

Sabit Hassan | Curriculum Vitae

Address: 14 Cable PL, Pittsburgh, 15213 PA

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

EDUCATION

PhD in Computer Science <i>Supervised by: Dr. Malihe Alikhani</i>	<i>University of Pittsburgh</i>	2021–
BSc in Computer Science <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 AND AWARDS

Outstanding Academic Achievement Award	<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

PhD Research <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">Developed discourse-aware models for supervised style transfer by integrating PDTB and RST frameworks with pretrained transformer models for better content preservation in style transferred text.Improved prosody in sign language generation by developing strategies grounded in linguistics for incorporating enhanced gloss representations within a novel Progressive Transformer model.		
Research Assistant <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">Built resources and trained SOTA ML/transformer models for Arabic offensive language, adult content, spam, and hate-speech detection, emotion analysis, dialect identification and analysis of COVID-19 related tweets.Studied efficacy of cross-lingual approaches for Spanish, Arabic, and English emotion detectionStudied importance of data diversity and linguistically aware segmentation for pretraining a SOTA Arabic BERT model from scratch		
Research Associate <i>MADAR Project, Supervised by: Prof. Nizar Habash and Asst. Teach. Prof. Houada Bouamor</i>	<i>Carnegie Mellon University Qatar, Doha, Qatar</i>	May 2018–Aug 2019
<ul style="list-style-type: none">Co-organized MADAR Shared Task on fine-grained Arabic Dialect Identification		
<i>Minicomplexity Project, Supervised by: Assoc. Teach. Prof. Christos Kapoutsis</i>		
<ul style="list-style-type: none">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 <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">Proposed clustering method that reduces training time of <i>Active Learning</i>.Proposed a stratification method that uses a learned relationship between feature vectors and classifier labels to dynamically allocate samples that reduce bias for active evaluation of classifiers.		

SERVICES

- **Co-organizer**, MADAR Shared Task on Fine-Grained Arabic Dialect Identification, 2019
- **Reviewer**: AAAI-22 Special Track on AI for Social Impact, SemEval 2020
- **Member**, Student Academic Committee, Academic Review Board, Carnegie Mellon University Qatar

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

COURSEWORK

- **CS 2731**: Introduction to NLP
- **CS 2637**: Foundations of HCI
- **CS 2710**: Foundation of AI
- **CS 2150**: Design & Analysis of Algorithms
- **11-785**: Introduction to Deep Learning
- **10-601**: Introduction to Machine Learning
- **15-451**: Algorithm Design and Analysis
- **15-440**: Distributed Systems
- **15-437**: Web Application Development
- **15-453**: Formal Language, Automata and Computability
- **21-484**: Graph Theory

TECHNICAL SKILLS

- **Programming**: Python, JavaScript, Java, C,
- **Machine Learning, Deep Learning, and NLP**: Tensorflow, Keras, Pytorch
- **Web Development**: Django, Flask, Angular JS, Bootstrap
- **Other tools**: ROS, Hadoop, **Languages**: English, Bengali

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
- 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] Inan, M. *, Yang, Z. *, **Hassan, S. ***, Quandt, L., & Alikhani, M. Modeling Intensification for Signed Language Generation: A Computational Approach. *ACL, 2022 (Findings)*
- [2] Mubarak, H., **Hassan, S.**, & Chowdhury, S. (2021) Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech. *NLE, 2022.*
- [3] **Hassan, S.**, Shaar, S., & Darwish, K. Cross-lingual Emotion Detection. *LREC, 2022*
- [4] Mubarak, H., **Hassan, S.**, Chowdhury, S. A., & Alam, F. ArCovidVac: Analyzing Arabic Tweets About COVID-19 Vaccination. *LREC, 2022.*
- [5] **Hassan, S.**, Mubarak, H., Abdelali, A., & Darwish, K. ASAD: Arabic Social Media Analytics and unDerstanding. *EACL 2021*
- [6] Mubarak, H., **Hassan, S.**, & Abdealali, A. Adult Content Detection on Arabic Twitter: Analysis and Experiments. *WANLP 2021*
- [7] Abdelali, A., Mubarak, H., Samih, Y., **Hassan, S.**, & Darwish, K. QADI: Arabic Dialect Identification in the Wild. *WANLP 2021*
- [8] Mubarak, H., & **Hassan, S.** (2021, April). ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic. *LOUHI, 2021*
- [9] Mubarak, H., Abdelali, A., **Hassan, S.**, & Darwish, K. Spam Detection on Arabic Twitter. *SocInfo, 2020*
- [10] **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* [2nd place in Arabic subtask A]
- [11] **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* [1st place in subtask A]
- [12] Mubarak, H., **Hassan, S.**, & Abdelali, A. Constructing a bilingual corpus of parallel tweets. *BUCC, 2020*
- [13] Bouamor, H., **Hassan, S.**, & Habash, N. The MADAR Shared task on Arabic Fine-Grained Dialect Identification. *WANLP, 2019. [co-organizer]*
- [14] Anabtawi, M., **Hassan, S.**, Kapoutsis, C., & Zakzok, M. An Oracle Hierarchy for Small One-way Finite Automata. *LATA, 2019*
- [15] **Hassan, S.**, Shaar, S., Raj, B., & Razak, S. (2018, December). Interactive Evaluation of Classifiers Under Limited Resources. *IEEE ICMLA, 2018*

Pre-prints:

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