient Multimodal Adaptation), a lightweight cross-modality module designed to efficiently fuse visual and textual encodings, generating instruction-aware visual representations for the language model. Our key contributions include: (1) an efficient early fusion mechanism that integrates vision and language representations with minimal added parameters (less than 0.2% increase in model size), (2) an in-depth interpretability analysis that sheds light on the internal mechanisms of the proposed method; (3) comprehensive experiments that demonstrate notable improvements on both specialized and general benchmarks for MLLMs.

LipSim: A Provably Robust Perceptual Similarity Metric

In this work, we demonstrate the vulnerability of the SOTA perceptual similarity metric based on an ensemble of ViT-based feature extractors to adversarial attacks. We then propose a framework to train a robust perceptual similarity metric called LipSim (Lipschitz Similarity Metric) with provable guarantees by leveraging 1-Lipschitz neural networks as backbone and knowledge distillation approach to distill the knowledge of the SOTA models. Finally, a comprehensive set of experiments shows the performance of LipSim in terms of natural and certified scores and on the image retrieval application.

R-LPIPS: An Adversarially Robust Perceptual Similarity Metric

In this work, we show that the LPIPS metric is sensitive to adversarial perturbation and propose the use of Adversarial Training to build a new Robust Learned Perceptual Image Patch Similarity (R-LPIPS) that leverages adversarially trained deep features. Based on an adversarial evaluation, we demonstrate the robustness of R-LPIPS to adversarial examples compared to the LPIPS metric. Finally, we showed that the perceptual defense achieved over LPIPS metrics could easily be broken by stronger attacks developed based on R-LPIPS.

EXPERIENCE

• **Research Assistant**, New York University (NYU), US Jan. 2023 - Jan. 2027

My research is focused on building robust and scalable multimodal perception systems that can operate in the real world.

Skills: Deep Learning, Vision Foundation Models, Multimodal Large Language Models (MLLM), Large Multimodal Model (LMM), Python, High-Performance Computing (HPC), Debugging, Collaborative Problem Solving, Jupyter Notebook, PyCharm, Git/GitHub, Unix Shell

• **Technical Team Lead**, Narvan Startup Studio, Iran Oct. 2021 - Jan. 2023

The technical team leader of a software development team. The goal was to develop a cryptocurrency exchange website and mobile application. I was in charge of developing the backend of the software with a team of backend developers. Our stack was composed of Python, Django, PostgreSQL, Sentry, and gRPC and we followed the microservices architecture pattern for our architecture.

Skills: Technical Engineering, Team Management, Python, Debugging, Collaborative Problem Solving, PostgreSQL, Redis, Elasticsearch, Jupyter Notebook, PyCharm, Git/GitHub, Unix Shell, Kafka, Django, Flask.

• **Data Engineer**, Narvan Startup Studio, Iran Aug. 2020 - Oct. 2021

I designed and implemented data pipelines using Kafka, as well as storing and retrieving data in both SQL-like databases, such as PostgreSQL, and NoSQL databases, including Elasticsearch, InfluxDB, and Redis. Additionally, I designed an NLP pipeline applied to Persian news, creating preprocessing components such as a tokenizer, lemmatizer, and normalizer using Python. Furthermore, I implemented and fine-tuned convolutional neural networks (CNNs) with residual connections for topic modeling and text classification tasks using PyTorch.