(647) 679-8032
torabian.alireza@gmail.com

Alireza Torabian

Machine Learning Engineer

1997alireza.github.io
1997alireza
alireza-torabian

As a machine learning researcher and software engineer with a strong math background, I have been involved with different projects in various areas including deep learning, computer vision, adversarial machine learning, and natural language processing. My current focus is on advancing large language models.

Experience

Huawei Technologies – R&D software engineer — Toronto, Canada — Jun. 2023 – Present

- Developed and optimized LLVM IR passes for loop transformations and vectorization in the BiSheng compiler.
- Contributed to upstream LLVM, Clang, and Flang with patches improving numerical workloads and language support.
- Analyzed and optimized scientific and ML workloads, ensuring SIMD utilization and memory efficiency.

York University – Machine learning researcher — Toronto, Canada — Jan. 2021 – Aug. 2023

- Established a structured and formal framework for machine learning calibration.
- Introduced a metric for calibration to achieve a less-biased evaluation.
- Published at XAI 2024.

Amirkabir University of Technology, NLP Lab – NLP researcher — Tehran, Iran — Jan. 2020 – Aug. 2020

- Implemented a real-time question-answering system based on a knowledge base in Python.
- A sequence-tagging model based on BERT is used for named entity recognition.
- SVM and CNN classification models used to classify questions achieved 96% accuracy and F1-score of 92.7%.

National University of Singapore, Trustworthy Machine Learning Lab – Computer vision researcher — Singapore — Jul. 2019 – Sep. 2019

- Designed a plugin that obscures images for increased privacy using adversarial attacks, with a 35% success rate.
- Performed facial recognition attack on FaceNet and face detection attack on SSD MobileNet V1 using PGD.
- Used image augmentations to attack black-box models increased success rate by 1.5x.

Diaalog – Deep learning R&D intern — Tehran, Iran — Jul. 2018 – Dec. 2018

- Developed a Persian chatbot using Python Tensorflow.
- Expanded dataset by clustering questions with LDA and using answers interchangeably.
- LSTM Seq2Seq model with Luong-style attention mechanism is used to generate answers.

Amirkabir University of Technology, Cognitive Robotics Lab – Research assistant — Tehran, Iran — Oct. 2016 – Sep. 2017

- Engineered an autonomous exploration algorithm for robots that won 2nd place in RoboCup 2017.
- Object detection task performed to detect victims using YOLO model achieved 99.7% accuracy.

Publications and Projects

Alireza Torabian, Ruth Urner, "Investigating Calibrated Classification Scores Through the Lens of Interpretability", **XAI 2024**, Malta

Finetuning LLM models

- Instruction-tuning Llama-2 using QLoRA to improve performance in instruction-following tasks. *Github repo*

Retrieval-Augmented Generation (RAG)

- Implemented a various RAG techniques including RAPTOR, Self RAG, Corrective RAG, Graph RAG, and Adaptive Retrieval.
- Frameworks: LangChain and LlamaIndex.

Alternative Actor and Co-Star Suggestion Using a Graph Autoencoder Model

- Applied graph autoencoder LoNGAE to actor network using Keras, achieving 99.46% accuracy in reconstructing the graph. *Github repo*
- An alternative actor is found by searching the latent space using a K-d tree.
- A co-star is suggested according to the predicted weights from the autoencoder model.

Skills

Languages: Python, Java, C, C++, JavaScript
ML Tools: TensorFlow, PyTorch, OpenCV, PEFT, LangChain, Hugging Face, Transformers, Keras, Numpy, Pandas, Scikit-learn, JAX
Other Tools: CUDA, Git, Unix shell, Jupyter

Concepts: GenAI, NLP, LLM, RAG, Computer Vision, Pattern Recognition, Object Detection, Data Augmentation, Adversarial Attack, Unsupervised Learning, GNN
Math: ML Theory, Stats & Prob, Signal Proc., Stochastic Processes, Convex Optimization, Quantum Computing

Education

York University — Toronto, Canada — Jan. 2021 – Aug. 2023
M.Sc. in Computer Science, *GPA: A+*

Amirkabir University of Technology (Tehran Polytechnic) — Tehran, Iran — Sep. 2015 – Aug. 2020
B.Sc. in Computer Engineering, *GPA: 3.9/4*