

Wajdi Zaghouni, PhD

Associate Professor of AI and Media Analytics with over 20 years of experience at the intersection of artificial intelligence, language, and society in the Middle East and North Africa. Research focuses on large language models, generative AI, and digital platforms, with extensive publications and international impact.

Phone: (+974) 33454992

E-Mail: wajdi.zaghouni@northwestern.edu

Google Scholar: https://scholar.google.com/citations?user=Jayj_cAAAAJ&hl=en

Research Group: <https://sites.google.com/view/dh4ai/home>

Citizenship: Canadian

Citations: 4200 ; **h-index** 35; **i10-index:** 83; Scopus Field-Weighted Citation Impact: 2.47

Research Interests

Digital Humanities, Natural Language Processing, AI, Social Media Analytics, Computational Social Sciences, Corpus Linguistics, Language Resources and Evaluation.

Education

Doctorate in Natural Language Processing, University of Paris X Nanterre La Defense, 2015

M.A in Linguistics, University of Montreal, 2009.

B.A in Computational Linguistics, (Minor in Computer Science, Major in Linguistics), the University of Quebec in Montreal, **2002**

Academic Experience

8/2024 – Today: **Associate Professor in AI and Media Analytics**, Communication Program, Northwestern Qatar

7/2023 – 07/2024: **Associate Professor**, Digital Humanities Program, College of Humanities and Social Sciences, **Hamad Bin Khalifa University** (Qatar).

4/2019 – 8/2021: **PhD Program Coordinator**, College of Humanities and Social Sciences, **Hamad Bin Khalifa University** (Qatar).

1/2018 – 06/2024: **Assistant Professor**, Digital Humanities Program, College of Humanities and Social Sciences, **Hamad Bin Khalifa University** (Qatar).

1/2015 – 12/2017: **Postdoctoral Research Associate**, **Carnegie Mellon**

University (Qatar). 9/2012 - 12/2014: **Senior Research Associate and project coordinator**, Computer Science Program, **Carnegie Mellon University** (Qatar).

2/2006 - 6/2011: **Senior Researcher**, the Linguistic Data Consortium, **University of Pennsylvania**, Philadelphia (USA).

3/2005 – 1/2006: **Computational Linguist** at the Language Technology group of the Joint Research Center (JRC), Ispra (Italy) with Dr. Ralf Steinberger. Worked mainly to create an Arabic Named Entities recognition module for the **Europe News Monitor** (EMM News Explorer and News Brief)

Consultancy/ Industry Experience

1/2009 - 8/2010: **Consultant** (Computational Linguist), Colorado University with **Dr. Martha Palmer**, Boulder, Colorado (USA)

12/2007 - 7/2008: **Consultant** in social media data analytics and information extraction and lexicon building, TEMIS, Paris (France)

9/2002- 6/2004: **Computational Linguist / Researcher** (Information Extraction) NSTEIN Technologies (Now OpenText Inc.), Montreal (Canada)

Research Grants

Funded - In Progress

Lead PI in : The Future of Digital Citizenship in Qatar: a Socio-Technical Approach, Funded by Qatar National Research Fund (October 31, 2022) (\$3,826,768.00), Funded - In Progress, Fall 2022, Co-PI Amal Al Malki, Co-PI Marc Jones, Co-PI Georgios Mikros, Program Coordinator Raian Ali

Lead PI in : MARSAD SP1, Funded by Qatar National Research Fund (October 31, 2022) (\$3,826,768.00), NPRP14C-37878-SP-465 by QRDI

PI in Critical Digital Literacy, Funded by Qatar National Research Fund (October 31, 2022) NPRP14C-37878-SP-465 by QRDI

PI in : ECMV01-0928-250034 Responsible AI and Large Language Models for Social Good in the Arab World: A Cross-Institutional Innovation Workshop. Funded by QRDI

PI in : UREP 32-0303-250332 Development of Unified Medical App for Patient-Centric Healthcare Management in Qataq, funded by QRDI.

PI in : LiLEC01-0302-250052 NexTutor: Pioneering Qatar's Next-Generation AI Learning Ecosystem Award Active PI

Lead PI in: MCSC02-0217-250013 The AI Sadaqah Journey: Explore, Give, Inspire.

Completed

Lead PI in : Resources and Applications for Detecting and Classifying Polarized and Hate Speech in Arabic Social Media, Funded by Qatar National Research Fund (May 1, 2021 - May 1, 2024), awarded January 18, 2021 (\$509,000.00), Funded - In Progress, Spring 2021, Program Coordinator Wajdi Zaghouani with Co-PI Anis Charfi, Co-PI Paolo Rosso, Co-PI Marc Jones, Co-PI Georgios Dimitropoulos, Co-PI Georgios Mikros, Co-PI Dr. Fadi Makki, Co-PI Prof. Joseph Cannataci, Co-PI Dr. Aitana Radu

PI in : Supporting Best Practices in Synchronous Conferencing Software-assisted Teaching: An EC Faculty Group Perspective, Funded by education City Innovation in Teaching Grant (ITG) - VCU, WCMQ and CHSS (August 31, 2022) (\$9,500.00), Funded - In Progress, Fall 2022, PI Wajdi Zaghouani with PI Robert Bianchi, PhD, Program Coordinator Alan Weber

PI in : Visual Exploration of Chain of Narrators in the Kutub Al-Sittah , Funded by Qatar National Research Fund (April 19, 2021 - April 19, 2023), awarded January 18, 2021 (\$600,000.00), Funded - In Progress, Spring 2021, Program Coordinator Dr. Younss Ait Mou with PI Jens Schneider, PI Wajdi Zaghouani

PI in : A Multimedial Education Tool for Qatar's History, Funded by HBKU Thematic Research Grant Program (January 1, 2021 - January 31, 2023), awarded December 15, 2020 (\$300,000.00), Funded - In Progress, Spring 2021, Program Coordinator Jens Schneider with PI Marc Jones, PI Wajdi Zaghouani, PI Georgios Mikros, PI Dena Al Thani, PI Marco Agus

PI in : Writing the Future: Metadiscourse and Voice in English and Arabic Argumentative Writings of Qatari University Students, Funded by QNRF (May 1, 2020 - May 31, 2022), awarded October 31, 2019 (\$516,000.00), Funded - In Progress, Fall 2019, Program Coordinator Abdelhamid Ahmed with PI Debra Myhill, PI Wajdi Zaghouani

PI in : Developing an HBKU Index of Social Cohesion: Exploration and Intervention at the State and Region Levels, Funded by HBKU (August 15, 2022 - August 15, 2022), (QAR700,000), Funded - In Progress, Fall 2022, Program Coordinator Abdolmehdi Riazi with PI Wajdi Zaghouani

Lead PI in : COVID-19 Misinformation and Social Networks Analysis in Qatar, Funded by HBKU Innovation Center (December 1, 2020 - August 31, 2021), awarded May 6, 2020 (\$100,000.00), Completed, Fall 2020, Program Coordinator Wajdi Zaghouani with Co-PI Marc Jones, Co-PI Georgios Mikros

PI in : A Personalized and Intelligent Digital Mental Health Platform for Qatar and the Arab world, Funded by Qatar National Research Fund (November 18, 2020 - July 30, 2022), awarded September 1, 2020 (\$395,720.00), Funded - In Progress, Spring 2020, Program Coordinator Mowafa Househ with PI Wajdi Zaghouani

Lead PI in : Proactive Profiling of Fake News Spreaders (Pro2Faker), Funded by Spanish Agency of Innovation belonging to the Ministry of Science (CDTI) (September 1, 2020 - September 30, 2022), awarded September 1, 2020, Completed, Fall 2020, PI Wajdi Zaghouani

Lead PI in : The First International Workshop on Arabic Big Data & AI, Funded by Qatar National Research Fund (April 7, 2020 - April 24, 2021), awarded April 7, 2020 (\$35,000.00), Completed, Spring 2020, PI Wajdi Zaghouani

Lead PI in : Building Multi-Arabic Dialect Resources, Funded by HBKU CHSS Faculty Internal Start-up Grant (May 1, 2019 - April 30, 2021) – (50,000 QAR).

PI in : Arabic Author Profiling for Cyber-Security, Funded by Qatar National Research Fund (February 4, 2017 - June 4, 2020), awarded March 23, 2016 (\$719,526.00), Completed, Fall 2018, Co-PI Wajdi Zaghouni with PI Anis Charfi, Co-PI Paolo Rosso, Co-PI Abdelmajid BEN HAMADOU

Published Datasets

- **Qatari Corpus of Argumentative Writing**. Linguistic Data Consortium. (2022)
<https://catalog.ldc.upenn.edu/LDC2022T04>
- **Arabic Treebank Part 1 Version 4.1**. LDC Catalog, LDC2010T13. Linguistic Data Consortium.
<https://catalog.ldc.upenn.edu/LDC2010T13>
- **Arabic Treebank: Part 2 v 3.1**. LDC2011T09. ISBN 1-58563-590-1. 2011. Linguistic Data Consortium
<https://catalog.ldc.upenn.edu/LDC2011T09>
- **Arabic Treebank: Part 3 v 3.2**. 2010. LDC Catalog No. : LDC2010T08. Linguistic Data Consortium
<https://catalog.ldc.upenn.edu/LDC2010T08>
- **The Qatar Arabic Language Bank**
<https://camel.abudhabi.nyu.edu/qalb-shared-task-2015/>
- **The ARAP Corpus**
<https://arap.qatar.cmu.edu/>
- **The MADAR Corpus**
<https://sites.google.com/nyu.edu/madar/>

Publications (Peer Reviewed)

Monographs and Book Chapters

- *Shurafa C.*, Darwish K., **Zaghouni W.** (2020) Political Framing: US COVID19 Blame Game. In: Aref S. et al. (eds) Social Informatics. SocInfo 2020. Lecture Notes in Computer Science. Vol 12467. Springer.
https://doi.org/10.1007/978-3-030-60975-7_25
- **Wajdi Zaghouni**. Le repérage automatique des entités nommées dans la langue arabe. Book published by Les Éditions Universitaires Européennes. 156p. ISBN 9786131565953, Germany, 2011.
- J. Veronis; Hamon, C. Ayache; R. Belmouhoub, O. Kraif, D. Laurent, T.M.H. Nguyen, N. Semmar, F. Stuck, **W. Zaghouni**. (2008). La campagne d'évaluation ARCADE II. In Chaudiron, S. & Choukri, K. (Eds.) L'évaluation des technologies de traitement de la langue (pp 47-69). Paris: Hermes Science Publications, IC2 Cognition Collection. ISBN 978-2-7462-1992-2.

Journal Articles

- Al-Boinin, H., Vatanasakdakul, S., & Zaghouni, W. (2025). Gender, Culture, and Social Media: Exploring Women's Adoption of Social Media Entrepreneurship in Qatari Society. *Administrative Sciences*, 15(3), 89.
- Bashir, Muhammad Huzaifa; M. Azmi, Aqil; Nawaz, Haq; **Zaghouni, Wajdi**; Diab, Mona; Al-Fuqaha, Ala (2022) Arabic Natural Language Processing for Qur'anic Research: A Systematic Review. Article accepted to appear in *Artificial Intelligence Review (AIRE)*, in press (Accepted). Preprint. <https://doi.org/10.36227/techrxiv.14381891.v1>. **Journal Impact Factor: 9.58**
- Arfan Ahmed, Nashva Ali, Mahmood Alzubaidi, **Wajdi Zaghouni**, Alaa Abd-alrazaq, Mowafa Househ (2022). Arabic Chatbot Technologies: A Scoping Review, published in *Computer Methods and Programs in Biomedicine Update*. Volume 2, 2022, 100057. ISSN 2666-9900,
- <https://doi.org/10.1016/j.cmpbup.2022.100057>. **Journal Impact Factor: 5.42**
- Arfan Ahmed, Nashva Ali, Mahmood Alzubaidi, **Wajdi Zaghouni**, Alaa A Abd-alrazaq, Mowafa Househ (2022). Freely Available Arabic Corpora: A Scoping Review, *Computer Methods and Programs in Biomedicine Update*, Volume 2, 2022, 100049, ISSN 2666-9900, <https://doi.org/10.1016/j.cmpbup.2022.100049>. **Journal Impact Factor: 5.42**
- Rangel, F., Rosso, P., **Zaghouni, W.**, & Charfi, A. (2020). Fine-grained analysis of language varieties and demographics. *Natural Language Engineering*, 26(6), 641-661. doi:10.1017/S1351324920000108. **Journal Impact Factor: 1.841**
- Rosso, P, Rangel, F, Farías, IH, Cagnina, L, **Zaghouni, W**, Charfi, A. A survey on author profiling, deception, and irony detection for the Arabic language. *Linguistics and Language Compass*. 2018; 12:e12275. <https://doi.org/10.1111/lnc3.12275>. **Journal Impact Factor: 2.0**
- **Zaghouni, W.** (2018). A Large-Scale Social Media Corpus for the Detection of Youth Depression (Project Note). *Procedia Computer Science*, 142, 347–351. **Journal Impact Factor: 2.27**
- **Zaghouni, W.**, Hawwari, A., Diab, M. et al. AMPN: a semantic resource for Arabic morphological patterns. *International Journal of Speech Technology*. 19, 281–288 (2016). <https://doi.org/10.1007/s10772-015-9331-3>. **Journal Impact Factor: 1.80**
- **Wajdi Zaghouni**. 2012. RENAR: A Rule-Based Arabic Named Entity Recognition System. *ACM Transactions on Asian Language Information Processing* 11, 1, Article 2 (March 2012), 13 pages. <https://doi.org/10.1145/2090176.2090178>. **Journal Impact Factor: 1.41**
- Pouliquen Bruno, Ralf Steinberger, Camelia Ignat, Irina Temnikova, Anna Widiger, **Wajdi Zaghouni** & Jan Žižka (2005). Multilingual person name recognition and transliteration. *Journal CORELA - Cognition*,

Published Conference Papers (Peer Reviewed)

2026

Wajdi Zaghouani. Accountable Human-AI Deliberation with LLMs: Scaling Collective Intelligence through Symbiotic Scaffolding. In Proceedings of the 2nd Workshop on Language-driven Deliberation Technology (DELITE 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani. Beyond the Black Box: Ethical and Theoretical Grounding in Affective Computing. In Proceedings of the Workshop on Computational Affective Science (CAS 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani. The Generator–Eraser Paradox: Community Guidelines for Responsible LLM-Assisted Dialect Resource Creation. In Proceedings of the Dialect Resources Workshop (DialRes 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani. High Resource Bias in AI-Driven Neology: Structural Inequality in Lexical Innovation. In Proceedings of the Workshop on Neology and Large Language Models (NeoLLM 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani. Grounding Information Disorder in NLP: A Theoretical and Operational Framework. In Proceedings of the Workshop on Information Disorder (INDOR 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani. Cultural Adaptation in Large Language Models for Political Discourse. In Proceedings of the Political Natural Language Processing Workshop (Political NLP 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani. Toward Cognitive Alignment in Large Language Models: Integrating Linguistic Theory and Human Data. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani. Toward Responsible and Epistemically Grounded Multilingual LLMs for Computational Social Science and Humanities. In Proceedings of the Workshop on Large Language Models for Social Sciences and Humanities (LLM4SSH 2026) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani, Kholoud Khalil Aldous and Isra Fejzullaj. AlbanianLLMSafety: A Safety Evaluation Dataset for Large Language Models in Albanian. In Proceedings of the SIGUL 2026 Workshop (Special Interest Group on Under-resourced Languages) at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani, Shimaa Amer Ibrahim, Aruzhan Muratbek, Olzhasbek Zhakenov and Adiya Akhmetzhanova. KZ-SafetyPrompts: A Kazakh Safety Evaluation Prompt Dataset for Large Language Models. In Proceedings of the SIGUL 2026 Joint Workshop with ELE, EURALI and DCLRL at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani, Kholoud Khalil Aldous and Yicheng Gao. Beyond English and Evasion: A Human-Annotated Multi-Domain Benchmark for High-Stakes LLM Safety Evaluation in Chinese. In Proceedings of the RESOURCEFUL 2026 Workshop at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Wajdi Zaghouani, Mabrouka Bessghaier and Kais Attia. Nakba Discourse 2025: A Bilingual Social Media Dataset for Collective Trauma Analysis. In Proceedings of Nakba-NLP 2026: The 2nd International Workshop on Nakba Narratives as Language Resources at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Kholoud Khalil Aldous, Md Rafiul Biswas, Mabrouka Bessghaier, Shimaa Amer Ibrahim, Kais Attia and Wajdi Zaghouani. StanceNakba Shared Task: Actor and Topic-Aware Stance Detection in Public Discourse. In Proceedings of Nakba-NLP 2026: The 2nd International Workshop on Nakba Narratives as Language Resources at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Ashhadul Islam, Md Rafiul Biswas, Samir Brahim Belhaouari and Wajdi Zaghouani. Pushing Boundaries at NakbaVirality: Recursive Prompt Improvement for Multimodal Virality Classification. In Proceedings of Nakba-NLP 2026: The 2nd International Workshop on Nakba Narratives as Language Resources at the Language Resources and Evaluation Conference (LREC 2026). Palma de Mallorca, Spain, 11–16 May 2026.

Zaghouani, W.; Biswas, M. R.; Bessghaier, M.; Ibrahim, S. A.; & Mikros, G. (2026). *ClimateChat-300K: A Multi-Modal Facebook Dataset for Understanding Diverse Perspectives in Climate Communication.* In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026).*

Sharqawi, E. A.; & Zaghouni, W. (2026). *AraHopeCorpus: Annotation Guidelines and Dataset for Hope Speech in Arabic Social Media Crisis Discourse*. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026)*.

Zaghouni, W.; Bessghaier, M.; Biswas, M. R.; & Ibrahim, S. A. (2026). *Audience Engagement with Arabic Women's Social Empowerment and Wellbeing: A Decadal Corpus*. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026)*.

Ali Al-Athba, A.; & Zaghouni, W. (2026). *Cohesion-6K: An Arabic Dataset for Analyzing Social Cohesion and Conflict in Online Discourse*. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026)*.

Zaghouni, W.; Ibrahim, S. A.; & Bouamor, H. (2026). *ArabDiscrim: A Decade-Long Arabic Facebook Corpus on Racism and Discrimination*. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026)*.

Zaghouni, W.; Attia, K.; Biswas, M. R.; & Eryani, F. (2026). *ArPoMeme: An Annotated Arabic Multimodal Dataset for Political Ideology and Polarization*. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026)*.

Zaghouni, W.; Ibrahim, S. A.; Bouamor, H.; & Bessghaier, M. (2026). *JobArabi: An Arabic Corpus and Analysis of Job Announcements from Social Media*. In *Proceedings of the 15th Language Resources and Evaluation Conference (LREC 2026)*.

Bessghaier, M.; Biswas, M. R.; Ibrahim, S.; & Zaghouni, W. (2026). *A Multi-Task Learning Framework for Modeling Engagement and Topic-Sensitive Responses in Arabic Women's Discourse*. In *Findings of the Association for Computational Linguistics: EACL 2026*.

Zhetpysbayev, R.; & Zaghouni, W. (2026). *Eterna: An AI-Powered Journaling System for Contextual Memory Preservation and Digital Legacy Creation*. In *Companion Proceedings of the 2026 ACM International Conference on Intelligent User Interfaces (IUI 2026), Demo Track*.

Zaghouni, W.; Shlkamy, E. S.; & Bessghaier, M. (2026). *From Posts to Pressure: An Arabic Dataset about Stress and Mental-Health Monitoring*. In *Proceedings of the 2nd Workshop on NLP for Languages Using Arabic Script (AbjadNLP 2026)*.

2025

Darwish, K.; Ali, A.; Abu Farha, I.; Touileb, S.; Zitouni, I.; Abdelali, A.; Al-Ghamdi, S.; Alkhereyf, S.; Zaghouni, W.; Khalifa, S.; AlKhamissi, B.; Almatham, R.; Hamed, I.; Alyafeai, Z.; Alowisheq, A.; Inoue, G.; Mrini, K.; & Alshammari, W. (Eds.). (2025). *Proceedings of The Third Arabic Natural Language Processing Conference*. Association for Computational Linguistics.

Zaghoani, W., & Biswas, M. R. (2025). *HopeEmo: A Bilingual Social Media Corpus for Emotion and Hope Speech Analysis*. *Impulses and Approaches to Computer-Mediated Communication*, 33.

Darwish, K.; Ali, A.; Abu Farha, I.; Touileb, S.; Zitouni, I.; Abdelali, A.; Al-Ghamdi, S.; Alkhereyf, S.; Zaghouni, W.; Khalifa, S.; AlKhamissi, B.; Almatham, R.; Hamed, I.; Alyafeai, Z.; Alowisheq, A.; Inoue, G.; Mrini, K.; & Alshammari, W. (Eds.). (2025). Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks. Association for Computational Linguistics.

Biswas, M. R.; Bessghaier, M.; Alam, F.; & Zaghouni, W. (2025). MarsadLab at AraGenEval Shared Task: LLM-Based Approaches to Arabic Authorship Style Transfer and Identification. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 88–93). Association for Computational Linguistics.

Bessghaier, M.; Ibrahim, S.; Biswas, M. R.; & Zaghouni, W. (2025). MarsadLab at AraHealthQA: Hybrid Contextual–Lexical Fusion with AraBERT for Question and Answer Categorization. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 233–238). Association for Computational Linguistics.

Ibrahim, S.; Biswas, M. R.; Bessghaier, M.; & Zaghouni, W. (2025). MarsadLab at BAREC Shared Task 2025: Strict-Track Readability Prediction with Specialized AraBERT Models on BAREC. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 274–279). Association for Computational Linguistics.

Bashiti, A.; Aljabari, A.; Hamoud, H. K.; Biswas, M. R.; Shalash, B. M.; Jarrar, M.; Zaraket, F.; Mikros, G.; Asgari, E.; & Zaghouni, W. (2025). ImageEval 2025: The First Arabic Image Captioning Shared Task. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 376–389). Association for Computational Linguistics.

Zaghouni, W.; Biswas, M. R.; Bessghaier, M.; Ibrahim, S.; Mikros, G.; Hasnat, A.; & Alam, F. (2025). MAHED Shared Task: Multimodal Detection of Hope and Hate Emotions in Arabic Content. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 560–574). Association for Computational Linguistics.

Biswas, M. R.; Attia, K.; Ibrahim, S.; Bessghaier, M.; & Zaghouni, W. (2025). MarsadLab at NADI Shared Task: Arabic Dialect Identification and Speech Recognition using ECAPA-TDNN and Whisper. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 752–756). Association for Computational Linguistics.

Biswas, M. R.; Ibrahim, S.; Attia, K.; Alam, F.; & Zaghouni, W. (2025). MarsadLab at PalmX Shared Task: An LLM Benchmark for Arabic Culture and Islamic Civilization. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 818–824). Association for Computational Linguistics.

Boucekif, A.; Rashwani, S.; Mohamed, E. S. A.; Alkhatib, M.; Sbahi, H.; Gaben, S.; Zaghouni, W.; Erbad, A.; & Ghaly, M. (2025). QIAS 2025: Overview of the Shared Task on Islamic Inheritance Reasoning and Knowledge Assessment. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 851–860). Association for Computational Linguistics.

Ruggeri, F.; Muti, A.; Korre, K.; Struß, J. M.; Siegel, M.; Wiegand, M.; Alam, F.; Biswas, M. R.; Zaghouni, W.; Nawrocka, M.; Ivasiuk, B.; Razvan, G.; & Mihail, A. (2025). Overview of the CLEF-2025 CheckThat! Lab task 1 on subjectivity in news articles. Working Notes of CLEF 2025 (pp. 681–694).

Bessghaier, M.; Biswas, M. R.; Dhouib, A.; & Zaghouni, W. (2025). MarsadLab at TAQEEM 2025: Prompt-Aware Lexicon-Enhanced Transformer for Arabic Automated Essay Scoring. In Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks (pp. 998–1002). Association for Computational Linguistics.

Biswas, M. R.; Ibrahim, S.; Bessghaier, M.; & Zaghouni, W. (2025). Evaluation of Pretrained and Instruction-Based Pretrained Models for Emotion Detection in Arabic Social Media Text. In Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing: Natural Language Processing in the Generative AI Era (pp. 158–165). INCOMA Ltd., Shoumen, Bulgaria.

Zaghouni, W.; & Biswas, M. R. (2025). EmoHopeSpeech: An Annotated Dataset of Emotions and Hope Speech in English and Arabic. In Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing: Natural Language Processing in the Generative AI Era (pp. 1406–1412). INCOMA Ltd., Shoumen, Bulgaria.

Zaghouni, W.; & Biswas, M. R. (2025). An Annotated Corpus of Arabic Tweets for Hate Speech Analysis. In Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing: Natural Language Processing in the Generative AI Era (pp. 1413–1419). INCOMA Ltd., Shoumen, Bulgaria.

Biswas, M. R.; & Zaghouni, W. (2025). Enhancing Arabic Dialectal Sentiment Analysis through Advanced Data Augmentation Techniques. In Proceedings of the Shared Task on Sentiment Analysis for Arabic Dialects (pp. 24–28). INCOMA Ltd., Shoumen, Bulgaria.

Ibrahim, S. A.; Bessghaier, M.; & Zaghouni, W. (2025). Ahasis Shared Task: Hybrid Lexicon-Augmented AraBERT Model for Sentiment Detection in Arabic Dialects. In Proceedings of the Shared Task on Sentiment Analysis for Arabic Dialects (pp. 29–34). INCOMA Ltd., Shoumen, Bulgaria.

Al Heraki, H.; & Zaghouni, W. (2025). Analyzing Digital Polarization on Hijab: A Dataset of Annotated YouTube Comments. In Proceedings of the Nineteenth International AAI Conference on Web and Social Media (ICWSM 2025) (pp. 2350–2360). AAI Press.

Suwaileh, R.; Hasanain, M.; Hubail, F.; Zaghouni, W.; & Alam, F. (2025). ThatiAR: Subjectivity Detection in Arabic News Sentences. In Proceedings of the Nineteenth International AAI Conference on Web and Social Media (ICWSM 2025) (pp. 2587–2602). AAI Press.

Zaghouni, W.; & Biswas, M. R. (2025). EmoHopeSpeech: An Annotated Dataset of Emotions and Hope Speech in English and Arabic. arXiv preprint arXiv:2505.11959.

Zaghouani, W.; & Biswas, M. R. (2025). An Annotated Corpus of Arabic Tweets for Hate Speech Analysis. arXiv preprint arXiv:2505.11969.

Biswas, M. R.; Alam, F.; & Zaghouani, W. (2025). MARSAD: A Multi-Functional Tool for Real-Time Social Media Analysis. arXiv preprint arXiv:2512.01369.

Al-Sharafi, A. G.; Khader, M. M.; Ahmed, M.; Al-Sioufy, M. H.; Zaghouani, W.; & Al-Zawqari, A. (2025). A hybrid annotation model for Arabic argumentative debate corpus. In Proceedings of the International Conference on Arabic Language Processing (pp. 97–113). Springer Nature Switzerland.

Bensalem, I.; Grubišić, I.; El Goual, A.; Rosso, P.; Charfi, A.; & Zaghouani, W. (2025). Stance Detection in Arabic Dialects: Preliminary Experiments. In Proceedings of the International Conference on Arabic Language Processing (pp. 139–150). Springer Nature Switzerland.

Al-Absi, H. R., Kurup, D. G., Daoud, A., Schneider, J., Zaghouani, W., Al Marri, S. M. H., & Mou, Y. A. (2025). Geo-spatial and Temporal Analysis of Hadith Narrators. In International Congress on Information and Communication Technology (pp. 343-352). Singapore: Springer Nature Singapore.

Daoud, A.; Kurup, D. G.; Al-Absi, H. R.; Schneider, J.; Zaghouani, W.; Al Marri, S. M. H. M.; & Mou, Y. A. (2025). A Webtool for Hadith Authorship and Verification with Scholarly Expertise.

Kurup, D. G.; Daoud, A.; Schneider, J.; Zaghouani, W.; Al Marri, S. M. H. M.; Al-Absi, H. R.; & Mou, Y. A. (2025). Exploring Semantic Hadith Overlap Across Topics. In Proceedings of the International Conference on Arabic Language Processing (pp. 127–138). Springer Nature Switzerland.

Shurafa, C.; & Zaghouani, W. (2025). Corpus Analysis of COVID-19 Related Loneliness on Twitter. In Communications in Computer and Information Science (vol. 2340) (pp. 80–93). Springer, Cham.

Zaghouani, W.; Seljan, S.; Dunder, I.; Yahiaoui, R.; & Al-Adwan, A. (2025). Creating a Multilingual Dataset in Arabic and Croatian from Sports Videos Through a Data Processing Pipeline Combining ASR and MT. In Proceedings of the International Conference on Arabic Language Processing (pp. 182–195). Springer Nature Switzerland.

2024

Charfi, A., Atalla, A., Akasheh, R., Bessghaier, M., & Zaghouani, W. (2024). A Web-Based Hate Speech Detection System for Dialectal Arabic. *DATA 2024*, 478–485.

Ahmed, A. M.; Zhang, X.; Rezk, L. M.; & Zaghouani, W. (2024). Building an annotated L1 Arabic/L2 English bilingual writer corpus: The Qatari corpus of argumentative writing (QCAW). *Corpus-Based Studies across Humanities*, 1(1), 183–215. De Gruyter.

Al-Absi, H. R.; Kurup, D. G.; Daoud, A.; Schneider, J.; Zaghouni, W.; Al Marri, S. M. H. M.; & Mou, Y. A. (2024). Al-Ahadeeth: A Visualization Tool of the Hadiths' Chain of Narrators. In International Conference on Human-Computer Interaction (pp. 3–8). Springer Nature Switzerland, Cham.

Al Emadi, M. M.; & Zaghouni, W. (2024).

Emotional Toll and Coping Strategies: Navigating the Effects of Annotating Hate Speech Data. In Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies @ LREC-COLING 2024 (pp. 66–72).

Al Emadi, M.; ElMesselmani, J.; Bermak, L.; Abdullah, G.; Sharqawi, E.; Jrad, A.; Zouabi, Z.; & Zaghouni, W. (2024).

Narrative Navigators at FIGNEWS 2024 Shared Task: New Frontiers in Bias and Propaganda Annotation Techniques. In Proceedings of the Second Arabic Natural Language Processing Conference (pp. 548–554).

Alam, F.; Biswas, M. R.; Shah, U.; Zaghouni, W.; & Mikros, G. (2024).

Propaganda to Hate: A Multimodal Analysis of Arabic Memes with Multi-agent LLMs. In International Conference on Web Information Systems Engineering (pp. 380–390). Springer Nature Singapore.

Amina Daoud, D. Kurup, H. Al Absi, J. Schneider, W. Zaghouni, S. Al Marri, & Y. Ait Mou (2024). A Webtool for Hadith Authorship and Verification with Scholarly Expertise. In ICIW 2024: The Nineteenth International Conference on Internet and Web Applications and Services (pp. [https–www](https://www)).

Biswas, M. R.; Abir, A. T.; & Zaghouni, W. (2024).

Nullpointer at CheckThat! 2024: Identifying Subjectivity from Multilingual Text Sequence. arXiv preprint arXiv:2407.10252.

Biswas, M. R.; Shah, Z.; & Zaghouni, W. (2024).

MemeMind at ArAIEval Shared Task: Spotting Persuasive Spans in Arabic Text with Persuasion Techniques Identification. arXiv preprint arXiv:2408.04540.

Charfi, A.; Besghaier, M.; Akasheh, R.; Atalla, A.; & Zaghouni, W. (2024).

Hate Speech Detection with ADHAR: A Multi-Dialectal Hate Speech Corpus in Arabic. Frontiers in Artificial Intelligence, 7, Article 1391472.

Charfi, A.; Ben-Sghaier, M.; Atalla, A. S. R.; Akasheh, R.; Al-Emadi, S.; & Zaghouni, W. (2024). MARASTA: A Multi-Dialectal Arabic Cross-Domain Stance Corpus.

In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 11060–11069).

Charfi, A.; Bessghaier, M.; Atalla, A.; Akasheh, R.; & Zaghouni, W. (2024).

A Web-Based Hate Speech Detection System for Dialectal Arabic.

Charfi, A.; Bessghaier, M.; ... Al-Emadi, S.; & Zaghouni, W. (2024).

Stance Detection in Arabic with a Multi-Dialectal Cross-Domain Stance Corpus. Social Network Analysis and Mining, 14(1), 161. Springer.

Hasanain, M.; Hasan, M. A.; Ahmed, F.; Suwaileh, R.; Biswas, M. R.; Zaghouni, W.; & Alam, F. (2024). ArAIEval Shared Task: Propagandistic Techniques Detection in Unimodal and Multimodal Arabic Content. arXiv preprint arXiv:2407.04247.

Hasanain, M.; Suwaileh, R.; Weering, S.; Li, C.; Caselli, T.; Zaghouni, W.; Barrón-Cedeño, A.; Nakov, P.; & Alam, F. (2024).

Overview of the CLEF-2024 CheckThat! Lab Task 1 on Check-Worthiness Estimation of Multigenre Content. In 25th Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024) (pp. 276–286). CEUR-WS.

Jones, M. O.; & Zaghouni, W. (2024).

Misogynistic Disinformation on Arabic Twitter: A Case Study of Online Harassment Against Women Journalists in the Middle East. In DH BENELUX 2024.

Khader, M. M.; Al-Sharafi, A.; Al-Sioufy, M. H.; Zaghouni, W.; & Al-Zawqari, A. (2024).

Munazarat 1.0: A Corpus of Arabic Competitive Debates. In Proceedings of the 6th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT) @ LREC-COLING 2024 (pp. 20–30).

Laabar, S.; & Zaghouni, W. (2024).

Multi-Dimensional Insights: Annotated Dataset of Stance, Sentiment, and Emotion in Facebook Comments on Tunisia's July 25 Measures. In Proceedings of the Second Workshop on NLP for Political Sciences @ LREC-COLING 2024 (pp. 22–32).

Mohamed, A.; & Zaghouni, W. (2024). Expression of Depression Among Arab Twitter Users Using Arabic Corpus Analysis.

Procedia Computer Science, 244, 76–85. Elsevier.

Shah, U.; Biswas, M. R.; Agus, M.; Househ, M.; & Zaghouni, W. (2024).

MemeMind at ArAIEval Shared Task: Generative Augmentation and Feature Fusion for Multimodal Propaganda Detection in Arabic Memes. In Proceedings of the Second Arabic Natural Language Processing Conference (pp. 467–472).

Shestakov, A.; & Zaghouni, W. (2024). Analyzing Conflict Through Data: A Dataset on the Digital Framing of Sheikh Jarrah Evictions.

In Proceedings of the Second Workshop on Natural Language Processing for Political Sciences @ LREC-COLING 2024 (pp. 55–67).

Shurafa, C.; & Zaghouni, W. (2024).

Sentiment Analysis and Emotion Annotation of a Large-Scale Arabic YouTube Trauma Corpus. In 2024 11th International Conference on Behavioural and Social Computing (BESC) (pp. 1–7). IEEE.

Struß, J. M.; Ruggeri, F.; Barrón-Cedeño, A.; Alam, F.; Dimitrov, D.; Galassi, A.; Pachov, G.; Koychev, I.; Nakov, P.; & Siegel, M. (2024).

Overview of the CLEF-2024 CheckThat! Lab Task 2 on Subjectivity in News Articles. In CEUR Workshop Proceedings, 3740 (pp. 287–298). CEUR-WS.

Suwaileh, R.; Hasanain, M.; Hubail, F.; Zaghouni, W.; & Alam, F. (2024).

ThatiAR: Subjectivity Detection in Arabic News Sentences. arXiv preprint arXiv:2406.05559.

Xiao, Y.; Bouamor, H.; & Zaghouni, W. (2024).

Chinese Offensive Language Detection: Current Status and Future Directions. arXiv preprint arXiv:2403.18314.

Zaghouani, W.; Ahmed, A.; Zhang, X.; & Rezk, L. (2024).

QCAW 1.0: Building a Qatari Corpus of Student Argumentative Writing.
In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 13382–13394).

Zaghouani, W.; Mubarak, H.; & Biswas, M. R. (2024).

So hateful! Building a multi-label hate speech annotated Arabic dataset.
In Proceedings of LREC-COLING 2024 (pp. 15044–15055).

Zaghouani, W.; Jarrar, M.; Habash, N.; Bouamor, H.; Zitouni, I.; Diab, M.; El-Beltagy, S. R.; & AbuOdeh, M. (2024). The fignews shared task on news media narratives.
arXiv preprint arXiv:2407.18147.

2023

Bashir, M. H.; Azmi, A. M.; Nawaz, H.; Zaghouani, W.; Diab, M.; Al-Fuqaha, A.; & Qadir, J. (2023). Arabic natural language processing for Qur'anic research: A systematic review. *Artificial Intelligence Review*, 56(7), 6801–6854. Springer Netherlands, Dordrecht.

Bashir, M. H.; Azmi, A. M.; Nawaz, H.; Zaghouani, W.; Diab, M.; Al-Fuqaha, A.; & Qadir, J. (2023). Author Correction: Arabic natural language processing for Qur'anic research: A systematic review. *Artificial Intelligence Review*, 56(11), 13951–13952. Springer Netherlands, Dordrecht.

Bianchi, R.; Weber, A.; Yyelland, B.; Ghanam, R.; Kittaneh, K.; & Zaghouani, W. (2023). Synchronous conferencing software-assisted teaching & learning: Preliminary lessons from an inter-institutional faculty group in Qatar.
In INTED2023 Proceedings (pp. 5739–5744). IATED.

Biswas, M. R.; Islam, A.; Shah, Z.; Zaghouani, W.; & Belhaouari, S. B. (2023). Can ChatGPT be your personal medical assistant? In Proceedings of the 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–5). IEEE.

Biswas, M. R.; Mohsen, F.; Shah, Z.; & Zaghouani, W. (2023).

Potentials of ChatGPT for annotating vaccine-related tweets.
In Proceedings of the 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–6). IEEE.

Galassi, A.; Ruggeri, F.; Barrón-Cedeño, A.; Alam, F.; Caselli, T.; Kutlu, M.; Struß, J. M.; Antici, F.; Hasanain, M.; Köhler, J.; & others (2023).

Overview of the CLEF-2023 CheckThat! Lab: Task 2 on subjectivity in news articles.
In 24th Working Notes of the Conference and Labs of the Evaluation Forum (CLEF-WN 2023) (pp. 236–249). CEUR-WS.

Alam, F.; Barrón-Cedeño, A.; Cheema, G. S.; Shahi, G. K.; Hakimov, S.; Hasanain, M.; Li, C.; Míguez, R.; Mubarak, H.; Zaghouani, W.; & others (2023).

Overview of the CLEF-2023 CheckThat! Lab Task 1 on check-worthiness of multimodal and multigenre content.

Barrón-Cedeño, A.; Alam, F.; Galassi, A.; Da San Martino, G.; Nakov, P.; Elsayed, T.; Azizov, D.; Caselli, T.; Cheema, G. S.; Haouari, F.; & others (2023).

Overview of the CLEF-2023 CheckThat! Lab on checkworthiness, subjectivity, political bias, factuality, and authority of news articles and their source. In International Conference of the Cross-Language Evaluation Forum for European Languages (pp. 251–275). Springer Nature, Switzerland.

Hasanain, M.; Alam, F.; Mubarak, H.; Abdaljalil, S.; Zaghouni, W.; Nakov, P.; Martino, G. D. S.; & Freihat, A. A. (2023).

ArAIEval shared task: Persuasion techniques and disinformation detection in Arabic text. arXiv preprint arXiv:2311.03179.

Islam, A.; Biswas, M. R.; Zaghouni, W.; Belhaouari, S. B.; & Shah, Z. (2023).

Pushing boundaries: Exploring zero-shot object classification with large multimodal models. In Proceedings of the 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1–5). IEEE.

Magnossão de Paula, A. F.; Bensalem, I.; Rosso, P.; & Zaghouni, W. (2023).

Transformers and ensemble methods: A solution for hate speech detection in Arabic languages. arXiv preprint arXiv:2303.09823.

Sawaf, H.; El-Beltagy, S. R.; Zaghouni, W.; Magdy, W.; Abdelali, A.; Tomeh, N.; Farha, I. A.; Habash, N.; Khalifa, S.; Keleg, A.; & others (2023).

Proceedings of ArabicNLP 2023.

Yyelland, B.; Weber, A. S.; Bianchi, R.; Zaghouni, W.; Kittaneh, K.; Ghanam, R.; Fejzullaj, S.; Iqbal, H.; & Niaz, A. (2023).

Privacy concerns during remote emergency learning in higher education in Qatar. International Association for Development of the Information Society (IADIS).

2022

Damith Premasiri, Tharindu Ranasinghe, **Wajdi Zaghouni**, Ruslan Mitkov . DTW at Qur'an QA 2022: Utilizing Transfer Learning with Transformers for Question Answering in a Low-resource Domain. In Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022). **H-Index: 49**

Nakov, Preslav and Barron Cedeno, Alberto and Da San Martino, Giovanni and Alam, Firoj and Struss, Julia Maria and Mandl, Thomas and Miguez, Ruben and Caselli, Tommaso and Kutlu, Mucahid and **Zaghouni, Wajdi** (2022). The CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection. In **Springer**, European Conference on Information Retrieval, pages 416- 428.

De Paula, Angel Felipe Magnossao and Rosso, Paolo and Bensalem, Imene and **Zaghouni, Wajdi** (2022). UPV at the Arabic Hate Speech 2022 Shared Task: Offensive Language and Hate Speech Detection using Transformers and Ensemble Models. In Proceedings of the OSACT 2022, LREC. **H- Index: 49**

Nakov, Preslav and Barron-Cedeno, Alberto and Da San Martino, Giovanni and Alam, Firoj and Kutlu, Mucahid and **Zaghouani, Wajdi** and Li, Chengkai and Shaar, Shaden and Mubarak, Hamdy and Nikolov, Alex (2022). Overview of the CLEF-2022 CheckThat! lab task 1 on identifying relevant claims in tweets

Nakov, Preslav and Barr'on-Cedeno, Alberto and da San Martino, Giovanni and Alam, Firoj and Struss, Julia Maria and Mandl, Thomas and Miguez, Ruben and Caselli, Tommaso and Kutlu, Mucahid and **Zaghouani, Wajdi** (2022). Overview of the CLEF--2022 CheckThat! Lab on Fighting the COVID- 19 Infodemic and Fake News Detection. **Springer**, International Conference of the Cross-Language Evaluation Forum for European Languages, Pages 495-520

2021

- Shaar, Shaden and Alam, Firoj and Martino, Giovanni Da San and Nikolov, Alex and **Zaghouani, Wajdi** and Nakov, Preslav and Feldman, Anna (2021). Findings of the NLP4IF-2021 shared tasks on fighting the COVID-19 infodemic and censorship detection
- Alam, Firoj and Shaar, Shaden and Dalvi, Fahim and Sajjad, Hassan and Nikolov, Alex and Mubarak, Hamdy and Martino, Giovanni Da San and Abdelali, Ahmed and Durrani, Nadir and Darwish, Kareem and **Wajdi Zaghouani** (2021). Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society. Findings of EMNLP-2021. **H-Index 154** (Second Best NLP Conference)
- *Al-Boinin, Hamda* and **Zaghouani, Wajdi** and Vatanasakdakul, Savanid Nui (2021). Women Micro- Entrepreneurs in Qatar: Motivation, Challenges, and Social Media Opportunities. In Proceedings of the 8th International Conference on Behavioral and Social Computing (BESC)

2020

- *Abbes, I., Zaghouani, W., , El-Hardlo, O., & Achour, F.* (2020). DAICT: A Dialectal Arabic Irony Corpus Extracted From Twitter. In *Proceedings of the 12th Edition of its Language Resources and Evaluation Conference*. Language Resources and Evaluation Conference. **H-Index: 49**
- *Al-Mulla, S., Zaghouani, W., &* (2020). Building a Qatari Heritage Expressions Corpus. In *The 4th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Task on Offensive Language Detection. International Conference on Language Resources and Evaluation (LREC- 2020)*. **H-Index: 49**
- Schlippe, Tim and *Alessai, Shaimaa* and El-Taweel, Ghanimeh and Wolfel, Matthias and **Zaghouani, Wajdi** (2020). Visualizing Voice Characteristics with Type Design in Closed Captions for Arabic. In Proceedings of the International Conference on Cyberworlds (CW), 2020

2019

- *Sawaly, D., Zaghouani, W., , & Kaufer, D.* (2019). A Voice for Education: A Digital Analysis of Sheikha Moza's Public Statements, 2009–2018. In *International*

Conference on Social Informatics SocInfo 2019: Social Informatics (p. pp 239–252). Springer, Cham.

- Charfi, A., **Zaghouani, W.**, Hassan Mehdi, S., & Mohamed, E. (2019). A Fine-Grained Annotated Multi- Dialectal Arabic Corpus. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*.
- ACL.
- Rangel, F., Rosso, P., Charfi, A., **Zaghouani, W.**, Ghanem, B., & Sánchez-Junquera, J. (2019). Overview of the track on author profiling and deception detection in Arabic. In *Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019)*. *CEUR Workshop Proceedings*. In: *CEUR-WS.org*,. Association for Computing Machinery (ACM).
- Rangel, F., Rosso, P., Charfi, A., **Zaghouani, W.**, Ghanem, B., & Sánchez-Junquera, J. (2019). On the Author Profiling and Deception Detection in Arabic shared task at FIRE. In *Proceedings of the 11th Forum for Information Retrieval Evaluation*. Association for Computing Machinery (ACM).
- Rangel, F., & Rosso, P., A. Charfi, **Zaghouani, W.** (2019). Detecting Deceptive Tweets in Arabic for Cyber-Security. In *Proceedings of the 17th IEEE International Conference on Intelligence and Security Informatics (ISI)*.

2018

- Ossama Obeid, Salam Khalifa, Nizar Habash, Houda Bouamor, Wajdi Zaghouani and Kemal Oflazer (2018) MADARi: A Web Interface for Joint Arabic Morphological Annotation and Spelling Correction. (LREC 2018, Miyazaki, Japan (7-12 May 2018). **H-Index: 49**
- Bouamor, H., Habash, N., Salameh, M., Zaghouani, W., Rambow, O., Abdulrahim, D. (2018). The MADAR Arabic dialect corpus and lexicon. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC- 2018)*. **H-Index: 49**
- Nizar Habash, Salam Khalifa, Fadhl Eryani, Owen Rambow, Dana Abdulrahim, Alexander Erdmann, Reem Faraj, Wajdi Zaghouani, Houda Bouamor, Nasser Zalmout, Sara Hassan, Faisal Al shargi, Sakhar Alkhereyf, Basma Abdulkareem, Ramy Eskander, Mohammad Salameh and Hind Saddiki (2018) Unified Guidelines and Resources for Arabic Dialect Orthography (LREC 2018, Miyazaki, Japan (7-12 May 2018). **H-Index: 49**
- Alberto Barrón-Cedeno, Tamer Elsayed, Reem Suwaileh, Lluís Màrquez, Pepa Atanasova, Wajdi Zaghouani, Spas Kyuchukov, Giovanni Da San Martino, Preslav Nakov (2018). Overview of the CLEF- 2018 CheckThat! Lab on Automatic Identification and Verification of Political Claims. Task 2: Factuality. In *CEUR Workshop Proceedings*.
- Preslav Nakov, Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Lluís Màrquez, Wajdi Zaghouani, Pepa Atanasova, Spas Kyuchukov & Giovanni Da San Martino (2018). Overview of the CLEF-

2018 checkthat! lab on automatic identification and verification of political claims. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

- Atanasova, P., Barron-Cedeno, A., Elsayed, T., Suwaileh, R., Zaghouni, W., Kyuchukov, S., Nakov, P. (2018). Overview of the CLEF-2018 CheckThat! Lab on automatic identification and verification of political claims. Task 1: Check-worthiness.
- Sawsan AlQahtani, Mona Diab and Wajdi Zaghouni (2018) ARLEX: A Large Scale Comprehensive Lexical Inventory for Modern Standard Arabic. In *Proceedings of the OSACT3 Workshop, LREC 2018, Miyazaki, Japan (7-12 May 2018)*. **H-Index: 49**
- Zaghouni, W., & Charfi, A. (2018). AraP-Tweet: A large multi-dialect twitter corpus for gender, age and language variety identification. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA). **H-Index: 49**
- Zaghouni, W., & Charfi, A. (2018). Guidelines and Annotation Framework for Arabic Author Profiling. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA). **H-Index: 49**

2016 and 2017

- Wajdi Zaghouni. Language Technologies for Social Media. INFuture 2017, Zagreb, Croatia
- Wajdi Zaghouni, Abdelati Hawwari, Sawsan Alqahtani, Houda Bouamor, Mahmoud Ghoneim, Mona Diab and Kemal Oflazer. Using Ambiguity Detection to Streamline Linguistic Annotation, In *Proceedings of Coling Workshop "Computational Linguistics for Linguistic Complexity" (CL4LC)*, Osaka Japan, December 2016
- Wajdi Zaghouni, Nizar Habash, Houda Bouamor, Ossama Obeid, Sawsan Alqahtani, Mona Diab and Kemal Oflazer. 2016. Filtering Dialectal Arabic Text in Two Large Scale Annotation Projects. The 2nd Workshop on Noisy User-generated Text (W-NUT), December 11 2016, Osaka, Japan.
- Wajdi Zaghouni, Ahmed Abdelali, Francisco Guzman and Hassan Sajjad. 2016.
- Normalizing Mathematical Expressions to Improve the Translation of Educational Content. In *Proceedings of the AMTA 2016 Workshop Semitic Machine Translation (SeMaT) Collocated with EMNLP 2016 Workshops on November 1st, 2016 Austin, Texas, USA*. **H-Index 154 (Second Best NLP Conference)**
- Wajdi Zaghouni and Dana Awad. 2016. Toward an Arabic Punctuated Corpus: Annotation Guidelines and Evaluation. In *Proceedings of The*

2nd workshop on Arabic Corpora and Processing Tools 2016 Theme: Social Media. **H-Index: 49**

- Zaghouni, Wajdi, Houda Bouamor, Abdelati Hawwari, Mona Diab, Ossama Obeid, Mahmoud Ghoneim, Sawsan Alqahtani, Kemal Oflazer. 2016. Guidelines and Framework for a Large Scale Arabic Diacritized Corpus. In Proceedings of the International Conference on Language Resources and Evaluation (LREC'2016). **H-Index: 49**
- Irina Temnikova, Zaghouni Wajdi, Stephan Vogel, Nizar Habash. 2016. Applying the Cognitive Machine Translation Evaluation Approach to Arabic. In Proceedings of the International Conference on Language Resources and Evaluation (LREC'2016). **H-Index: 49**
- Zaghouni Wajdi, Nizar Habash, Ossama Obeid, Behrang Mohit, Houda Bouamor, Kemal Oflazer. 2016. Building an arabic machine translation post-edited corpus: Guidelines and annotation. In Proceedings of the International Conference on Language Resources and Evaluation (LREC'2016). **H-Index: 49**
- Ossama Obeid, Houda Bouamor, Zaghouni, Wajdi, Mahmoud Ghoneim, Abdelati Hawwari, Sawsan Alqahtani, Mona Diab, Kemal Oflazer. 2016. MANDIAC: A Web-based Annotation System For Manual Arabic Diacritization. In Proceedings of The 2nd Workshop on Arabic Corpora and Processing Tools 2016 Theme: Social Media.

2015

- E.A. Draffan, Mike Wald, Nawar Halabi, Ouadie Sabia, Wajdi Zaghouni, Amatullah Kadous, Amal Idris, Nadine Zeinoun, David Banes, Dana Lawand. Generating acceptable Arabic Core Vocabularies and Symbols for AAC users. In Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies, Interspeech 2015, Dresden, Germany, 6-10 September 2015. **H-Index 105** (Third Best NLP Conference)
- Zaghouni, Wajdi, Nizar Habash, Houda Bouamor, Alla Rozovskaya, Behrang Mohit, Abeer Heider and Kemal Oflazer. Correction Annotation for Non-Native Arabic Texts: Guidelines and Corpus. In Proceedings of the 9th Linguistic Annotation Workshop, co-located with NAACL in Denver, Colorado, USA, 2015. **H-Index 105** (Third Best NLP Conference)
- Houda Bouamor, Wajdi Zaghouni, Mona Diab, Ossama Obeid, Kemal Oflazer, Mahmoud Ghoneim and Abdelati Hawwari. A Pilot Study on Arabic Multi-Genre Corpus Diacritization. 2015. In Proceedings of the ACL 2015 Workshop on Arabic Natural Language Processing (ANLP), Beijing, China, July 2015.
- Wajdi Zaghouni, Taha Zerrouki and Amar Balla. SAHSOH@QALB-2015 Shared Task: A Rule-Based Correction Method of Common Arabic Native and Non-Native Speakers' Errors. The Second QALB Shared Task on Automatic Text Correction for Arabic. In Proceedings of the ACL 2015 Workshop on Arabic Natural Language Processing (ANLP), Beijing, China, July 2015. **H-Index 169** (Best NLP Conference)

- Alla Rozovskaya; Houda Bouamor; Nizar Habash; Wajdi Zaghouani; Ossama Obeid; Behrang Mohit. The Second QALB Shared Task on Automatic Text Correction for Arabic. In Proceedings of the ACL 2015 Workshop on Arabic Natural Language Processing (ANLP), Beijing, China, July 2015. **H-Index 169** (Best NLP Conference)

2014

- Behrang Mohit; Alla Rozovskaya; Nizar Habash; Wajdi Zaghouani; Ossama Obeid. The First QALB Shared Task on Automatic Text Correction for Arabic. In Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), Doha, Qatar, October 2014. **H-Index 154** (Second Best NLP Conference)
- Serena Jeblee; Houda Bouamor; Wajdi Zaghouani; Kemal Oflazer. CMUQ@QALB-2014: An SMT-based System for Automatic Arabic Error Correction. In Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), Doha, Qatar, October 2014. **H-Index 154** (Second Best NLP Conference)
- Wajdi Zaghouani, Behrang Mohit, Nizar Habash, Ossama Obeid, Nadi Tomeh and Kemal Oflazer. 2014. Large-scale Arabic Error Annotation: Guidelines and Framework. in the Proceedings of the International Conference on Language Resources and Evaluation (LREC'2014), Rejkavik, Iceland, 26-31 May 2014. **H-Index: 49**
- Wajdi Zaghouani and Kais Dukes. 2014. Can Crowdsourcing be used for Effective Annotation of Arabic? In Proceedings of the International Conference on Language Resources and Evaluation (LREC'2014), Rejkavik, Iceland, 26-31 May 2014. **H-Index: 49**
- Wajdi Zaghouani. 2014. Critical Survey of the Freely Available Arabic Corpora. In Proceedings of the International Conference on Language Resources and Evaluation (LREC'2014), OSACT Workshop. Rejkavik, Iceland, 26-31 May 2014. **H-Index: 49**
- Ossama Obeid, Wajdi Zaghouani, Behrang Mohit, Nizar Habash, Kemal Oflazer and Nadi Tomeh. A Web-based Annotation Framework For Large-Scale Text Correction. In Proceedings of IJCNLP'2013, Nagoya, Japan. **H-Index: 37**

2013

- Abdelaati Hawwari, Mohsen Rashwan and Wajdi Zaghouani. 2013. A Lexical Semantic Resource for Quranic Morphological Patterns. The International conference for the development of Quranic studies. <http://www.quranicconferences.com/> . Riyadh, KSA. 16-20 February 2013.
- Wajdi Zaghouani. Arabic Natural Language Processing and the Future. In proceedings of the CECTAL'13, Montreal, Canada. Sept 26th 2013.
- Hawwari, A.; Zaghouani, W.; O'Gorman, T.; Badran, A.; Diab, M., "Building a Lexical Semantic Resource for Arabic Morphological

Patterns," Communications, Signal Processing, and their Applications (ICCSPA), 2013, vol., no., pp.1,6, 12-14 Feb. 2013.

2012

- Wajdi Zaghouni, Abdelati Hawwari and Mona Diab. 2012. A Pilot PropBank Annotation for Quranic Arabic. In Proceedings of the first workshop on Computational Linguistics for Literature, NAACL-HLT 2012, Montreal, Canada. **H-Index 105** (Third Best NLP Conference)
- Mohammed Maamouri, Wajdi Zaghouni, Violetta Cavalli-Sforza, Dave Graff and Mike Ciul. 2012. Developing ARET: An NLP-based Educational Tool Set for Arabic Reading Enhancement. In Proceedings of The 7th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT 2012, Montreal, Canada. **H-Index 105** (Third Best NLP Conference)

2002 until 2011

- Wajdi Zaghouni. 2012. Étude sur la composition des noms de personnes dans la langue arabe. In proceedings of the 25th Journées de linguistique de Laval. 9-11 March 2011, Laval University , Québec, Canada.
- Wajdi Zaghouni , Mona Diab , Aous Mansouri, Sameer Pradhan and Martha Palmer.2010. The Revised Arabic PropBank. In proceedings of the 4th Linguistic Annotation workshop ACL held in Uppsala. July 15-16 2010.
- Eric Atwell, Kais Dukes, Abdul-Baquee Sharaf, Nizar Habash, Bill Louw,Bayan Abu Shawar,Tony McEnery,Wajdi Zaghouni, Mahmoud El- Haj. 2010.Understanding the Quran: a new Grand Challenge for Computer Science and Artificial Intelligence. In Grand Challenges in Computing Research for 2010 and beyond. part of ACM-BCS Visions of Computer Science conference. 13-16 April 2010, Edinburgh University
- Mohamed Maamouri, Ann Bies, Seth Kulick, Wajdi Zaghouni, Dave Graff and Mike Ciul. 2010. From Speech to Trees: Applying Treebank Annotation to Arabic Broadcast News. In Proceedings of LREC 2010, Valetta, Malta, May 17-23, 2010. **H-Index: 49**
- Wajdi Zaghouni, Bruno Pouliquen, Mohamed Ebrahim and Ralf Steinberger .2010.Adapting a resource-light highly multilingual Named Entity Recognition system to Arabic. In Proceedings of LREC 2010, Valetta, Malta, May 17-23, 2010. **H-Index: 49**
- Wajdi Zaghouni. 2010. L'intégration d'un outil de repérage d'entités nommées pour la langue arabe dans un système de veille. Session Démo, TALN 2010, Montréal, 19-23 juillet 2010.
- Mona Diab, Aous Mansouri, Martha Palmer, Olga Babko-Malaya,Wajdi Zaghouni, Ann Bies, Mohammed Maamouri. A Pilot Arabic Propbank; LREC 2008, Marrakech, Morocco, May 28-30, 2008. **H-Index: 49**

- Bruno Pouliquen, Marco Kimler, Ralf Steinberger, Camelia Ignat, Tamara Oellinger, KenBlackler, FlavioFuart, Wajdi Zaghoulani, Anna Widiger, Ann-Charlotte Forslund, Clive Best (2006). Geocoding multilingual texts: Recognition, Disambiguation and Visualisation. Proceedings of the (LREC'2006), pp. 53-58. Genoa, Italy, 24-26 May 2006. **H-Index: 49**
- Yun-Chuang Chiao, Olivier Kraif, Dominique Laurent, Thi Minh Huyen Nguyen, Nasredine Semmar, François Stuck, Jean Véronis, Wajdi Zaghoulani (2006). Evaluation of multilingual text alignment systems: the ARCADE II project. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006). Genoa, Italy, 24-26 May 2006. **H-Index: 49**
- **Zaghoulani Wajdi**. "AUTO-ÉVAL: VERS UN MODÈLE D'ÉVALUATION AUTOMATIQUE DES TEXTES". In Proceedings of Actes du colloque CESLA colloquim, Montréal, UQAM, 2002.

Research Grants Evaluation

- **QNRF** High School Research Experience Program (HSREP), **Panel expert reviewer** for five consecutive years: the editions of 2019; 2020; 2021, 2022, and 2023.
- **QCRI Internships awards, invited Expert Judge (Sept 2022)**
- **National Science Foundation (NSF), USA**, expert reviewer for the Arabic Language Resources area.
- **Grant reviewer** for the American Association for the Advancement of Science (AAAS) - 2016
- **Grant reviewer** for the KACST International Grant Competition on Arabic Natural Language Processing (2017)

Research Group

Given the funding obtained in recent years (6 active grants), I created in 2020 the **Data For Social Good (D4SG)** Research Group¹. The group is currently expanding. Currently, the group is composed of a research group head, 2 affiliate researchers, 1 Post-Doc, 4 Research Assistants, 5 PhD students and 5 MA Students.

Honors/ Awards

- o **Winner of the Early Career Researcher Award** (2nd place) , A HBKU wide award that recognizes outstanding researchers earlier on in their career and acknowledges their excellent research work and accomplishments. September 2022.
- o **Best Paper Award** for Qur'an QA shared task 2022. The 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022).
- o **Best project Award** from The Ministry of Development Planning and Statistics (MDPS). Meetings of the minds, April 2017, Carnegie Mellon University. The Arabic author profiling for cyber-security

- o **Best Poster Award.** Meetings of the minds, Carnegie Mellon University. The QALB Project: Building Resources and Systems for the Automatic Correction of Arabic Text.
- o **Award:** 105,000\$ CAD. Joseph-Armand Bombardier Scholarships. Distinctive Canadian Program for Doctoral Scholarships (2011-2014)

Journal Editorial Board

- I was recently invited to join the Editorial Board of **Nature Journal** for the Humanities & Social Sciences Communications category. First published in 1869, Nature is the world's leading multidisciplinary science journal.
<https://www.nature.com/palcomms/editorialboard>
- **Topic Editor** of Countering Fake News and Hate Speech on Social Media Platforms,
Frontiers in Artificial Intelligence Journal
<https://www.frontiersin.org/research-topics/43248/countering-fake-news-and-hate-speech-on-social-media-platforms>
- **Editorial Board member** of the **Cambridge University Press Journal** on Natural Language Engineering (**Q1**) – **Since 2020 Until Today**
<https://www.cambridge.org/core/journals/natural-language-engineering/information/about-this-journal/editorial-board>
- **Associate Editor** for the ACM Journal on **Transactions on Asian and Low-Resource Language Information Processing (TALIP)**– **Since 2018 – Today**. The ACM TALLIP publishes high-quality original archival papers, survey papers, and technical notes in the areas of computation and processing. <https://dl.acm.org/journal/tallip/editorial-board>
- **Associate Editor** for the Frontiers in Artificial Intelligence - Natural Language Processing – **Since 2021 Until Today**. Frontiers in Artificial Intelligence publishes rigorously peer-reviewed research on the cutting-edge, disruptive technological revolution of Artificial Intelligence (AI).
<https://www.frontiersin.org/journals/artificial-intelligence/editors>

Selected Invited Talks and Panels

- Challenges in Analyzing Data Produced in Arabic at the CIRS Roundtable on Big Data in the Middle East: January 19 & 20, 2020
- Social Media Analytics and Religious Text, Frankfurt University (Goethe), 22 October 2020
- Language Technologies for Social Media. Invited talk at the InFuture Conference in Zagreb, Croatia (2017)
- DA/MSA lexicon development. First Workshop on Arabic Dialect Technologies, NYU ABU DHABI, 14 to 16 November 2016
- Social Entrepreneurship: Improving Literacy in the Arab World Through Ebooks for Children. ROTA Empower Workshops 2016.

- Survey of the Freely Available Arabic Corpora. Qatar Computing Research Institute, Arabic Language Technologies Round Table, Doha, Qatar, April 23 2014.
- RENAR: A Rule-Based Arabic Named Entity Recognition System. Invited talk at the Carnegie Mellon University NLP Round Table., December 2013.
- Building a Lexical Semantic Resource for Arabic Morphological Patterns. 2013. Invited Oral presentation at the Arabic Natural Language Processing session at the International Conference on Communications, Signal Processing and their Applications (ICCSIPA'13). Sharjah, UAE.
- The QALB Project. Guest talk at the 8th Assistive Technology Research Forum at MADA Center, Doha, Qatar, November 26th 2013
- Arabic Natural Language Processing and the Future. Invited by the CECTAL'13, Montreal, Canada. Sept 26th 2013.
- Recent advance of PropBank Annotation for Arabic. Invited presentation at the UQAM - AECSL Round Table, Montreal, Canada, December 2013

Selected Invited Tutorials

- Tutorial on Mobile Application Creation, 6 May 2019 (QCRI Research Complex)
- Fact-Checking in Social Media 101, Zagreb University, Croatia (Nov 2017)
- Data Mining for Humanities for Social Media (Nov 2017), a graduate seminar delivered to 24 PhD students from Zagreb University, Croatia
- Semantic Annotation (2 hours tutorial), University of Quebec in Montreal Linguistics Colloquium (December 2012)
- Introduction to Corpus Linguistics (3 hours tutorial), the University of Quebec in Montreal Students Linguistics Colloquium (May 2011)
- Arabic Corpus Linguistics: Tutorial (2 days), December 2011, Faculty of Humanities, Kairouan University
- Unix commands tutorial for Natural Language Processing (with Dave Graff). The Linguistic Data Consortium (LDC), The University of Pennsylvania. March 2009, Philadelphia, PA

Media Outreach and Interviews

- Media Coverage for the International Workshop on Arabic Big Data and AI to be held at QNL auditorium on May 11-12 2022 and organized by Wajdi Zaghouani and posted in Qatar Living, Lusail News, Fanar Media and Al Arab.
- Tools and Applications to help us fight the COVID-19. Al Araby TV. (2020)
- COVID-19 and the Job Market. Qatar TV. (2020)
- MES Conference Presentation and Discussion. Al-Sharq.com (2020)
- COVID-19 and Fake News Detection. Www.alarab.qa. (2020)
- MES Conference Presentation and Discussion. Qatar TV. (2020)
- Digital Humanities in MENA region. Al Watan (2018)

International Conference and Workshop Organization

- **Chair** of the First International Workshop on Arabic Big Data & AI (IWABigDAI), May 11 and 12, 2022

- **Co-organizer and publicity chair** of the 2022 Conference on Empirical Methods in Natural Language Processing (Abu Dhabi, December 8th 2022)
- **Co-organizer** of the CLEF-2022 CheckThat! Lab on fighting the COVID-19 infodemic and fake news detection (5 - 8 September 2022, Bologna – Italy)
- Co-organizer of the WANLP-2022, The Seventh Arabic Natural Language Processing Workshop co-locate with EMNLP 2022 (Abu Dhabi, December 8th 2022)
- Co-organizer of the SHARED TASK ON PROPAGANDA DETECTION IN ARABIC, co-located with EMNLP in Dec 8, 2022, Abu Dhabi, UAE
- Co-organizer, MESD - MES Conference 2022,. Digital Humanities Panel.
- Program Chair, 8th IEEE International Conference on Behavioral and Social Computing (BESC), (University), 2021
- **Co-organizer** of the Sixth Arabic Natural Language Processing Workshop Co-located Online with EACL 2021, Apr 19, 2021
- Co-organizer of the first SHARED TASK ON SARCASM AND SENTIMENT DETECTION IN ARABIC, co- located with WANLP 2021: The Sixth Workshop for Arabic Natural Language Processing / EACL Apr 19, 2021
- **Co-organizer** of The Fifth International Arabic Natural Language Processing Workshop , ACL 2020, Barcelona, Spain, Dec 12, 2020
- **Chair and Organizer** of the Workshop on Digital Humanities and Computational Linguistics in the Middle East, April 20, 21 2020
- **Co-organizer** of Author Profiling and Deception Detection in Arabic shared task, The 11th meeting of Forum for Information Retrieval Evaluation 2019 held in Kolkata, India,
- Co-organizer of the International Fake News Detection Contest (QICC), held from October 1-3, 2019.
- **Co-organizer** of The Fourth International Arabic Natural Language Processing Workshop ,
- ACL 2019, Florence, Italy, July 28-Aug 2, 2019
- **Chair and organizer** of the First International Workshop on Social Media Analysis in the Arab World co- located with the Social Informatics Conference (SocInfo 2019) on November 18, 2019
- **Co-organizer** of the CLEF-2018, international evaluation LAB on Automatic Identification and Verification of Claims in Political Debates held on Sept 2018 in Avignon, France.
- **Conference Area Chair** (Resources and Evaluation) of the 55th annual meeting of the Association for Computational Linguistics (ACL) in Vancouver, Canada. July 30 - August 4 2017.
- **Co-organizer** of the Third Arabic Natural Language Processing Workshop co-located with EACL 2017, Valencia, Spain, Apr 3, 2017
- **Co-organizer** of the 2nd Arabic NLP workshop co-located with ACL- IJCNLP 2015, Beijing, China.
- **Co-organizer** and member of the Shared task committee in the first Shared Task: Automatic Arabic Error Correction co-located with the WANLP (ACL-IJCNLP) 2015, Beijing, China
- **Co-organizer** of the 1st Arabic NLP Workshop co-located with EMNLP 2014 in Doha, Qatar
- **Co-organizer** and member of the Shared task committee in the first Shared Task: Automatic Arabic Error Correction co-located with the WANLP (EMNLP) 2014 in Doha, Qatar
- **Chair** of the CEC-TAL 2013 Conference, Tunisia

Service to the Academic Community

- Program Committee Member (Reviewer) - RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING, biennial RANLP conference, 2011, 2013, 2015 and 2021
- Program Committee Member (Reviewer) - The 58th Annual Meeting of the Association for Computational Linguistics (ACL)
- Program Committee Member (Reviewer) - The 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019
- Program Committee Member (Reviewer) - Artificial Intelligence and Natural Language Conference, AINL: Artificial Intelligence and Natural Language Conference, 2019
- Program Committee Member (Reviewer International Arabic Natural Language Processing Workshop, 2022, 2021, 2020, 2019, 2017, 2015, and 2014
- Program Committee Member (Reviewer) - 7th International Conference Arabic Language Processing ICALP , ICALP , ICALP (ex-CITALA), 2019
- Standing Reviewer for Computer Speech & Language Journal, Computer Speech & Language Journal
- Standing Reviewer for Natural Language Engineering Journal, Natural Language Engineering Journal
- Standing Reviewer for The Journal of Natural Language Engineering (2013, 2015, 2016, 2017, and 2018)
- Regular Reviewer for The Language Resources and Evaluation journal (2012, 2013, 2014, 2015, 2016, 2017 and 2019)
- Standing Reviewer for The Journal of King Saud University- Computer and Information Sciences (2013, 2014, 2016)
- Standing Reviewer for The International Journal of Speech Technologies (2011, 2014,2015, 2016 and 2017)
- Standing Reviewer for The Information Processing and Management Journal (2017)
- Standing Reviewer for The Journal of Information Science (2017)
- Area Chair at the The 55th annual meeting of the Association for Computational Linguistics (ACL) for the Language Resources and Evaluation Area, 2017
- PC member of the 2017 European Association for Computational Linguistics 2017, Valencia, Spain.
- Session Chair of the Arabic Natural Language Processing session at the 13th International Conference on Communications Signal Processing and their Applications, 2013, Sharjah, UAE.
- PC member of the LREC 2012 workshop on religious text
- PC member of Language Resources and Evaluation Conference (LREC): 2022, 2020, 2018, 2016, 2014, 2012, 2010 and 2008.

Courses Created and Taught

- **GEN/AI and NLP for Media Innovation (Fall 2024, Northwestern Qatar)**

- **Intro to Social Computing (Spring 2025, Northwestern Qatar)**
- **Data Sciences and Text Analytics (Spring 2025, Northwestern Qatar)**
- **Methods in Digital Humanities** (Graduate course), HBKU (Spring 2018, Fall 2018, Fall 2019, Fall 2020, Fall 2021, and Fall 2022)
- **Coding for Humanities**, HBKU (Spring 2022, Spring 2023)
- **Digital Communication and Media**, HBKU (Graduate course), (Spring 2018)
- **Emerging Technologies and Applications**, HBKU (Fall 2018)
- **Independent Research Project**, HBKU (Fall 2018, Fall 2019, Fall 2019, Fall 2020, Fall 2020, Fall 2021, Fall 2022)
- **Doctoral Independent Study**, HBKU (Spring 2021, Spring 2022)
- **Doctoral Pro Seminar in Digital Humanities**, HBKU (Spring 2020)

Graduate Students Supervision

- **MA Students Supervised: 27**
- **PHD Students Supervised: 4**

University Service

- AI Arabic Strategy Experience Committee (HBKU President Office)
With Preslav Nakov and Jens Schneider, (University)
- Shared Governance Committee (CHSS Representative)
- PhD Program Committee, (College)
- AI in Education City Steering Committee, (University)
- MA Program in MES Curriculum Committee, (College)
- CHSS Student Life Committee (College of Humanities and Social Sciences)
- CHSS Academic Curriculum, Assessment, and Accreditation Committee, (College of Humanities and Social Sciences)
- Ph.D. Program Admission and Scholarships Committee, (College of Humanities and Social Sciences)
- College Senate Committee, (College of Humanities and Social Sciences)
- Digital Humanities M.A Students Admission Committee, (Middle Eastern Studies Department)
- CHSS-CMU-Q 4+1 Committee, (Middle Eastern Studies Department)
- Student Recruitment Trip (China), (University), (University) Ph.D. Student Recruitment Trip in Beijing, Shagnhai and Hangzu. Nov 20 - Nov 28 2019.
- Student Recruitment Trip (Oman), (University)
- Student Recruitment Event (Qatar University), (University)

Professional Membership

- **Member of the advisory committee** of the Arabic Natural Language Processing Workshop WANLP.
<https://sites.google.com/view/wanlp2022/organizers?authuser=0>

- SIGARAB (ACL Special Interest Group on Arabic NLP)

Member of the advisory board: <https://sites.google.com/nyu.edu/wanlp-archive/sigarab>

- ELRA (European Language Resources Association)

- ACL (Association for Computational Linguistics)
- SIGLEX (ACL Special Interest Group on Lexicons)
- SIGSEM (ACL Special Interest Group on Computational Semantics)
- AAAI (American Association for Artificial Intelligence)
- Canadian Language Industry Association (CLIA).

Selected Conference and Panel Presentations

- DTW at Qur'an QA 2022: Utilizing Transfer Learning with Transformers for Question Answering in a Low-resource Domain. In Proceedings of the 5th Workshop on Open-Source Arabic Corpora and

Processing Tools (OSACT5) at the 13th Language Resources and Evaluation Conference (LREC 2022), Marseille (France) June 20-25, 2022.

- Findings of the NLP4IF-2021 shared tasks on fighting the COVID-19 infodemic and censorship detection, 6 June 2021 (Mexico)
- BUILDING A QATARI HERITAGE EXPRESSIONS CORPUS. The 4th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Task on Offensive Language Detection. Language Resources and Evaluation Conference. Online, on May 11-16, 2020
- DAICT: A Dialectal Arabic Irony Corpus Extracted From Twitter. Proceedings of the 12th Edition of its Language Resources and Evaluation Conference. Language Resources and Evaluation Conference. Online, on May 11-16, 2020
- On the Author Profiling and Deception Detection in Arabic shared task at FIRE. Proceedings of the 11th Forum for Information Retrieval Evaluation. Association for Computing Machinery (ACM). Kolkata, India (12 Dec 2019)
- A Fine-Grained Annotated Multi-Dialectal Arabic Corpus. Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019). (Sophia, Bulgaria on 2019, September)
- Detecting Deceptive Tweets in Arabic for Cyber-Security. Proceedings of the 17th IEEE International Conference on Intelligence and Security Informatics (ISI). Nov. 9-10, 2020 held as a virtual event
- Building Dialectal Arabic Tools and Resources. Computational Study of Culture: Digital Humanities Workshop. Doha Institute, 19, October 2019.
- Integration of MT technology into AV web- based environment. 10th Annual International Translation. Doha, Qatar, (2019, March)
- Can The Tweets of Vulnerable Groups Provide Reliable Windows into Their Vulnerability? A Corpus Analysis. International Conference on Web and Social Media. Munich, Germany. 19-20 October 2019.
- The Detection of Radicalization in Social Media. Rethinking prevention through digital technology. Sfax, Tunisia: The Digital Research Center of Sfax, Tunisia. 25 December 2018.

- Fake News and Irony Detection. Rethinking prevention through digital technology. The Digital Research Center of Sfax, Tunisia: The Digital Research Center of Sfax, Tunisia. 26 December 2018
- Madari: A web interface for joint Arabic morphological annotation and spelling correction. LREC 2018 - 11th International Conference on Language Resources and Evaluation. Japan, Miyazaki. 7-12 May 2018, Miyazaki (Japan)
- AraP-Tweet: A large multi-dialect twitter corpus for gender, age and language variety identification. LREC 2018 - 11th International Conference on Language Resources and Evaluation. 7-12 May 2018, Miyazaki (Japan)
- La construction d'une ressource lexicale pour la langue arabe. May 9th 2013. ACFAS Colloquium, Laval University, Quebec.
- Developing ARET: An NLP-based Educational Tool Set for Arabic Reading Enhancement. Oral presentation at the 7th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT 2012, Montreal, Canada.
- A Pilot PropBank Annotation for Quranic Arabic. Oral presentation at the first workshop on Computational Linguistics for Literature, NAACL-HLT 2012, Montreal, Canada.
- A resource-light Arabic Named Entity Recognition system . Georgetown University Round Table 2010. Arabic Language and Linguistics, March 12- 14 2010.
- **Detection of person names and their translations in multilingual news**, Colloque Traitement Lexicographique des noms propres, Tours, 24 March 2005.
- Annotation Guidelines For Non-native Arabic Text In The Qatar Arabic Language Bank. In Qatar Foundation Annual Research Conference (ARC 20Doha, Qatar. 18-19 Nov 2014.
- The QALB Project. In Qatar Foundation Annual Research Conference (ARC 2013), Doha, Qatar. 24- 25 Nov 2013.
- The QALB Project: Building Resources and Systems for the automatic Correction of Arabic Text. Best Poster at the 2013 Meetings of the Mind, Carnegie Mellon University, Doha, Qatar.

Past Research Projects

- Multi-Arabic Dialect Applications and Resources (MADAR): the main objective of this project is to develop a suite of four novel multi-dialectal resources which will be used to conduct original research in two applications that are valuable enabling technologies necessary to support future research in Arabic NLP.
- The Optimal Diacritization for readability and comprehension project (OPTDIAC): An Optimal Diacritization Scheme for Arabic Orthographic Representation. The
- overarching objective is to improve Arabic NLP in general as well as improve readability and comprehension rates for Arabic text

- Qatar Arabic Language Bank (QALB): This project aims to build a large corpus of manually corrected Arabic text for building automatic correction tools for Arabic text.
- WikiTea Open Data Project: The goal of the project is to create techniques, tools and resources for enhancement and expansion of Arabic Wikipedia through Statistical Machine Translation.
- AL-Kitaab Digital Book Project: The goal of this project is to create a Digital version of the Al- Kitaab book in order to develop a Reading comprehension tool for Arabic learners.
- Arabic Symbol Dictionary For AAC Users: The goal of the project is the development of an Arabic Symbol Dictionary for Augmentative and Alternative Communication (AAC) users, as well as those who may benefit from the use of symbols to enhance literacy skills.
- Fairuz Arabic Historical Dictionary: The goal of the project is to create unique historical dictionary of the Arabic language. The project is a collaboration between the University of Pennsylvania and Bibliotheca Alexandria.
- The Georgetown Digital Dictionary of Iraqi: I was involved with Mohammed Maamouri at the Linguistic Data Consortium in the digitization and the development of the Iraqi lexicon for the project with Georgetown University Press.
- The Georgetown Digital Dictionary of Moroccan Arabic: I was involved with Mohammed Maamouri at the Linguistic Data Consortium in the digitization and the development of the Iraqi lexicon for the project with Georgetown University Press.
- Early Warning Intelligence Solutions (EWIS) : The project aims to build advanced analytics solutions for warnings about outbreaks of diseases and viral infections.
- The Penn Arabic PropBank Digital Corpus Project: a corpus that is annotated with verbal propositions and their arguments.
- The Penn Arabic TreeBank Digital Corpus Project: a parsed text corpus that annotates the syntactic structure according the LDC guidelines.
- The Nstein E-Publishing Tool: The Arabic NConcept Extractor: I helped in building the Arabic Concept Extractor with Nstein Technologies, a provider of computer- aided indexing(CAI) solutions to thee-publishing industry.
- The Europe Media Monitor (EMM) Digital News Explorer Project: I helped in building the Arabic News Explorer system using the JRC developed technology to automatically generate daily news summaries, allowing users to see the major news stories in various languages for any specific day and to compare how the same events have been reported in the media written in different languages.
- The Arabic News Medical Media Alert System: I was heavily involved in building the Arabic Medical Terminology used in the MEDISYS system which is a media monitoring system providing event-based surveillance to rapidly identify potential public health threats using information from media reports.
- The PropBank Annotation of Quran: In this pilot project, we leverage our knowledge and experience from our involvement in the APB project. All the Quranic Arabic PropBank annotations will be made freely available for research purposes.

- The Pilot Crowdsourcing experiment for the Open Quran Project:
In this pilot project, I teamed with Kais Duker in order to do a Crowdsourcing experiment to annotate the Quranic Arabic Corpus.

Edited Proceedings

- Guandong Xu, Raian Ali, Wajdi Zaghrouani, Xiaohui Tao, Lin Li. 2021. Proceedings of 2021 8th IEEE International Conference on Behavioural and Social Computing (BESC). DOI: 10.1109/BESC53957.2021.9635463, Publisher: IEEE, Publication date: 17 December 2021. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9635573>

The WANLP, is a premier and annual international event endorsed by the SIGARAB, the Special Interest Group of the Association for Computational Linguistics for researchers concerned with all aspects of Arabic NLP. The H-index of the WANLP is: **H-29**. Each year, the WANLP is co-located with one of the main ACL conferences such as ACL, EACL and EMNLP.

<https://scholar.google.com/citations?user=LGzh8jYAAAAJ>

- Houda Bouamor, Hend Al Khalifa, Fethi Bougares, Kareem Darwish, Owen Rambow, Ahmed Abdelali, Nadi Tomeh, Salam Khalifa and **Wajdi Zaghrouani** (Editors). 2022. Proceedings of the Seventh Arabic Natural Language Processing Workshop. **Forthcoming / To be Published In December 2022** by the Association for Computational Linguistics.

Website: <https://sites.google.com/view/wanlp2022/organizers?authuser=0>

- Nizar Habash, Houda Bouamor, Hazem Hajj, Walid Magdy, **Wajdi Zaghrouani**, Fethi Bougares, Nadi Tomeh, Ibrahim Abu Farha, Samia Touileb (Editors). 2021. Proceedings of the Sixth Arabic Natural Language Processing Workshop. Published by the Association for Computational Linguistics. <https://aclanthology.org/2021.wanlp-1>.
- Zitouni, Imed and Abdul-Mageed, Muhammad and Bouamor, Houda and Bougares, Fethi and El-Haj, Mahmoud and Tomeh, Nadi and **Zaghrouani, Wajdi**. 2020. Proceedings of the Fifth Arabic Natural Language Processing Workshop. Published by the Association for Computational Linguistics. <https://aclanthology.org/2020.wanlp-1>
- Wassim El-Hajj, Lamia Hadrich Belguith, Fethi Bougares, Walid Magdy, Imed Zitouni, Nadi Tomeh, Mahmoud El-Haj, **Wajdi Zaghrouani** (Editors). 2019. Proceedings of the Fourth Arabic Natural Language Processing Workshop. Published by the Association for Computational Linguistics. <https://aclanthology.org/W19-46>
- Nizar Habash, Mona Diab, Kareem Darwish, Wassim El-Hajj, Hend Al-Khalifa, Houda Bouamor, Nadi Tomeh, Mahmoud El-Haj, **Wajdi Zaghrouani** (Editors). Proceedings of the Third Arabic Natural Language Processing Workshop. Published by the Association for Computational Linguistics. <https://aclanthology.org/volumes/W17-13/>

- **Wajdi Zaghouani** (Editor), 2013. Proceedings of the CEC-TAL 2013 : 1ier colloque pour les Étudiants Chercheurs en Traitement Automatique du Langage Naturel et ses application
URL: <http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=28049©ownerid=44435>

Languages

English, French, Arabic and Italian