LLM Alignment for the Arabs: A Homogenous Culture or Diverse Ones?

Amr Keleg
Institute for Language, Cognition and Computation
School of Informatics, University of Edinburgh
a.keleg@sms.ed.ac.uk

Abstract

Large language models (LLMs) have the potential of being useful tools that can automate tasks and assist humans. However, these models are more fluent in English and more aligned with Western cultures, norms, and values. Arabic-specific LLMs are being developed to better capture the nuances of the Arabic language, as well as the views of the Arabs. Yet, Arabs are sometimes assumed to share the same culture. In this position paper, I discuss the limitations of this assumption and provide preliminary thoughts for how to build systems that can better represent the cultural diversity within the Arab world. The invalidity of the cultural homogeneity assumption might seem obvious, yet, it is widely-adopted in developing multilingual and Arabic-specific LLMs. I hope that this paper will encourage the NLP community to be considerate of the cultural diversity within various communities speaking the same language.

1 Introduction

Even in the global world we live in, people residing in different parts of the world nourish different ideas, have different interests, and face different challenges. These differences can be too extreme to the extent that people could be considered to be living in totally distinct worlds (Sapir, 1929, p. 209 as cited in Bird, 2024, p. 3). For instance, Kirk et al. (2024) found that US participants questioned Large Language Models (LLMs) about abortion more than non-US ones. People from different regions can also have different perceptions of the same topic, as exemplified by English speakers from the US, UK, Singapore, Kenya, and South Africa disagreeing on what counts as Hate Speech (Lee et al., 2024). All these differences could be attributed to the cultural diversity among various communities across the world.

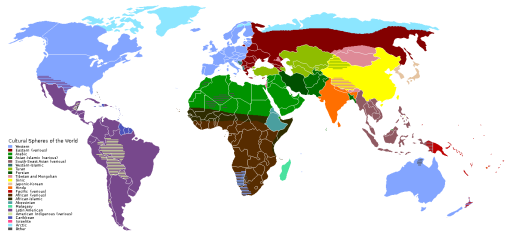
A major step in developing the current LLMs is aligning their responses to the users' needs. With the popularized one-model-fits-all paradigm, it is challenging to build models that can produce personalized responses that appeal to people of different demographics (Kirk et al., 2024). Current models tend to generate responses that better match the expectations of Western users (Cao et al., 2023; Naous et al., 2024; Wang et al., 2024; AlKhamissi et al., 2024; Ryan et al., 2024; Mihalcea et al., 2024). Moreover, the views of Arabs—one group of many underrepresented non-Western communities—tend to be ignored,[1] sometimes unconsciously and other times with deliberate intent, putting people of these communities at a higher risk of discrimination (Alimardani and Elswah, 2021; Shahid and Vashistha, 2023; Magdy et al., 2025).

Arabic is privileged by having (a) a community of Arab NLP experts (Habash and Vogel, 2014; Habash et al., 2015, 2017; El-Hajj et al., 2019; Zitouni et al., 2020; Habash et al., 2021; Bouamor et al., 2022; Sawaf et al., 2023; Habash et al., 2024; Al-Khalifa et al., 2020, 2022, 2024; El-Haj et al., 2019; Ezzini et al., 2025), and (b) interest backed by funding from some Arab countries to build Arabic-focused models that serve its speakers. Jais (Sengupta et al., 2023), AceGPT (Huang et al., 2024), Allam (Bari et al., 2024), and Fanar (Fanar Team et al., 2025) are Arabic-centric LLMs developed in 2023 and 2024. While earlier models like Jais focused on better modeling the linguistic features of Arabic, AceGPT, ALLaM, and Fanar are marketed as models that better align to the *Arabic/Arab Culture*.

It is well-known that local varieties of Dialectal Arabic (DA) exist in different Arabic-speaking regions, in addition to a standardized variety (MSA)

[1] Despite the attempt of curating model alignment data from different multi-cultural demographics, the PRISM Alignment dataset (Kirk et al., 2024) had only 51 participants (out of 1,500) who reported that they reside in the Middle East, out of which 47 reside in Israel, 2 in Turkey, and 1 in each of Sudan and Kuwait. Moreover, only 14 participants self-reported themselves as *Middle Eastern/ Arab*.

Proceedings of the 3rd Workshop on Cross-Cultural Considerations in NLP (C3NLP 2025), pages 1–9



(a) Culture Areas: Zones of high cultural overlap due to shared geography and long-term contact (Source: VividMaps)

Figure 1: A visualization of the areas with substantial cultural similarities as in (Bird, 2024, p. 3). Arabic speakers (green color) are grouped into a single region.

that is generally perceived as a shared variety across the Arabic-speaking communities. (Habash, 2010). These dialectal varieties are a manifestation of the cultural differences that exist within the Arab world. However, the notion of a single *Arabic Culture* only focuses on the shared values and norms among the Arabs, marginalizing any regional differences between them. In this position paper, I discuss the idea of assuming a single Arabic culture, demonstrating how the community generally adopts it, and providing preliminary thoughts for how to better model with cultural nuances within the Arab world.

2 Arabs - a Single or Multiple Cultures?

Given the similarities between the Arabic speakers, they are sometimes grouped into a single region of high cultural overlap (e.g., Figure 1). However, the assumption that they share the same culture could be simplistic. Conceptually, there are multiple ways to define who an Arab is. A broader definition is that *anyone having Arabic as their native language is an Arab*. Accordingly, there are more than 420 million Arabs distributed across the Arab region (Bergman and Diab, 2022), a large proportion of which reside in North Africa and the Arabian peninsula. I discuss two contrasting extreme views of the Arabic culture:[2]

View #1 - One Culture Arab nationalism is an ideology that started to gain traction in the 20th century, with the goal of unifying the Arab countries under a single goal, fostering economic cooperation between them. Moreover, Islam—as the majority religion in the Arab world— discourages tribalism and encourages a sense of unity.[3]

View #2 - Multiple Unrelated Cultures A contrasting ideology fosters the notion of local national identities, focusing on what makes these identities different from other Arab nations. Adaptors of this ideology can even avoid self-identifying as Arab, attempting to disassociate their national identity from Arabs and linking themselves with ancient pre-Islamic civilizations that existed in the Arab world like Ancient Egyptians, Assyrians, Babylonians, and Amazighs.

It is worth mentioning that the distinction between *Arabic Culture* and *Arab Culture* in English—the former linking the culture to the Arabic language, while the latter links it to Arabs—does not exist in Arabic, as both are termed الثقافة العربية. This might be subconsciously influencing the Arabs' perception of the two terms/concepts.

3 How is the Arabic Culture Currently Represented?

On surveying more than 90 papers related to cultural representation in LLMs, Adilazuarda et al. (2024) found that none of the papers explicitly mention how they operationalize the concept of a culture. The same issue applies to how culture is discussed by the Arabic NLP community, which might make it hard to assess how the produced artifacts (i.e., models and datasets) are culturally representative.[4] Hence, I taxonomize the datasets into three different categories according to their intended use as follows:

Classical Task-specific Datasets The community widely acknowledges the presence of different varieties of DA, with many datasets having samples from multiple dialects to model this linguistic variation (Mubarak et al., 2017; Alsarsour

[2]These contrasting views are manifested in the Wikipedia articles for عرب (Arab), in which the MSA and the Moroccan Arabic versions are more representative of the first view, while the Egyptian Arabic version link the Arabs to the Gulf and Levantine countries while excluding Egypt and the other North African countries. **Note:** The links to عرب (Arab) article in MSA, Moroccan and Egyptian Arabic respectively: ar.wikipedia.org/wiki/%D8%B9%D8%B1%D8%A8, ary.wikipedia.org/wiki/%D9%84%D8%B9%D8%B1%D8%A8, arz.wikipedia.org/wiki/%D8%B9%D8%B1%D8%A8

[3]Christianity is another religion that is adopted by a significant minority of Arabs (e.g., in Lebanon and Egypt). Moreover, Arab Jews used to be a vital part of Arab societies until the 20th century (Atta, 2023), and are still a minority in some countries like Morocco.

[4]Notably, AlKhamissi et al. (2024) provide a comprehensive discussion of what a culture is.

	English Translation	Arabic
Instruction	Suggest men's clothing for a family gathering	أقترح ملابس رجالية تناسب اجتماع عائلي
Choice (A)	Casual pants and a T-shirt	بنطلون كاجوال وتيشيرت
Choice (B)	Shorts and a polo shirt	شورت وتيشيرت بولو
Choice (C)	Formal shirt and pants	قميص وبنطلون رسمي
Choice (D)	Jellabiya and ghutra	جلابية وغترة
Answer	Choice (D)	
Instruction	I ate Kabsa using	اكلت الكبسة باستخدام
Choice (A)	a fork	الشوكة
Choice (B)	a spoon	الملعقة
Choice (C)	my hand	يدي
Choice (D)	a knife	السكين
Answer	Choice (C)	

Table 1: Two cherry-picked examples of edited instructions with multiple choices from *CIDAR-MCQ-100*. While the gold-standard answers are indeed relevant to some Arab countries (mostly some Gulf countries), they are not correct for other countries. **Note:** I provide the English translations for clarity.

et al., 2018; Ousidhoum et al., 2019; Chowdhury et al., 2020; Abu Farha and Magdy, 2020; Alturayeif et al., 2022). Given that dialects are signs of cultural diversity (Falck et al., 2012 as cited in Singh et al., 2024), this implies that such diversity might be modeled in the datasets. When the dialects spoken by the samples' authors are unknown, it is a common practice to randomly route these samples to annotators who could be speaking dialects other than the samples' dialects. *This assumes that Arabic is a monolith language, and disregards the cultural differences between its speakers.*[5]

Two independent papers found that Arabic-speaking annotators are harsher in labeling hate speech (Bergman and Diab, 2022), and less capable of identifying sarcasm (Abu Farha and Magdy, 2022), on annotating samples written in dialects that the annotators do not speak. On analyzing 15 publicly available datasets covering 5 different tasks, and having samples from multiple dialects that were randomly routed to annotators, Keleg et al. (2024) found that the interannotator agreement scores decreased as the level of dialectness of the samples increased. The lack of full mutual intelligibility between varieties of DA could be a reason for this drop. *However, cultural nuances form another plausible cause.* Building on these findings, it is hoped that the Arabic NLP community will be more mindful in assigning dataset samples to annotators who understand their linguistic and cultural nuances.

Less Subjective Culture Understanding Benchmarks Country-level sample curation was used to allow for capturing the cultural diversity in the Arab world. Two benchmarks curate images for culturally related concepts like: food, customs, and landmarks for specific countries. *CVQA* (Romero et al., 2024) has about 300 images related to Egypt, that were manually curated and are accompanied by QA pairs. Conversely, *Henna* (Alwajih et al., 2024) has 10 images from each of 11 Arab countries, accompanied by automatically generated image captions.

Similarly, *ArabicMMLU* (Koto et al., 2024) consists of multiple-choice questions (MCQs) in MSA covering different subjects, that were sourced from the school exams of 8 different Arab countries.[6] *AraDICE-Culture* (Mousi et al., 2025) has 180 MCQs from 6 different Arab countries (30 each) that were manually curated. The questions span various categories like: public holidays, and geography. *DLAMA (Arab-West)* (Keleg and Magdy, 2023) has Wikidata factual triplets from 20 predicates, equally balanced between Arab countries

[5] On analyzing the errors of a hate-speech detection model, Keleg et al. (2020) found two Egyptian Arabic quotes from films that were used sarcastically, yet labeled as hate speech. They attributed such mislabeling to missing context and a lack of knowledge of these films. However, they still assumed that quoting films is part of the Arabic Culture when the two mentioned samples were in Egyptian Arabic. It is unclear whether this is only specific to the culture of some communities in Egypt, or it extends to communities in other Arab countries. Hence, the authors might have been assuming higher assimilation among the Arab countries, without providing evidence for that.

[6] *ArabicMMLU*'s authors acknowledge the data is not equally representative of the different countries.

and a comparable set of Western countries. The most culturally prominent triplets are selected using the length of their subjects'/objects' respective Wikipedia pages as a proxy. *Cultural ArabicMTEB* (Bhatia et al., 2024) contains 1,000 queries automatically synthesized using Command-R+ from Wikipedia articles related to multiple categories such as: history, local movies, and food items for 20 different Arab countries. Lastly, *BLEnD* (Myung et al., 2024) has 1,000 MCQs about everyday knowledge of the cultures existing in Algeria.

Values Alignment Datasets Surprisingly, all Arabic-specific LLMs but ALLaM and Fanar perform alignment only using Supervised Fine-Tuning (SFT), with datasets that are either machine-translated or repurposed from task-specific datasets.

CIDAR (Alyafeai et al., 2024) is the first open Arabic instruction-tuning dataset composed of manually localized instruction/output pairs, edited by speakers of different varieties of Arabic.[7] While the authors focused on localizing person names and country names in the dataset to Arabic ones, which makes most of them culturally representative, few outputs are still biased by the annotator's views/country of origin, as exemplified in Table 1.

This issue is much more prominent in the *Arabic Cultural Value Alignment (ACVA)* benchmark (Huang et al., 2024), which is introduced to evaluate the alignment of different LLMs to *the Arabic Culture* (Huang et al., 2024; Bari et al., 2024). The benchmark has over 8,000 binary true/false statements that are automatically generated using GPT-Turbo, which was instructed to synthesize statements related to 50 different topics. Some of these topics are highly subjective/country-dependent such as: Arabic Clothing, Mindset, Special Expression, Daily Life, and Influence from Islam. Figure 2 lists two examples of non-inclusive statements, *which are a result of assuming a single homogenous Arabic culture.*

An example statement for *Communications*:
في الثقافة العربية، يمكن للرجال والنساء الجلوس معًا في الكافيهات.
In the Arabic culture, men and women can sit together in cafes.

Verdict: False (خطا).

An example statement for *Influnce From Islam*:
العرب يعتبرون الإسلام جزءًا لا يتجزأ من هويتهم الثقافية.
Arabs consider Islam an integral part of their cultural identity.

Verdict: True (صح).

Figure 2: Two statements from the ACVA benchmark showcasing misrepresentation of the cultural nuances within the Arab world. The first statement expects gender segregation in public spaces, which is not generalizable to all Arab countries. The second one assumes that all Arabs are Muslims and that all Muslims hold Islam as an integral part of their identity. Adding a quantifier like "معظم العرب (Most Arabs)" would make the statement more precise and less controversial.

4 Recommendations

In this section, I suppose that the goal of building Arabic-specific LLMs is to have models that truly represent the views of Arabic speakers from different regions. Following the discussion and the examples in §3, it is clear that assuming a single Arabic culture is not inclusive of the cultural diversity within the Arab world. Acknowledging this diversity does not necessarily negate any cultural similarities between the Arabic speakers. In contrast, it provides a more inclusive view of them.

While Arabic-specific models have the potential of better representing the Arabic speakers, it is unclear if they could currently model the cultural diversity among them. Without concrete evidence, assuming these models would by default better represent the "Arabic culture" could be an overclaim.

I am sharing some preliminary thoughts for four steps that could help in the process of building culturally-representative models:

Step #1 - Improving the Diversity of the Research Teams A first step is to ensure that the research teams responsible for building the models are representative of the different regions of the Arab World. Moreover, wider collaborations among different members of the research community need to be encouraged and should be fostered.

Step #2 - Understanding the Topics of Interest of the Speakers across the Arab World Many AI systems are developed without a clear vision of what they solve and how they would serve the needs of their users (Mihalcea et al., 2024). Given that people from different regions engage differently with LLMs (Kirk et al., 2024), we should start identifying the topics of interest of Arabic speakers from different regions, especially that their views were excluded in building the PRISM dataset (Kirk

[7] *CIDAR*'s creators acknowledge that the responses could be biased by the views of the different dataset contributors.

et al., 2024) on which the aforementioned finding is based. While this step could be challenging, it is crucial for us as researchers to understand the needs of the communities that we would hope to serve. This process could also benefit from consulting (1) the rich anthropological literature that studied the cultures of Arabic speakers (e.g., Deeb and Winegar, 2012), and (2) the recommendations from the Human-Computer Interaction (HCI) field for designing surveys and tools to understand the Arabic speakers' needs.

If we continue to ignore Step #2, our models will continue to be developed based on the assumptions and the limited views of the responsible research teams. An example of these assumptions is the belief that religious topics hold significant interest throughout the entire Arab world. Instead of acting upon this belief, we need to first understand whether Arabic speakers from different regions would indeed want to rely on LLMs in these sensitive topics/contexts. Doing so would allow for identifying the contexts in which the LLMs should engage in religious topics, if any, which in turn could help in controlling the dangers of shipping public-facing models that engage in religious discussions (Keleg and Magdy, 2022; Alyafeai et al., 2024).

Step #3 - Identifying the Languages/Varieties that Arabic Speakers Use on Engaging with Technologies On adapting the ArabicMMLU dataset to Moroccan Arabic (also known as Moroccan Darija), Shang et al. (2024) discarded the samples that they deemed as "too technical" and "beyond the user's needs" for an LLM that generates responses in Darija. This again indicates that researchers have some preconceived assumptions on the users' needs and the language varieties they would generally use to engage with the different technological systems.

In order to determine the language or variety that Arabic speakers would use when interacting with technology, we can first draw insights from the lessons of Blaschke et al.'s (2024) study, which analyzed the German users' preferences for having their local varieties supported as inputs or outputs of different language technologies such as virtual assistants and machine-translation systems.

However, the new study needs to also acknowledge that a non-negligible portion of the Arabic speakers in some regions are bilingual. Hence, English and French can be more preferred in different regions over using Standard Arabic or the regional local variety of Arabic to interact with technology in specific contexts. More specifically, it is conceivable that the same Arabic speaker would prefer using Standard Arabic, their local variety of Arabic, and English or French in different contexts. Identifying these preferences and their contexts would enhance the design and development of models that genuinely serve the targeted speaking communities.

Step #4 - Collecting More Inclusive Alignment Data There is a clear need for collecting alignment and preference data to improve the Arabic-specific LLMs. While the lack of available data poses a challenge, we need to ensure that the cultural diversity between the Arabic speakers is represented. Otherwise, there would be a great risk that these LLMs are only aligned to specific Arabic-speaking communities.

5 Conclusion

Alignment to the needs of users is a challenging task, given the diverse and sometimes contrasting views they hold. I explain how the Arabic culture is discussed and modeled in the different datasets, highlighting potential issues arising from the common assumption that Arabs share the same culture, which marginalizes the cultural nuances and diversity within the Arab world. Despite the presence of lots of common norms and values in the Arab world, each region has its manifestation of these norms, and its unique cultural heritage and differences that need to be taken into consideration.

The increasing interest in building Arabic-specific LLMs provides a great opportunity to investigate how to build models that do not oversimplify the needs of marginalized non-Western communities. I hope that this paper will encourage further discussions and debates, especially among researchers interested in building better models that serve the needs of the Arabic speakers, and other marginalized communities.

Limitations

I hope that a better understanding of the needs of the Arabs from diverse regions across the Arab World would allow for designing and building models that are more suited to their needs. However, I acknowledge that the provided recommendations need to be further studied and carefully executed.

Acnowledgments

I am grateful to Merham Keleg for attentively listening to the preliminary arguments that led to this paper. I also thank SMASH for their feedback on an earlier version of the paper. Special thanks to Walid Magdy, Björn Ross, Maria Walters, and Xue Li for their valuable comments and suggestions. Lastly, I really appreciate the anonymous reviewers' insightful feedback.

This work was supported by the UKRI Centre for Doctoral Training in Natural Language Processing, funded by the UKRI (grant EP/S022481/1) and the University of Edinburgh, School of Informatics.

References

Ibrahim Abu Farha and Walid Magdy. 2020. From Arabic sentiment analysis to sarcasm detection: The ArSarcasm dataset. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 32–39, Marseille, France. European Language Resource Association.

Ibrahim Abu Farha and Walid Magdy. 2022. The effect of Arabic dialect familiarity on data annotation. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 399–408, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Muhammad Farid Adilazuarda, Sagnik Mukherjee, Pradhyumna Lavania, Siddhant Shivdutt Singh, Alham Fikri Aji, Jacki O'Neill, Ashutosh Modi, and Monojit Choudhury. 2024. Towards measuring and modeling "culture" in LLMs: A survey. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15763–15784, Miami, Florida, USA. Association for Computational Linguistics.

Hend Al-Khalifa, Kareem Darwish, Hamdy Mubarak, Mona Ali, and Tamer Elsayed, editors. 2024. *Proceedings of the 6th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT) with Shared Tasks on Arabic LLMs Hallucination and Dialect to MSA Machine Translation @ LREC-COLING 2024*. ELRA and ICCL, Torino, Italia.

Hend Al-Khalifa, Tamer Elsayed, Hamdy Mubarak, Abdulmohsen Al-Thubaity, Walid Magdy, and Kareem Darwish, editors. 2022. *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection*. European Language Resources Association, Marseille, France.

Hend Al-Khalifa, Walid Magdy, Kareem Darwish, Tamer Elsayed, and Hamdy Mubarak, editors. 2020. *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*. European Language Resource Association, Marseille, France.

Mahsa Alimardani and Mona Elswah. 2021. Digital orientalism: #SaveSheikhJarrah and Arabic content moderation. *SSRN Scholarly Paper*.

Badr AlKhamissi, Muhammad ElNokrashy, Mai Alkhamissi, and Mona Diab. 2024. Investigating cultural alignment of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12404–12422, Bangkok, Thailand. Association for Computational Linguistics.

Israa Alsarsour, Esraa Mohamed, Reem Suwaileh, and Tamer Elsayed. 2018. DART: A large dataset of dialectal Arabic tweets. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Nora Saleh Alturayeif, Hamzah Abdullah Luqman, and Moataz Aly Kamaleldin Ahmed. 2022. Mawqif: A multi-label Arabic dataset for target-specific stance detection. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 174–184, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Fakhraddin Alwajih, El Moatez Billah Nagoudi, Gagan Bhatia, Abdelrahman Mohamed, and Muhammad Abdul-Mageed. 2024. Peacock: A family of Arabic multimodal large language models and benchmarks. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12753–12776, Bangkok, Thailand. Association for Computational Linguistics.

Zaid Alyafeai, Khalid Almubarak, Ahmed Ashraf, Deema Alnuhait, Saied Alshahrani, Gubran Abdulrahman, Gamil Ahmed, Qais Gawah, Zead Saleh, Mustafa Ghaleb, Yousef Ali, and Maged Al-shaibani. 2024. CIDAR: Culturally relevant instruction dataset for Arabic. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 12878–12901, Bangkok, Thailand. Association for Computational Linguistics.

Zubaydah Mohamed Atta. 2023. *اليهود في العالم العربي Jews in the Arab World*. Elain Publishing House.

M Saiful Bari, Yazeed Alnumay, Norah A. Alzahrani, Nouf M. Alotaibi, Hisham A. Alyahya, Sultan AlRashed, Faisal A. Mirza, Shaykhah Z. Alsubaie, Hassan A. Alahmed, Ghadah Alabduljabbar, Raghad Alkhathran, Yousef Almushayqih, Raneem Alnajim, Salman Alsubaihi, Maryam Al Mansour, Majed Alrubaian, Ali Alammari, Zaki Alawami, Abdulmohsen Al-Thubaity, Ahmed Abdelali, Jeril Kuriakose, Abdalghani Abujabal, Nora Al-Twairesh, Areeb Alowisheq, and Haidar Khan. 2024. ALLaM: Large language models for Arabic and English. *Preprint*, arXiv:2407.15390.

A. Bergman and Mona Diab. 2022. Towards responsible natural language annotation for the varieties of Arabic. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 364–371, Dublin, Ireland. Association for Computational Linguistics.

Gagan Bhatia, El Moatez Billah Nagoudi, Abdellah El Mekki, Fakhraddin Alwajih, and Muhammad Abdul-Mageed. 2024. Swan and ArabicMTEB: Dialect-aware, Arabic-centric, cross-lingual, and cross-cultural embedding models and benchmarks. *Preprint*, arXiv:2411.01192.

Steven Bird. 2024. Must NLP be extractive? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14915–14929, Bangkok, Thailand. Association for Computational Linguistics.

Verena Blaschke, Christoph Purschke, Hinrich Schuetze, and Barbara Plank. 2024. What do dialect speakers want? a survey of attitudes towards language technology for German dialects. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 823–841, Bangkok, Thailand. Association for Computational Linguistics.

Houda Bouamor, Hend Al-Khalifa, Kareem Darwish, Owen Rambow, Fethi Bougares, Ahmed Abdelali, Nadi Tomeh, Salam Khalifa, and Wajdi Zaghouani, editors. 2022. *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid).

Yong Cao, Li Zhou, Seolhwa Lee, Laura Cabello, Min Chen, and Daniel Hershcovich. 2023. Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 53–67, Dubrovnik, Croatia. Association for Computational Linguistics.

Shammur Absar Chowdhury, Hamdy Mubarak, Ahmed Abdelali, Soon-gyo Jung, Bernard J. Jansen, and Joni Salminen. 2020. A multi-platform Arabic news comment dataset for offensive language detection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6203–6212, Marseille, France. European Language Resources Association.

Lara Deeb and Jessica Winegar. 2012. Anthropologies of Arab-majority societies. *Annual Review of Anthropology*, 41(Volume 41, 2012):537–558.

Mahmoud El-Haj, Paul Rayson, Eric Atwell, and Lama Alsudias, editors. 2019. *Proceedings of the 3rd Workshop on Arabic Corpus Linguistics*. Association for Computational Linguistics, Cardiff, United Kingdom.

Wassim El-Hajj, Lamia Hadrich Belguith, Fethi Bougares, Walid Magdy, Imed Zitouni, Nadi Tomeh, Mahmoud El-Haj, and Wajdi Zaghouani, editors. 2019. *Proceedings of the Fourth Arabic Natural Language Processing Workshop*. Association for Computational Linguistics, Florence, Italy.

Saad Ezzini, Hamza Alami, Ismail Berrada, Abdessamad Benlahbib, Abdelkader El Mahdaouy, Salima Lamsiyah, Hatim Derrouz, Amal Haddad Haddad, Mustafa Jarrar, Mo El-Haj, Ruslan Mitkov, and Paul Rayson, editors. 2025. *Proceedings of the 4th Workshop on Arabic Corpus Linguistics (WACL-4)*. Association for Computational Linguistics, Abu Dhabi, UAE.

Oliver Falck, Stephan Heblich, Alfred Lameli, and Jens Südekum. 2012. Dialects, cultural identity, and economic exchange. *Journal of Urban Economics*, 72(2):225–239.

Fanar Team, Ummar Abbas, Mohammad Shahmeer Ahmad, Firoj Alam, Enes Altinisik, Ehsannedin Asgari, Yazan Boshmaf, Sabri Boughorbel, Sanjay Chawla, Shammur Chowdhury, Fahim Dalvi, Kareem Darwish, Nadir Durrani, Mohamed Elfeky, Ahmed Elmagarmid, Mohamed Eltabakh, Masoomali Fatehkia, Anastasios Fragkopoulos, Maram Hasanain, Majd Hawasly, Mus'ab Husaini, Soon-Gyo Jung, Ji Kim Lucas, Walid Magdy, Safa Messaoud, Abubakr Mohamed, Tasnim Mohiuddin, Basel Mousi, Hamdy Mubarak, Ahmad Musleh, Zan Naeem, Mourad Ouzzani, Dorde Popovic, Amin Sadeghi, Husrev Taha Sencar, Mohammed Shinoy, Omar Sinan, Yifan Zhang, Ahmed Ali, Yassine El Kheir, Xiaosong Ma, and Chaoyi Ruan. 2025. Fanar: An Arabic-centric multimodal generative AI platform. *Preprint*, arXiv:2501.13944.

Nizar Habash, Houda Bouamor, Ramy Eskander, Nadi Tomeh, Ibrahim Abu Farha, Ahmed Abdelali, Samia Touileb, Injy Hamed, Yaser Onaizan, Bashar Alhafni, Wissam Antoun, Salam Khalifa, Hatem Haddad, Imed Zitouni, Badr AlKhamissi, Rawan Almatham, and Khalil Mrini, editors. 2024. *Proceedings of The Second Arabic Natural Language Processing Conference*. Association for Computational Linguistics, Bangkok, Thailand.

Nizar Habash, Houda Bouamor, Hazem Hajj, Walid Magdy, Wajdi Zaghouani, Fethi Bougares, Nadi Tomeh, Ibrahim Abu Farha, and Samia Touileb, editors. 2021. *Proceedings of the Sixth Arabic Natural Language Processing Workshop*. Association for Computational Linguistics, Kyiv, Ukraine (Virtual).

Nizar Habash, Mona Diab, Kareem Darwish, Wassim El-Hajj, Hend Al-Khalifa, Houda Bouamor, Nadi Tomeh, Mahmoud El-Haj, and Wajdi Zaghouani, editors. 2017. *Proceedings of the Third Arabic Natural Language Processing Workshop*. Association for Computational Linguistics, Valencia, Spain.

Nizar Habash and Stephan Vogel, editors. 2014. *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*. Association for Computational Linguistics, Doha, Qatar.

Nizar Habash, Stephan Vogel, and Kareem Darwish, editors. 2015. *Proceedings of the Second Workshop on Arabic Natural Language Processing*. Association for Computational Linguistics, Beijing, China.

Nizar Y Habash. 2010. *Introduction to Arabic natural language processing*. Morgan & Claypool Publishers.

Huang Huang, Fei Yu, Jianqing Zhu, Xuening Sun, Hao Cheng, Song Dingjie, Zhihong Chen, Mosen Alharthi, Bang An, Juncai He, Ziche Liu, Junying Chen, Jianquan Li, Benyou Wang, Lian Zhang, Ruoyu Sun, Xiang Wan, Haizhou Li, and Jinchao Xu. 2024. AceGPT, localizing large language models in Arabic. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8139–8163, Mexico City, Mexico. Association for Computational Linguistics.

Amr Keleg, Samhaa R. El-Beltagy, and Mahmoud Khalil. 2020. ASU_OPTO at OSACT4 - offensive language detection for Arabic text. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 66–70, Marseille, France. European Language Resource Association.

Amr Keleg and Walid Magdy. 2022. SMASH at qur'an QA 2022: Creating better faithful data splits for low-resourced question answering scenarios. In *Proceedinsg of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection*, pages 136–145, Marseille, France. European Language Resources Association.

Amr Keleg and Walid Magdy. 2023. DLAMA: A framework for curating culturally diverse facts for probing the knowledge of pretrained language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6245–6266, Toronto, Canada. Association for Computational Linguistics.

Amr Keleg, Walid Magdy, and Sharon Goldwater. 2024. Estimating the level of dialectness predicts inter-annotator agreement in multi-dialect Arabic datasets. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 766–777, Bangkok, Thailand. Association for Computational Linguistics.

Hannah Rose Kirk, Alexander Whitefield, Paul Rottger, Andrew M. Bean, Katerina Margatina, Rafael Mosquera-Gomez, Juan Ciro, Max Bartolo, Adina Williams, He He, Bertie Vidgen, and Scott Hale. 2024. The PRISM alignment dataset: What participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models. In *Advances in Neural Information Processing Systems*, volume 37, pages 105236–105344. Curran Associates, Inc.

Fajri Koto, Haonan Li, Sara Shatnawi, Jad Doughman, Abdelrahman Sadallah, Aisha Alraeesi, Khalid Almubarak, Zaid Alyafeai, Neha Sengupta, Shady Shehata, Nizar Habash, Preslav Nakov, and Timothy Baldwin. 2024. ArabicMMLU: Assessing massive multitask language understanding in Arabic. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 5622–5640, Bangkok, Thailand. Association for Computational Linguistics.

Nayeon Lee, Chani Jung, Junho Myung, Jiho Jin, Jose Camacho-Collados, Juho Kim, and Alice Oh. 2024. Exploring cross-cultural differences in English hate speech annotations: From dataset construction to analysis. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4205–4224, Mexico City, Mexico. Association for Computational Linguistics.

Walid Magdy, Hamdy Mubarak, and Joni Salminen. 2025. Who should set the standards? Analysing censored Arabic content on Facebook during the Palestine-Israel conflict. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA. Association for Computing Machinery.

Rada Mihalcea, Oana Ignat, Longju Bai, Angana Borah, Luis Chiruzzo, Zhijing Jin, Claude Kwizera, Joan Nwatu, Soujanya Poria, and Thamar Solorio. 2024. Why AI is WEIRD and should not be this way: Towards AI for everyone, with everyone, by everyone. *Preprint*, arXiv:2410.16315.

Basel Mousi, Nadir Durrani, Fatema Ahmad, Md. Arid Hasan, Maram Hasanain, Tameem Kabbani, Fahim Dalvi, Shammur Absar Chowdhury, and Firoj Alam. 2025. AraDiCE: Benchmarks for dialectal and cultural capabilities in LLMs. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 4186–4218, Abu Dhabi, UAE. Association for Computational Linguistics.

Hamdy Mubarak, Kareem Darwish, and Walid Magdy. 2017. Abusive language detection on Arabic social media. In *Proceedings of the First Workshop on Abusive Language Online*, pages 52–56, Vancouver, BC, Canada. Association for Computational Linguistics.

Junho Myung, Nayeon Lee, Yi Zhou, Jiho Jin, Rifki Putri, Dimosthenis Antypas, Hsuvas Borkakoty, Eunsu Kim, Carla Perez-Almendros, Abinew Ali Ayele, Victor Gutierrez Basulto, Yazmin Ibanez-Garcia, Hwaran Lee, Shamsuddeen H Muhammad, Kiwoong Park, Anar Rzayev, Nina White, Seid Muhie Yimam, Mohammad Taher Pilehvar, Nedjma Ousidhoum, Jose Camacho-Collados, and Alice Oh. 2024. BLEnD: A benchmark for llms on everyday knowledge in diverse cultures and languages. In *Advances in Neural Information Processing Systems*, volume 37, pages 78104–78146. Curran Associates, Inc.

Tarek Naous, Michael J Ryan, Alan Ritter, and Wei Xu. 2024. Having beer after prayer? measuring cultural bias in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16366–16393, Bangkok, Thailand. Association for Computational Linguistics.

Nedjma Ousidhoum, Zizheng Lin, Hongming Zhang, Yangqiu Song, and Dit-Yan Yeung. 2019. Multilingual and multi-aspect hate speech analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4675–4684, Hong Kong, China. Association for Computational Linguistics.

David Romero, Chenyang Lyu, Haryo Wibowo, Santiago Góngora, Aishik Mandal, Sukannya Purkayastha, Jesus-German Ortiz-Barajas, Emilio Cueva, Jinheon Baek, Soyeong Jeong, Injy Hamed, Yong Zheng-Xin, Zheng Wei Lim, Paula Silva, Jocelyn Dunstan, Mélanie Jouitteau, David LE MEUR, Joan Nwatu, Ganzorig Batnasan, Munkh-Erdene Otgonbold, Munkhjargal Gochoo, Guido Ivetta, Luciana Benotti, Laura Alonso Alemany, Hernán Maina, Jiahui Geng, Tiago Timponi Torrent, Frederico Belcavello, Marcelo Viridiano, Jan Christian Blaise Cruz, Dan John Velasco, Oana Ignat, Zara Burzo, Chenxi Whitehouse, Artem Abzaliev, Teresa Clifford, Gráinne Caulfield, Teresa Lynn, Christian Salamea-Palacios, Vladimir Araujo, Yova Kementchedjhieva, Mihail Mihaylov, Israel Azime, Henok Ademtew, Bontu Balcha, Naome A. Etori, David Adelani, Rada Mihalcea, Atnafu Lambebo Tonja, Maria Cabrera, Gisela Vallejo, Holy Lovenia, Ruochen Zhang, Marcos Estecha-Garitagoitia, Mario Rodríguez-Cantelar, Toqeer Ehsan, Rendi Chevi, Muhammad Adilazuarda, Ryandito Diandaru, Samuel Cahyawijaya, Fajri Koto, Tatsuki Kuribayashi, Haiyue Song, Aditya Khandavally, Thanmay Jayakumar, Raj Dabre, Mohamed Imam, Kumaranage Nagasinghe, Alina Dragonetti, Luis Fernando D'Haro, Niyomugisha Olivier, Jay Gala, Pranjal Chitale, Fauzan Farooqui, Thamar Solorio, and Alham Aji. 2024. CVQA: Culturally-diverse multilingual visual question answering benchmark. In *Advances in Neural Information Processing Systems*, volume 37, pages 11479–11505. Curran Associates, Inc.

Michael J Ryan, William Held, and Diyi Yang. 2024. Unintended impacts of LLM alignment on global representation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16121–16140, Bangkok, Thailand. Association for Computational Linguistics.

E. Sapir. 1929. The status of linguistics as a science. *Language*, 5(4):207–214.

Hassan Sawaf, Samhaa El-Beltagy, Wajdi Zaghouani, Walid Magdy, Ahmed Abdelali, Nadi Tomeh, Ibrahim Abu Farha, Nizar Habash, Salam Khalifa, Amr Keleg, Hatem Haddad, Imed Zitouni, Khalil Mrini, and Rawan Almatham, editors. 2023. *Proceedings of ArabicNLP 2023*. Association for Computational Linguistics, Singapore (Hybrid).

Neha Sengupta, Sunil Kumar Sahu, Bokang Jia, Satheesh Katipomu, Haonan Li, Fajri Koto, William Marshall, Gurpreet Gosal, Cynthia Liu, Zhiming Chen, Osama Mohammed Afzal, Samta Kamboj, Onkar Pandit, Rahul Pal, Lalit Pradhan, Zain Muhammad Mujahid, Massa Baali, Xudong Han, Sondos Mahmoud Bsharat, Alham Fikri Aji, Zhiqiang Shen, Zhengzhong Liu, Natalia Vassilieva, Joel Hestness, Andy Hock, Andrew Feldman, Jonathan Lee, Andrew Jackson, Hector Xuguang Ren, Preslav Nakov, Timothy Baldwin, and Eric Xing. 2023. Jais and Jais-chat: Arabic-centric foundation and instruction-tuned open generative large language models. *Preprint*, arXiv:2308.16149.

Farhana Shahid and Aditya Vashistha. 2023. Decolonizing content moderation: Does uniform global community standard resemble utopian equality or western power hegemony? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery.

Guokan Shang, Hadi Abdine, Yousef Khoubrane, Amr Mohamed, Yassine Abbahaddou, Sofiane Ennadir, Imane Momayiz, Xuguang Ren, Eric Moulines, Preslav Nakov, Michalis Vazirgiannis, and Eric Xing. 2024. Atlas-Chat: Adapting large language models for low-resource Moroccan Arabic dialect. *Preprint*, arXiv:2409.17912.

Shivalika Singh, Angelika Romanou, Clémentine Fourrier, David I. Adelani, Jian Gang Ngui, Daniel Vila-Suero, Peerat Limkonchotiwat, Kelly Marchisio, Wei Qi Leong, Yosephine Susanto, Raymond Ng, Shayne Longpre, Wei-Yin Ko, Madeline Smith, Antoine Bosselut, Alice Oh, Andre F. T. Martins, Leshem Choshen, Daphne Ippolito, Enzo Ferrante, Marzieh Fadaee, Beyza Ermis, and Sara Hooker. 2024. Global MMLU: Understanding and addressing cultural and linguistic biases in multilingual evaluation. *Preprint*, arXiv:2412.03304.

Wenxuan Wang, Wenxiang Jiao, Jingyuan Huang, Ruyi Dai, Jen-tse Huang, Zhaopeng Tu, and Michael Lyu. 2024. Not all countries celebrate thanksgiving: On the cultural dominance in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6349–6384, Bangkok, Thailand. Association for Computational Linguistics.

Imed Zitouni, Muhammad Abdul-Mageed, Houda Bouamor, Fethi Bougares, Mahmoud El-Haj, Nadi Tomeh, and Wajdi Zaghouani, editors. 2020. *Proceedings of the Fifth Arabic Natural Language Processing Workshop*. Association for Computational Linguistics, Barcelona, Spain (Online).