

CURRICULUM VITAE

PERSONAL INFORMATION

Ehsaneddin Asgari, Ph.D.
Munich, Germany
✉ [asgari \[at\] berkeley \[dot\] edu](mailto:asgari[at]berkeley[dot]edu)
🖥️ <https://llp.berkeley.edu/asgari/>
in <https://linkedin.com/in/ehsaneddinasgari/>

INTERESTS

🏠 Natural Language Processing, Multi-model NLP, Digital Humanities, 🌐 Bioinformatics

EDUCATION



University of California, Berkeley, CA, USA.

🎓 Doctor of Philosophy in Applied Science and Technology, 2019.
Designated Emphasis in Computational Data Science and Engineering.

PhD Thesis : “**Life Language Processing** : Deep Learning-based Language-agnostic Processing of Proteomics, Genomics/Metagenomics, and Human Languages”.

🎓 Master of Science in Applied Science and Technology, 2016.



Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland.

🎓 Master of Science in Computer Science, 2014.



Massachusetts Institute of Technology (MIT), MA, USA, 2013 -2014.

Master's Thesis : “**Automatic Story Merging using Topic Modeling** : A Case Study on the Story of Moses”. Supervised by Mark Finlayson at the Genesis Group (PI : P. Winston).
Visiting student at the Computer Science and Artificial Intelligence Laboratory (CSAIL).



Sharif University of Technology, Tehran, Iran.

🎓 Bachelor of Science in Software Engineering, 2011.

POSITIONS



AI Innovation Center, Volkswagen Group, Munich, Germany.

- Natural Language Processing Lead, Jan. 2023 - present.
- Natural Language Processing Data scientist, Jan. 2019 - Dec. 2022.



Helmholtz Center for Infection Research, Braunschweig, Germany.

- Post-doctoral Researcher at Prof. Alice McHardy's Group, Aug. 2019 - present.
- Research Fellow at Prof. Alice McHardy's Group, Jun. 2017 - Aug. 2019.



Center for Information and Language Processing, Munich, Germany.

- Research Fellow at Prof. Hinrich Schütze's Group, Feb. 2017 - Dec. 2018.
- Intern at Prof. Hinrich Schütze's Group, Jun. 2014 - Aug. 2014.



ABB Corporate Research, Zurich, Switzerland.

- Intern : Data-mining for the mining industry, Jun. 2012 - Oct. 2012



Advanced Digital Sciences Center of UIUC, Singapore, Singapore.

- Intern : Machine learning researcher Jun. 2010 - Sep. 2010

AWARDS AND HONORS

- 🎖️ 72nd Lindau Nobel Laureate Meeting Alumni (Nominated by Helmholtz Research Center). 2023
- 🎖️ Selected for the ISCB Fellowship at the ISMB. 2020
- 🎖️ Ranked 2nd in Semantic Evaluation (SemEval) Shared-Task 1. 2020
- 🎖️ Ranked 1st and 2nd in two out of three tasks in The Critical Assessment of protein Function Annotation algorithms (CAFA) 3.14. 2019
- 🎖️ Awarded Google Scholarship for Machine Learning Summer School (LxMLS), Lisbon. 2018
- 🎖️ Best Paper Award at the Multilingual and Cross-lingual Methods in NLP at the NAACL-HLT Workshop, Sponsored by Google, San Diego, USA. 2016
- 🎖️ Selected Submission at the ICML Computational Biology workshop, New York, USA. 2016
- 🎖️ Awarded Google Travel Grant for attending International Conference on Machine Learning (ICML) 2016, New York, USA. 2016
- 🎖️ Awarded Graduate Division Fellowship, University of California, Berkeley. 2014-2016
- 🎖️ Awarded Outstanding Student Certificate and Prize by Sharif University President. 2008
- 🎖️ Scholarship for undergraduate studies, Iranian National Foundation of Elites. 2007-2011
- 🎖️ Gold medal in the Iranian National Olympiad in Literature and Linguistics. 2007

PROFESSIONAL ACTIVITIES & SERVICES

- 🏠 Association for Computational Linguistics (ACL) and NLP Journals
Area Chair for the ACL Conferences
- ACL 2021, 2022, 2023
- EMNLP 2023

Reviewer for the ACL Conferences

- EMNLP 2017-2023
- ACL 2017-2023
- SEM-Eval 2021-2023
- Computational Approaches to Historical Language Change Workshop 2023
- NAACL 2019-2021
- SEM* 2020-2021
- ACL Student Research Workshop 2020-2021
- Typology for Polyglot NLP Workshop 2019-2020

Editor for Frontiers in Artificial Intelligence (NLP Section) Since 2023

Scientific reviewer for Agence Nationale de la Recherche proposals, France 2021

Reviewer for Natural Language Engineering Journal

Reviewer for Digital Scholarship in the Humanities

**Reviewing for Bioinformatics Related Journals/Conferences**

- Nature Communications
- Bioinformatics Journal
- PLOS Computational Biology Journal
- Briefings in Bioinformatics
- BMC Bioinformatics Journal
- IEEE/ACM Transactions on Computational Biology and Bioinformatics
- International Joint Conference on Artificial Intelligence (IJCAI) 2019-2020
- Intelligent Systems for Molecular Biology (ISMB) 2018-2019

🎖️ **A Member of the Admission Committee, Data Science M.Sc. program** 2016

ADVISING
EXPERIENCES

🎓 PhD Co-advising

- **Giorgos Kallergis** (Helmholtz Center, Co-advised with Alice McHardy) *2021-Present*
Language Modeling of Chemicals
- **Andrew Dickson** (UC Berkeley, Co-advised with M.R. Mofrad) *2020-Present*
Protein Language Modeling (in continuation of my PhD work)

🎓 Masters Co-advising

- **Alireza Sahebi** (Co-advised with A. Sharifi) *2021-2023*
Machine Learning Approaches for the Prediction of Pathogenicity in Genome Variations
- **Yongxin Huang** (LMU Munich/VW, Co-advised with Hinrich Schütze) *2022*
Clarifying QA Generation for Machine Learning Model Construction
- **Nicolai-David Bernard Ruhnau** (LMU Munich, Co-advised with Hinrich Schütze) *2018*
Distributional Semantics for the Computational Analysis of Food Corpora.

👤 Mentor of working students at Volkswagen Group, Data-Lab, Munich. *2022-present*

Bachelor Students :

- Shuqi Wang (2023) - Weak Supervision for Text Anonymization.
- Mobina Pournemat (2023) - Automatic Persian Text Editing.
- Arshan Dalili (2023) - Visual Word Sense Disambiguation.
- Soroush Nasiriany (2016) - NLP for Mental Health Prediction.
- Kaveh Karbasi (2016) - Benchmarking Protein-to-Protein Interaction Prediction Methods.
- Amine Mezzour (2015) - Deep Convolutional Network for RNA Structure Prediction.
- Sam Njie (2014) - Gene Ontology Exploring Package.

TEACHING
EXPERIENCES

- **Machine Learning Seminar** (HZI, BIOMEDAS Program) *Spring 2022, Spring 2023*
Contributing Lecturer : Deep Representation Learning.
- **Natural Language Processing** (Online weekend seminars) *Fall 2020-Present*
Volunteer Lecturer. NLP at the graduate level.
- **Deep Learning Seminar** (TU Braunschweig, with A. McHardy) *2019*
Contributing Lecturer and Designer of the Seminar.

🎓 Graduate Teaching Assistant

- **Data Structures and Algorithms** (University of California, Berkeley) *2016*
Information School (ISchool) (Summer 2016)
- **Programming languages** (University of California, Berkeley) *Spring 2016*
Python, Matlab, and Java Instructor
Self-Paced Center of EECS Department.

🎓 Undergraduate Teaching Assistant

- **Digital Signal Processing** (Sharif U. of Technology) *Spring 2011*
- **Probability and Statistics** (Sharif U. of Technology) *Fall 2010*
- **Theory of Languages and Automata** (Sharif U. of Technology) *Fall 2010*
- **Scientific and Technical Presentations** (Sharif U. of Technology) *Spring 2009-Spring 2011*
- **Web Programming** (Sharif U. of Technology) *Spring 2011*

SKILLS


- **Programming Languages** : Python, Java, C++, and R.
- **ML/DL Frameworks** HuggingFace, PyTorch, Spacy&nltk, Scikit-learn, Pandas, Elastic.
- **Languages** :
Persian (Native), English (Fluent), French (B1), German (B1), Arabic (Linguistic knowledge).

PUBLICATION

Year 2023


*A Dickson, *E Asgari, MH Alice C, M Mohammad RK

 *Go bench : Shared-hub for universal benchmarking of machine learning-based protein functional annotations.*

 **Bioinformatics 39 (2), 1367-4811, 2023.**

L März, E Asgari, F Braune, F Zimmermann, B Roth

 *XPASC : Measuring Generalization in Weak Supervision.*


 **Accepted at Natural Language Engineering Journal.**

Z Taghavi, PH Naeini, MAS Javaheri, S Gooran, E Asgari, HR Rabiee, H Sameti

 *A CLIP-Based Approach for Comparing Cross-modality and Unimodality in Visual Word Sense Disambiguation.*

 **Proc. of the The 17th International Workshop on Semantic Evaluation, 2023.**

O Ghahroodi, MAS Javaheri, D Dastgheib, MS Baghshah, MH Rohban, E Asgari

 *A Class-Token Attention-based Model for Human Value Detection.*

 **Proc. of the The 17th International Workshop on Semantic Evaluation, 2023.**

A Sadeghi, R Alipour, K Taeb, P Morassafar, N Salemahim, E Asgari

 *A Multilingual Transformer Language Model-based Approach for the Detection of News Genre, Framing and Persuasion Techniques.*

 **Proc. of the The 17th International Workshop on Semantic Evaluation, 2023.**


O Ghahroodi, SA Dalili, S Mesforoush, E Asgari

 *Prompt Generation for Visual Word Sense Disambiguation.*

 **Proc. of the The 17th International Workshop on Semantic Evaluation, 2023.**


P Münch, R Mreches, XY To, HA Gündüz, J Moosbauer, S Klawitter, Z Deng, G Robertson, M Rezaei, E Asgari, E Franzosa, C Huttenhower, B Bischl, A McHardy, M Binder

 *A platform for deep learning on (meta) genomic sequences.*


 **Under Review**


N Mirzakhmedova, J Kiesel, M Alshomary, M Heinrich, N Handke, X Cai, B Valentin, D Dastgheib, O Ghahroodi, M Sadraei, E Asgari, L Kawaletz, H Wachsmuth, B Stein

 *The Touché 23-ValueEval Dataset for Identifying Human Values behind Arguments.*

 **arXiv preprint arXiv :2301.13771, 2023.**


C Ma, A ImaniGooghari, H Ye, E Asgari, H Schütze

 *Taxi1500 : A Multilingual Dataset for Text Classification in 1500 Languages.*

 **arXiv preprint arXiv :2305.08487, 2023.**

Year 2022

SK Hotop, S Reimering, A Shekhar, E Asgari, U Beutling, C Dahlke, et al.

 *Peptide microarrays coupled to machine learning reveal individual epitopes from human antibody responses with neutralizing capabilities against SARS-CoV-2.*

 **Emerging microbes & infections 11 (1), 1037-1048, 2022.**

D Dastgheib, E Asgari



 *Keyword-based Natural Language Premise Selection for an Automatic Mathematical Statement Proving.*

 **Proc. of TextGraphs-16 : Graph-based Methods for Natural Language, 2022.**

S Mirzababaei, AH Kargaran, H Schütze, E Asgari



-  [Hengam : An Adversarially Trained Transformer for Persian Temporal Tagging.](#)
 **Proc. of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (AAACL), 2022.**

SH Alavian, A Satvaty, S Sabouri, E Asgari, H Sameti



-  [Docalog : Multi-document Dialogue System using Transformer-based Span Retrieval.](#)
 **Proc. of the Second DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering, 2022.**

Year 2021



A Bahai, E Asgari, MRK Mofrad, A Kloetgen, AC McHardy

-  [EpitopeVec : linear epitope prediction using deep protein sequence embeddings.](#)
 **Bioinformatics 37 (23), 4517-4525, 2021.**



E Nourani, E Asgari, AC McHardy, MRK Mofrad

-  [TripletProt : deep representation learning of proteins based on siamese networks.](#)
 **IEEE/ACM Transactions on Computational Biology and Bioinformatics 19 (6), 2021.**

L März, E Asgari, F Braune, F Zimmermann, B Roth



-  [KnowMAN : Weakly Supervised Multinomial Adversarial Networks.](#)
 **Empirical Methods in Natural Language Processing (EMNLP), 2021.**

Asgari E.



-  [Method, Computer Program and Apparatus for Relating Text Units.](#)
 **European Patent 20206180.0.6 - 1231, January 2021.**

Year 2020



A Khaledi, A Weimann, M Schniederjans, E Asgari, TH Kuo, A Oliver, ...

-  [Predicting antimicrobial resistance in Pseudomonas aeruginosa with machine learning-enabled molecular diagnostics.](#)
 **EMBO molecular medicine 12 (3), e10264, 2020.**



E Asgari, C Ringlstetter, H Schütze

-  [EmbLexChange at SemEval-2020 Task 1 : Unsupervised embedding-based detection of lexical semantic changes.](#)
 **Proc. of the Fourteenth Workshop on Semantic Evaluation, 201-207, 2020.**



E Asgari, F Braune, B Roth, C Ringlstetter, MRK Mofrad

-  [UniSent : Universal Adaptable Sentiment Lexica for 1000+ Languages.](#)
 **Proc. of The 12th Language Resources and Evaluation Conference (LREC), 2020.**


E Asgari, PC Münch, TR Lesker, AC McHardy, MRK Mofrad


-  [Data-driven Variable-length Segmentation of Biological Sequences : Applications in Metagenomics and Proteomics.](#)
 **NeurIPS - Computational Biology Workshop, 2020.**

M Aldawsari, E Asgari, MA Finlayson


-  [Story Fragment Stitching : The Case of the Story of Moses.](#)
 **1st Workshop on Artificial Intelligence for Narratives (AI4N) at the International Conference on Artificial Intelligence (IJCAI 2020).**


Asgari E.

 *Method, Computer Program and Apparatus for Detecting a Semantic Change of a Word between Domains.*

 **European Patent 20190280.6 - 1231, November 2020.**


E Asgari, MJ Sabet, P Dufter, C Ringlstetter, H Schütze

 *Subword sampling for low resource word alignment.*

 **arXiv preprint arXiv :2012.11657, 2020.**


M Ahmadi, MR Jahed-Motlagh, E Asgari, AT Rahmani, AC McHardy


 *ToxVec : Deep Language Model-Based Representation Learning for Venom Peptide Classification.*

 **bioRxiv, 2020.09. 29.319046, 2020.**


Year 2019


N Zhou, Y Jiang, TR Bergquist, AJ Lee, BZ Kacsóh, AW Crocker, et al.

 *The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens.*


 **Genome biology 20 (1), 1-23, 2019.**


E Asgari, AC McHardy, MRK Mofrad

 *Probabilistic variable-length segmentation of protein sequences for discriminative motif discovery (DiMotif) and sequence embedding (ProtVecX).*

 **Scientific reports 9 (1), 3577, 2019.**


E Asgari, PC Münch, TR Lesker, AC McHardy, MRK Mofrad

 *DiTaxa : nucleotide-pair encoding of 16S rRNA for host phenotype and biomarker detection.*


 **Bioinformatics 35 (14), 2498-2500, 2019.**


E Asgari, MRK Mofrad

 *Deep genomics and proteomics : Language model-based embedding of biological sequences and their applications in bioinformatics.*


 **Leveraging Biomedical and Healthcare Data, 167-181, 2019.**


A Ehsaneddin, CMH Alice, RK Mohammad

 *Probabilistic variable-length segmentation of protein sequences for discriminative motif discovery (DiMotif) and sequence embedding (ProtVecX).*

 **Scientific Reports 9 (1), 3577, 2019.**


E Asgari, N Poerner, AC McHardy, MRK Mofrad


 *DeepPrime2Sec : deep learning for protein secondary structure prediction from the primary sequences.*

 **BioRxiv, 705426, 2019.**

Year 2018


E Asgari, K Garakani, AC McHardy, MRK Mofrad

 *MicroPheno : Predicting environments and host phenotypes from 16S rRNA gene sequencing using a k-mer based representation of shallow sub-samples.*

 **Bioinformatics 34 (13), i32-i42, 2018.**


Z Jahed, D Fadavi, UT Vu, E Asgari, GWG Luxton, MRK Mofrad

 *Molecular insights into the mechanisms of SUN1 oligomerization in the nuclear envelope.*

 **Biophysical journal 114 (5), 1190-1203, 2018.**


H Adel, E Asgari, H Schütze

 *Overview of character-based models for natural language processing.*

 *Lecture Notes in Computer Science (LNCS)*, book series (LNCS, volume 10761) 2018, 3–16.

Year 2017

E Asgari, H Schütze


 *Past, present, future : a computational investigation of the typology of tense in 1000 languages.*

 *In Proc. of the Empirical Methods on Natural Language Processing (EMNLP), 2017.*

Highlighted by MIT Tech. Review Magazine & LMU Research Magazine

E Asgari, A Sanaei


 *Measuring Countries' Human Rights Positions in UN Universal Periodic Review.*

 **American Political Science Association (APSA) annual meeting, 2017.**

Year 2016


E Asgari, MRK Mofrad

 *Comparing Fifty Natural Languages and Twelve Genetic Languages Using Word Embedding Language Divergence (WELD) as a Quantitative Measure of Language Distance.*

 **In Proc. of the NAACL-HLT Workshop on Multilingual Methods in NLP, 2016. Best Paper Award**


E Asgaria, S Nasiriany, M Mofrad

 *Text Analysis and Automatic Triage of Posts in a Mental Health Forum.*

 **In Proc. of the NAACL-HLT Workshop on Computational Linguistics and Clinical Psychology : From Linguistic Signal to Clinical Reality, 2016.**

H Schuetze, H Adel, E Asgari

 *Nonsymbolic Text Representation.*


 **arXiv preprint arXiv :1610.00479, 2016.**

Originally was a short paper written by the first author at the EACL 2017.

Year 2015

E Asgari, MRK Mofrad


 *Continuous distributed representation of biological sequences for deep proteomics and genomics.*

 **PloS one 10 (11), e0141287, 2015.**

Year 2014

M Neshati, D Hiemstra, E Asgari, H Beigy


 *Integration of scientific and social networks.*

 **World wide web 17, 1051-1079, 2014.**

Year 2013


E Asgari, JC Chappelier

 *Linguistic Resources & Topic Models for the Analysis of Persian Poems.*


 **In Proc. of the NAACL-HLT Workshop on Computational Linguistics for ... , 2013.**


E Asgari, M Ghassemi, MA Finlayson

 *Confirming the themes and interpretive unity of Ghazal poetry using topic models.*


 **Neural Information Processing Systems (NIPS) Workshop for Topic Models, 2013.**

M Neshati, E Asgari, D Hiemstra, H Beigy

 *A Joint Classification Method to Integrate Scientific and Social Networks.*

 **Advances in Information Retrieval, 122-133, 2013.**


RESEARCH
EXPERIENCES


 **Recent Industrial Project Topics** AI Innovation Center, Volkswagen *Jan. 2019 - present*


 Large language model-based chat bot solution

 Large language model-based data annotation

 Sentiment and aspect analysis


 Text classification pipeline


 Automatic event extraction

 **Post-doctoral at HZI** Machine learning for infection research *Aug. 2019 - present*

(i) Anti-bacterial resistance prediction and analysis, (ii) Host phenotype prediction from microbial communities, and (iii) Deep chemical and genome language modeling.


PI : Prof. A. McHardy

 **Visiting at LMU-Munich** Deep language processing

 Character-level and subword-level NLP for 1000 languages. *Feb. 2017 - Jan. 2019*

 Detection of new trends in scientific communities. *Jun. 2014 - Aug. 2014*

PI : Prof. H. Schütze

 **Collab. with UC Berkeley Near Eastern Studies Dep.** Ancient scripts *Feb. 2017-Jan. 2019*
Statistical language processing for cuneiform.

PI : Prof. N. Veldhuis and Dr. Eduardo Escobar

  **Life Language Processing at UCB** Protein/microbial informatics *Aug. 2014 - Present*


Introducing language-model based representation of biological sequences in my PhD thesis and currently co-supervising the new PhD students of the lab.

PI : Prof. M. Mofrad

 **Visiting at MIT-CSAIL** Story Analysis *Sep. 2013 - Mar. 2014*


Structural and semantic analysis of narratives using, particularly developing an algorithm for automatic merging of stories.

PI : Prof. P. Winston and Prof. M. Finlayson

 **Visiting at MIT-Harvard** EEG Processing *May 2013 - Sep. 2013*

Neural signal processing for sleep phase detection.

PI : Prof. E. Brown

 **Intern at Audiovisual Communications Lab at EPFL** ECG Processing *Jan 2013 - May. 2013*


Research assistant : Efficient real-time ECG processing on mobile phones in collab. with Qualcomm.

PI : Prof. M. Vetterli

 **Intern at Artificial Intelligence Lab at EPFL** Literature analysis *Aug. 2012 - Feb. 2013*

Computational semantic analysis of Persian poetry in 1000 years.

PI : Dr. J-C. Chappelier

 **Intern at Non-Invasive BCI Lab at EPFL** EOG/EEG Processing *Dec. 2011 - Jul. 2012*


Eye-tracking supervision for brain-computer interface.

PI : Prof. J. Milan

 **AI Lab at Sharif Uni.** Expert Finding *Jul. 2011 - Jan. 2012*

Enriching expert finding using social network matching.

PI : Dr. Hamid Beigy

 **Intern at the ADSC of UIUC** Computer Vision *Jul. 2010 - Sep. 2010*

Efficient Visual Computing for Interactive Vision Applications on Mobile Devices.

PI : Prof. M. Winslett