

Sabit Hassan | Curriculum Vitae

Address: 247 Republic St, Pittsburgh, 15211 PA

Mobile: +1 412 9839715 | E-mail: sabithassan64@gmail.com

EDUCATION

PhD in Computer Science *University of Pittsburgh* Expected Jan, 2025
Dissertation: Leveraging Active Learning for Safe, Inclusive, and Democratic Conversational AI
Committee: Malihe Alikhani (Co-Chair), Jacob Biehl (Co-Chair), Diane Litman, Xiaowei Jia

BSc in Computer Science *Carnegie Mellon University* 2014–2018
GPA: 3.79/4.0, Minors: Mathematics and History
Thesis: Interactive Evaluation and Training of Classifiers under Limited Resources,
Advisors: Bhiksha Raj and Saquib Razak

HONORS AND AWARDS

Computer Science Dept. Fellowship	<i>University of Pittsburgh</i>	Aug 2023 – Apr 2024
Alexa Prize Taskbot Challenge 3rd Place Award	<i>Amazon</i>	Sep 2023
Alexa Prize Taskbot Challenge Grant	<i>Amazon</i>	Jan 2023
Outstanding Academic Achievement	<i>Carnegie Mellon University Qatar</i>	May 2018
University Honors	<i>Carnegie Mellon University</i>	May 2018
SCS College Honors	<i>Carnegie Mellon University</i>	May 2018
History Honors	<i>Phi Alpha Theta</i>	May 2018
Carnegie Mellon Dean's List	<i>Carnegie Mellon University</i>	Apr 2015–Apr 2018

RESEARCH

Doctoral Research *University of Pittsburgh, Pittsburgh, PA* Aug 2021–present
Computer Science Department, Supervised by: Dr. Malihe Alikhani

ISABEL: IncluSive And collaBorative Alexa [Funded by Amazon]

- Developed a safety framework and dataset with knowledge distillation and social media data
- Proposed a novel search algorithm leveraging knowledge distillation and clustering

AI Moderation for Safer Online Space [Funded by DARPA]

- Developed a discourse-aware framework for counterspeech generation
- Integrated RST and PDTB frameworks with LLMs for better content preservation during style-transfer
- Curated a large-scale multilingual dataset from Reddit for AI moderation

Active Learning for NLP

- Developed a novel dynamic clustering-based active learning algorithm which is robust against bias
- Formulated effective active learning for generative task for the first time

Other Projects:

- Improved prosody in sign language generation by incorporating intensity-enhanced GLOSS representations within a novel Progressive Transformer model.
- Modeled socio-affective and cognitive engagement in patient-nurse conversation

Research Assistant *Qatar Computing Research Institute, Doha, Qatar* Oct 2019–Jun 2021
Arabic Language Technologies Group, Supervised by: Dr. Kareem Darwish

- Trained SOTA ML/transformer models and collected datasets for Arabic dialect identification, offensive language, adult content, spam, and hate-speech detection..
- Analyzed efficacy of cross-lingual approaches for Spanish, Arabic, and English emotion detection
- Analyzed the role of data diversity and segmentation for pretraining a SOTA Arabic BERT model

Research Associate *Carnegie Mellon University Qatar, Doha, Qatar* May 2018–Aug 2019
MADAR Project, Supervised by: Prof. Nizar Habash and Asst. Teach. Prof. Houda Bouamor

- Co-organized MADAR Shared Task on fine-grained Arabic Dialect Identification

Minicomplexity Project, Supervised by: Assoc. Teach. Prof. Christos Kapoutsis

- Characterized Polynomial-size Oracle Hierarchy for small one-way finite automata and proved key properties of the hierarchy, such as strictness, upper, and lower bounds of levels of the hierarchy.

SCS Honors Undergraduate Thesis *Carnegie Mellon University* Aug 2017–Apr 2018
School of Computer Science, Supervised by: Prof. Bhiksha Raj and Assoc. Teach. Prof. Saquib Razak

- Proposed a clustering-based method that reduces training time for *Active Learning*.
- Proposed a novel clustering-based algorithm for interactive evaluation of classifiers.

TEACHING

Guest Lectures

- Active Learning for Democratizing Safety and Fairness in LLMs **Northeastern University**, Mar 22, 2024
- Causal inference in LLMs, **Northeastern University**, Mar 19, 2024
- Machine Translation, **University of Pittsburgh**, Nov 15, 2023

Graduate Teaching Assistant *University of Pittsburgh, Pittsburgh, PA* Aug 2021–Aug 2024

- CS 1501: Algorithms and Data Structures 2, CS 2731: Introduction to Natural Language Processing, CS 1520: Introduction to Web Applications

Course Assistant *Carnegie Mellon University Qatar, Doha, Qatar* Jan 2016–May 2018

- CMUQ Freshman Edge Summer Program, 15–251: Great Theoretical Ideas in Computer Science, 15–210: Parallel and Sequential Data Structures and Algorithms, 15–112: Fundamentals of Programming and Computer Science, 21–127: Concepts of Mathematics

Summer Internship Mentor *Qatar Computing Research Institute, Doha, Qatar* May 2020–Jul 2020

- Mentored interns on the use of machine learning in NLP tasks

SERVICES

- **Co-organizer**, MADAR Shared Task on Fine-Grained Arabic Dialect Identification, 2019
- **Program Committee/Reviewer**: ARR, Journal of NLP, ACL, EACL, EMNLP, IJCNLP–AAACL, IEEE MM
- **Member**, Student Academic Committee, Academic Review Board, Carnegie Mellon University Qatar

PROJECTS

FARASA *Qatar Computing Research Institute, Doha, Qatar* 2020–2021
Arabic Language Technologies Group, Supervised by: Dr. Kareem Darwish

- Assisted in frontend development of FARASA toolkit (<https://farasa.qcri.org/>)
- Implemented API quota allocation and managed SQL database of requests and users

ASAD *Qatar Computing Research Institute, Doha, Qatar* 2020–2021
Arabic Language Technologies Group, Supervised by: Dr. Kareem Darwish

- Deployed SVM/BERT models for text classification within the social media analysis toolkit ASAD (<https://asad.qcri.org/>)
- Developed initial frontend and backend for ASAD.

Early Warning Simulation *Qatar Computing Research Institute, Doha, Qatar* 2020–2021
Arabic Language Technologies Group, Supervised by: Dr. Kareem Darwish

- Developed a system that crawls newspapers articles and identifies antagonism level against Qatar
- Developed frontend and managed MongoDB database of newspapers articles

PUBLICATIONS [\[Google Scholar\]](#)

Published/Accepted:

* denotes equal contribution

- [1] Sicilia, A., Asano, Y., Atwell, K., Cheng, Q., Gupta, D., **Hassan, S.** Inan, M., Nwogu, J., Sharma, P., Alikhani, M. *ISABEL: An Inclusive and Collaborative Task-Oriented Dialogue System*. **Amazon Science, 2023**. [3rd place award in Alexa Prize Taskbot Challenge]
- [2] **Hassan, S.** & Alikhani, M. DisCGen: A Framework for Discourse-Aware Counterspeech Generation. **IJCNLP-AACL, 2023**
- [3] **Hassan, S.** & Alikhani, M. D-CALM: A Dynamic Clustering-based Active Learning Approach for Mitigating Bias. **ACL FINDINGS, 2023**
- [4] Wang, Y., Donovan, H.A.S., **Hassan, S.**, Alikhani, M. MedNgage: A Dataset for Understanding Engagement in Patient-Nurse Conversations. **ACL FINDINGS, 2023**
- [5] Ye, M., Sikka, K., Atwell, K., **Hassan, S.**, Divakaran, A., Alikhani, M. Multilingual Content Moderation and Challenges: A Case Study of Reddit. **EACL, 2023**
- [6] Mubarak, H., **Hassan, S.**, & Chowdhury, S. Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech. **NLE, 2023**.
- [7] Atwell, K.*, **Hassan, S.***, Alikhani, M. APPDIA: A Discourse-aware Transformer-based Style Transfer Model for Offensive Social Media Conversations. **COLING, 2022**
- [8] Inan, M.*, Yang, Z.*, **Hassan, S.***, Quandt, L., & Alikhani, M. Modeling Intensification for Signed Language Generation: A Computational Approach. **ACL FINDINGS, 2022**
- [9] **Hassan, S.**, Shaar, S., & Darwish, K. Cross-lingual Emotion Detection. **LREC, 2022**
- [10] Mubarak, H., **Hassan, S.**, Chowdhury, S. A., & Alam, F. ArCovidVac: Analyzing Arabic Tweets About COVID-19 Vaccination. **LREC, 2022**.
- [11] **Hassan, S.**, Mubarak, H., Abdelali, A., & Darwish, K. ASAD: Arabic Social Media Analytics and unDerstanding. **EACL 2021**
- [12] Mubarak, H., **Hassan, S.**, & Abdealali, A. Adult Content Detection on Arabic Twitter: Analysis and Experiments. **WANLP 2021**
- [13] Abdelali, A., Mubarak, H., Samih, Y., **Hassan, S.**, & Darwish, K. QADI: Arabic Dialect Identification in the Wild. **WANLP 2021**
- [14] Mubarak, H., & **Hassan, S.** (2021, April). ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic. **LOUHI, 2021**
- [15] Mubarak, H., Abdelali, A., **Hassan, S.**, & Darwish, K. Spam Detection on Arabic Twitter. **SocInfo, 2020**
- [16] **Hassan, S.**, Samih, Y., Mubarak, H., & Abdelali, A. ALT at SemEval-2020 task 12: Arabic and English offensive language identification in social media. **SemEval, 2020**
- [17] **Hassan, S.**, Samih, Y., Mubarak, H., Abdelali, A., Rashed, A., & Chowdhury, S. A. ALT submission for OSACT shared task on offensive language detection. **OSACT, 2020** [Ranked 1st in subtask A]
- [18] Mubarak, H., **Hassan, S.**, & Abdelali, A. Constructing a bilingual corpus of parallel tweets. **BUCC, 2020**
- [19] Bouamor, H., **Hassan, S.**, & Habash, N. The MADAR Shared task on Arabic Fine-Grained Dialect Identification. **WANLP, 2019**. [co-organizer]
- [20] Anabtawi, M., **Hassan, S.**, Kapoutsis, C., & Zakzok, M. An Oracle Hierarchy for Small One-way Finite Automata. **LATA, 2019**
- [21] **Hassan, S.**, Shaar, S., Raj, B., & Razak, S. (2018, December). Interactive Evaluation of Classifiers Under Limited Resources. **IEEE ICMLA, 2018**

Pre-prints:

- [22] Abdelali, A., **Hassan, S.**, Mubarak, H., Darwish, K., & Samih, Y. (2021) Pre-Training BERT on Arabic Tweets: Practical Considerations. *arXiv preprint arxiv:2102.10684*

Abstracts:

- [23] **Hassan, S.***, Atwell, KJ.*, Alikhani, M. Studying the Effect of Moderator Biases on the Diversity of Online Discussions: A Computational Cross-linguistic Study. **CogSci, 2022**
- [24] Inan, M.*, Zhong, Y.*, **Hassan, S.***, Quandt, L., Alikhani, M. Learning cognitive and linguistic prosodic categories for automatic cross-lingual sign language understanding. **CogSci, 2022**