

Romina Oji

Department of Computer and Information Science (IDA), Linköping University, Linköping, Sweden
+46 13 28 13 13 | romina.oji@liu.se | github.com/rominaoji | linkedin.com/rominaoji | rominaoji.github.io/

Education

Linköping University

PhD in Computer Science

Linköping, Sweden

Aug 2024 – Aug 2029

- **Thesis:** TrustLLM: Democratizing trustworthy and factual large language model technology for Europe
- **Thesis Advisor:** Professor Marco Kuhlmann [↗](#) and Dr. Marcel Bollmann [↗](#)

University of Tehran

M.Sc. in Computer Engineering

Tehran, Iran

Sep 2019 – Sep 2022

- **GPA:** 17.91/20 - 3.77/4 (Overall)
- **Thesis:** Non-Word and Real-Word Error Detection and Correction Using Deep Learning
- **Thesis Advisor:** Dr. Hesham Faili [↗](#)

Dr. Shariaty Technical and Vocational University

B.Sc. in Computer Engineering

Tehran, Iran

Sep 2014 – Jul 2018

- **GPA:** 18.77/20 - 3.87/4 (Overall)
- **Ranked 2nd** in cumulative GPA among all entrants of 2014, software engineering, computer engineering group

Publications

- **Oji R**, Dousti M.J, Faili H. Using a Pre-Trained Language Model for Context-Aware Error Detector and Corrector in Persian language [↗](#)
- Vali A , Razavi SZ, Borhanifard Z, **Oji R**. Cost-Effective Development of Custom Wake Word Detection Models for Low-Resource Languages in Embedded Devices [↗](#)
- **Oji R**, Taghizadeh N, Faili H. PerSpellData: An Exhaustive Parallel Spell Dataset For Persian. NSURL 2021. 2021:24. [↗](#)
- **Oji R**, Razavi SF, Dehsorkh SA, Hariri A, Asheri H, Hosseini R. ParsiNorm: A Persian Toolkit for Speech Processing Normalization. In 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS) 2021 Dec 29 (pp. 1-5). IEEE. [↗](#)

Research Experience

Context Based Spell Correction Using Deep Learning

Jan 2020 - Sep 2022

Advisor: Dr. Hesham Faili [↗](#)

- Generating parallel dataset (word-level and sentence-level) of all types of errors in Persian language
- contextualized BERT-based model for error detection
- Unsupervised contextualized BERT-based model for error correction

Conversational Call Center

Mar 2021 - Jan 2023

Advisor: Dr. Reshad Hosseini [↗](#)

- A Speech processing tool called ParsiNorm
- Implementing and training a new BERT model to be used in speech applications
- A Persian phonemizer which gets sentences or words and returns standard phones
- Text-based and voice-based hate and dissatisfaction detection of a call center calls

Honors And Awards

- **Best Paper Award**, 7th conference on Signal Processing and Intelligent Systems (**ICSPIS 2021**) 2021
- **Ranked 2nd** in cumulative GPA among all entrants of 2014, **software engineering, computer engineering group**, among more than 30 students, Dr. Shariaty University, Tehran, Iran. 2018
- Head leader of **open-source association** in Shariaty university. 2016

Work Experience

Ussistant(Diginext) [↗](#)

Jan 2023 - Oct 2023

Natural Language Processing Researcher and Engineer

- Developing chatbots using LLMs with LangChain
- Crawling websites and generating FAQ Bot using LLMS
- Research and develop different word-embedding for extracting the most similar text to a Sentence
- Research and develop a Persian text-to-speech model
- Chit-chat Part of a smart assistant(speaker)

Hara AI [↗](#)

Mar 2021 - Jan 2023

Machine Learning Researcher

- Research and develop a call center using text-to-speech and speech-to-text models
- A chat-bot for flight reservation company

Vira Afzar Adon

Mar 2021 - Sep 2022

Machine Learning Researcher

- Developing commercial persian spell checker based on edit distance and neural language model
- User based error suggestion and correction

Rayan Eghtesad Novin Bank

Sep 2017 - Aug 2018

Full-stack web developer

- Developing back-end and UI of a website for presenting investment funds
- Developing mobile app

Online Courses

- Generative AI with Large Language Models, DeepLearning.AI [↗](#) Aug 2023
- Deep Learning Specialization(5 Courses), DeepLearning.AI [↗](#) July 2023
- Machine Learning, Coursera [↗](#) May 2022

Projects

Method Name Prediction

Jan 2024

Python, PyTorch, Hugging Face

- Finetuned CodeT5 models from Hugging Face on a dataset of Python Numpy functions, sourced from the project's GitHub repository, to accurately predict method names based on their function body code.)

Voice Emotion Classification

Jun 2022

Python, PyTorch, Pandas

- Fine tuning YamNet for emotion classification on Shemo(Sharif Emotional Dataset)

Hate Speech Recognition

April 2022

Python, PyTorch, Pandas

- Dataset gathering based on swear words in Twitter and OpenSubtitle dataset
- Designing and fine tuning a BERT model for hate detection

Textual Entailment

Feb 2022

Python, PyTorch, Pandas

- Fine tuning different BERT models(monolingual and multilingual) on ParsiNLU and SNLI entailment dataset

Real-time Message Analysis Using Telegram Data

Jul 2021

Python, MongoDB, Apache Cassandra, Elasticsearch, Apache Spark, Apache Kafka, Docker, Pandas

- Real-time data collecting from different Telegram channels and preprocessing gathered data, and extracting hashtags and key-words.
- Using Apache Cassandra and Elasticsearch for data storage and retrieval.
- Using SparkML to predict next post's time of a channel and predict hashtags based on written message.

A descriptive project to analyse Iranian currency changes

Dec 2020

Python, Matplotlib, Seaborn, Beautiful Soup, Pandas, NumPy

- Data gathering based on scraping the central Bank of Iran website
- Performing explanatory data analysis, visualizing data from different aspects.

Implementing Different GAN Models

Jun 2020

Python, TensorFlow, Pandas, NumPy

- Implementing diverse GAN models based on their paper implementations such as DCGAN, Cycle GAN, Conditional GAN.

Traffic Sign Recognition

May 2020

Python, Keras, Pandas, NumPy, Matplotlib

- Designing and Implementing a CNN-based Model for Traffic Sign Recognition

Teaching Experience

Trustworthy and Interpretable AI

Winter 2023

Teaching Assistant | Instructor: Dr. Mohammad Amin Sadeghi [↗](#) and Dr. Mostafa Tavassolipour [↗](#) University of Tehran

Big Data

Winter 2022

Head Teaching Assistant | Instructor: Dr. Mohammad Javad Dousti [↗](#) University of Tehran

Deep Learning with Applications

Winter 2022

Teaching Assistant | Instructor: Dr. Reshad Hosseini [↗](#) University of Tehran

Neural Network and Deep Learning

Fall 2021, Winter 2021, Fall 2020

Head Teaching Assistant | Instructor: Dr. Ahmad Kalhor [↗](#) University of Tehran

Natural Language Processing

Winter 2021 - Winter 2023

Teaching Assistant | Instructor: Dr. Hesham Faili [↗](#) and Dr. Yadollah Yaghoobzadeh [↗](#) University of Tehran

Fundamental Models in NLP

Fall 2021

Teaching Assistant | Instructor: Dr. Yadollah Yaghoobzadeh [↗](#) University of Tehran

Statistical Inference

Fall 2021, Winter 2021-2023

Teaching Assistant | Instructor: Dr. Behnam Bahrak [↗](#) and Dr. Abdol-hossein Vahabie [↗](#) University of Tehran

Intelligence Information Retrieval

Fall 2020

Teaching Assistant | Instructor: Dr. Azadeh Shakeri [↗](#) University of Tehran

Technical Skills

Programming Languages

Python, R, Bash

Machine Learning Frameworks, Libraries, and Tools

PyTorch, Keras, TensorFlow, Scikit-learn, NumPy, Pandas, NLTK, SparkML, Fairseq, OpenNMT, MLflow

Databases, Datastores, and Tools

MongoDB, Apache Cassandra, Elasticsearch, Apache HBase, ClickHouse, SQL SERVER

Big Data Platforms and Tools

Apache Spark, Apache Kafka, Apache Hadoop, Neo4j

Visualization Tools

Matplotlib, Seaborn, Plotly, Bokeh, Geoplotlib

CI/CD Tools

Git, Gitflow

Language Skills

English (TOEFL Overall:91, Reading:23, Listening:22, Speaking:22, Writing: 24), **Persian** (Native)