Romina Oji

Department of Electrical and Computer Engineering, University of Tehran, AmirAbad, Tehran, Iran

+98 919 122 5933 | romina.oji.ro@gmail.com | github.com/rominaoji | linkedin.com/romina-oji

EDUCATION

University of Tehran
Tehran, Iran

M.Sc. Student in Computer Engineering
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. Heshaam Faili ↗

Dr. Shariaty Technical and Vocational University
Tehran, Iran

B.Sc. in Computer Engineering
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. Virastman: A Persian Spell and Contextual Error Detection and Correction. (To Be Submitted TALLIP 2022)
- **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. In2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS) 2021 Dec 29 (pp. 1-5). IEEE.

RESEARCH INTERESTS

Natural Language Processing, Computer Vision, Deep Learning, Theory of Machine Learning, Neuroscience

RESEARCH EXPERIENCE

Context Based Spell Correction Using Deep Learning
Jan 2020 - Present

Advisor: Dr. Heshaam Faili ↗

- 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 - Present

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*
- **Ranked 101th** Among More Than 12,000 Participants in the Nationwide University Entrance Exam Known as Konkur for M.Sc Degree in Computer Engineering — *2019*
- **Ranked Top 2 Percent** Among More Than 250,000 Participants in the Nationwide University Entrance Exam Known as Konkur for B.Sc Degree in Computer Engineering   . — *2014*

Work Experience

Hara ⬈   *Mar 2021 - Present*
Machine Learning Engineer
- Research and Develop a Call Center using Text-to-Speech and Speech-to-Text Models
- Train Persian Bert
- Persian Text-to-Speech Normalization
- A Chat-Bot for Flight Reservation Company
- Hate Speech Detection
- Persian Phonemizer

Vira Adon Afzar   *Mar 2021 - Sep 2022*
Machine Learning Researcher
- Developing Commercial Persian Spell Checker Based on Edit Distance and Language Model
- Defining Rules to Auto-Complete Errors
- Defining Rules to Reduce the Count of Suggestions of Wrong Written Words and Showing high Probable Correct Forms of It
- 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

Teaching Experience

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 2022, Winter 2021
Teaching Assistant | Instructor: Dr. Heshaam 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
Teaching Assistant | Instructor: Dr. Behnam Bahrak ⬈   *University of Tehran*

Intelligence Information Retrieval   Fall 2020
Teaching Assistant | Instructor: Dr. Azadeh Shakery ⬈   *University of Tehran*

Projects

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

Chatbot for flight Reservation   *Sep 2021*
Python, PyTorch, Pandas, Rasa
- Pre-Processing ATIS Dataset by Data Balancing
- A NLU model which Defines Slots and Fill the Correct Slots in the Template Sentences to Answer the Questions of Customers

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.

Traffic Sign Recognition
May 2020

Python, Keras, Pandas, NumPy, Matplotlib
- Designing and Implementing a CNN-based Model for Traffic Sign Recognition

Sentiment Analysis and Email Spam Detection
Apr 2020

Python, Keras, Pandas, NumPy, NLTK
- Designing, Training and Fine Tuning a Deep Learning Model, Which Uses Bert and Elmo as Word Embedding.
- Designing and Training Naive Bayes and Logistic Regression Models For Sentiment and Spam Detection.

TECHNICAL SKILLS

Programming Languages

Python, R, Bash, Java, C++

Machine Learning Frameworks, Libraries, and Tools

Pytorch, Keras, TensorFlow, Sckit-learn, NumPy, Pandas, NLTK, SparkML, Fairseq, OpenNMT, MLflow, Rasa

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

Container Platforms

Docker

Operating Systems

Ubuntu, Windows

Other Tools and Skills

LATEX, MS Office, Web Scraping

LANGUAGE SKILLS

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

REFERENCES

References, Further information, and Proofs are available upon Request