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   *University of Pittsburgh*   2021–
Supervised by: Dr. Malihe Alikhani

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,
Supervised by: Prof. Bhiksha Raj and Assoc. Teach. Prof. Saquib Razak

HONORS AND AWARDS

Outstanding Academic Achievement Award	*Carnegie Mellon University Qatar*	May 2018
University Honors	*Carnegie Mellon University*	May 2018
SCS College Honors	*Carnegie Mellon University*	May 2018
History Honors	*Phi Alpha Theta*	May 2018
Carnegie Mellon Dean's List	*Carnegie Mellon University*	Apr 2015–Apr2018

RESEARCH

Graduate Student Researcher   *University of Pittsburgh, Pittsburgh, PA*   Aug 2021–
Computer Science Department, Supervised by: Dr. Malihe Alikhani
- Developed discourse-aware style-transfer models by integrating PDTB and RST frameworks with pretrained transformer models for better content preservation.
- Improved prosody in sign language generation by incorporating intensity-enhanced GLOSS representations within a novel Progressive Transformer model.

Research Assistant   *Qatar Computing Research Institute, Doha, Qatar*   Oct 2019–Jun 2021
Arabic Language Technologies Group, Supervised by: Dr. Kareem Darwish
- 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 detection
- Studied importance of data diversity and linguistically aware segmentation for pretraining a SOTA Arabic BERT model from scratch

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.

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

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] Atwell, K.*, **Hassan, S.***, Alikhani, M. APPDIA: A Discourse-aware Transformer-based Style Transfer Model for Offensive Social Media Conversations. *COLING, 2022*

[2] Inan, M.*, Yang, Z.*, **Hassan, S.***, Quandt, L., & Alikhani, M. Modeling Intensification for Signed Language Generation: A Computational Approach. *ACL FINDINGS, 2022*

[3] Mubarak, H., **Hassan, S.**, & Chowdhury, S. Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech. *NLE, 2022.*

[4] **Hassan, S.**, Shaar, S., & Darwish, K. Cross-lingual Emotion Detection. *LREC, 2022*

[5] Mubarak, H., **Hassan, S.**, Chowdhury, S. A., & Alam, F. ArCovidVac: Analyzing Arabic Tweets About COVID-19 Vaccination. *LREC, 2022.*

[6] **Hassan, S.**, Mubarak, H., Abdelali, A., & Darwish, K. ASAD: Arabic Social Media Analytics and unDerstanding. *EACL 2021*

[7] Mubarak, H., **Hassan, S.**, & Abdealali, A. Adult Content Detection on Arabic Twitter: Analysis and Experiments. *WANLP 2021*

[8] Abdelali, A., Mubarak, H., Samih, Y., **Hassan, S.**, & Darwish, K. QADI: Arabic Dialect Identification in the Wild. *WANLP 2021*

[9] Mubarak, H., & **Hassan, S.** (2021, April). ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic. *LOUHI, 2021*

[10] Mubarak, H., Abdelali, A., **Hassan, S**., & Darwish, K.Spam Detection on Arabic Twitter. *SocInfo, 2020*

[11] **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*

[12] **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]

[13] Mubarak, H., **Hassan, S.**, & Abdelali, A. Constructing a bilingual corpus of parallel tweets. *BUCC, 2020*

[14] Bouamor, H., **Hassan, S.**, & Habash, N. The MADAR Shared task on Arabic Fine-Grained Dialect Identification. *WANLP, 2019*. **[co-organizer]**

[15] Anabtawi, M., **Hassan, S.**, Kapoutsis, C., & Zakzok, M. An Oracle Hierarchy for Small One-way Finite Automata. *LATA, 2019*

[16] **Hassan, S.**, Shaar, S., Raj, B., & Razak, S. (2018, December). Interactive Evaluation of Classifiers Under Limited Resources. *IEEE ICMLA, 2018*

Pre-prints:

[17] 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:

[18] **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*

[19] 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*