Mohammad Sadegh Rasooli
Principal Applied Scientist
Microsoft, Mountain View, CA
rasooli@cs.columbia.edu

EMPLOYMENT

- Principal applied scientist, **Microsoft**, Mountain View, CA, August 2021-Now.
  - Principal applied scientist, August 2024-Now (former Senior applied scientist).
  - Speech and language modeling managed by Jinyu Li, Shawn Chang and Michael Levit.

    1) **Production streaming ASR model for Microsoft Teams**: implementation, model training and its components including ONNX conversion, latency and performance evaluation. 2) **LM for speech** based on PyTorch that handles large-scale training data.

- Postdoctoral researcher, **University of Pennsylvania**, January 2020- August 2021.
  - Supervisor: Chris Callison-Burch. Project: Low-resource neural machine translation.

- Research scientist, **Facebook AI**, Menlo Park, CA, October 2018-January 2020.
  - Member of the language and translation technologies (LATTE) team (Facebook AI Applied Research) managed by Dmitriy Genzel and Necip Fazil Ayan.
  - Machine translation for Facebook translation including the Pytorch-Translate code.

- Research assistant, Department of Computer Science, **Columbia University**, 2012-2018.
  - Supervised by Michael Collins focusing on cross-lingual transfer of NLP systems.

- Research intern, **Microsoft AI and Research**, Sunnyvale, CA, Summer 2017.
  - Mentor: Sarangarajan Parthasarathy, Project: entity-aware LM for ASR systems.

- Research intern, **Google Research**, New York, NY, Summer 2015.
  - Mentors: Kuzman Ganchev and Emily Pitler, Project: Multitask learning for semantic role labeling.

- Research intern, **Yahoo Research**, New York, NY, Summer 2014.
  - Mentor: Joel Tetreault, Project: Grammatical error correction.

- Research intern, **Nuance Communications**, Sunnyvale, CA, Summer 2013.
  - Mentor: Joel Tetreault, Project: Disfluency correction of spoken text.

- Research manager, Dadegan Research Group, Tehran, Iran, 2010-2012.
  - The Persian Dependency Treebank project. Managing 12 linguists and 4 engineers.

- Research engineer, Noor, Tehran, Iran, 2009-2010.
  - Manager: Omid Kashefi, Project: The Virastyar spell checker for Persian.

EDUCATION

- Ph.D., Computer Science, **Columbia University**, 2012-2018.
  - Advisor: Michael Collins.
  - *Thesis*: Cross-Lingual Transfer of Natural Language Processing Systems

- MSc. (Ph.D. track), Department of Computer Science, **Columbia University**, 2012-2014.

- MSc., Artificial Intelligence, Iran University of Science and Technology, 2009-2012.

- BSc., Software Engineering, Iran University of Science and Technology, 2005-2009.

SELECTED AND RECENT PUBLICATIONS

Full list of publications at Google Scholar page with about 1080 citations (h-index=19): `http://scholar.google.com/citations?user=UVTYgJIAAAAJ`

1. Sreyan Ghosh, **Mohammad Sadegh Rasooli**, Michael Levit, Peidong Wang, Jian Xue, Dinesh Manocha, Jinyu Li. Failing Forward: Improving Generative Error Correction for ASR with Synthetic Data and Retrieval Augmentation, arXiv:2410.13198, 2024.

2. Ajay Patel, Bryan Li, **Mohammad Sadegh Rasooli**, Noah Constant, Colin Raffel, and Chris Callison-Burch. Bidirectional Language Models Are Also Few-shot Learners, ICLR 2023.

3. Michael Levit, Sarangarajan Parthasarathy, Cem Aksoylar, **Mohammad Sadegh Rasooli**, and Shuangyu Chang. External language model integration for factorized neural transducers, arXiv:2305.17304, 2023.

4. Bryan Li, **Mohammad Sadegh Rasooli**, Ajay Patel, and Chris Callison-burch. Multilingual Bidirectional Unsupervised Translation through Multilingual Finetuning and Back-Translation, LoResMT, pp. 16–31, 2023.

5. **Mohammad Sadegh Rasooli**, Pegah Safari, Amirsaeid Moloodi, and Alireza Nourian. The Persian Dependency Treebank Made Universal. LREC 2022.

6. Daniel Khashabi, et al. ParsiNLU: A Suite of Language Understanding Challenges for Persian. TACL, Vol. 9, 2021.

7. **Mohammad Sadegh Rasooli**, Chris Callison-Burch, and Derry Wijaya. "Wikily" Neural Machine Translation Tailored to Cross-Lingual Tasks, EMNLP 2021.

8. Nikzad Khani, Isidora Chara Tourni, **Mohammad Sadegh Rasooli**, Chris Callison-Burch and Derry Tanti Wijaya. Cultural and Geographical Influences on Image Translatability of Words across Languages, NAACL 2021.

9. Maryam Aminian, **Mohammad Sadegh Rasooli**, and Mona Diab. Multitask Learning for Cross-Lingual Transfer of Broad-coverage Semantic Dependencies. EMNLP 2020.

10. **Mohammad Sadegh Rasooli**, and Michael Collins. Low-Resource Syntactic Transfer with Unsupervised Source Reordering, NAACL 2019.

11. **Mohammad Sadegh Rasooli**, and Sarangarajan Parthasarathy. Entity-Aware Language Model as an Unsupervised Reranker, INTERSPEECH 2018.

12. **Mohammad Sadegh Rasooli**, Noura Farra, Axinia Radeva, Tao Yu, and Kathleen McKeown. Cross-lingual sentiment transfer with limited resources. Machine Translation, Vol. 32, PP. 143-165, 2018.

13. **Mohammad Sadegh Rasooli** and Michael Collins. Cross-Lingual Syntactic Transfer with Limited Resources. TACL, Vol. 5, 2017.

14. **Mohammad Sadegh Rasooli** and Michael Collins. Density-Driven Cross-Lingual Transfer of Dependency Parsers. EMNLP 2015.

15. **Mohammad Sadegh Rasooli**, Thomas Lippincott, Nizar Habash and Owen Rambow. Unsupervised Morphology-Based Vocabulary Expansion. ACL 2014.

16. **Mohammad Sadegh Rasooli** and Joel Tetreault. Non-Monotonic Parsing of Fluent Umm I mean Disfluent Sentences. EACL 2014.

17. **Mohammad Sadegh Rasooli** and Joel Tetreault. Joint Parsing and Disfluency Detection in Linear Time, EMNLP 2013.

18. **Mohammad Sadegh Rasooli**, Manouchehr Kouhestani, and Amirsaeid Moloodi. Development of a Persian Syntactic Dependency Treebank, NAACL–HLT 2013.

STUDENT AND INTERN MENTORSHIP

- Microsoft intern (2024): Sreyan Ghosh from the University of Maryland College Park.
    - Boosting Generative Error Correction for ASR with Synthetic Data and Retrieval-Augmented Generative Error Correction.
- Co-mentor for Microsoft intern (2022): Bryan Li from the University of Pennsylvania.
    - Contextually biasing language models using attention over memory.

PROFESSIONAL SERVICES

Reviewer for NLP conferences and Journals since 2013 including TACL, ACL, NAACL, EACL, EMNLP, AACL, LRE, ACM TALIP, and Elsevier Computer Speech and Language.

Implemented and released the Yara Dependency Parser during Yahoo internship.

One of the contributors to the Universal Dependencies project.

TECHNICAL SKILLS

Python, Java, , C#, C++, Pytorch and HuggingFace (Transformers, PEFT, and accelerate).

TEACHING EXPERIENCE

- Co-instructor for Michael Collins' NLP course, Columbia University, Fall 2017.
- TA for Michael Collins' NLP course, Columbia University, Fall 2013 and Spring 2015.