

Nedjma Djouhra Ousidhoum

Email OusidhoumN@cardiff.ac.uk
Webpage https://nedjmaou.github.io
Twitter @nedjmaou
Bluesky nedjmaou-nlp
LinkedIn Nedjma Ousidhoum

About Me

I am a Lecturer (Assistant Professor) at Cardiff University and a Visiting Academic at the University of Cambridge. I work on NLP and CSS in multilingual settings and for low-resource language, in addition to toxic content detection and evaluation, automated fact-checking, and related tasks. I am interested in how access to multilingual, diverse, and structured data sources via machine learning models and language technologies can help people learn and address practical problems such as fact-checking and moderation.

Work Experience

June 2023-now Lecturer (Assistant Professor) at the School of Computer Science and Informatics, Cardiff University.

April 2021-June 2023 Postdoctoral Researcher at the University of Cambridge (**Advisor** Prof. Andreas Vlachos).

April 2014-August 2014 Postgraduate intern at the Hong Kong University of Science and Technology (HKUST).

December 2012-April 2014 Research Assistant at the University of Science and Technology Houari Boumedienne, Algiers, Algeria.

Education

September 2014-August 2021 Ph.D. at the Hong Kong University of Science and Technology, HKUST (*Thesis defended on July 5th 2021*).

- **Advisors** Dr. Yangqiu Song (*Supervisor*), Prof. Dit-Yan Yeung (*Co-supervisor*)
- My thesis was on multilingual toxic content detection and evaluation.
- I started working on hate speech detection and related topics in September 2018.
- From 2014 to 2018, I was working on statistical machine translation and syntactic parsing.
- I took classes in Machine Learning, Advanced Algorithms, HCI, Knowledge Discovery and other topics.

September 2010- June 2012 Master of Science in Software Engineering at the University of Science and Technology Houari Boumedienne, Algiers, Algeria.

September 2007- June 2010 Bachelor of Science in Computer Science at the University of Science and Technology Houari Boumedienne, Algiers, Algeria.

Teaching and Supervisions

2024-2025 Supervisor of 4 BSc students at Cardiff University.

Spring 2025 Instructor of Emerging Technologies (CM3202).

Autumn 2024 Instructor (Module Leader) of Machine Learning for NLP (CMT122, Master's in NLP class).

Autumn 2024 and ongoing Co-supervisor of Yuefeng Shi (PhD student, co-supervised by Jose Camacho Colados).

Spring 2024 Instructor (Module Leader) of Emerging Technologies (CM3202).

2023-2024 Supervisor of 7 MSc students and 4 BSc students at Cardiff University.
Autumn 2022 Supervisor of Object Oriented Programming at Fitzwilliam College, University of Cambridge.
Summer 2022 Supervisor of three projects at the Online Summer Programme at Pembroke College, University of Cambridge.
Spring 2021 Co-Supervisor of Zhangdie (Moy) Yuan's (MPhil student), University of Cambridge.
Spring 2018 Teaching Assistant of IT Entrepreneurship at HKUST.
Spring 2017 Teaching Assistant of Introduction to Natural Language Processing at HKUST.
Spring 2016 Teaching Assistant of Multimedia Computing at HKUST.
Autumn 2015 Teaching Assistant of the C Programming Bridging Course at HKUST.

Academic Services and Activities

- Student Volunteers Chair at EMNLP 2025.
- Senior Area Chair at ACL 2025 (Multilingualism and Cross-Lingual NLP Track).
- Area Chair at COLING 2025 (Ethics and Bias Track).
- Action Editor at ACL Rolling Review (and Area Chair at EMNLP 2024).
- Reviewer for Computational Linguistics and OSNEM (journals).
- Area Chair (Less-Resourced/Endangered/Less-studied Languages track) at LREC-COLING 2024.
- Area Chair (Ethics in NLP) at EACL 2024.
- Senior Area Chair (NLP Applications track) at EMNLP 2023 (**Outstanding Senior Area Chair Award**).
- Reviewer for ACL Rolling Review, ACL 2022, NAACL 2022, NEATClass workshop at ICWSM, EMNLP 2022, EMNLP 2021, ACL 2021, NAACL 2021.
- Diversity, Equity and Inclusion Chair at ACL 2023.
- Diversity, Equity and Inclusion committee member (financial accessibility chair) at NAACL 2022.
- Organiser of the North African in NLP affinity group social at EMNLP 2020, COLING 2020, EACL 2021, NAACL 2021, ACL 2021, NAACL 2022.
- Organizer of the North African in ML affinity group workshop at NeurIPS 2022.
- Secondary reviewer for COLING 2020, AAAI 2019.

Publications

Nedjma Ousidhoum, Meriem Beloucif, Saif M. Mohammad: Building Better: Avoiding Pitfalls in Developing Language Resources when Data is Scarce (*ArXiv Preprint, under review* <https://arxiv.org/abs/2410.12691>).

Naba Rizvi, Harper Strickland, Daniel Gitelman, Tristan Cooper, Alexis Morales-Flores, Michael Golden, Aekta Kallepalli, Akshat Alurkar, Haaset Owens, Saleha Ahmedi, Isha Khirwadkar, Imani Munyaka, **Nedjma Ousidhoum**: AUTALIC: A Dataset for Anti-AUTistic Ableist Language In Context (*ArXiv Preprint, under review* <https://arxiv.org/abs/2410.16520>).

Dimosthenis Antypas, Christian Arnold, Jose Camacho-Collados, **Nedjma Ousidhoum**, Carla Perez-Almendros: Words as Trigger Points in Social Media Discussions (*ArXiv Preprint, under review* <https://arxiv.org/abs/2405.10213>).

Genta Indra Winata, Frederikus Hudi, Patrick Amadeus Irawan, David Anugraha, Rifki Afina Putri, Yutong Wang, Adam Nohejl, Ubaidillah Ariq Prathama, **Nedjma Ousidhoum**, Afifa Amriani, Anar Rzayev, Anirban Das, Ashmari Pramodya, Aulia Adila, Bryan Wilie, Candy Olivia Mawalim, Ching Lam Cheng, Daud Abolade, Emmanuele Chersoni, Enrico Santus, Fariz Ikhwantri, Garry Kuwanto, Hanyang Zhao, Haryo Akbarianto Wibowo, Holy Lovenia, Jan Christian Blaise Cruz, Jan Wira Gotama Putra, Junho Myung, Lucky Susanto, Maria Angelica Riera Machin, Marina Zhukova, Michael Anugraha, Muhammad Farid Adilazuarda, Natasha Santosa, Peerat Limkonchotiwat, Raj Dabre, Rio Alexander Audino, Samuel Cahyawijaya, Shi-Xiong Zhang,

Stephanie Yulia Salim, Yi Zhou, Yinxuan Gui, David Ifeoluwa Adelani, En-Shiun Annie Lee, Shogo Okada, Ayu Purwarianti, Alham Fikri Aji, Taro Watanabe, Derry Tanti Wijaya, Alice Oh, Chong-Wah Ngo: **WORLDCUISINES**: A Massive-Scale Benchmark for Multilingual and Multicultural Visual Question Answering on Global Cuisines (*Accepted at NAACL2025*).

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, David Ifeoluwa Adelani, Ibrahim Said Ahmad, Saminu Mohammad Aliyu, Nelson Odhiambo Onyango, Lilian DA Wanzare, Samuel Rutunda, Lukman Jibril Aliyu, Esubalew Alemneh, Oumaima Hourrane, Hagos Tesfahun Gebremichael, Elyas Abdi Ismail, Meriem Beloucif, Ebrahim Chekol Jibril, Andiswa Bukula, Rooweither Mabuya, Salomey Osei, Abigail Oppong, Tadesse Destaw Belay, Tadesse Kebede Guge, Tesfa Tegegne Asfaw, Chiamaka Ijeoma Chukwuneke, Paul RÄüttger, Seid Muhie Yimam, **Nedjma Ousidhoum**: AfriHate: A Multilingual Collection of Hate Speech and Abusive Language Datasets for African Languages (*Accepted at NAACL2025*).

Junho Myung, Nayeon Lee, Yi Zhou, Jiho Jin, Rifki Afina Putri, Dimosthenis Antypas, Hsuvas Borkakoty, Eunsu Kim, Carla Perez-Almendros, Abinew Ali Ayele, Victor Gutierrez-Basulto, Yazmin Ibanez-Garcia, Hwaran Lee, Shamsuddeen Hassan Muhammad, Kiwoong Park, Anar Sabuhi Rzayev, Nina White, Seid Muhie Yimam, Mohammad Taher Pilehvar, **Nedjma Ousidhoum**, Jose Camacho-Collados, Alice Oh: **BLEnD**: A Benchmark for LLMs on Everyday Knowledge in Diverse Cultures and Languages (*NeurIPS 2024 Benchmarks and Datasets Track, Best non-archival paper award at the C3NLP Workshop, co-located with ACL2024*).

Nedjma Ousidhoum, Shamsuddeen Hassan Muhammad, Mohamed Abdalla, Idris Abdulmumin, Ibrahim Said Ahmad, Sanchit Ahuja, Alham Fikri Aji, Vladimir Araujo, Meriem Beloucif, Christine De Kock, Oumaima Hourrane, Manish Shrivastava, Tamar Solorio, Nirmal Surange, Krishnapriya Vishnubhotla, Seid Muhie Yimam, Saif M. Mohammad: SemEval-2024 Task 1: Semantic Textual Relatedness for African and Asian Languages (*SemEval 2024, co-located with NAACL2024*). (**Best Task Description Paper Honourable Mention**)

Nedjma Ousidhoum, Shamsuddeen Hassan Muhammad, Mohamed Abdalla, Idris Abdulmumin, Ibrahim Said Ahmad, Sanchit Ahuja, Alham Fikri Aji, Vladimir Araujo, Abinew Ali Ayele, Pavan Baswani, Meriem Beloucif, Chris Biemann, Sofia Bourhim, Christine De Kock, Genet Shanko Dekebo, Oumaima Hourrane, Gopic-hand Kanumolu, Lokesh Madasu, Samuel Rutunda, Manish Shrivastava, Tamar Solorio, Nirmal Surange, Hailegnaw Getaneh Tilaye, Krishnapriya Vishnubhotla, Genta Winata, Seid Muhie Yimam, Saif M. Mohammad: SemRel2024: A Collection of Semantic Textual Relatedness Datasets for 13 Languages (*Findings of ACL 2024*).

Michael Schlichtkrull, **Nedjma Ousidhoum**, Andreas Vlachos: The Intended Uses of Automated Fact-Checking Artefacts: Why, How and Who (*Findings of EMNLP 2023*).

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, **Nedjma Ousidhoum**, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'ïd Ahmad, Meriem Beloucif, Saif M. Mohammad, Sebastian Ruder, Oumaima Hourrane, Pavel Brazdil, Felermino DÄārio MÄārio AntÄānio Ali, Davis David, Salomey Osei, Bello Shehu Bello, Falalu Ibrahim, Tajuddeen Gwadabe, Samuel Rutunda, Tadesse Belay, Wendimu Baye Messelle, Hailu Beshada Balcha, Sisay Adugna Chala, Hagos Tesfahun Gebremichael, Bernard Opoku, Steven Arthur: Afrisenti: A Twitter Sentiment Analysis Benchmark for African Languages (*EMNLP 2023*).

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelani, Ibrahim Said Ahmad, **Nedjma Ousidhoum**, Abinew Ali Ayele, Saif Mohammad, Meriem Beloucif, Sebastian Ruder: SemEval-2023 task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval)(*SemEval 2023 Shared Task, co-located with ACL 2023*).

Nedjma Ousidhoum,Zhangdie Yuan, Andreas Vlachos: Varifocal Question Generation for Fact-checking. (*EMNLP 2022*).

Nedjma Ousidhoum, "On the Importance and Challenges of the Experimental Design of Multilingual Toxic Content Detection". *Ph.D. Thesis*. The Hong Kong University of Science and Technology.

Nedjma Ousidhoum, Xinran Zhao, Tianqing Fang, Yangqiu Song, Dit-Yan Yeung: Probing Toxic Content in Large Pre-Trained Language Models (*ACL-IJCNLP 2021*).

Nedjma Ousidhoum, Yangqiu Song, Dit-Yan Yeung: Comparative Evaluation of Label-Agnostic Selection Bias in Multilingual Hate Speech Datasets (*EMNLP 2020*).

Nedjma Ousidhoum, Zizheng Lin, Hongming Zhang, Yangqiu Song, Dit-Yan Yeung: Multilingual and Multi-Aspect Hate Speech Analysis (*EMNLP 2019*).

Nedjma Ousidhoum, Nacéra Bensaou: Towards The Refinement of the Arabic Soundex (*NLDB 2013*).

Nedjma Ousidhoum, Asma Bensalah, Nacéra Bensaou: A New Classical Arabic Soundex Algorithm (*CIT 2012*).

Talks

July 2024 NLP for Low-resource Languages flash talk (*Multilingualism in the Era of Artificial Intelligence Workshop*).

July 2024 NLP for Low-resource Languages Discussion (*CollaborativeNLP Workshop*).

March 2024 On Benchmarking and Building resources: The inevitable and the preventable pitfalls. *Queen Mary University of London (Seminar)*.

May 2023 What Is Needed Vs What is Built in NLP: Toxic Language Detection and Automated Fact-checking Models As Use Cases. *Uppsala NLP (Seminar)*.

May 2023 What Is Needed Vs What is Built in NLP: Toxic Language Detection and Automated Fact-checking Models As Use Cases. *CohereAI Community Talks*.

May 2023 Expectations vs. Reality: Doing Multilingual Toxic Content Detection in NLP. *Aston Institute of Forensic Linguistics (Seminar)*.

January 2023 What Is Needed Vs What is Built in NLP: Toxic Language Detection and Automated Fact-checking Models As Use Cases (*Cardiff NLP Seminar*).

December 2022 Being a Researcher in Arabic NLP. *Panel discussion at WiNLP Workshop (co-located with EMNLP 2022)*

May 2022 Arabic Toxic Content Detection in NLP. Arabic AI and Toxic (*Online Content Detection Panel discussion at IWABigDAI*).

June 2022 On Multilingual Toxic Content in NLP (*Keynote at LateRAISSE Workshop collocated with LREC 2022*).

May 2022 Title *Arabic Toxic Content Detection in NLP*. Panel discussion at IWABigDAI conference.

February 2022 Expectations vs. Reality: Lessons Learned from Working on Toxic Content Detection in NLP. Language Technology Group Seminar (Hamburg Universität).

January 2022 Expectations vs. Reality: Lessons Learned from Working on Toxic Content Detection in NLP. *Cambridge NLIP Seminar (University of Cambridge)*.

September 2021 Expectations vs. Reality: Lessons Learned from Working on Toxic Content Detection in NLP. *MilaNLP Group Seminar (Bocconi University, Italy)*.

August 2021 Challenges in Toxic Content Detection. *Language and Multimodal AI Lab (LAMA) Group Seminar (Imperial College London)*.

November 2020 Normalizing the Experimental Design of Multilingual Hate Speech Detection. *Group Seminar of the Digital Technologies Research Center (National Research Council, Canada)*.

References

Prof. Andreas Vlachos (Postdoc advisor).

Dr. Yangqiu Song (PhD supervisor).

(Additional references can be provided upon request.)