Masoud Jalili Sabet
Volkswagen AG, Data Lab

✉ jalili.masoud@gmail.com • in masoudjalili • ⊕ masoudjs
⑧ Masoud Jalili Sabet • 🐦 _masoudjalili

Research Interests

○ LLM Chatbots ○ Explainability for VLMs ○ Multilinguality in LLMs

Education

Ph.D. in Computer Science **LMU Munich**
Supervisor: Prof. Hinrich Schütze *2018 – 2022*
Thesis: Multilingual Representations and Models for Improved Low-Resource Language Processing

M.Sc. in Computer Engineering (Software) **University of Tehran**
GPA - 17.54/20 (3.72/4.0) *2013 – 2016*
Thesis: Proposing a Translation Model using Spectral Embedding of Lexicon

B.Sc. in Computer Engineering (Software) **Iran University of Science & Tech, Tehran**
GPA - 16.09/20 (3.27/4.0) *2009 – 2013*

Experience

AI Expert (ML Scientist) **Volkswagen AG**
I work in the AI Innovation group and provide solutions using recent NLP models and LLMs *2022 – Present*
Manager: Christoph Ringlstetter

Research Intern (Remote, London) **Meta AI (Facebook AI)**
Project: Retriever-Reader models for open-domain question answering systems. *06/2020 – 12/2020*
Supervisor: Majid Yazdani

Software Eng. Intern, (London, UK) **Ads Ranking Team, Meta (Facebook)**
Project: Modeling epistemic uncertainty in Ads ranking models. *07/2019 – 10/2019*
Supervisors: Hastagiri Vanchinathan, Damien Lefortier

Research Intern (Trento, Italy) **HLT, FBK**
Project: Development of an open-source tool called TMOP for cleaning Translation Memories. *07/2015 – 10/2015*
Supervisors: Marco Turchi, Matteo Negri

Data Scientist **Cafe Bazaar**
I made an automated system to extract metrics using PySpark & Apache Airflow. *2016 – 2017*

Part-time Software Developer (C++/Java) **Yara Information Technology**
I worked on game engines and developed mobile games for iOS and Android. *2009 – 2011*

Selected Publications [Google Scholar][DBLP]

PatchCraft: Learning a Universal Patch to Manipulate the Attention of VLMs **CVPR** 2024
Jalili Sabet M.*, Rezaei R.*, Khakzar A., Rueckert D., Torr P., Gu J. *(under review)*

Glot500: Scaling Multilingual Corpora and Language Models to 500 Languages **ACL** 2023
Imani A., Lin P., Kargaran A., Severini S., Jalili Sabet M., Kassner N., Ma C., Schmid H.,
Martins A., Yvon F., Schütze H. **–Area Chair Award–**

Graph-Based Multilingual Label Propagation for Low-Resource Part-of-Speech Tagging, **EMNLP** 2022
Imani A., Severini S., Jalili Sabet M., Yvon F., Schütze H.

Graph Algorithms for Multiparallel Word Alignment **EMNLP** 2021
Imani A.*, Jalili Sabet M.*, Senel L.K., Dufter P., Yvon F., Schütze H.

ParCourE: A Parallel Corpus Explorer for a Massively Multilingual Corpus **ACL Demo** 2021
Imani A., Jalili Sabet M., Dufter P., Cysouw M., Schütze H.

SimAlign: High Quality Word Alignments Without Parallel **EMNLP Findings** 2020
Training Data Using Static and Contextualized Embeddings
Jalili Sabet M.*, Dufter P.*, Yvon F., Schütze H.

CaMEL: Case Marker Extraction without Labels   **ACL** 2022
Weissweiler L., Hofmann V., <u>Jalili Sabet M.</u>, Schütze H.

Graph Neural Networks for Multiparallel Word Alignment   **ACL Findings** 2022
Imani A., Senel L.K., <u>Jalili Sabet M.</u>, Yvon F., Schütze H.

Automatic Translation Memory Cleaning   **Machine Translation Journal** 2017
Negri M., Ataman D., <u>Jalili Sabet M.</u>, Turchi M., Federico M.
(DOI: 10.1007/s10590-017-9191-5)

An Unsupervised Method for Automatic Translation Memory Cleaning   **ACL** 2016
<u>Jalili Sabet M.</u>, Negri M., Turchi M., Barbu E.

TMop: a Tool for Unsupervised Translation Memory Cleaning   **ACL Demo** 2016
<u>Jalili Sabet M.</u>, Negri M., Turchi M., de Souza JGC., Federico M.

Improving Word Alignment of Rare Words with Word Embeddings   **Coling** 2016
<u>Jalili Sabet M.</u>, Faili H., Haffari G.

Honors and Awards

Travel Grant for Google NLP Summit   **Google**
For participating in Google's 3rd NLP Summit   *Zurich, Switzerland - 2019*

Travel Grant for LxMLS   **Google**
For participating in the 8th Lisbon Machine Learning School (LxMLS)   *Lisbon, Portugal - 2018*

COLING Conference Registration Grant
For participating in the 26th Conference on Computational Linguistics   *Osaka, Japan - 2016*

Ranked 16th in IEEEXtreme Programming Competition   International Competition - *2013*

Ranked 4th in RoboCup, 2D Soccer Simulation league   *2013*

Ranked [8th, 7th, 5th, 7th] in ACM ICPC Regional Contests   *2009 - 2012*

Selected Projects

ParCourE [GitHub Link]   **CIS, LMU, 2021**
ParCourE is a WebApp to browse a word aligned multiparallel corpus.

SimAlign [GitHub Link]   **CIS, LMU, 2020**
*SimAlign is a high-quality word alignment tool that uses static and contextualized
embeddings and does not require parallel training data.*

TMOP [GitHub Link]   **FBK, 2015**
TMop is an open-source software designed for cleaning and maintaining a Translation Memory.

TehranNama (Android application)   **2011 - 2013**
*Our team developed an offline map application for the city of Tehran at the request of the
Tehran municipality. The map was rendering Geo-spatial data and I was in charge of the offline
search and path-finding modules.*

Technical Skills

Programming: Python, C/C++, Java

ML Tools: PyTorch, TensorFlow, Keras, HuggingFace Transformers

Other Tools: Numpy, Scikit-learn, PySpark, Pandas

References

Prof. Hinrich Schütze (LMU Munich)   Marco Turchi (Zoom)
Christoph Ringlstetter (Volkswagen)   Majid Yazdani (Meta AI)