

EACL 2026

**The 19th Conference of the European Chapter of the
Association for Computational Linguistics**

**Proceedings of the 2nd Workshop on
NLP for Languages Using Arabic Script
(AbjadNLP 2026)**

Edited by

Mo El-Haj, Paul Rayson, Mustafa Jarrar,
Ignatius Ezeani, Saad Ezzini, Sina Ahmadi,
Amal Haddad Haddad, Cynthia Amol,
Ahmad Abdelali, Shadi Abudalfa

March 28, 2026
Rabat, Morocco

©2026 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-361-6

Preface

It is my great pleasure to welcome you to the second edition of the AbjadNLP Workshop, continuing our mission to advance research and applications for languages that use the Arabic script in Natural Language Processing (NLP). Building on the success of our inaugural workshop at COLING 2025 in Abu Dhabi, this year's event at EACL 2026 in Rabat serves as a vital platform to unite researchers and practitioners addressing the linguistic, cultural, and computational challenges inherent to these languages.

The interest in AbjadNLP this year has been fantastic. We received 93 submissions, accepting 21 papers for oral presentation, 19 posters, and 33 shared task papers. This represents a remarkable growth from our first edition and demonstrates the increasing recognition of Arabic-script languages within the NLP community. The submissions span a rich variety of languages, including Arabic, Urdu, Persian, Kurdish, and several Ajami languages, reflecting the linguistic and cultural diversity of the communities we serve.

A highlight of this year's workshop is the inclusion of four shared tasks, which have attracted significant participation and innovative approaches:

- **AbjadGenEval:** Abjad AI Generated Text Detection Shared Task for Languages Using Arabic Script
- **AbjadStyleTransfer:** Authorship style transfer (text generation), transforming a formal input text into the style of a specified author while preserving semantics
- **AbjadAuthorID:** Abjad Authorship Identification (multiclass classification), identifying the author of a given text excerpt across diverse genres and periods
- **AbjadMed:** Medical text classification in Arabic (multi-class classification), advancing Arabic NLP in the medical domain

These shared tasks have fostered collaborative research and provided valuable benchmarks for evaluating progress in understudied areas of Arabic-script language processing.

The contributions to this workshop continue to showcase remarkable diversity, addressing both fundamental NLP challenges and domain-specific applications. Research spans core NLP tasks such as morphology, tokenisation, and language modelling, as well as applications in machine translation, speech technologies, and social media analysis. We are particularly pleased to see growing attention to low-resource and marginalised languages, including various Ajami traditions across Africa, and the development of resources and tools that support long-term capacity building for these communities.

Many thanks to everyone who submitted, reviewed, and organised, and to the wider community for the strong engagement. Your contributions have made this workshop possible and continue to strengthen our collaborative community dedicated to Arabic-script languages in NLP.

Looking forward to inspiring presentations, engaging discussions, and seeing you in Rabat.

Mo El-Haj
General Chair
AbjadNLP 2026 Workshop

Organizing Committee

General Chair

- Mo El-Haj, VinUniversity, Vietnam; Lancaster University, UK

Programme Chairs

- Paul Rayson, Lancaster University, UK
- Mustafa Jarrar, HBKU / Birzeit University
- Ignatius Ezeani, Lancaster University, UK
- Saad Ezzini, KFUPM, Saudi Arabia

Publication Chair

- Sina Ahmadi, University of Zurich, Switzerland

Publicity Chairs

- Amal Haddad Haddad, University of Granada, Spain
- Cynthia Amol, Maseno University, Kenya

Shared-Task Chairs

- Ahmad Abdelali, Humain, Saudi Arabia
- Shadi Abudalfa, KFUPM, Saudi Arabia

Advisory Committee

- Nizar Habash, New York Abu Dhabi, UAE
- Miriam Butt, University of Konstanz, Germany
- Ruslan Mitkov, Lancaster University, UK

Table of Contents

<i>ArabicDialectHub: A Cross-Dialectal Arabic Learning Resource and Platform</i> Salem Lahlou	1
<i>Rethinking Polarity Detection: When BPE Fails Across Scripts</i> Manodyna K H and Luc De Nardi	6
<i>Optimizing What We Trust: Reliability-Guided QUBO Selection of Multi-Agent Weak Framing Signals for Arabic Sentiment Prediction</i> Rabab Alkhalifa	15
<i>Optimizer Choice and Calibration for QARiB on Arabic-Script Social Media Offensive Language Detection</i> Auda Elshokry and Mohammed Alhanjouri	26
<i>Tarab: A Multi-Dialect Corpus of Arabic Lyrics and Poetry</i> Mo El-Haj	37
<i>LLM-to-Speech: A Synthetic Data Pipeline for Training Dialectal Text-to-Speech Models</i> Ahmed Khamis and Hesham Ali Ahmed	47
<i>HCMUS_PrompterXPrompter at AbjadMed: When Classification Meets Retrieval: Taming the Long Tail in Arabic Medical Text Classification</i> Duy Minh Dao Sy, Trung Kiet Huynh, Nguyen Dinh Ha Duong, Nguyen Chi Tran, Phu Quy Nguyen Lam and Hoa Pham Phu	55
<i>KazakhOCR: A Synthetic Benchmark for Evaluating Multimodal Models in Low-Resource Kazakh Script OCR</i> Henry Gagnier, Sophie Gagnier and Ashwin Kirubakaran	60
<i>Seeing Words Differently: Visual Embeddings for Robust English-Arabic Machine Translation</i> Mahdi Alshaikh Saleh and Irfan Ahmad	66
<i>Character-Level Transformer for Tajik–Persian Transliteration with a Parallel Lexical Corpus</i> Arabov Mullosharaf Kurbonovich	75
<i>Arabic Dialect Translation with Small LLMs: Enhancing through Reasoning-Oriented Reinforcement Learning</i> Sohaila Abdulsattar and Keith Ross	84
<i>MedArabs at AbjadMed: Arabic Medical Text Classification via Data- and Algorithm-Level Fusion</i> Amrita Singh	100
<i>GATech at AbjadMed: Bidirectional Encoders vs. Causal Decoders: Insights from 82-Class Arabic Medical Classification</i> Ahmed Khamis	105
<i>Orthographic Robustness of Persian Named Entity Recognition Models</i> Henry Gagnier and Sophie Gagnier	110
<i>ArabicMedicalBERT-QA-82 at AbjadMed: Fighting Class Imbalance in Arabic Medical Text Classification</i> Gleb Shanshin	115

<i>KvochurHegel at AbjadMed: Combining LDAM Loss and Adversarial Training for Arabic Medical Question-Answer Classification</i>	
Minh-Hoang Le	120
<i>baellouf at AbjadMed: Efficient Fine-tuning with All-Linear LoRA for Arabic Medical QA Classification</i>	
Abdallah Khallouf	124
<i>Supachoke at AbjadMed: Enhancing Arabic Medical Text Classification Using Fine-Tuned AraBERT</i>	
Thanh Phu Nguyen, Tuan Thai Huy Nguyen Cu, Son Thai Pham and Tri Duy Ho Nguyen	127
<i>REIGNITE at AbjadMed: Imbalance-Aware Fine-Tuning of Pretrained Arabic Transformers for Arabic Medical Text Classification Task</i>	
Nahid Montasir Rifat and Foyez Ahmed Dewan	132
<i>Tashkees-AI at AbjadMed 2026: Flat vs. Hierarchical Classification for Fine-Grained Arabic Medical QA</i>	
Fatimah Mohamed Emad Eldin	137
<i>MetaSwarm at AbjadMed: Forensic Optimization and Class-Balanced Discovery for Medical Diglossia in Abjad Scripts</i>	
Rahul Jaisy	144
<i>QurSci-Onto: A Hierarchical Ontology and Dataset for Scientific Exegesis in the Quran</i>	
Ibad-ur-Rehman Rashid, Junaid Hussain and Sadam Al-Azani	149
<i>AjamiMorph: Zero-Annotation Morphological Discovery for Hausa Ajami via Multi-Method Consensus</i>	
Soumedhik Bharati, Shibam Mandal, PRITHWISH GHOSH, Swarup Kr Ghosh and Sayani Mondal	166
<i>Morphological Feature Extraction for Fine-Grained Sorani Kurdish Dialect Identification: A Hybrid Transformer-Linguistic Approach</i>	
Soumedhik Bharati, Shibam Mandal, Subham Majumdar, Swarup Kr Ghosh and Sayani Mondal	172
<i>Olga Snissarenko at AbjadMed: Arabic Clinical Text Classification with AraBERT: Results from the AbjadMed Shared Task</i>	
Olga Snissarenko	177
<i>From Classical to Contemporary: Evolutionary Analysis & Classification of Urdu Poetry</i>	
NOOR FATIMA, HASAN FARAZ KHAN and IRFAN AHMAD	182
<i>Alkhalil Corpus: An Open-Source Thematic and Lemmatized Corpus for Modern Standard Arabic</i>	
Samir BELAYACHI and Azzeddine Mazroui	192
<i>Enhancing Urdu Sentiment Classification through Instruction-Tuned LLMs and Cross-Lingual Transfer</i>	
HASAN FARAZ KHAN, NOOR FATIMA and IRFAN AHMAD	198
<i>Back-of-the-Book Index Automation for Arabic Documents</i>	
Nawal Haidar, Ahmad Kashmar and Fadi Zaraket	208
<i>Improving on State-of-the-Art Models for Sentiment Analysis on Saudi-English Code-Switching Text</i>	
Samaher Alghamdi, Paul Rayson and Reem Alotibi	218

<i>OMAN-SPEECH: A Multi-Layer Annotated Speech Corpus for Omani Arabic Dialects</i>	
Rayyan S. Al Khadhuri, Firas Al Mahrouqi, Salim Al Mandhari, Amir Azad Al-Kathiri, Omar Said Alshahri, Ghassab Mansoor Alsaqr, Badri Abdulhakim Mudhsh and Tarek Fatnassi	229
<i>Hala Technical Report Building Arabic-Centric Instruction & Translation Models at Scale</i>	
Hasan Abed Al Kader Hammoud, Mohamad Bilal Zbib and Bernard Ghanem	236
<i>Arabic Citation Parsing using Part of Speech and Named Entity Recognition</i>	
youssef KAROUT, Hadi Hammoud and Fadi Zaraket.	245
<i>DeformAR: A Visual Analytics Framework for Evaluation of Arabic Named Entity Recognition</i>	
Ahmed Mustafa Younes	253
<i>A Corpus-Based Investigation of Contemporary Arabic Dialects Using the SADA Corpus</i>	
Ghada Alfattni	276
<i>HACS-TL: Cross-Script Transfer Learning for Hausa Ajami Hate Speech Detection Using Transformer-Based Architecture</i>	
Abdulkadir Shehu Bichi, Muqaddar Ali, Prashant Sharma and Ismail Dauda Abubakar	287
<i>Code-Switching as a Safety Failure Mode in Large Language Models: An Empirical Study of Roman Urdu across English, Mixed, and Transliteration-Only Inputs</i>	
Waleed Jamil and Saima Rafi	295
<i>QAMAR: A New Fully Verified and Accurate Quranic Arabic Morphological Analysis Resource.</i>	
SARA FAQIHI, Karim Bouzoubaa, Rachida Tajmout and Driss Namly	301
<i>AraModernBERT: Transtokenized Initialization and Long-Context Encoder Modeling for Arabic</i>	
Omar elshehy, Omer Nacar, Abdelbasset Djamai, Muhammed Ragab, Khlood AL Jallad and Mona Abdelazim.	313
<i>Parameter-Efficient Adaptation of Self-Supervised Models for Arabic Speech Recognition</i>	
Wafa MOHAMMED ALSHEHRI, Wasfi G. AL-KHATIB and Mohammad Ismail Amro	322
<i>Current state of LLMs for Arabic dialectal machine translation</i>	
Josef Jon, Rawan Bondok and Ondřej Bojar	329
<i>A Hybrid Confidence-Aware Framework for Arabic Toxicity Detection in Social Media</i>	
Fawzia Zaal Alanazi, Asma Mohammed Alamri, Arwa Bin Saleh and Abdullah I. Alharbi	364
<i>Arabic-Adapted One-Step Speech-to-Diacritized ASR: Evaluation and Error Analysis</i>	
Osamah A. I. Abduljalil, Dalal Ali, Razan A. Bajaman and Abdullah I. Alharbi	371
<i>GATech at AbjadGenEval Shared Task: Multilingual Embeddings for Arabic Machine-Generated Text Classification</i>	
Ahmed Khamis	380
<i>AraLingBench: A Human-Annotated Benchmark for Evaluating Arabic Linguistic Capabilities of Large Language Models</i>	
Mohamad Bilal Zbib, Hasan Abed Al Kader Hammoud, Ammar Mohanna, Nadine Rizk, Fatima Karnib, Sina Moukaled and Bernard Ghanem	385

<i>REGLAT at AbjadMed: Handling Imbalanced Arabic Medical Text Classification via Hierarchical KNN-MLP Architecture</i>	
Ahmed Megahed Fetouh, Mohammed Rahmath, Omer Dawood, Mariam Labib, Nsrin Ashraf and Hamada Nayel	394
<i>Murabaa: A comprehensive Resource Platform for Arabic Morphology</i>	
Karim Bouzoubaa, driss namly, hamid jihad, rachida tajmout, jamal ezzouaine and hakima khamar	398
<i>Sujith Kanakkassery at AbjadMed: Imbalance-Aware Transformer Fine-tuning for Arabic Medical Text Classification</i>	
Sujith Kanakkassery	408
<i>A Knowledge Graph Based Diagnostic Framework for Analyzing Hallucinations in Arabic Machine Reading Comprehension</i>	
Najwa Abdullah AlGhamdi, Sadam Al-Azani, Kwabena Nuamah and Alan Bundy	413
<i>From Posts to Pressure: An Arabic Dataset about Stress and Mental-Health Monitoring</i>	
Wajdi Zaghouni, Eman Sedqy shlkamy and Mabrouka Bessghaier	422
<i>HCMUS_TheFangs at AbjadGenEval Shared Task: Weighted Layer Pooling with Attention Fusion for Arabic AI-Generated Text Detection</i>	
Duy Minh Dao Sy, Nguyen Chi Tran, Trung Kiet Huynh, NGUYEN LAM PHU QUY, Pham Phu Hoa and Nguyen Dinh Ha Duong	433
<i>HCMUS_The Fangs at AbjadStyleTransfer Shared Task: Learning to Query Style, Contrastive Representations for Zero-Shot Arabic Authorship Style Transfer</i>	
Duy Minh Dao Sy, Trung Kiet Huynh, Nguyen Chi Tran, NGUYEN LAM PHU QUY, Pham Phu Hoa and Nguyen Dinh Ha Duong	438
<i>U-RoCX: An xLSTM based Approach to AI-Generated Urdu Text Detection</i>	
Rabee Adel Al-Qasem	443
<i>HCMUS_PrisonDilemma at AbjadAuthorID Shared Task: Less is More with Base Models</i>	
Trung Kiet Huynh, Duy Minh Dao Sy, Nguyen Chi Tran, Pham Phu Hoa, NGUYEN LAM PHU QUY and Truong Bao Tran	448
<i>U-MIRAGE: Benchmarking Chain-of-Thought Reasoning for Urdu Medical QA</i>	
Ali Faheem, Faizad Ullah, Muhammad Hammad, Ahmed Hassan, Muhammad Sohaib Ayub and Asim Karim	453
<i>XLMR-Urdu at AbjadGenEval Shared Task: A Data-Centric Transformer-Based Approach for AI-Generated Urdu Text Detection</i>	
Mohannad Mohammad Hendi	461
<i>LoRAD: Low-Resource AI-Generated Text Detection with XLM-RoBERTa</i>	
Ali Zain	468
<i>A Stylometric and Statistical Pipeline for Urdu AI-Generated Text Classification</i>	
saeed A. Anabtawi	472
<i>QalamID at AbjadAuthorID Shared Task: Morphology Matters, A Hybrid Ensemble for Arabic Authorship Attribution</i>	
Youssef Zaghloul	476

<i>Kashif-AI at AbjadGenEval Shared Task: A Transformer-based Approach for Arabic AI-Generated Text Detection</i>	
Fatimah Mohamed Emad Eldin	483
<i>NileUn at AbjadGenEval Shared Task: Contrastive Learning with Stacking Ensemble for Efficient Arabic AI-Generated Text Detection</i>	
Mohamed Hussein Mohamed, Shrouk Shalaby and Nesreen Mohamed	489
<i>REGLAT at AbjadGenEval: Multi-Model Ensemble Approach for Arabic AI-Generated Text Detection</i>	
Mariam Labib Francies, Nsrin Ashraf, Ahmed Megahed Fetouh and Hamada Nayel	493
<i>AyahVerse at AbjadGenEval Shared Task: Monolingual Precision and Cross-Lingual Analysis in Perso-Arabic AI Detection</i>	
Fizza Nawaz, Ibad-ur-Rehman Rashid, Uswa Abid and Junaid Hussain	497
<i>AbjadMed: Arabic Medical Text Classification at AbjadNLP 2026</i>	
Pranav Gupta, Niranjana Kumar M, Balaji Nagarajan, Imed Zitouni and Mo El-Haj	506
<i>Uslub at AbjadAuthorID Shared Task: A Comparative Analysis of Traditional Machine Learning and Transformer-Based Models for Authorship Attribution in Arabic and Urdu</i>	
Shahad Alsuhaibani and Mohamed Alkaoud	515
<i>Arabic Author Attribution Using Transformer-Based Models: Insights from the AbjadAuthorID Shared Task</i>	
ghader kurdi	520
<i>R-R at AbjadAuthorID Shared Task: A Fine-Tuned Approach for Kurdish Authorship Identification</i>	
Rania Azad M. San Ahmed and Rebwar M. Nabi	525
<i>AbjadGenEval: Abjad AI Generated Text Detection Shared Task for Languages Using Arabic Script at AbjadNLP 2026</i>	
Saad Ezzini, Irfan Ahmad, Salmane Chafik, Shadi Abudalfa, Mo El-Haj, Ahmed Abdelali, Mustafa Jarrar, Nadir Durrani, Hassan Sajjad and Farah Adeeba	530
<i>AbjadAuthorID: Authorship Identification for Arabic-Script Languages at AbjadNLP 2026</i>	
Shadi Abudalfa, Saad Ezzini, Ahmed Abdelali, Mustafa Jarrar, Mo El-Haj, Nadir Durrani, Hassan Sajjad, Farah Adeeba and Sina Ahmadi	538
<i>AbjadStyleTransfer: Authorship Style Transfer for Arabic-Script Languages at AbjadNLP 2026</i>	
Shadi Abudalfa, Saad Ezzini, Ahmed Abdelali, Mustafa Jarrar, Mo El-Haj, Nadir Durrani, Hassan Sajjad and Farah Adeeba	545

Program

AbjadNLP 2026 Workshop

March 28, 2026

9:00–9:15	Welcome and Opening Remarks Overview of the shared tasks <i>Mo El-Haj, Ahmad Abdelali, Shadi Abudalfa</i>
9:15–10:00	Keynote Address <i>Violetta Cavalli-Sforza, Al Akhawayn University, Ifrane, Morocco</i>
10:00–10:30	Oral Presentations (Session 1) Paper presentations
10:30–11:00	Coffee Break
11:00–13:00	Oral Presentations (Session 2) Paper presentations
13:00–14:00	Lunch Break
14:00–15:30	Oral Presentations (Session 3) Paper presentations
15:30–16:30	Poster Session Including EACL coffee break (15:30–16:00)
16:30–17:30	Oral Presentations (Session 4) Paper presentations
17:30–17:45	Closing Remarks

