Sepehr Rafiei

Personal info

Name: Sepehr Rafiei
Email: sepehr.rafiei.k@ut.ac.ir - sep.rafiei.k@gmail.com
Phone number: (98) 9372477115
Github Linkedin Personal Website Google Scholar

Education

TU Dortmund 2024 - Present
Graduate student of M.Sc. in Automation and Robotics

B.Sc. University of Tehran 2017 - 2022
Undergraduate student of B.Sc. in Software Engineering
Total GPA: 17.83/20(3.80/4) **Last 2 years: 18.77/20(3.96/4)** Rank in **top 15%** among more than 100 student

- Thesis: Presenting a quality estimation model of English to Farsi machine translator using transfer learning. Under the supervision of Prof Heshaam Faili.

Shahid Ejei (NODET) 2013 - 2017
Diploma in Mathematics and Physics Discipline
NODET (National Organization for Development of Exceptional Talents)
GPA: 19.41/20

Research Interests

- Natural Language Processing
- Contrastive Learning
- Deep Learning
- Cognitive Science
- Machine Learning
- Information Retrieval

Related Course

- Introduction to the computing system and programming(20/20)
- Parallel Computing (19/20)
- Design and analysis of algorithms(20/20)
- Artificial Intelligence(20/20)
- Computer Networks (19/20)
- Advanced Graph Algorithm(17.3/20)
- Formal Language and Automata (18.7/20)

Publications

Evaluating the Preservation of Linguistic Cues in Synthetic Data Generation by Large Language Models for Cognitive Impairment Detection
Ali Zolnour, Hossein AzadMaleki, Masoud Khani, AmirSajjad Taleban, James Noble, Farid Hosseini, Samin Mahdizadeh, Sepehr Rafiei, Abdol-Hossein Vahabie, Yadollah Yaghoobzadeh, Maryam Zolnoori. Submitted to the Journal of the American Medical Informatics Association: JAMIA

Speaker type identification in audio recorded patient-nurse verbal communication in clinical setting
Maryam Zolnoori, Sasha Vergez, Sridevi Sridharan, Ali Zolnour, Sepehr Rafiei, Kathryn Bowles, Zoran Kostic, Maxim Topaz. "Speaker type identification in audio recorded patient-nurse verbal communication in a clinical setting." AMIA Annual Conference, New Orleans, LA. [Podium presentation] (2023).

Clustering of Urban Traffic Patterns by K-Means and Dynamic Time Warping: Case Study
Sadegh Etemad, Raziyeh Mosayebi, Tadeh Alexani Khodavirdian, Elahe Dastan, Amir Salari Telmadarreh, Mohammadreza Jafari, Sepehr Rafiei. "Clustering of Urban Traffic Patterns by K-Means and Dynamic Time Warping: Case Study." International Conference on Artificial Intelligence and Smart Vehicle (2023).

Presenting a quality estimation model of English to Farsi machine translator using transfer learning
Jafari Harandi, Mohammad Hossein, Fateme Azadi, Sepehr Rafiei, Hesham Faili, and Mohammad Javad Dousti. "Presenting a quality estimation model of English to Farsi machine translator using transfer learning." Language and Linguistics 18, no. 35 (2022): 71-86.

Research Experience

Research Assistant at Columbia School of Nursing 2022 - present

- Working on Data Augmentation and Text-Generation employing LLM for clinical data set focused on Alzheimer's data. Under the supervision of **Maryam Zolnoori**.

Research Assistant at NLP Lab University of Tehran 2022 - present

- Working on Quality Estimation (QE) of Machine Translation from English to Persian.Under the supervision of **Heshaam Faili**.

Researcher at Snapp Data Science Chapter 2022 - present

- Working on clustering and analyzing traffic pattern time series data using Dynamic Time Warping (DTW) and K-Means methods and using these clusters in different map applications. Under the supervision of **Sadegh Etemad**.

Selected Work experience

Snapp! - Data Scientist July 2021 - July 2024
Data Science Chapter-Map

- In Snapp!, I was a member of a highly talented, vibrant team that develops searching services (Geo-coding) for maps.
- In the Address Experience team we were responsible for developing a search service for maps that tries to recommend the most relevant places to users' queries using different **NLP approaches**, and **LTR models like LambdaMart.**
- Worked on **Named-Entity-Recognition** service using **Transformer-based models** and **pre-trained language models LLM** used for address tokenization and entity tagging

- Design and implement a new evaluation framework for search **Information Retrieval (IR) metrics** (like MRR, MAP) which reduces the time of A/B testing by a large margin.

Snapp! - Software Search Engineer Dec 2020 - July 2024
Map Team-Address Experience

- Design and implement **Map Search (Geocoding)** core services to respond to highly concurrent user requests with low response time and low resource usage.
- Working on large-scale stream processing service with Apache Flink designed to aggregate and transform service logs and also calculate IR evaluation metrics for geocoding service.
- Worked on IR query weight optimization using **linear programming**.

Alibaba Travels - Software Engineer July 2020 - Dec 2020
R&D Team

- In the R&D team we were responsible for a highly efficient Rate-limiting software named Zebel which tries to analyze users' traffic with **statistical approaches** and **regular user traffic trends** block most web scrapers, and DDoS attacks and reduces the traffic by a large margin.

Honors

Top 15% Ranking Univesity of Tehran 2021
*Ranked in the **top 15%** among over 100 Computer Engineering students. Awarded a scholarship by the Supporter Foundation.*

Best Impact Project AI2 (Allen Institute) 2021
Best NLP Hackathon Impact project focused on toxic comment detection using weak supervision labeling

FOE Award - First year highest GPA 2017
*Certificate of **First year Highest GPA** in my major between more than 60 students at the University Of Tehran*

Nationwide University Entrance Exam 2018
*Rank in **top 1%** among more than 140,000 participant*

Participated in the second stage of Physics Olympiad 2017
Participated in the second stage of nationwide Physics Olympiad exam

Snapp Go-up Programming Contest 2021
***Rank 1st** in Snapp GO programming contest among more than 900 participants*

Notable Projects

Hermes - Graph Database
Snapp! Address Experience Team

Create a graph database for Geo-spatial data designed to do Shortest Path algorithm on nodes by perfoming **A*(Dijkstra's algorithm with heuristic)**.

Diagnosing COVID19 and Penumonia using Neural Networks
Artificial Intelligence Course, University of Tehran

Developed a **CNN** neural network to diagnose COVID-19 and Pneumonia using CT scans images.

Personal skills

Programming Languages	Experienced in Python, C/C++, Golang Working knowledge in Java, Bash-Scripting, Verilog.
Software Engineering	Experienced in different object-oriented design patterns, software development methodologies such as Agile and DevOps.
Tools	Experienced in Docker, Gitlab CI/CD, Redis(+streams), Kubernetes, Prometheus(+grafana), Elasticsearch(ELK stack), Kafka, G-RPC.
Languages	Persian (native) English Full professional proficiency TOEFL: 100