Amirhossein Abaskohi *September 19, 2000*

Masters Student Majoring in Computer Science

🌐 Website | ✉ aabaskoh@cs.ubc.ca | ⌨ AmirAbaskohi

in amirhossein-abaskohi | 🌐 Google Scholar

Education

University Of British Columbia	Sept 2023 - Now

M.Sc in Computer Science

Department of Computer Science

- **GPA:** A+
- Member of UBC NLP Group

University Of Tehran	Sept 2018 - Apr 2023

B.Sc in Computer Engineering — *Ranked top 10% in class*

Department of Electrical and Computer Engineering

- **Cumulative GPA:** 17.94/20.00 (3.8/4) [Last Two Years GPA: 18.42 (4/4)] (Faculty Average: 15.10/20.00)

Research Interests

- Natural Language Generation
- Cross-lingual NLP
- Ethics and Fairness in ML
- NLP for Common-sense Reasoning
- Multi-modal Reasoning
- Human Centered AI

Publications

FM^{2}DS: Few-Shot Multimodal Multihop Data Synthesis with Knowledge Distillation for Question Answering — 2024

Amirhossein Abaskohi, Spandana Gella, Giuseppe Carenini, Issam H. Laradji
Submitted to CVPR 205

BigDocs: An Open and Permissively-Licensed Dataset for Training Multimodal Models on Document and Code Tasks — 2024

Juan A. Rodriguez, Xiangru Jian, Siba Smarak Panigrahi, Tianyu Zhang, Aarash Feizi, Abhay Puri, Akshay Kalkunte, Francois Savard (First Author Equal Contribution), **Amirhossein Abaskohi** (**Second Author Equal Contribution**), Ahmed Masry, Perampalli Shravan Nayak, Mahsa Massoud, Rabiul Awal, Pierre-André Noël, Mats L. Richter, Saverio Vadacchino, Shubham Agarwal, Sanket Biswas, Ying Zhang, Sathwik Tejaswi Madhusudhan, João Monteiro, Krishnamurthy (Dj) Dvijotham, Torsten Scholak, Nicolas Chapados, Sean Hughes, Tamer Özsu, Aishwarya Agrawal, Marco Pedersoli, Christopher Pal, Perouz Taslakian, David Vazquez, Issam H. Laradji, Spandana Gella, Sai Rajeswar Mudumba
RBFM@NeurIPS2024

InsightBench: Evaluating Business Analytics Agents Through Multi-Step Insight Generation — 2024

Gaurav Sahu, Abhay Puri, Juan Rodriguez, **Amirhossein Abaskohi**, Mohammad Chegini, Alexandre Drouin, Perouz Taslakian, Valentina Zantedeschi, Alexandre Lacoste, David Vazquez, Nicolas Chapados, Christopher Pal, Sai Rajeswar Mudumba, Issam Hadj Laradji
arXiv preprint arXiv:2407.06423

Benchmarking Large Language Models for Persian: A Preliminary Study Focusing on ChatGPT — 2024

Amirhossein Abaskohi, Sara Baruni, Mostafa Masoudi, Nesa Abbasi, Mohammad Hadi Babalou, Ali Edalat, Sepehr Kamahi, Samin Mahdizadeh Sani, Nikoo Naghavian, Danial Namazifard, Pouya Sadeghi, and Yadollah Yaghoobzadeh
LREC-COLING 2024

Beyond Words: A Multimodal and Multilingual Exploration of Persuasion in Memes — 2024

Amirhossein Abaskohi[*], Amirhossein Dabiriaghdam[*], Lele Wang, Giuseppe Carenini
International Workshop on Semantic Evaluation co-located with NAACL 2024

Can LLMs be Lateral Thinkers? — 2024

Pouya Sadeghi [*] **Amirhossein Abaskohi**[*], Yadollah Yaghoobzadeh
International Workshop on Semantic Evaluation co-located with NAACL 2024

Paraphrasing-Guided Data Augmentation for Contrastive Prompt-Based Few-Shot Fine-Tuning — 2023

Amirhossein Abaskohi, Yadollah Yaghoobzadeh, Sascha Rothe
Annual Meeting of the Association for Computational Linguistics (ACL) 2023

A Large-Scale Analysis of Persian Tweets Regarding Covid-19 Vaccination 2023

Taha ShabaniMirzaei, Houmaan Chamani, **Amirhossein Abaskohi**, Zhivar Sourati Hassan Zadeh, Behnam Bahrak
Springer's Social Network Analysis and Mining Journal

PEACH: Pre-Training Sequence-to-Sequence Multilingual Models for Translation with Semi-Supervised Pseudo-Parallel Document Generation 2023

Amirhossein Abaskohi*, Alireza Salemi*, Sara Tavakoli, Yadollah Yaghoobzadeh, Azadeh Shakeri
Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT) co-located with EACL 2023

Persian Emotion Detection using ParsBERT and Imbalanced Data Handling Approaches 2022

Amirhossein Abaskohi, Nazanin Sabri, Behnam Bahrak
arXiv preprint arXiv:2211.08029

A Comparative Analysis of Sarcasm Detection Using Generative-based and Mutation-based Data Augmentation 2022

Amirhossein Abaskohi, Arash Rasouli*, Tanin Zeraati*, Behnam Bahrak
International Workshop on Semantic Evaluation co-located with NAACL 2022

Automatic Speech recognition for Speech Assessment of Persian Preschool Children 2022

Amirhossein Abaskohi, Fatemeh Mortazavi, Hadi Moradi
arXiv 2203.12886

Research Experience

Graduate research assistant at the University of British Columbia MARCH 2024 - Now
Under Supervision of Prof. Giuseppe Carenini

I work on multimodal reasoning over long and multiple documents. My work includes developing methods for multimodal multi-hop question answering and topic modeling.

Research Intern at ServiceNow Research JUN 2024 - Now
Under Supervision of Prof. Issam H. Laradji and Under Supervision of Prof. Giuseppe Carenini

I work on foundational large language models and advanced data analysis, utilizing LLMs to explore and interpret complex datasets. This involves designing methods to harness the capabilities of these models for deeper insights and more effective data understanding.

Graduate research assistant at the University of British Columbia JAN 2024 - DEC 2024
Under Supervision of Prof. Giuseppe Carenini, Prof. Brad Hackinen, and Prof. Francesco Trebbi

This collaborative endeavor involves Western University's Ivey School and the University of California, Berkeley, focusing on regulatory documents. Our objective is to examine how various writing styles utilized in rule comments affect the responses received from regulators.

Graduate research assistant at the University of British Columbia JAN 2024 - DEC 2024
Under Supervision of Prof. Giuseppe Carenini, Prof. Brad Hackinen, and Prof. Francesco Trebbi

This collaborative endeavor involves Western University's Ivey School and the University of California, Berkeley, focusing on regulatory documents. Our objective is to examine how various writing styles utilized in rule comments affect the responses received from regulators.

Undergraduate research assistant at the University of Tehran JUL 2022 - MAR 2023
Under Supervision of Prof. Yadollah Yaghoobzadeh and Dr. Sascha Rothe

As part of this project, I delved into the few-shot fine-tuning of smaller language models such as RoBERTa. This involved employing paraphrase generation techniques with large language models like GPT-3, as well as leveraging contrastive learning methods.

Undergraduate research assistant at the University of Tehran NOV 2021 - JUN 2022
Under Supervision of Prof. Azadeh Shakery and Prof. Yadollah Yaghoobzadeh

I was working on creating pseudo-translation documents for the pre-training stage of a multilingual transformer-based language model. We mapped sentences from one language to another by using bilingual dictionaries with distinct monolingual corpora. The quality of mapped sentences is improved by training a model reordering to enhance the quality of sentences for a language.

Undergraduate research assistant SEPT 2021 - JUN 2022
Under Supervision of Prof. Pedram Rooshenas

During this remote research experience, I developed a VAE-based generative model to simultaneously produce a picture and its caption. We had two goals: 1) Checking the effect of text and image fusion in the latent space on the quality of generative models, and 2) Using the model to create artificial image captioning datasets.

Undergraduate research assistant at the University of Tehran
AUG 2021 - MAR 2023

Under Supervision of Prof. Behnam Bahrak

During my tenure at the University of Tehran's Data Analytics Lab, I've undertaken various projects. These include binary sarcasm classification, analyzing Persian tweets amidst the Covid pandemic in Iran, and Persian emotion detection.

Undergraduate research assistant at the University of Tehran
NOV 2020 - SEPT 2021

Under Supervision of Prof. Hadi Moradi

We created an AI-driven cognitive assessment platform tailored for preschoolers to gauge their school readiness. Initially, I focused on the backend development of the website, implementing an AI-powered testing system. Subsequently, we proposed a novel pre-training objective for the Wav2Vec model to enhance the ASR system used in speech-based tests, achieving a cutting-edge model for automatic speech recognition in Persian.

Volunteering

- Artificial Intelligence and Machine Learning Mentor in Summer School 2021   SUMMER 2021
- Freelance Content Creation on Medium   JUL 2021-PRESENT
- Freelance Content Creation on YouTube   JUL 2021-PRESENT
- Working on The Backend API System of Cognitive Tests For Preschool Children Project   NOV 2020-FEB 2021

Honors and Awards

Mitacs Fellowship
Jun 2024

I was honored to receive the Mitacs Fellowship, which provided me with a unique opportunity to collaborate with industry partners on cutting-edge research projects. This experience allowed me to apply my academic knowledge to real-world challenges, enhancing my skills in problem-solving and innovation.

UBC International Tuition Award
Feb 2023

Recipient of the UBC International Tuition Award, recognizing academic excellence and providing financial support to eligible international graduate students at UBC.

Admission to the University of British Columbia Computer Science department
Feb 2023

Successfully admitted to the Computer Science department at the University of British Columbia, recognized as one of the top computer science departments globally. This achievement reflects academic excellence and underscores the rigorous standards met to gain acceptance into a prestigious institution known for its cutting-edge research and innovative programs.

Supporter Foundation of University of Tehran Scholarship
Sept 2020

Awarded to the top 500 students among 35000 students in the university.

Best Undergraduate Project Award
Sept 2022

My project on "Pre-Training Sequence-to-Sequence Multilingual Models for Translation with Semi-Supervised Pseudo-Parallel Document Generation" has been awarded as the best undergraduate project in the 16th Project Day competition of ECE school of the University of Tehran.

Admission to University of Tehran
Sept 2018

Ranked 46th (regional rank), and 1125th (national rank) among 144,437 participants in the Iranian Nationwide University Entrance Exam for Mathematics and Physics discipline.

Iran's Elites Foundation Membership and Financial Support
Nov 2021

After the Automatic Speech Recognition for Preschool Children project, I have been awarded a 1-year scholarship from this foundation.

Skills

Programming Languages: Python, C\C++, SQL, C#, Java, Dart, JavaScript, GNU Octave, MATLAB, Verilog HDL

Technologies: Git, Docker, Jupyter, Google Colab, LaTeX

Tools: Tensorflow, Matplotlib, Pytorch, Jupyter, CUDA, Django, Flask, GraphQL, Apache Airflow, PowerBI

Operating Systems: Linux (Ubuntu, Parrot, Kali), Windows, Cloudera

Work Experience

Back-end Developer   Nov 2022 - Oct 2023

Mate Machine Inc

- **Skills used:** ASP.NET, MicoService, Docker, CI/CD, EntityFramework, GRPC, RabbitMq, Grafana, DataDog, PowerBI, Swagger, Agile, Scrum

Back-end Developer   Dec 2022 - May 2023

Ramand

- **Skills used:** ASP.NET, RabbitMq, Dapper, Agile, Scrum, Grafana

Machine Learning Researcher   Sept 2021 - Jan 2023

University of Tehran Science and Technology Park

- **Skills used:** Python, Research, Tensorflow, PyTorch, NLP, Data Analysis

Full-stack developer internship   Jun 2022 - Sept 2022

System Group

- **Skills used:** ASP.NET, CI/CD, Angular

Financial data science remote internship   Aug 2021 - Oct 2021

World Data Science Institute

- **Skills used:** Python, Tensorflow, Research, Project Leadership, Flask, Django, GraphQL

Back-end Developer   Jun 2020 - Mar 2021

Idea Varzan System

- **Skills used:** ASP.NET, Swagger, SeqLog, Agile, Scrum, Grafana, Dart