Mohammad Arvan

(312) 8682677 mo.arvan@gmail.com Chicago, IL mo-arvan.github.io linkedin.com/in/mo-arvan

Education

University of Illinois at Chicago (UIC) Expected Early 2024
Ph.D., Computer Science GPA: 4.0/4.0
- Thesis: Machine Learning and Open Science: On Risks and Challenges

Qazvin Islamic Azad University (QIAU) July 2016
B.Sc., Software Engineering Ranked 2nd in class of 180 students

Skills

- Libraries and Technologies: PyTorch, NumPy, Pandas, Huggingface Transformers, PyTorch Lightning, Numba, TensorFlow, Keras, JAX, scikit-learn, SciPy, Streamlit, Matplotlib, Spacy, NLTK, Docker, Kubernetes, AWS, GCP, Azure, Linux/Unix, ssh, Git

- Languages: Python, C++, SQL, Octave/MATLAB, C#, Java, R

Research Experience

Research Assistant Aug 2018 to Present
UIC Natural Language Processing Laboratory Chicago, IL

- Implemented statistical analysis and containerization best practices in machine learning, leading the adoption of MLOps methodologies to enhance project workflows, ensure scientific report reproducibility, and improve evaluation and reporting processes.
- Conducted comprehensive reviews and analyses of scholarly publications, meticulously identifying discrepancies in source code, data collection, and evaluation methodologies. Uncovered and addressed critical issues, including data leakage, unsupported claims, and flawed experimental designs.
- Pioneered research in applying Large Language Models (LLMs) to unstructured clinical notes, establishing a robust pipeline for detecting houseless populations, cardio-oncology, cervical cancer screening, and COVID-19 symptom analysis at the University of Illinois Hospital.
- Designed a proof-of-concept for heterogeneous compute unit-based distributed training of neural networks in unreliable connections, showcasing a paradigm applicable to training extremely large networks through volunteer computing.
- Implemented, trained, and evaluated LSTM and transformer models for sequence processing across various tasks, such as language modeling, machine translation, and mathematical question answering.
- Attained an 80% accuracy rate in retrieval-augmented question answering on the SQuAD dataset through the implementation of transfer learning (BERT) and utilization of corpus-level information retrieval.
- Provided technical leadership and mentorship to junior team members, nurturing their development in data science and machine learning, evidenced by successful experiment executions and written reports.

Skills: Foundation Models, Generative AI, Distributed Training, Quantization, Transformer, Linear Algebra, Calculus, Probability, Deep Learning, Neural Networks, Fine-tuning, AI/ML Frameworks.

Research Assistant Aug 2013 to Aug 2015
Mechatronic Research Laboratory Iran

- Successfully implemented AI algorithms, including path planning, motion planning, and behavior control, in robotics, leading to a victory in an international robotic competition.
- Engineered features and assessed AdaBoost, SVM, Decision Trees, and Random Forests for classification, achieving a 90% accuracy rate in object detection.

Skills: Machine Learning, Computer Vision, Data Analysis, Problem Solving, Rapid Growth and Learning, C++, Artificial Intelligence

Work Experience

Data Engineer Aug 2016 to Aug 2018

Kara Intelligent System Iran

- Created interactive visualization dashboards using Tableau, translating data into actionable insights that empowered stakeholders to make informed, data-driven decisions and significantly enhanced operational processes.
- Accelerated data processing time by 70% through benchmarking, profilers, and debugging methods.
- Engaged in cross-functional collaborations to pinpoint key business and performance requirements, translating them into precise technical specifications and ensuring data accuracy and availability for analytics.

Skills: Data Visualization, Data Cleaning, API Development, Cloud Computing, Statistical Analysis.

Co-Founder and Developer Aug 2015 to Aug 2016

Indooria Startup Iran

- Headed the design and development of a highly scalable software application, achieving a 50% reduction in computational costs.
- Leveraged programming, mathematics, and statistics expertise to boost performance by 20%.
- Effectively communicated solutions to a diverse audience of both technical and non-technical stakeholders, ensuring comprehensive understanding and alignment across all involved parties.

Skills: Data Structures, Software Architecture, Object-Oriented Programming, Technical Leadership, Agile Development

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)

Honors and Awards

- Finalist in 2024 Three Minute Thesis Competition (3MT). University of Illinois at Chicago
- Recipient of the 2020 **Provost's Graduate Research Award.** University of Illinois at Chicago ($5000)

- Innovative User Interface Award, **RoboCup World Championship**, Rescue Robot League. Hefei, China, 2015
- Ranked 1st, **RoboCup World Championship**, Rescue Robot League. Hefei, China, 2015

Presentations

- M. Arvan, "Stochasticity in Deep Learning Evaluation: Challenges and Solutions," UIC NLP Lab Meeting, Oct 19, 2023.
- M. Arvan, "Challenges in Reproducibility of Human Evaluations in NLP," UIC NLP Lab Meeting, Sep 21, 2023.
- M. Arvan, "On Evaluation and Reproducibility," Invited presentation in UIC Natural Language Processing Course, Fall 2023.
- M. Arvan, "Machine Learning and Open Science: On Risks and Challenges," UIC NLP Lab Meeting, May 4, 2023.
- M. Arvan, "AI Democratization: Towards Training Billion Parameter Models Using The Power of the Crowd," UIC NLP Lab Meeting, Apr 19, 2022.
- M. Arvan, "Reproducibility: On Why and How," Invited presentation in UIC Natural Language Processing Course, Nov 2022.
- M. Arvan, "Reproducibility in Computational Linguistics: Is Source Code Enough?" UIC NLP Lab Meeting, Oct 21, 2022.
- M. Arvan, "Reproducibility in NLP Research: A Deep Dive into the Evaluation of Neural Text Simplification Models," UIC NLP Lab Meeting, Sep 16, 2022.

Community Service

- Served as Evaluation and Resources track reviewer for **EMNLP 2023**
- Served as reviewer for the ReScience **C** journal.

Certificate

- "Fundamentals of Accelerated Computing with CUDA C/C++," Issued by **NVIDIA Deep Learning Institute**, 2023