

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 |
| Kuzneski Innovation Cup Award | <i>Big Idea Center</i> | Nov 2022 |
| 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* Aug 2021–

- Generative active learning, generative AI safety, dialogue systems, embodied agents,
- Sign language generation, content moderation, contextual ASR, NLP in healthcare

Research Assistant *Qatar Computing Research Institute* Oct 2019–Jun 2021

- Pre-trained language model ([QARiB](#))
- Dialect identification, offensive language, cross-lingual emotion detection
- Social Media Analytics Suite ([ASAD](#)), NLP processing tool ([Farasa](#))

Research Associate *Carnegie Mellon University Qatar* May 2018–Sep 2021

- Co-organized MADAR shared task on Arabic dialect identification
- Characterized oracle hierarchy for small one-way finite automata

TEACHING

Teaching Fellow *University of Pittsburgh* Jan 2025–

- Course instructor for CS 0012: Introduction to Computing for Humanities

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* Aug 2021–Aug 2024

- Intro to Natural Language Processing, Algorithms and Data Structures, Intro to Web Applications

Teaching Assistant

Carnegie Mellon University

Jan 2016–May 2018

- Freshman Edge Summer Program, Great Theoretical Ideas in CS, Parallel and Sequential Data Structures and Algorithms, Fundamentals of Programming and CS, Concepts of Mathematics

Summer Internship Mentor

Qatar Computing Research Institute

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, ACL, EACL, EMNLP, IJCNLP-AACL, Journal of NLP, IEEE MM, Heilyon, ICARD
- **Member**, Student Academic Committee, Academic Review Board, Carnegie Mellon University Qatar

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

Published/Accepted:

* denotes equal contribution

- [1] **Hassan, S.**, Chung, H., Tan, Z., & Alikhani, M. *Coherence-driven Multimodal Safety Dialogue for Embodied Agents*. **AAMAS, 2025**
- [2] **Hassan, S.**, Sicilia, A., & Alikhani, M. *An Active Learning Framework for Inclusive Generation by Large Language Models*. **COLING, 2025**
- [3] Asano, Y., **Hassan, S.**, Sharma, P., Sicilia, A., Atwell, K., Litman, D., & Alikhani, M. *Contextual ASR Error Handling with LLMs Augmentation for Goal-Oriented Conversational AI*. **COLING, 2025**
- [4] **Hassan, S.**, Sicilia, A., & Alikhani, M. *Active Learning for Robust and Representative LLM Generation in Safety-Critical Scenarios*. **Customizable NLP @ EMNLP 2024 [Outstanding Paper Award]**
- [5] 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. [Alexa Prize Taskbot Challenge Award]**
- [6] **Hassan, S.** & Alikhani, M. *DisCGen: A Framework for Discourse-Aware Counterspeech Generation*. **IJCNLP-AACL, 2023**
- [7] **Hassan, S.** & Alikhani, M. *D-CALM: A Dynamic Clustering-based Active Learning Approach for Mitigating Bias*. **ACL FINDINGS, 2023**
- [8] Wang, Y., Donovan, H.A.S., **Hassan, S.**, Alikhani, M. *MedNgage: A Dataset for Understanding Engagement in Patient-Nurse Conversations*. **ACL FINDINGS, 2023**
- [9] Ye, M., Sikka, K., Atwell, K., **Hassan, S.**, Divakaran, A., Alikhani, M. *Multilingual Content Moderation and Challenges: A Case Study of Reddit*. **EACL, 2023**
- [10] Mubarak, H., **Hassan, S.**, & Chowdhury, S. *Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech*. **NLE, 2023**.
- [11] Atwell, K.*, **Hassan, S.***, Alikhani, M. *APPDIA: A Discourse-aware Transformer-based Style Transfer Model for Offensive Social Media Conversations*. **COLING, 2022**
- [12] Inan, M.*, Yang, Z.*, **Hassan, S.***, Quandt, L., & Alikhani, M. *Modeling Intensification for Signed Language Generation: A Computational Approach*. **ACL FINDINGS, 2022**
- [13] **Hassan, S.**, Shaar, S., & Darwish, K. *Cross-lingual Emotion Detection*. **LREC, 2022**
- [14] **Hassan, S.**, Mubarak, H., Abdelali, A., & Darwish, K. *ASAD: Arabic Social Media Analytics and unDerstanding*. **EACL 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 at OSACT Shared Task on Offensive Language Detection*. **OSACT, 2020 [Ranked 1st in subtask A]**

- [18] Bouamor, H., **Hassan, S.**, & Habash, N. *The MADAR Shared task on Arabic Fine-Grained Dialect Identification*. **WANLP, 2019. [co-organizer]**
- [19] Anabtawi, M., **Hassan, S.**, Kapoutsis, C., & Zakzok, M. *An Oracle Hierarchy for Small One-way Finite Automata*. **LATA, 2019**
- [20] **Hassan, S.**, Shaar, S., Raj, B., & Razak, S. *Interactive Evaluation of Classifiers Under Limited Resources*. **IEEE ICMLA, 2018**

Pre-prints:

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

Abstracts:

- [22] **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**
- [23] 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**