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*	2021–2026
Supervised by: Asst. Prof. Malihe Alikhani		(expected)

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–Apr 2018

RESEARCH

PhD Research   *University of Pittsburgh, Pittsburgh, PA*   Aug 2021–
Computer Science Department, Supervised by: Asst. Prof. Malihe Alikhani

- Improved sign language generation by enhancing GLOSS representation with intensity modifiers [15]

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

- Trained and deployed SVM and BERT models for emotion, sentiment, offensiveness, hate speech, spam, adult content, and dialect identification [1] (https://asad.qcri.org/).
- Collected novel datasets and trained ML/DL models for offensiveness [14] and adult content detection [2], dialect identification [3]. tweets related to Covid-19 [4] and spam detection [5].
- Trained ensemble of ML/DL models for Arabic (**2nd** place) and English offensiveness detection at SemEval 2020 [6] and Arabic offensiveness detection at OSACT 2020 [7] (**1st** place).
- Automatically collected a corpus of English-Arabic parallel tweets [8].
- Trained SOTA Arabic/Spanish emotion detection models using transformers and studied efficacy of different cross-lingual approaches [12].
- Trained Arabic BERT model from scratch and performed quantitative analysis of key factors[13].

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

- Collected a dataset of tweets (21 countries) and trained ML/DL models for Arabic dialect ID.
- Summarized findings of MADAR Shared Task on fine-grained Arabic dialect ID at WANLP 2019 [9].

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 [10].

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 clustering method for reducing training time of *Active Learning.*
- Proposed a stratification method that uses a learned relationship between feature vectors and classifier labels and dynamic sample allocation methods that reduce bias in *Active Testing* [11].

Summer Research *Carnegie Mellon University Qatar, Doha, Qatar* May 2017–Aug2017
Qatar Student Initiated Undergraduate Research Project, Supervised by: Bhiksha Raj
- Studied sampling methods for reducing cost of classifier evaluation and human labeling.

Independent Studies *Carnegie Mellon University* Jan 2016–Apr 2018
Minicomplexity, Supervised by: Assoc. Teach. Prof. Christos Kapoutsis
- Studied general complexity theory and subsequently focused on complexity of finite automata

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, Carnegie Mellon University Qatar
- **Member,** Academic Review Board, Carnegie Mellon University Qatar

TEACHING

Graduate Teaching Assistant *University of Pittsburgh, Pittsburgh, PA* Aug 2021–
Led recitations, held office hours, and graded homework for:
- Algorithms and Data Structures 2 (CS 1501)

Course Assistant *Carnegie Mellon University Qatar, Doha, Qatar* Jan 2016–May 2018
Held office hours and graded homework for the following courses:
- 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
- **11-785:** Introduction to Deep Learning
- **10-601:** Introduction to Machine Learning
- **15-451:** Algorithm Design and Analysis
- **15-440:** Distributed Systems
- **CS 2710:** Foundation of AI
- **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

Road Surface Identification   *Carnegie Mellon University*   Spring 2018
Course: Introduction to Deep Learning (11–785)
- Collected a dataset of mobile sensor data while driving over different road surfaces
- Proposed a hybrid HMM–LSTM model for road surface identification

Remote Method Invocation Library   *Carnegie Mellon University*   Fall 2017
Course: Distributed Systems (15–440)
- Implemented concurrent Remote Method Invocation library over a distributed file system.

PUBLICATIONS

Published/Accepted:

[1] **Hassan, S.,** Mubarak, H., Abdelali, A., & Darwish, K. (2021, April). ASAD: Arabic Social Media Analytics and unDerstanding. In *Proceedings of the Demonstrations of the 16th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 113–118)

[2] Mubarak, H., **Hassan, S.,** & Abdealali, A. (2021, April). Adult Content Detection on Arabic Twitter: Analysis and Experiments. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop (pp. 136–144)*

[3] Abdelali, A., Mubarak, H., Samih, Y., **Hassan, S.,** & Darwish, K. (2021, April). QADI: Arabic Dialect Identification in the Wild. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop* (pp. 1–10)

[4] Mubarak, H., & **Hassan, S.** (2021, April). ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis* (pp. 1–6)

[5] Mubarak, H., Abdelali, A., **Hassan, S**., & Darwish, K. (2020, October). Spam Detection on Arabic Twitter. In *International Conference on Social Informatics* (pp. 237–251). Springer, Cham.

[6] **Hassan, S.,** Samih, Y., Mubarak, H., & Abdelali, A. (2020, December). ALT at SemEval-2020 task 12: Arabic and English offensive language identification in social media. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation* (pp. 1891–1897). [**2nd** place in Arabic subtask A]

[7] **Hassan, S.,** Samih, Y., Mubarak, H., Abdelali, A., Rashed, A., & Chowdhury, S. A. (2020, May). ALT submission for OSACT shared task on offensive language detection. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection* (pp. 61–65).[**1st** place in subtask A]

[8] Mubarak, H., **Hassan, S.,** & Abdelali, A. (2020, May). Constructing a bilingual corpus of parallel tweets. In *Proceedings of the 13th Workshop on Building and Using Comparable Corpora* (pp. 14–21).

[9] Bouamor, H., **Hassan, S.,** & Habash, N. (2019, August). The MADAR Shared task on Arabic Fine-Grained Dialect Identification. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop* (pp. 199–207). **[co-organizer]**

[10] Anabtawi, M., **Hassan, S.,** Kapoutsis, C., & Zakzok, M. (2019, March). An Oracle Hierarchy for Small One-way Finite Automata. In *International Conference on Language and Automata Theory and Applications* (pp. 57–69). Springer, Cham.

[11] **Hassan, S.,** Shaar, S., Raj, B., & Razak, S. (2018, December). Interactive Evaluation of Classifiers Under Limited Resources. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 173–180). IEEE.

Pre-print/Under Review:

[12] **Hassan, S.,** Shaar, S., & Darwish, K. (2021). Cross-lingual Emotion Detection. *arXiv preprint arXiv:2106.06017*

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

[14] Mubarak, H., **Hassan, S.,** & Chowdhury, S.(2021) Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech (Submitted to special issue of Natural Language Engineering)

[15] Inan, M.*, Yang, Z.*, **Hassan, S.*,** Quandt, L., & Alikhani, M..(2021) Modeling Intensification for Signed Language Generation: A Computational Approach (submitted to ACL 2022)

* denotes equal contribution