# CURRICULUM VITAE



Awards and Honors	Ranked 1st in Semantic Evaluation (SemEval) Shared-Task, Multilingual Detection of H Techniques in Memes.	Persuasion 2024.
	$\bigotimes$ 72 <sup>2d</sup> Lindau Nobel Laureate Meeting Alumni (Nominated by Helmholtz Research Center	). 2023
	$\bigotimes$ Selected for the ISCB Fellowship at the ISMB.	2020
	$\bigotimes$ Ranked 2nd in Semantic Evaluation (SemEval) Shared-Task 1.	2020
	$\bigotimes$ Ranked $1^{st}$ and $2^{nd}$ in two out of three tasks in The Critical Assessment of protein Funct tation algorithms (CAFA) 3.14.	ion Anno- 2019
	$\bigotimes$ 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 NA. Workshop, Sponsored by Google, San Diego, USA.	ACL-HLT 2016
	$\bigotimes$ Selected Submission at the ICML Computational Biology workshop, New York, USA.	2016
	Awarded Google Travel Grant for attending International Conference on Machine Learnin 2016, New York, USA.	ng (ICML) 2016
	leave the second state of	2014-2016
	$\bigotimes$ Awarded Outstanding Student Certificate and Prize by Sharif University President.	2008
	$\bigotimes$ Scholarship for undergraduate studies, Iranian National Foundation of Elites.	2007-2011
	$\bigotimes$ Gold medal in the Iranian National Olympiad of Literature and Linguistics.	2007
Professional Activities & Services	<ul> <li>Association for Computational Linguistics (ACL) and NLP Journals</li> <li>Senior Area Chair for the ACL Conferences</li> <li>ACL</li> <li>Efficient Natural Language &amp; Speech Processing at NeuRIPS</li> </ul>	2025 2024
	Area Chair for the ACL Conferences	
	• EMNLP • NAACL	2021-2025 2023-2025 2023-2025 2024-2025
	Reviewer for the ACL Conferences	
	<ul> <li>ACL</li> <li>SEM-Eval</li> <li>Computational Approaches to Historical Language Change Workshop</li> <li>NAACL</li> <li>SEM*</li> <li>ACL Student Research Workshop</li> </ul>	2017-2023 2017-2023 2021-2023 2023 2019-2021 2020-2021 2020-2021 2019-2020
	Editor for Frontiers in Artificial Intelligence (NLP Section) Scientific reviewer for Agence Nationale de la Recherche proposals, France Reviewer for Natural Language Engineering Journal Reviewer for Digital Scholarship in the Humanities	Since 2023 2021
	<ul> <li>Reviewing for Bioinformatics Related Journals/Conferences</li> <li>Nature Communications</li> <li>Bioinformatics Journal</li> <li>PLOS Computational Biology Journal</li> <li>Briefings in Bioinformatics</li> <li>BMC Bioinformatics Journal</li> <li>IEEE/ACM Transactions on Computational Biology and Bioinformatics</li> </ul>	

International Joint Conference on Artificial Intelligence (IJCAI)
Intelligent Systems for Molecular Biology (ISMB)

Advising	➢ PhD Co-advising				
EXPERIENCES	• Mohammad Parsa (UC Berkeley, Co-advised with M.R. Mofrad) Evalution of Protein Language Models.	2024-Present			
	• Fatemeh Shaterian (Sharif Uni., Co-advised with A. Fazli) Code LLMs and their Evaluations.	2024-Present			
	• <b>Giorgos Kallergis</b> (Helmholtz Center, Co-advised with Alice McHardy) Language Modeling of Chemicals.	2021-Present			
	• Andrew Dickson (UC Berkeley, Co-advised with M.R. Mofrad) Protein Language Modeling (in continuation of my PhD work)	2020-Present			
	➢ Masters Co-advising				
	• Omid Ghahroodi (currently working with QCRI) Language Models for Hierarchical Classifications.	2021-2024			
	• M.A. Sadraei (currently working with QCRI) (Co-advised with H. Ra Continual Learning of Language Models.	abiee) 2021-2024			
	• Alireza Sahebi (Co-advised with A. Sharifi) Machine Learning Approaches for the Prediction of Pathogenicity in Genome	2021-2023 e Variations			
	• Yongxin Huang (LMU Munich/VW, Co-advised with Hinrich Schütze) Clarifying QA Generation for Machine Learning Model Construction	2022			
	<ul> <li>Nicolai-David Bernard Ruhnau (LMU Munich, Co-advised with Hinrich Distributional Semantics for the Computational Analysis of Food Corpora.</li> </ul>	A Schütze) 2018			
	▲ Mentoring at QCRI Summer Internship Program (The team ranked 3 <sup>rd</sup> ) Mapping the Evolution of Scientific Legacy. Summer 2024				
	<ul> <li>♦ Malak Ibrahim (CMU-Q) ♦ Anas Madkoor (QU) ♦ Talha Javad (Doha Science and Technology)</li> <li>♦ Myesha Hoque (QU) ♦ Fajr Amar (QU) ♦ Arwa Eleradi (QU) ♦ Bryson Jandawa (Princeton)</li> </ul>				
	$\clubsuit$ Mentor of working students at Volkswagen Group, Data-Lab, Munich.	2022-present			
	<ul> <li>Bachelor Students :</li> <li>Mahdi Abootorabi (2023) - Text localization in Audio.</li> <li>Shuqi Wang (2023) - Weak Supervision for Text Anonymization.</li> <li>Mobina Pournemat (2023) - Automatic Persian Text Editing.</li> <li>Arshan Dalili (2023) - Visual Word Sense Disambiguation.</li> <li>Soroush Nasiriany (2016) - NLP for Mental Health Prediction.</li> <li>Kaveh Karbasi (2016) - Benchmarking Protein-to-Protein Interaction Prediction Methods.</li> <li>Amine Mezzour (2015) - Deep Convolutional Network for RNA Structure Prediction.</li> <li>Sam Njie (2014) - Gene Ontology Exploring Package.</li> </ul>				
Teaching Experiences	• Machine Learning Seminar (HZI, BIOMEDAS Program) Spring Contributing Lecturer : Deep Representation Learning.	g 2022, Spring 2023			
	<ul> <li>Natural Language Processing/LLMs (Online seminars) Volunteer Lecturer.</li> </ul>	Fall 2020-Fall 2023			
	• <b>Deep Learning Seminar</b> (TU Braunschweig, with A. McHardy) Contributing Lecturer and Coordinator.	2019			
	➢ Graduate Teaching Assistant				
	• Data Structures and Algorithms (University of California, Berkeley) Information School (ISchool) (Summer 2016)	2016			
	• <b>Programming languages</b> (University of California, Berkeley) Python, Matlab, and Java Instructor Self-Paced Center of EECS Department.	Spring 2016			
Skills	<ul> <li>Programming Languages Python, Java, C++, and R.</li> <li>ML/DL Frameworks HuggingFace, PyTorch, Spacy&amp;nltk, Scikit-learn, Pa</li> <li>Languages Persian (Native), English (Fluent), French (B1), German (B1), Arabian (B1</li></ul>	ks HuggingFace, PyTorch, Spacy&nltk, Scikit-learn, Pandas, Elastic.			

PUBLICATION

Fanar Team (My contributions was in Arabic Tokenizer, Arabic/Multimodal Benchmarking)
Fanar : An Arabic-Centric Multimodal Generative AI Platform

© preprint arXiv :2501.13944 (2025).

Kallergis, Georgios, Ehsannedin Asgari, Behrooz Azarkhalili, Martin Empting, Anna Hirsch, and Alice McHardy.

ChemLM : Domain adaptable language modeling of chemical compounds identifies potent pathoblockers for Pseudomonas aeruginosa.

Accepted at Communications Chemistry 2025.

Mohammad-Mahdi Abootorabi and Ehsaneddin Asgari

**CLASP** : Contrastive Language-Speech Pretraining for Multilingual Multimodal Information Retrieval.

Accepted at The European Conference on Information Retrieval (ECIR), 2025.

Ma, Chunlan, Ayyoob ImaniGooghari, Haotian Ye, Renhao Pei, Ehsaneddin Asgari, and Hinrich Schuetze

Taxi1500 : A multilingual dataset for text classification in 1500 languages.

Solution Accepted at NAACL, 2025.

Mahshid Dehghani, Amirahmad Shafiee, Ali Shafiei, Neda Fallah, Farahmand Alizadeh, Mohammad Mehdi Gholinejad, Hamid Behroozi, Jafar Habibi, Ehsaneddin Asgari

Emo3D : Metric and Benchmarking Dataset for 3D Facial Expression Generation from Emotion Description

Solution Accepted at NAACL (Findings), 2025.

## Year 2024

P. Zahraee; E. Asgari.

**u** TuringQ : Benchmarking AI Comprehension in Theory of Computation

In Findings of the Association for Computational Linguistics : EMNLP 2024.

M.A. Sadraei; E. Asgari; A. McHardy; H.R. Rabiee.

SuperPos-Prompt : Enhancing Soft Prompt Tuning of Language Models with Superposition of Multi Token Embeddings

Sefficient Natural Language & Speech Processing at NeurIPS 2024.

E. Asgari; Y. Elkheir.

**Morphologically-Aware Tokenizer** 

**VU.S.** Provisional Patent Application No. 63/679,403.

O. Ghahroodi, M. Nouri, M. Sanian, A. Sahebi, D. Dastgheib, \*E. Asgari, \*M. Soleymani, \*M. Rohban.
Khayyam Challenge (PersianMMLU) : Is Your LLM Truly Wise to The Persian Language ?
COLM 2024.

Hu, Kaixin, Fernando Meyer, Zhi-Luo Deng, Ehsaneddin Asgari, Tzu-Hao Kuo, Philipp C. M $\tilde{A}^1_4$ nch, and Alice C. McHardy.

Second System 2 (1997) Second System 2 (1997)

 $\blacksquare$  Briefings in Bioinformatics 25.3 (2024) : bbae206.

Nailia Mirzakhmedova, Johannes Kiesel, Milad Alshomary, Maximilian Heinrich, Nicolas Handke, Xiaoni Cai, Valentin Barriere, Doratossadat Dastgheib, Omid Ghahroodi, MohammadAli SadraeiJavaheri, Ehsaneddin Asgari, Lea Kawaletz, Henning Wachsmuth, and Benno Stein.

The Touche 23 ValueEval Dataset for Identifying Human Values behind Arguments In Proc. of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italia.

MohammadAli SadraeiJavaheri, Ehsaneddin Asgari, and Hamid Reza Rabiee.

Transformers for Bridging Persian Dialects : Transliteration Model for Tajiki and Iranian Scripts

◊ In Proc. of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italia.

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.* 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.
- 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

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.

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 (AACL), 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

Epitope Vec : 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. *Furopean 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 learningenabled 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.

Separate 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.

S NeurIPS - Computational Biology Workshop, 2020.

M Aldawsari, E Asgari, MA Finlayson

**L** 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 Kacsoh, 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.

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 bio. 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.

E Asgari, JC Chappelier

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

In Proc. of the NAACL-HLT Workshop on Comp. Ling. for Literature, 2013.

- E Asgari, M Ghassemi, MA Finlayson
  - Confirming the themes and interpretive unity of Ghazal poetry using topic models.
     Neural Information Processing Systems 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
- Industrial Project Topics AI Innovation Center, Volkswagen Jan. 2019 2023
   Large language model-based chat bot solution
   Large language model-based data annotation tool
   Sentiment and aspect analysis
- $\bigotimes$  Text classification pipeline
- $\bigotimes$  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

 Image: Visiting at LMU-Munich
 Deep language processing

 Image: Optimize Character-level and subword-level NLP for 1000 languages.
 Feb. 2017 - Jan. 2019

 Image: Optimize Character-level and subword-level NLP for 1000 languages.
 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/mircobial 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
 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
 Computational semantic analysis of Persian poetry in 1000 years.
 PI : Dr. J-C. Chappelier

Intern at Non-Invasive BCI Lab at EPFL EOG/EEG Proc Eye-tracking supervision for brain-computer interface. PI : Prof. J. Milan	cessing Dec. 2011 - Jul. 2012
<b>AI Lab at Sharif Uni.</b> Expert Finding Enriching expert finding using social network matching. <b>PI : Dr. Hamid Beigy</b>	Jul. 2011 - Jan. 2012
<ul> <li>Intern at the ADSC of UIUC Computer Vision</li> <li>Efficient Visual Computing for Interactive Vision Applications on</li> <li>PI : Prof. M. Winslett</li> </ul>	Jul. 2010 - Sep. 2010 Mobile Devices.