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 – June 2022 (Expected)*
- **GPA:** 17.91/20 - 3.77/4 (Overall)
- **Thesis:** Real-Word Error Detection and Correction
- **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

Hazrate Zahra High School   Tehran, Iran
Diploma in Mathematics and Physics Discipline   *2010 - 2014*

PUBLICATIONS

- **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, Big Data Analytics, Computer Vision, Deep Learning, Theory of Machine Learning

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**.   *2016-2017*
- **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*

RESEARCH EXPERIENCE

Real-Word (Context Based) Spell Correction Using Deep Learning   *Jan 2019 - Present*
Advisor: Dr. Heshaam Faili ⬈
- Using logs of user error correction of the Virastar to generate an exhaustive dataset that contains real-Word and non-word errors and covering all types of spelling errors that can happen in Persian.
- Train a machine learning model to correct non-word errors using logistic regression with selected features.
- Train a Deep learning model using a neural language model for context-sensitive error detection and correction in the Persian language.

Training Modified Bert for the Persian Language   *Mar 2021 - Present*
Advisor: Dr. Reshad Hosseini ⬈
- Data gathering from different resources and reprocess each data based on their properties
- Implement a new method to improve the quality of Bert in Persian that we can use in speech applications using acoustic models
- Using Bert in different Persian tasks such as POS tagger

Work Experience

Hara ⬈ *March 2021 - Present*

Machine Learning Engineer
- Research and Develop a call center using Text-to-Speech and Speech-to-Text
- Train Persian Bert
- Persian Text-to-Speech Normalization

Vira Adon Soft *March 2021 - Present*

Machine Learning Engineer
- 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
- Improving performance of the spell checker based on user data logs which are highly frequent

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 ⬈ *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

Real-time Message Analysis Using Telegram Data *Jul 2021*

Python, MongoDB, Apache Cassandra, Elasticsearch, Apache Spark, Apache Kafka, Docker, Pandas
- Collect data from Telegram channels and preprocessing the data, and extracting hashtags and key-words.
- Using Apache Cassandra and Elasticsearch for Data storage and retrieval.
- Using SparkML to generate time prediction model and logistic regression model to predict hashtags in the messages.

A Descriptive Project to Analyse Iranian Currency Changes *Dec 2020*

Python, Matplotlib, Seaborn, Beautiful Soup, Pandas, NumPy
- Gathered data from the Central Bank of Iran website, perform 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.

Part-of-Speech tagger   *Mar 2020*

Python, Keras, Pandas, NumPy, Matplotlib, scikit-learn, NLTK
- Real-time data-intensive face recognition platform.
- Analyze offline/online video streams (e.g., Youtube) and tag videos by participants in videos.

Sentiment Analysis and Email Spam Detection   *Apr 2020*

Python, Keras, Pandas, NumPy, NLTK, Keras
- Designing and training a deep learning model, which uses Bert and Elmo as word embedding and fine tuning.
- Designing and training Naive Bayes and Logistic Regression models to predict each sentence sentiment.

Predicting Air Pollution   *Apr 2020*

Python, Keras, Pandas
- Time-series prediction using RNN neural network

Paradigmatic and Syntagmatic word Association   *Nov 2019*

Python, NLTK, scikit-learn

SELECTED COURSES

- Natural Language Processing (19.4/20)
- Neural Networks (18.44/20)
- Massive Data Analysis and Systems (19.6/20)
- Statistical Inference (17.3/20)
- Data Analysis (18.75/20)
- Discrete Mathematics (19.5/20)
- Advanced Programming (18/20)
- Database (19.25)
- Artificial Systems Expert Systems (19.4/20)
- Operating System (19/20)
- Algorithm Design (19/20)
- Data Structure (19/20)
- Microprocessors and Assembly Language (19.25/20)
- Principals of Compiler Design (20/20)
- Machine Language Theory (20/20)
- Logic Circuits (19.75/20)
- Computer Networks (18/20)
- Engineering Mathematics (19.5/20)

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

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

Container Platforms
Docker

Operating Systems
Ubuntu, Windows

Other Tools and Skills
LATEX, MS Office, Web Scraping

LANGUAGE SKILLS

English (TOEFL Expected at Jan 2022), **Persian** (Native)

REFERENCES

References, Further information, and Proofs are available upon Request