

# Ehsaneddin Asgari

◇ **Email:** asgari [at] berkeley [dot] edu ◇ **Home page:** <https://llp.berkeley.edu/asgari/>

- EDUCATION
- ◇ **University of California, Berkeley, CA, USA**
    - Doctor of Philosophy in Applied Science and Technology (August 2014 - August 2019)  
*PhD Thesis:* “**Life Language Processing:** Deep Learning-based Language-agnostic Processing of Proteomics, Genomics/Metagenomics, and Human Languages”  
Designated Emphasis in Computational Data Science and Engineering
    - Master of Science in Applied Science and Technology (2014 - 2016)
  - ◇ **Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland**  
School of Computer and Communication Science (IC)
    - Master of Science in Computer Science (2011 - 2014)
  - ◇ **Massachusetts Institute of Technology (MIT), MA, USA**  
Computer Science and Artificial Intelligence Laboratory (CSAIL)  
*Master’s Thesis:* “Topic Model Story Merging”  
Supervised by Mark Finlayson/Patrick Winston (2013 - 2014)
  - ◇ **Sharif University of Technology, Tehran, Iran**
    - Bachelor of Science in Software Engineering (2007 - 2011)

- R&D EXPERIENCES
- ◇ **Helmholtz Center for Infection Research, Braunschweig, Germany**
    - Research Fellow (June 2017 - Present)
  - ◇ **ABB Corporate Research, Zurich, Switzerland**
    - Intern: Data-mining for the mining industry (Summer 2012)
  - ◇ **Advanced Digital Sciences Center**  
The Research Center of University of Illinois at Urbana-Champaign in Singapore
    - Intern: Machine learning researcher (July 2010 - September 2010)

## PUBLICATIONS ◇ **Journal Papers**

1. Asgari A, Poerner N, McHardy C. A, Mofrad.  
DeepPrime2Sec: Deep Learning for Protein Secondary Structure Prediction from the Primary Sequences.  
Submitted to *Bioinformatics Journal*.
2. A. Moiseeva, A. Friedrich, E. Asgari, and H. Schütze.  
A Benchmark for Trend and Antitrend Detection.  
Under Review at *PLOS One Journal*.
3. Khaledi A\*, Weimann A\*, Schniederjans M\*, Asgari E, Kuo T, Gabriel O, Kola A, Gastmeier P, Hogardt M, Jonas D, Mofrad MRK, Bremges A, McHardy AC, Hussler S.  
Fighting antimicrobial resistance in *Pseudomonas aeruginosa* with machine learning-enabled molecular diagnostics.  
Under revisions at *EMBO Molecular Medicine*, available at bioRxiv (2019): 643676.
4. Zhou, N. et al. The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens.  
Under Review at *Genome Biology*, available at bioRxiv 653105 (2019). doi:10.1101/653105.
5. E. Asgari, A. C. McHardy, and M. R. Mofrad,  
Probabilistic variable-length segmentation of protein sequences for discriminative motif discovery (DiMotif) and sequence

- embedding (ProtVecX).  
*Scientific Reports Journal*, 9.1 (2019): 3577.
6. E. Asgari, P. C. Münch, T. R. Lesker, A. C. McHardy, and M. R. Mofrad,  
 DiTaxa: Nucleotide-pair encoding of 16S rRNA sequences for host phenotype and biomarker detection.  
*Bioinformatics Journal*, bty954, 2018.
  7. E. Asgari, K. Garakani, A. C. McHardy, M.R.K. Mofrad,  
 MicroPheno: Predicting environments and host phenotypes from 16S rRNA gene sequencing using a k-mer based  
 representation of shallow sub-samples.  
*Bioinformatics Journal*, Volume 34, Issue 13, 1 July 2018, Pages i32i42.
  8. Z. Jahed, D. Fadavi, U. Vu, E. Asgari, G. Luxton, M. Mofrad,  
 Molecular Insights into Mechanism of SUN1 Assembly in the Nuclear Envelope.  
*Biophysical Journal*, Volume 114, Issue 5, p11901203, 13 March 2018.
  9. E. Asgari, M.R.K. Mofrad,  
 Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics.  
*PLOS One Journal* 10.11: e0141287, 2015.
  10. M. Neshati, D. Hiemstra, E. Asgari, H. Beigy,  
 Integration of Scientific and Social Networks.  
*World Wide Web Journal Springer (WWWJ)*, 2014.

PUBLICATIONS ◇ **Book Chapters**

1. E. Asgari and M. R. K Mofrad,  
 Deep Genomics and Proteomics: Language Model-Based Embedding of Biological Sequences and Their Applications in Bioinformatics.  
*Leveraging Biomedical and Healthcare Data* (pp. 167-181) 2019. Academic Press.
2. 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) 2017, 3–16.

PUBLICATIONS ◇ **Peer-reviewed Conference/Workshop Papers/Patents**

1. M. Aldawsari, E. Asgari, and M. Finlayson  
 Stitching Together Story Fragments: The Case of the Story of Moses.  
 Under Review at **EMNLP 2019**.
2. E. Asgari, F. Braune, B. Roth, C. Ringlstetter, and M. Mofrad  
 UniSent: Universal Adaptable Sentiment Lexica for 1000+ Languages.  
 Under Review at **CoNLL 2019**.
3. E. Asgari, H. Schütze  
 Past, Present, Future: A Computational Investigation of the Typology of Tense in 1000 Languages.  
 In Proceedings of the *Empirical Methods on Natural Language Processing (EMNLP)*, Copenhagen, Denmark, September 2017 {**Highlighted by MIT Tech. Review Magazine & LMU Research Magazine**}
4. E. Asgari, M.R.K. Mofrad  
 Comparing Fifty Natural Languages and Twelve Genetic Languages Using Word Embedding Language Divergence as a Quantitative Measure of Language Distance.  
 In Proceedings of the *NAACL-HLT Workshop on Multilingual Methods in NLP*, CA, June 2016 {**Best Paper Award**}
5. E. Asgari, M.R.K. Mofrad

Word Vectors of Biological Sequences and Their Applications in Bioinformatics.

*ICML Workshop on Computational Biology*, NY, June 2016 {**Selected Poster for Google Travel Award**}

6. E. Asgari, S. Nasiriany, M.R.K. Mofrad  
Text Analysis and Automatic Triage of Posts in a Mental Health Forum.  
In Proc. of the *NAACL-HLT Workshop on Comp. Linguistics and Clinical Psychology*, CA, June 2016.
7. M. Neshati, E. Asgari, D. Hiemstra, H. Beigy  
A Joint Classification Method For Scientific and Social Network Integration.  
In Proc. of *the European Conference on Information Retrieval (ECIR)*, Moscow, March 2013.
8. E. Asgari, J-C. Chappelier  
Linguistic Resources and Topic Models for the Analysis of Persian Poems.  
In Proceedings of the *NAACL-HLT Workshop on Computational Linguistics for Literature*, GA, June 2013.
9. E. Asgari, M. Ghassemi, M. Finlayson  
Confirming the Themes and Interpretive Unity of Ghazal Poetry Using Topic Models.  
In Proc. of the *NIPS Workshop on Topic Models*, NV, December 2013.
10. Hamid R. Zarandi, A. Fattaholmanan, A. Vakilian, E. Asgari, Mohammad R. Besharati  
Hodhod, Automatic Hardware Description Generator.  
*IRI Patent 64177*, March 2010.

#### PUBLICATIONS ◇ Preprints

1. E. Asgari, A. Sanaei  
Measuring Countries Human Rights Positions in Universal Periodic Review.  
Available at SSRN, <http://dx.doi.org/10.2139/ssrn.3029031>,  
Presented at the *American Political Science Association* annual meeting 2017, San Francisco.
2. H. Schütze, H. Adel, E. Asgari  
Nonsymbolic Text Representation.  
*arXiv preprint arXiv:1512.00397* (2017).  
*Originally was a short paper written by the first author at the EACL 2017*

- RESEARCH EXPERIENCES ◇ PhD thesis: **Life Language Processing**, Deep Learning-based Language-agnostic Processing of Proteomics, Genomics/Metagenomics, and Human Languages, *Molecular cell biomechanics laboratory* (PI: Prof. M.R.K. Mofrad), *UC Berkeley*, CA, USA (August 2014 - present)
- ◇ Visiting researcher: Machine learning for infection research (Projects: Anti-bacterial resistance prediction and analysis, Host phenotype prediction from microbial communities, and Deep learning for sequence imputation) *Bioinformatic group* (PI: Prof. A. McHardy), *Helmholtz Centre for Infection Research*, Braunschweig, Germany (June 2017 - Present)
- ◇ Visiting researcher: Character-level and subword-level deep language processing for super-parallel corpora of 1000 languages, *Deep Language Processing group* (PI: Prof. H. Schütze), *The Center for Information and Language Processing, LMU*, Munich, Germany (From June 2014 [continuously from February 2017] - January 2019)
- ◇ Part-time collaborator: Statistical language processing for cuneiform, *Near Eastern Studies Department* (PI: Prof. N. Veldhuis and Dr. Eduardo Escobar), *UC Berkeley*, CA, USA (February 2016 - January 2017)
- ◇ Intern researcher: Data analysis for detection of new trends in scientific communities, Deep Language Processing group (PI: Prof. H. Schütze), *The Center for Information and Language Processing, LMU*, Munich, Germany (June 2014 - August 2014)

- ◇ Visiting researcher: Structural and semantic analysis of narratives, *Computer Science and Artificial Intelligence Laboratory* (PIs: Prof. M. Finlayson and Prof. P. Winston), *Massachusetts Institute of Technology (MIT)*, MA, USA (September 2013 - March 2014)
- ◇ Visiting researcher: Neural signal processing for sleep phase detection, *Department of Brain and Cognitive Sciences* (PI: Prof. E. Brown), *MIT-Harvard*, MA, USA (May 2013 - September 2013)
- ◇ Research assistant: Efficient real-time ECG processing on mobile phones, *Audiovisual Communications Lab (PI: Prof. Vetterli)*, *EPFL*, Lausanne, Switzerland (January 2013 - May 2013)
- ◇ Research assistant: Eye-tracking supervision for brain-computer interface, *Non-Invasive BCI Lab* (PI: Prof. J. Milan), *EPFL*, Lausanne, Switzerland (December 2011 - July 2012)
- ◇ Research assistant: Computational semantic analysis of Persian poetry, *Artificial Intelligence Laboratory* (Dr. J-C. Chappelier), *EPFL*, Lausanne, Switzerland (August 2012 - February 2013)
- ◇ Research assistant: Enriching expert finding using social network matching, *Sharif University of Technology*, Tehran, Iran (July 2011 - September 2012)
- ◇ B.Sc. thesis: An analytical comparison of manifold learning algorithms, *Digital Media Lab (PI: Prof. H. Rabiee)*, *Sharif University of Technology*, Tehran, Iran (January 2011 - May 2011)
- ◇ Intern researcher: Efficient Visual Computing for Interactive Vision Applications on Mobile Devices, *Advanced Digital Sciences Center of University of Illinois at Urbana-Champaign (PI: Prof. M. Winslett)*, Singapore (July 2010 - September 2010)

#### AWARDS AND HONORS

- ◇ Ranked 1st and 2nd in two out of three tasks in The Critical Assessment of protein Function Annotation algorithms (CAFA) 3.14, 2019.
- ◇ Received German permanent residency as a highly skilled researcher, 2019.
- ◇ Awarded Deserve Scholarship, October 2018.
- ◇ Awarded Google Scholarship for **Machine Learning Summer School (LxMLS)**, Lisbon 2018.
- ◇ **Best Paper Award** Sponsored by **Google** at the Workshop on Multilingual and Cross-lingual Methods in Natural Language Processing at NAACL-HLT 2016, San Diego, USA.
- ◇ Selected among the **best submissions** at the **Computational Biology** workshop at **ICML 2016**, New York, USA.
- ◇ Awarded **Google Travel Grant** for attending International Conference on Machine Learning (ICML) 2016, New York, USA.
- ◇ **Awarded Graduate Division Fellowship, University of California, Berkeley**, 2014 - 2016.
- ◇ Ranked 1<sup>st</sup> in **Persian Literature and Linguistics Research**, High Council of Iranian Affairs Abroad, 2013.
- ◇ Awarded Summer Scholarship from ADSC (Advanced Digital Science Center), Research Center of the **University of Illinois at Urbana-Champaign** in Singapore, Summer 2010.
- ◇ Awarded **Outstanding Student** Certificate and Prize by Sharif University President, 2008.
- ◇ **Gold medal** in the Iranian National Olympiad in **Literature and Linguistics**, 2007.
- ◇ Four years grant for undergraduate studies from the **Iranian National Foundation of Elites**.

PROFESSIONAL  
ACTIVITIES &  
SERVICES

- ◇ **Committee Member** for
  - **Typology for Polyglot NLP** ACL workshop
  - Annual Meeting of the Association for Computational Linguistics (**ACL** 2019)
  - North American Chapter of the Association for Computational Linguistics (**NAACL** 2019)
  - Empirical Methods in Natural Language Processing **EMNLP** (2017, 2018, 2019)
  - The ACL 2019 Student Research Workshop (SRW)
  - ACL Workshop on Subword and Character level models in NLP **SCLeM** (2017 and 2018)
- ◇ **Reviewer** for:
  - **Natural Language Engineering Journal**
  - **Bioinformatics Journal**
  - **PLOS Computational Biology Journal**
  - **IEEE/ACM Transactions on Computational Biology and Bioinformatics**
  - International Joint Conference on Artificial Intelligence (**IJCAI** 2019)
  - Intelligent Systems for Molecular Biology **ISMB** (2018)
- ◇ **A Member of the Admission Committee**, Data Science M.Sc. program, UC Berkeley (2016)
- ◇ **Research Advising/Coadvising**
  - Andrew Dickson (January 2019 - present) - Deep learning for protein functional annotation
  - Nicolai-David Bernard Ruhnau (2018) - Distributional Semantics for the Computational Analysis of Food Corpora
  - Vikrum Nibber (2017) - Deep learning for Protein Structure Prediction.
  - Soroush Nasiriany (2016) - Mental Health Prediction Using NLP.
  - Kaveh Karbasi (2016) - Benchmarking Protein-to-Protein Interaction Prediction Methods.
  - Amine Mezzour (2015) - Deep Convolutional Network for RNA Structure Prediction.
  - Sam Njie (2014) - Working with Gene Ontology.
- ◇ **Student Volunteer**
  - Empirical Methods on Natural Language Processing, Copenhagen, Denmark (September 2017)
  - North American Chapter of the Association for Computational Linguistics Conference (NAACL):
    - NAACL-HLT 2016, San Diego, CA (June 2016)
    - NAACL-HLT 2013, Atlanta, GA (June 2013)

TEACHING  
EXPERIENCE

- ◇ **Graduate Student Instructor**, University of California, Berkeley
  - Data Structures and Algorithms: *Information School (ISchool)* (Summer 2016)
  - Python, Matlab, and Java Instructor: *Self-Paced Center of EECS Department* (Spring 2016)
- ◇ Undergraduate Teaching Assistant, Sharif University of Technology, (2008 - 2011)
  - Probability and Statistics (Fall 2010) & Digital Signal Processing (Spring 2011)
  - Scientific and Technical Presentations (Spring 2009) & Web Programming (Spring 2011)

## SKILLS

- ◇ **Programming Languages**: Python, Java, C++, R, and Matlab.
- ◇ **Python Machine Learning Libraries/Frameworks**: PyTorch, Keras, NLTK, Scikit-learn, and Pandas.
- ◇ **Web/DB Technologies**: Languages: PHP, MYSQL, Javascript, CSS, and HTML; Frameworks: Symfony, JQuery, and Django.
- ◇ **Languages**: Persian (Native), English (Fluent), French (Intermediate), German (Intermediate), Arabic (Linguistic knowledge).