Mohammad Arvan

📍 Chicago, Illinois (IL) • ✉ mo.arvan@gmail.com
➤ mo-arvan.github.io • in linkedin.com/in/mo-arvan • ⌂ github.com/mo-arvan

EDUCATION

Ph.D., Computer Science
08/2018 - present | Chicago, IL

University of Illinois at Chicago (UIC)
- **Proposal**: Machine Learning and Open Science: On Risks and Challenges
- **GPA**: 4.0/4.0
- **Expected Graduation**: *Early 2024*
- **Courses**: Machine Learning, Natural Language Processing, Algorithmic Fairness, Language and Vision, Computer Algorithms

B.Sc., Software Engineering
08/2012 - 08/2016 | Qazvin, Iran

Qazvin Islamic Azad University (QIAU)
- Ranked 2nd in class of 180 students

PROFESSIONAL EXPERIENCE

Research Assistant
08/2018 – present | Chicago, IL

UIC Natural Language Processing Laboratory, Supervisor: Dr. Natalie Parde, Assistant Professor
- Identified and reported bugs and errors in the automatic and human evaluations of scientific papers by conducting 10 reproducibility studies. These issues encompass data leakage, unsupported claims and results, as well as flawed experimental design.
- Utilized and promoted best practices for improving open science values such as reproducibility, rigorous evaluation, and transparency in research by underlining the importance of the availability of self-contained research artifacts and the use of statistical tests to evaluate the significance of results.
- Led research, development, and evaluation of an efficient clinical information extractor pipeline from unstructured clinical notes using Llama-2 and other large language models. This pipeline was utilized for detecting houseless population, cardio-oncology, cervical cancer screening, and COVID-19 symptom analysis at the University of Illinois Hospital.

Data Engineer
08/2016 – 08/2018 | Tehran, Iran

Kara Intelligent System
- Led the extract, transform, and load (ETL) process for over 100 GB of data, establishing a high-performance warehouse for analysis with a 90% reduction in processing time, achieved through parallel processing and data streaming. The data was utilized for fault detection, diagnosis, quality control, and user experience improvement of the largest telecom company in Iran, with 20 million users.
- Optimized indexes and queries, resulting in a 70% reduction in data retrieval time from the tabular database, significantly improving overall system performance and responsiveness.
- Developed interactive visualization dashboards using Tableau and ASP.NET Core, enabling data-driven decision-making and empowering stakeholders with actionable insights.

Co-Founder and Developer
08/2015 – 08/2016 | Qazvin, Iran

Indooria Startup
- Pioneered the design and development of an innovative Android application for indoor navigation and localization utilizing Bluetooth beacons, leading to an enhanced user navigation experience.
- Implemented a pathfinding algorithm to enable users to find the shortest routes between two points on the map.
- Leveraged techniques for outlier detection of wireless signals and utilized Kalman filter to further improve localization accuracy, ensuring precise indoor positioning for users.

Research Assistant
08/2013 – 08/2016 | Qazvin, Iran

QIAU Mechatronic Research Laboratory, Supervisor: Dr. Farshid Najafi
- Implemented exploration, motion planning, obstacle avoidance, and behavior control, enabling autonomous navigation of a mobile robot in unstructured unknown indoor environments. This work achieved 1st place in the Rescue Robot League in Robocup 2015 competitions.
- Developed and evaluated a highly accurate classifier using AdaBoost and engineered features for detecting humans and other objects from 2-D range scans.

- Optimized code base for efficient execution on edge devices with limited computational resources, allowing for real-time object detection in resource-constrained environments

SKILLS

Languages

Python, C++, SQL, Octave/MATLAB, Java, C#

Libraries and Technologies

PyTorch, NumPy, Pandas, Huggingface Transformers, PyTorch Lightning, Numba, TensorFlow, Keras, OpenCV, scikit-learn, SciPy, Matplotlib, Spacy, NLTK, Docker, AWS, GCP, Azure, Linux/Unix, ssh, Git

PUBLICATIONS

- **Mohammad Arvan**, Natalie Parde. *Human Evaluation Reproduction Report for Data-to-text Generation with Macro Planning.* The 3rd Workshop on Human Evaluation of NLP Systems (**HumEval 2023**).
- **Mohammad Arvan**, A. Seza Doğruöz, Natalie Parde. *Investigating Reproducibility at Interspeech Conferences: A Longitudinal and Comparative Perspective*. The 24th INTERSPEECH Conference (**INTERSPEECH 2023**)
- Maja Popović, **Mohammad Arvan**, Natalie Parde, Anya Belz. *Exploring Variation of Results from Different Experimental Conditions*. The Findings of 61st Annual Meeting of the Association for Computational Linguistics (**ACL Findings 2023**)
- **Mohammad Arvan**, Mina Valizadeh, Parian Haghighat, Toan Nguyen, Heejin Jeong, Natalie Parde. *Linguistic Cognitive Load Analysis on Dialogues with an Intelligent Virtual Assistant*. The 45th Annual Meeting of the Cognitive Science Society (**CogSci 2023**)
- Anya Belza, Craig Thomson, Ehud Reiter, Gavin Abercrombie, Jose M. Alonso-Moral, **Mohammad Arvan**, Jackie Cheung, Mark Cieliebak, Elizabeth Clark, Kees van Deemter, Tanvi Dinkar, Ondrej Dušek, Steffen Eger, Qixiang Fang, Albert Gatt, Dimitra Gkatzia, Javier González-Corbelle, Dirk Hovy, Manuela Hürlimann, Takumi Ito, Emiel van Miltenburg, Chris van der Lee, John D. Kelleher, Filip Klubicka, Saad Mahamood, Margot Mieskes, Malvina Nissim, Natalie Parde, Ondrej Plátek, Verena Rieser, Pablo Mosteiro Romero, Joel Tetreault, Xiaojun Wan, Leo Wanner, Lewis Watson, Diyi Yang. *Missing Information, Unresponsive Authors, Experimental Flaws: The Impossibility of Assessing the Reproducibility of Previous Human Evaluations in NLP*. The Forth Workshop on Insights from Negative Results in NLP (**Insights 2023**)
- **Mohammad Arvan**, Luís Pina, Natalie Parde. *Reproducibility in Computational Linguistics: Is Source Code Enough?* The 2022 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2022**)
- Parian Haghighat, Toan Nguyen, Mina Valizadeh, **Mohammad Arvan**, Natalie Parde, Myunghee Kim, Heejin Jeong. *Effects of an Intelligent Virtual Assistant on Office Task Performance and Workload in a Noisy Environment*. **Applied Ergonomics**, 109, 103969.
- Parian Haghighat, Toan Nguyen, Mina Valizadeh, **Mohammad Arvan**, Natalie Parde, Myunghee Kim, and Heejin Jeong. *Human Interaction with Intelligent Virtual Assistant in a Noisy Environment*. The 66th Proceedings of the Human Factors and Ergonomics Society Annual Meeting (**HFES 66th**)
- **Mohammad Arvan**, Luís Pina, Natalie Parde. *Reproducibility of Exploring Neural Text Simplification Models: A Review.* The 15th International Natural Language Generation Conference (**INLG 2022**)
- Farshid Najafi, Mehdi Dadvar, Soheil Habibian, Alireza Hosseini, Hossein Haeri, **Mohammad Arvan**, Behzad Peykari, Hamed Bagheri. *RoboCup Rescue 2016 Team Description Paper MRL*. Robocup Rescue 2016 TDP Collection

HONORS AND AWARDS

- Recipient of the 2023 **Diversity and Inclusion Conference Award**. The Cognitive Science Society ($1000)
- Recipient of the 2020 **Provost's Graduate Research Award.** University of Illinois at Chicago ($5000)
- Ranked 3rd, **RoboCup World Championship**, Rescue Robot League. Nagoya, Japan, 2017
- Ranked 2nd, **RoboCup World Championship**, Rescue Robot League. Leipzig, Germany, 2016
- Innovative User Interface Award, **RoboCup World Championship**, Rescue Robot League. Hefei, China, 2015
- Ranked 1st, **RoboCup World Championship**, Rescue Robot League. Hefei, China, 2015
- Ranked 2nd, **RoboCup World Championship**, Rescue Robot League. João Pessoa, Brazil, 2014

COMMUNITY SERVICE

- Served as Evaluation and Resources track reviewer for **EMNLP 2023**