

Marzieh Fadaee

Amsterdam, North Holland

email: marzieh.f@gmail.com

URL: marziehf.github.io

PROFILES: [GOOGLE SCHOLAR](#), [GITHUB](#), [LINKEDIN](#)

Current position

NLP Research Lead, Zeta Alpha Vector

Working on designing and implementing an NLU-Powered Tool to explore knowledge and information in large amount of data.

Areas of Expertise

Machine Learning • Natural Language Understanding • Machine Translation • Information Retrieval • Distributional Semantics • Question Answering • Summarization • Named Entity Recognition • Word Sense Disambiguation • Information Extraction

Publications

- 2022 Luiz Bonifacio, Hugo Abonizio, Marzieh Fadaee, Rodrigo Nogueira. **InPars: Data Augmentation for Information Retrieval using Large Language Models**, *Conference of the Association for Computing Machinery Special Interest Group in Information Retrieval (ACM SIGIR)*. Madrid, Spain
- 2021 Luiz Henrique Bonifacio, Vitor Jeronymo, Hugo Queiroz Abonizio, Israel Campiotti, Marzieh Fadaee, Roberto Lotufo, Rodrigo Nogueira. **mMARCO: A Multilingual Version of the MS MARCO Passage Ranking Dataset**, *arXiv*.
- 2020 Marzieh Fadaee. **Understanding and enhancing the use of context for machine translation**, *PhD dissertation*.
- 2020 Marzieh Fadaee, Olga Gureenkova, Fernando Rejon-Barrera, Carsten Schnober, Wouter Weerkamp, Jakub Zavrel. **A New Neural Search and Insights Platform for Navigating and Organizing AI Research**, *Workshop on Scholarly Document Processing (SDP) at EMNLP*. Online
- 2020 Marzieh Fadaee, Christof Monz. **The Unreasonable Volatility of Neural Machine Translation Models**, *Workshop on Neural Generation and Translation (WNGT) at ACL*. Seattle, USA
- 2018 Marzieh Fadaee, Christof Monz. **Back-Translation Sampling by Targeting Difficult Words in Neural Machine Translation**, *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Brussels, Belgium
- 2018 Marzieh Fadaee, Arianna Bisazza, Christof Monz. **Examining the Tip of the Iceberg: A Data Set for Idiom Translation**, *International Conference on Language Resources and Evaluation (LREC)*. Miyazaki, Japan
- 2017 Marzieh Fadaee, Arianna Bisazza, Christof Monz. **Data Augmentation for Low-Resource Neural Machine Translation**, *Association for Computational Linguistics (ACL)*. Vancouver, Canada

- 2017 Marzieh Fadaee, Arianna Bisazza, Christof Monz. [Learning Topic-Sensitive Word Representations](#), *Association for Computational Linguistics (ACL)*. Vancouver, Canada
- 2013 Marzieh Fadaee, Hamidreza Ghader, Hesham Faili, Azadeh Shakery. [Automatic WordNet Construction Using Markov Chain Monte Carlo](#), *International Conference on Intelligent Text Processing and Computational Linguistics (CICLING)*, Samos, Greece.

Education

- 2014-2019 **PhD** in Artificial Intelligence, University of Amsterdam
Worked in Information and Language Processing Systems (ILPS) group on developing models to understand interesting phenomena of the natural language. My research includes exploiting paraphrase relations for improved translation modeling, adapting the embedding space to capture the ambiguity nature of the language using topics, translation of infrequent words with neural systems, and idiom translation.
- 2010-2013 **MSc** in Computer Engineering, University of Tehran
- 2004-2009 **BSc** in Computer Engineering, Sharif University of Technology

Experience

- 2020 - now **NLP Research Lead**, [Zeta Alpha Vector](#)
Leading the research team on designing and implementing an NLU-Powered Tool to explore knowledge and information in large amount of data.
- 2019 - 2020 **NLP/ML Research Engineer**, [Zeta Alpha Vector](#)
Working on designing and implementing an NLU-Powered Tool to explore knowledge and information in large amount of data.
- 2018 **Research Scientist Intern**, [Ebay CoreAI](#)
Worked on transfer learning by integrating universal contextualized embeddings for inter-domain and cross-lingual challenges.
- 2014-2019 **Research Assistant**, [University of Amsterdam](#)
Duties include modeling and implementation of modules of an in-house translation system, data preprocessing, running and analysis of experiments.
- 2013-2014 **Researcher and Developer**, [Faraazin Translation Project](#)
Worked on developing a Persian↔English translation system.
- 2010-2013 **Research Assistant**, University of Tehran
Worked in NLP Lab and IIS Lab on constructing a Semantic Network for the Persian language and improving retrieval results in Query Expansion and Cross-Lingual IR.
- 2007-2009 **Software Developer**, Iran Info-Tech Development Co.

Technical skills

PROGRAMMING LANGUAGES

Proficient in Java, Python, Perl, C#, and R
Familiar with C/C++, SQL, HTML, PHP

NLP PACKAGES

Moses, SRILM, Phrasal, word2vec, Stanford CoreNLP, ELMo, BERT, OpenNMT

MACHINE LEARNING

Hands-on experience with supervised (classification/regression algorithms) and unsupervised (clustering, language modeling) techniques. Experience scikit-learn. Deep learning using Torch and PyTorch on GPUs. Probabilistic models.

Teaching & supervision

2019-2020	Sergi Castella Sapé. <i>MSc AI University of Amsterdam</i> . Topic: Semantic Representations in Language Models. Co-supervised with Jakub Zavrel
2019-2020	Ioannis Tsiamas. <i>MSc AI University of Amsterdam</i> . Topic: Text summarization. Co-supervised with Jakub Zavrel
2019-2020	Raggar Jonkers. <i>MSc AI University of Amsterdam</i> . Topic: Extracting semantic patterns from scientific papers. Co-supervised with Jakub Zavrel
2017-2018	<i>University of Amsterdam</i> . Applied Language Technology Course. Masters Artificial Intelligence
2016-2017	<i>University of Amsterdam</i> . Applied Language Technology Course. Masters Artificial Intelligence
2015-2016	<i>University of Amsterdam</i> . Data Mining Course. Bachelors Information Science
2013-2014	<i>University of Tehran</i> . Natural Language Processing Course. Masters Artificial Intelligence
2012-2013	<i>University of Tehran</i> . Pattern Recognition Course. Masters Artificial Intelligence
2011-2012	<i>University of Tehran</i> . Advance Algorithm Course. Masters Artificial Intelligence

Reviewer / program committee

2022	ICLR, SustaiNLP, ACL rolling review, ASOC, IEEE TAI, ICML, ECNLP, NeurIPS
2021	ICLR, EACL, NAACL, IKT, ACL-IJCNLP, NeurIPS, ACL rolling review, SustaiNLP
2020	ACL, EMNLP, ACL-IJCNLP, LREC, NLP-OSS
2019	NAACL, ECNLP, ACL, TDS
2018	ACL, WMT
2017	WMT

Activities

2022	Invited speaker at SEA: Search Engines Amsterdam meetup
2020	Invited speaker at <i>Transformers at Work: 2nd Edition</i> , A Deep Learning for NLP Workshop, Amsterdam
2020	Invited speaker at <i>Transformers at Work</i> , A Deep Learning for NLP Workshop, Amsterdam
2018	Invited speaker at the Machine Learning meetup, Amsterdam Police
2018	Invited speaker at the CL Seminar Series of the ILIC, University of Amsterdam

2018	Presentation in Deep Learning and Reinforcement Learning Summer School (DLRL2018), Toronto
2018	Organizer and chair of the research group's weekly seminar at University of Amsterdam
2017	Chair of the research group's weekly meeting at University of Amsterdam
2015	Co-developer of Arabic-English and Chinese-English translation systems of the University of Amsterdam for NIST Open MT Evaluation
2015	Participation Google NLP Summit, Zürich
2014	Participation in MT Marathon, Trento

Grants & awards

2018	Travel grant for participating in Deep Learning and Reinforcement Learning Summer School, Toronto, Canada
2017	Travel grant for participating in Association for Computational Linguistics (ACL) conference, Vancouver, Canada
2015	Travel grant for participating in Google's NLP PhD Summit, Zürich, Switzerland
2011	Research grant from Research Institute of ICT (No. CS1391-4-19)
2004	Ranked in the top 0.1% in the national university exam, Iran
2003	Semifinalist in Iran National Olympiad in Informatics

Language skills

Persian (Farsi): Native language
 English: Fluent, TOEFL (July 2013) Internet-Based Test: 110/120
 Spanish: Proficient, Cervantes DELE Diploma in Spanish Level B2
 Dutch: Intermediate, Diploma Level A2
 Arabic: Basic