Bashar Alhafni

+1(203) 916-3867 ⋄ alhafni@nyu.edu ⋄ http://basharalhafni.com/

RESEARCH INTERESTS

Natural Language Processing, Deep Learning, Language Generation, Machine Translation, Grammatical Error Correction, Rewriting, Simplification, Style Transfer, and Personalization.

EDUCATION

New York University   New York, NY
Ph.D. in Computer Science   Aug 2020 – Present
Advisor: Nizar Habash

Thesis Committee: Ted Briscoe, Kyunghyun Cho, Mona Diab, Nizar Habash, He He, Julia Stoyanovich

University of Southern California   Los Angeles, CA
M.S. in Computer Science   Aug 2017 – May 2019
Advisors: Jonathan May, Nanyun Peng

University of Bridgeport   Bridgeport, CT
B.S. in Computer Science   Aug 2013 – May 2017
Minor in Mathematics

EXPERIENCE

Graduate Research Assistant   Abu Dhabi, UAE
CAMeL Lab, NYU Abu Dhabi   Sep 2019 – Present
- Developed state-of-the-art Arabic grammatical error correction and detection models across multiple domains [1]
- Defined the task of gender rewriting and built models to reduce gender bias in Arabic NLP systems [2], [6], [7], [10]
- Studied the effects of pretraining data size and language variants in Arabic pretrained language models [9]
- Implemented and integrated several open-source Arabic NLP tools as part of the CAMeL Tools toolkit [11]

Applied Research Scientist Intern   San Francisco, CA
Grammarly   May 2023 – Present
- Mentors: Vipul Raheja, Dhruv Kumar, Vivek Kulkarni
- Working on personalizing large language models and building multilingual text editing systems

Research Intern   New York, NY
Dataminr   May 2022 – Sep 2022
- Mentors: Joel Tetreault, Ke Zhang, Shihao Ran
- Created a novel benchmark for local crisis event timeline extraction and summarization [3]

Graduate Student Researcher   Los Angeles, CA
USC Information Sciences Institute   May 2018 – Aug 2019
- Worked on low-resource machine translation and handled out-of-vocabulary words using unprocessed dictionaries
- Developed the first neural baselines for event temporal relation extraction on two story narrative datasets [12]

Associate Software Engineer   East Hartford, CT
Pratt & Whitney   Nov 2017 – May 2018
- Implemented a python-based pipeline to migrate reported software defects to a new bug tracking system

Technology Summer Analyst   New York, NY
Morgan Stanley   Summer 2016 & 2017
- Developed a code-generator to create Java Classes dynamically based on a proprietary meta-data language
- Designed and implemented a widget using WPF in a trading application

Software Engineering Intern   Farmington, CT
United Technologies Corp   Summer 2015
- Organized a server refresh project and deployed test procedures to reduce security vulnerabilities

Publications (Google Scholar)

[1] **Advancements in Arabic Grammatical Error Detection and Correction: An Empirical Investigation.** **Bashar Alhafni**, Go Inoue, Christian Khairallah, Nizar Habash. EMNLP 2023.

[2] **The User-Aware Arabic Gender Rewriter. Bashar Alhafni**, Ossama Obeid, Nizar Habash. 1st Workshop on Gender-Inclusive Translation Technologies, EAMT 2023.

[3] **CrisisLTLSum: A Benchmark for Local Crisis Event Timeline Extraction and Summarization.** Hossein Rajaby Faghihi, **Bashar Alhafni**, Ke Zhang, Shihao Ran, Joel Tetreault and Alejandro Jaimes. Findings of EMNLP 2022.

[4] **Arabic Word-level Readability Visualization for Assisted Text Simplification.** Reem Hazim, Hind Saddiki, **Bashar Alhafni**, Muhamed Al Khalil, Nizar Habash. EMNLP 2022 (System Demonstration).

[5] **Shared Task on Gender Rewriting. Bashar Alhafni**, Ossama Obeid, Houda Boamour, Nizar Habash. 7th Arabic NLP Workshop, EMNLP 2022.

[6] **User-Centric Gender Rewriting. Bashar Alhafni**, Nizar Habash, Houda Bouamor. NAACL 2022.

[7] **The Arabic Parallel Gender Corpus 2.0: Extensions and Analyses. Bashar Alhafni**, Nizar Habash, Houda Bouamor. LREC 2022.

[8] **Zero-shot Cross-Linguistic Learning of Event Semantics.** Malihe Alikhani, Thomas Kober, **Bashar Alhafni**, Mert Inan, Shahab Raji, Yue Chen, Elizabeth Nielsen, Mark Steedman, Matthew Stone. INLG 2022.

[9] **The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models.** Go Inoue, **Bashar Alhafni**, Nurpeiis Baimukan, Nizar Habash, Houda Bouamor. 6th Arabic NLP Workshop, EACL 2021.

[10] **Gender-Aware Reinflection using Linguistically Enhanced Neural Models. Bashar Alhafni**, Nizar Habash, Houda Bouamor. 2nd Workshop on Gender Bias in NLP, COLING 2020.

[11] **CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing.** Ossama Obeid, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, **Bashar Alhafni**, Go Inoue, Fadhl Eryani, Alex Erdmann, Nizar Habash. LREC 2020.

[12] **Contextualized Word Embeddings Enhanced Event Temporal Relation Extraction for Story Understanding.** Rujun Han, Mengyue Liang, **Bashar Alhafni**, Nanyun Peng. 1st Workshop on Narrative Understanding, NAACL 2019.

Technical Skills

Programming Languages: Python, Java, C#, Bash, C++, SQL, Scala, JavaScript

Technologies: PyTorch, transformers, spaCy, NLTK, scikit-learn, NumPy, Git, LaTeX, Amazon Mechanical Turk

Natural Languages: Arabic, English, French (Basic)

Teaching Experience

NYU Abu Dhabi
- CS-3090: Research Seminar in Computer Science. Taught a tutorial on pretrained language models.   Fall 2020
- Taught a three-day crash course on Arabic NLP for researchers in Digital Humanities.   Spring 2020

University of Southern California
- CSCI-572: Information Retrieval and Web Search Engines. TA for Ellis Horowitz.   Fall 2018 – Spring 2019

University of Bridgeport
- CS-101/102: Introduction to Computer Science I & II. TA for Subrina Thompson.   Fall 2014 – Spring 2017

Professional Service

Local Organization Committee: EMNLP 2022

Reviewer: ARR, NAACL 2022, EMNLP 2023/22/21, ACL 2020, LREC 2020, WANLP 2023/22/21/20/19

Student Volunteer: NAACL 2019, EMNLP 2023

Honors & Awards

NYU Global PhD Fellowship: 2020 – 2024

Upsilon Pi Epsilon ACM Scholarship: 2017

University of Bridgeport Trustee Scholarship: 2013 – 2017

University of Bridgeport President's List: 2013 – 2017