

# Sabit Hassan | Curriculum Vitae

Address: 14 Cable PL, Pittsburgh, 15213 PA

Mobile: +1 412 9839715 | E-mail: [sabithassan64@gmail.com](mailto:sabithassan64@gmail.com)

---

## EDUCATION

<b>PhD in Computer Science</b> <i>Supervised by: Dr. Malihe Alikhani</i>	<i>University of Pittsburgh</i>	2021–
<b>BSc in Computer Science</b> <i>GPA: 3.79/4.0, Minors: Mathematics and History</i> <i>Thesis: Interactive Evaluation and Training of Classifiers under Limited Resources,</i> <i>Supervised by: Prof. Bhiksha Raj and Assoc. Teach. Prof. Saquib Razak</i>	<i>Carnegie Mellon University</i>	2014–2018

---

## HONORS

<b>Outstanding Academic Achievement</b>	<i>Carnegie Mellon University Qatar</i>	May 2018
<b>University Honors</b>	<i>Carnegie Mellon University</i>	May 2018
<b>SCS College Honors</b>	<i>Carnegie Mellon University</i>	May 2018
<b>History Honors</b>	<i>Phi Alpha Theta</i>	May 2018
<b>Carnegie Mellon Dean's List</b>	<i>Carnegie Mellon University</i>	Apr 2015–Apr 2018

---

## RESEARCH

<b>Doctoral Research</b> <i>Computer Science Department, Supervised by: Dr. Malihe Alikhani</i>	<i>University of Pittsburgh, Pittsburgh, PA</i>	Aug 2021–
<ul style="list-style-type: none"><li>Developed discourse-aware style-transfer models by integrating PDTB and RST frameworks with pretrained transformer models for better content preservation.</li><li>Improved prosody in sign language generation by incorporating intensity-enhanced GLOSS representations within a novel Progressive Transformer model.</li></ul>		
<b>Research Assistant</b> <i>Arabic Language Technologies Group, Supervised by: Dr. Kareem Darwish</i>	<i>Qatar Computing Research Institute, Doha, Qatar</i>	Oct 2019–Jun 2021
<ul style="list-style-type: none"><li>Trained SOTA ML/transformer models and collected datasets for Arabic dialect identification, offensive language, adult content, spam, and hate-speech detection..</li><li>Studied efficacy of cross-lingual approaches for Spanish, Arabic, and English emotion detection</li><li>Studied importance of data diversity and segmentation for pretraining a SOTA Arabic BERT model</li></ul>		
<b>Research Associate</b> <i>MADAR Project, Supervised by: Prof. Nizar Habash and Asst. Teach. Prof. Houda Bouamor</i>	<i>Carnegie Mellon University Qatar, Doha, Qatar</i>	May 2018–Aug 2019
<ul style="list-style-type: none"><li>Co-organized MADAR Shared Task on fine-grained Arabic Dialect Identification</li></ul> <i>Minicomplexity Project, Supervised by: Assoc. Teach. Prof. Christos Kapoutsis</i> <ul style="list-style-type: none"><li>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.</li></ul>		
<b>SCS Honors Undergraduate Thesis</b> <i>School of Computer Science, Supervised by: Prof. Bhiksha Raj and Assoc. Teach. Prof. Saquib Razak</i>	<i>Carnegie Mellon University</i>	Aug 2017–Apr 2018
<ul style="list-style-type: none"><li>Proposed a clustering-based method that reduces training time for <i>Active Learning</i>.</li><li>Proposed a novel clustering-based algorithm for interactive evaluation of classifiers.</li></ul>		

---

## TEACHING

- Graduate Teaching Assistant** *University of Pittsburgh, Pittsburgh, PA* Aug 2021–  
• Algorithms and Data Structures 2 (CS 1501)
- Course Assistant** *Carnegie Mellon University Qatar, Doha, Qatar* Jan 2016–May 2018  
• CMUQ Freshman Edge Summer Program  
• Great Theoretical Ideas in Computer Science (15–251)  
• Parallel and Sequential Data Structures and Algorithms (15–210)  
• Fundamentals of Programming and Computer Science (15–112)  
• Concepts of Mathematics (21–127)
- Summer Internship Mentor** *Qatar Computing Research Institute, Doha, Qatar* May 2020–Jul 2020  
Mentored interns on the use of machine learning in NLP tasks

---

## AWARDS AND GRANTS

- **Team Lead**, Kuzneski Innovation Cup – \$5K (prize)
- **Team Member**, Amazon Taskbot Challenge – \$250K (grant)

---

## SERVICES

- **Co-organizer**, MADAR Shared Task on Fine-Grained Arabic Dialect Identification, 2019
- **Reviewer**: EACL 2022, EMNLP 2022, AACL–22 Special Track on AI for Social Impact, SemEval 2020
- **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

### Published/Accepted:

\* denotes equal contribution

- [1] Ye, M., Sikka, K., Atwell, K., **Hassan, S.**, Divakaran, A., Alikhani, M. Multilingual Content Moderation and Challenges: A Case Study of Reddit. *EACL*, **2023**
- [2] Atwell, K. \*, **Hassan, S. \***, Alikhani, M. APPDIA: A Discourse-aware Transformer-based Style Transfer Model for Offensive Social Media Conversations. *COLING*, **2022**
- [3] Inan, M. \*, Yang, Z. \*, **Hassan, S. \***, Quandt, L., & Alikhani, M. Modeling Intensification for Signed Language Generation: A Computational Approach. *ACL FINDINGS*, **2022**

- [4] Mubarak, H., **Hassan, S.**, & Chowdhury, S. Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech. **NLE, 2022.**
- [5] **Hassan, S.**, Shaar, S., & Darwish, K. Cross-lingual Emotion Detection. **LREC, 2022**
- [6] Mubarak, H., **Hassan, S.**, Chowdhury, S. A., & Alam, F. ArCovidVac: Analyzing Arabic Tweets About COVID-19 Vaccination. **LREC, 2022.**
- [7] **Hassan, S.**, Mubarak, H., Abdelali, A., & Darwish, K. ASAD: Arabic Social Media Analytics and unDerstanding. **EACL 2021**
- [8] Mubarak, H., **Hassan, S.**, & Abdealali, A. Adult Content Detection on Arabic Twitter: Analysis and Experiments. **WANLP 2021**
- [9] Abdelali, A., Mubarak, H., Samih, Y., **Hassan, S.**, & Darwish, K. QADI: Arabic Dialect Identification in the Wild. **WANLP 2021**
- [10] Mubarak, H., & **Hassan, S.** (2021, April). ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic. **LOUHI, 2021**
- [11] Mubarak, H., Abdelali, A., **Hassan, S.**, & Darwish, K. Spam Detection on Arabic Twitter. **SocInfo, 2020**
- [12] **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**
- [13] **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]
- [14] Mubarak, H., **Hassan, S.**, & Abdelali, A. Constructing a bilingual corpus of parallel tweets. **BUCC, 2020**
- [15] Bouamor, H., **Hassan, S.**, & Habash, N. The MADAR Shared task on Arabic Fine-Grained Dialect Identification. **WANLP, 2019. [co-organizer]**
- [16] Anabtawi, M., **Hassan, S.**, Kapoutsis, C., & Zakzok, M. An Oracle Hierarchy for Small One-way Finite Automata. **LATA, 2019**
- [17] **Hassan, S.**, Shaar, S., Raj, B., & Razak, S. (2018, December). Interactive Evaluation of Classifiers Under Limited Resources. **IEEE ICMLA, 2018**

#### Pre-prints:

- [18] 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:

- [19] **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**
- [20] 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**