

Bashar Alhafni

+1(203) 916-3867 | alhafni@nyu.edu | basharalhafni.com | github.com/balhafni

RESEARCH INTERESTS

Natural Language Processing, Deep Learning, Controlled Language Generation, Grammatical Error Correction, Rewriting, Educational NLP, Personalization, Machine Translation, Simplification

EDUCATION

New York University New York, NY
Ph.D. in Computer Science Aug 2020 – May 2025

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

- Investigated the impact of dialect identification on enhancing text normalization for dialectal Arabic [4]
- Created the first manually annotated Arabic text simplification dataset for school-aged learners [6], [5], [15]
- Studied Arabic grammatical error detection and correction across multiple domains [3], [9]
- Defined the task of gender rewriting and built models to reduce gender bias in Arabic NLP systems [10], [12], [13]
- Studied the effects of pretraining data size and language variants in Arabic pretrained language models [18]
- Implemented and integrated several open-source Arabic NLP tools as part of the [CAMEL Tools](#) toolkit [20]

Visiting Researcher Abu Dhabi, UAE
MBZUAI June 2024 – Present

- Mentor: [Ted Briscoe](#)
- Research on Arabic grammatical error correction, automated essay scoring, and Arabic AI writing assistants [1], [2]

Applied Research Scientist Intern San Francisco, CA
Grammarly May 2023 – April 2024

- Mentors: [Vipul Raheja](#), [Vivek Kulkarni](#), [Dhruv Kumar](#)
- Created a novel benchmark for personalizing LLMs based on various fine-grained linguistic attributes [8]
- Developed a series of multilingual instruction-tuned LLMs for text editing tasks covering seven languages [7]

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 [16]

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 [21]

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

- [1] Kirill Chirkunov, **Bashar Alhafni**, Chatrine Qwaider, Nizar Habash, and Ted Briscoe. 2025. [ARWI: Arabic Write and Improve](#). In *Proceedings of the Fourth Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2025)*, New Mexico, United States. Association for Computational Linguistics. **Diversity Award**.
- [2] Chatrine Qwaider, **Bashar Alhafni**, Kirill Chirkunov, Nizar Habash, and Ted Briscoe. 2025. [Enhancing Arabic automated essay scoring with synthetic data and error injection](#). Preprint, arXiv:2503.17739.
- [3] **Bashar Alhafni** and Nizar Habash. 2025. [Enhancing text editing for grammatical error correction: Arabic as a case study](#). Preprint, arXiv:2503.00985.
- [4] **Bashar Alhafni**, Sarah Al-Towaity, Ziyad Fawzy, Fatema Nassar, Fadhl Eryani, Houda Bouamor, and Nizar Habash. 2024. [Exploiting dialect identification in automatic dialectal text normalization](#). In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 42–54, Bangkok, Thailand. Association for Computational Linguistics.
- [5] Juan Liberato, **Bashar Alhafni**, Muhamed Khalil, and Nizar Habash. 2024. [Strategies for Arabic readability modeling](#). In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 55–66, Bangkok, Thailand. Association for Computational Linguistics.
- [6] **Bashar Alhafni**, Reem Hazim, Juan David Pineros Liberato, Muhamed Al Khalil, and Nizar Habash. 2024. [The SAMER Arabic text simplification corpus](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16079–16093, Torino, Italia. ELRA and ICCL.
- [7] Vipul Raheja, Dimitris Alikaniotis, Vivek Kulkarni, **Bashar Alhafni**, and Dhruv Kumar. 2024. [mEDIT: Multilingual text editing via instruction tuning](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 979–1001, Mexico City, Mexico. Association for Computational Linguistics.
- [8] **Bashar Alhafni**, Vivek Kulkarni, Dhruv Kumar, and Vipul Raheja. 2024. [Personalized text generation with fine-grained linguistic control](#). In *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*, pages 88–101, St. Julians, Malta. Association for Computational Linguistics.
- [9] **Bashar Alhafni**, Go Inoue, Christian Khairallah, and Nizar Habash. 2023. [Advancements in Arabic grammatical error detection and correction: An empirical investigation](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 6430–6448*, Singapore. Association for Computational Linguistics.
- [10] **Bashar Alhafni**, Ossama Obeid, and Nizar Habash. 2023. [The user-aware Arabic gender rewriter](#). In *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies, pages 3–11*, Tampere, Finland. European Association for Machine Translation.
- [11] Hazem Ibrahim, Fengyuan Liu, Rohail Asim, Balaraju Battu, Sidahmed Benabderrahmane, **Bashar Alhafni**, Wifag Adnan, Tuka Alhanai, Bedoor AlShebli, Riyadh Baghdadi, Jocelyn J. Bélanger, Elena Beretta, Kemal Celik, Moumena Chaqfeh, Mohammed F. Daqaq, Zaynab El Bernoussi, Daryl Fougny, Borja Garcia de Soto, Alberto Gandolfi, and 20 others. 2023. [Perception, performance, and detectability of conversational artificial intelligence across 32 university courses](#). *Scientific Reports*, 13(1).
- [12] **Bashar Alhafni**, Nizar Habash, and Houda Bouamor. 2022. [User-centric gender rewriting](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 618–631, Seattle, United States. Association for Computational Linguistics.
- [13] **Bashar Alhafni**, Nizar Habash, and Houda Bouamor. 2022. [The Arabic parallel gender corpus 2.0: Extensions and analyses](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1870–1884, Marseille, France. European Language Resources Association.
- [14] **Bashar Alhafni**, Nizar Habash, Houda Bouamor, Ossama Obeid, Sultan Alrowili, Daliyah AlZeer, Kawla Mohamad Shnqiti, Ahmed Elbakry, Muhammad ElNokrashy, Mohamed Gabr, Abderrahmane Issam, Abdelrahim Qaddoumi, Vijay Shanker, and Mahmoud Zyate. 2022. [The shared task on gender rewriting](#). In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 98–107, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

- [15] Reem Hazim, Hind Saddiki, **Bashar Alhafni**, Muhamed Al Khalil, and Nizar Habash. 2022. [Arabic word-level readability visualization for assisted text simplification](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 242–249, Abu Dhabi, UAE. Association for Computational Linguistics.
- [16] Hossein Rajaby Faghihi, **Bashar Alhafni**, Ke Zhang, Shihao Ran, Joel Tetreault, and Alejandro Jaimes. 2022. [CrisisLTLSum: A benchmark for local crisis event timeline extraction and summarization](#). In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5455–5477, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- [17] Malihe Alikhani, Thomas Kober, **Bashar Alhafni**, Yue Chen, Mert Inan, Elizabeth Nielsen, Shahab Raji, Mark Steedman, and Matthew Stone. 2022. [Zero-shot cross-linguistic learning of event semantics](#). In *Proceedings of the 15th International Conference on Natural Language Generation*, pages 212–224, Waterville, Maine, USA and virtual meeting. Association for Computational Linguistics.
- [18] Go Inoue, **Bashar Alhafni**, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. 2021. [The interplay of variant, size, and task type in Arabic pre-trained language models](#). In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 92–104, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.
- [19] **Bashar Alhafni**, Nizar Habash, and Houda Bouamor. 2020. [Gender-aware reinflection using linguistically enhanced neural models](#). In *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, pages 139–150, Barcelona, Spain (Online). Association for Computational Linguistics.
- [20] Ossama Obeid, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, **Bashar Alhafni**, Go Inoue, Fadhil Eryani, Alexander Erdmann, and Nizar Habash. 2020. [CAMEL tools: An open source python toolkit for Arabic natural language processing](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 7022–7032, Marseille, France. European Language Resources Association.
- [21] Rujun Han, Mengyue Liang, **Bashar Alhafni**, and Nanyun Peng. 2019. [Contextualized word embeddings enhanced event temporal relation extraction for story understanding](#). In *Proceedings of the First Workshop on Narrative Understanding*, Minneapolis, Minnesota. Association for Computational Linguistics.

TEACHING EXPERIENCE

NYU Abu Dhabi

- Capstone Undergraduate Research Advising:
 - **Sarah Al-Towaity, Ziyad Fawzy, Fatema Nassar** Fall 2023 – Spring 2024
 - Dialectal Arabic Text Normalization
 - Publication at ArabicNLP 2024 [4]
 - **Juan Piñeros Liberato** Spring 2023 – Spring 2024
 - Arabic Readability Assessment
 - First-author publication at ArabicNLP 2024 [5]; publication at LREC-COLING 2024 [6]
 - **Reem Hazim** Spring 2022 – Spring 2024
 - Arabic Text Simplification
 - First-author publication at EMNLP 2022 [15]; publication at LREC-COLING 2024 [6]
- CS-2216: Introduction to NLP. Taught six lectures on neural networks for NLP Fall 2024
- 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

INVITED TALKS & PRESENTATIONS

- [1] **Controlled Natural Language Generation for Morphologically Rich Languages: The Case of Arabic.** Mohamed bin Zayed University of Artificial Intelligence (MBZUAI). April, 2025.
- [2] **Arabic NLP with CAMEL Tools.** Arabic NLP Winter School. January, 2025.
- [3] **Panelist: Starting and Excelling in Arabic NLP.** Arabic NLP Winter School. January, 2025.

- [4] **LLMs in Education: Novel Perspectives, Challenges, and Opportunities.** COLING 2025 Tutorial (Proposal, Website). January, 2025.
- [5] **Panelist: Arabic Culture and AI: From Preservation to Enrichment.** Arabian Days. December, 2024.
- [6] **Controlled Natural Language Generation for Morphologically Rich Languages: The Case of Arabic.** National Research Council Canada (NRC). September, 2024.
- [7] **Panelist: Challenges and Solutions in Arabic NLP.** King Salman Global Academy for Arabic. September, 2024.
- [8] **User-Centric Gender Rewriting.** Mohamed bin Zayed University of Artificial Intelligence (MBZUAI). May, 2023.
- [9] **Arabic NLP: Challenges and Solutions.** The Symposium on the Evolution of the Arabic Word. October, 2022.
- [10] **Panelist: Young Researchers in Arabic NLP.** The 7th Arabic Natural Language Processing Workshop (WANLP 2022). December, 2022.
- [11] **Gender-Aware Reinflection using Linguistically Enhanced Neural Models.** ETH Zürich. October, 2021.

TECHNICAL SKILLS

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

Technologies: PyTorch, transformers, spaCy, NLTK, scikit-learn, NumPy, Git, L^AT_EX, Amazon Mechanical Turk

Natural Languages: Arabic (Native), English (Native), French (Basic)

PROFESSIONAL SERVICE

Organization Committees: BEA 2025 (Shared Tasks and Tutorials Chair), ArabicNLP 2024 (Publicity Chair), EMNLP 2022 (Local Organizer)

Reviewer: ARR, COLING 2025, NAACL 2022, EMNLP 2023-2021, ACL 2020, LREC 2020, Arabic NLP 2024-2023, WANLP 2022-2019, GeBNLP 2024-2021

Student Volunteer: ACL 2024, EMNLP 2023, NAACL 2019

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