IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

Cultural Nuances in LLMs: Bridging the Gap in AI Understanding – A Survey

Shubham Shinde[1], Aarya Bhave[2], Soham Late[3], Sachi Patankar[4], Prof. G. V. Kale[5]
Department of Computer Engineering[1-4]
Professor, Department of Computer Engineering[5]
Pune Institute of Computer Technology, Pune, India

Abstract: *Large Language Models (LLMs) have the potential to enable universal AI interactions, but this goal is not yet fully achieved. A key issue is that LLMs often reflect cultural biases from Western, English-language training data. When models focus on just one culture, they can make mistakes or reinforce stereotypes, which limits their usefulness around the world. This paper surveys the current state of cultural bias in LLMs and explores how to measure, reduce, and better understand it. We review major studies and highlight a shift from static, knowledge-based evaluations to more dynamic and interactive benchmarking methods. The paper also looks at effective solutions, such as building large open-access multilingual models (like Aya [3]), creating synthetic data with multi-agent systems (such as CulturePark [5] and Cultural Palette [8]), using Retrieval-Augmented Generation for dynamic knowledge integration (like ValuesRAG [7]), and advanced prompting techniques (such as Anthropological Prompting [1]). These techniques have helped improve the cultural sensitivity and fairness in the LLMs. This paper aims to review the existing literature and provide guidelines for future research efforts towards designing fair, efficient, and effective AI systems..*

Keywords: Large Language Models (LLMs), Cultural Bias, Cultural Alignment, Multilingual AI, Cross-Cultural Understanding, AI Ethics.

I. INTRODUCTION

Large Language Models (LLMs) are changing the way we communicate and do research by producing text that closely resembles human writing. However, as these models are used worldwide, a key challenge is making sure they reflect the many cultures they serve. Culture shapes how people think and interact, but most LLMs are trained on data from Western, English-speaking sources. This often leads to a Western bias, which can misrepresent or ignore other cultures, reinforce stereotypes, and leave out important perspectives. Fixing this is not just a technical issue—it is also a moral one, as fair AI should work for everyone. This survey reviews recent research on how to measure and improve the cultural understanding of LLMs. It covers both data-driven and model-based methods for cultural alignment, looks at how prompting can help models give more culturally aware answers, and discusses key theories about culture in AI. As LLMs become more common, making them culturally fair is crucial for ethical use around the world. This paper summarizes the latest research, points out what is working, what still needs improvement, and where future progress may come. It is meant as a resource for researchers, developers, and policymakers who want to build AI that is smart, culturally aware, and fair. We focus on recent studies, grouping their methods, datasets, and findings to show how the field is moving forward. Instead of a broad overview, this paper gives a focused look at the most up-to-date research.

II. LITERATURE REVIEW

In this review paper, we have attempted to summarize the most recent literature on an important topic in Large Language Models, which is Cultural Bias. While the literature does not just provide us with isolated studies, there is a noticeable progression in the way researchers address the problem. In particular, the approach has progressed from simply assessing bias to developing benchmarks that measure how good language models perform in actual cultural

DOI: 10.48175/IJARSCT-35590

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

scenarios. On the other hand, alignment strategies have progressed from basic prompting techniques to more complex and data and model-driven ones.

A. The Evolution of Evaluation: From Static Probes to Dynamic Benchmarks

There is quite a bit of literature specifically addressing measurement, which assumes that efficient evaluation is a necessary precursor for alignment. The first generation of measurement approaches was simply a modification of existing large-scale sociological surveys, designed to evaluate the culturally static values incorporated into the models. For example, Tao et al. [4] used the World Values Survey (WVS) and the European Values Study (EVS) to identify cultural biases in GPT models, demonstrating a strong connection to Western values. AlKhamissi et al. [1] adopted a similar strategy, using WVS data from Egypt and the United States to assess culture alignment.

However, researchers soon realized that knowing cultural facts is not the same as being able to handle real cross-cultural interactions. This led to the development of dynamic, behavior-based benchmarks. Wu et al. [2] created SocialCC, a benchmark comprising LM agents that must navigate a specific cultural situation and achieve specific social goals. Other researchers developed datasets focused on cultural dimensions. For example, Cao et al. [10] created cuDialog from movie subtitles and labeled dialogues using Hofstede's cultural vectors to benchmark how well models reflect cultural realism.

In a similar manner, Huang and Yang [12] developed the CALI dataset using Natural Language Inference (NLI) style scenarios based on cultural norms in the United States and India to demonstrate cultural differences in understanding. Lastly, an important direction is the creation of localized benchmarks, as evident from the conclusion that we must address the failure of Western-centric benchmarks. Tomar et al. [11] created BharatBBQ, a multilingual benchmark for the Indian socio-cultural context, and documented biases that were missing when using more general tools.

B. Strategies for Alignment: Data, Models, and Prompts

In response to the biases identified by these benchmarks, researchers have proposed interventions at every stage of the LLM pipeline.

1) Data-Centric Approaches

Multiple methods have been proposed that focus on engineering the training data. The Aya Model [3] is a large-scale initiative that aims to develop an openly available multilingual model that was curated and subsequently fine-tuned on a massive dataset of 101 languages. Other works focus on developing new cultural data. Li et al. [5] developed an economical course of action, CultureLLM, through the use of semantic augmentation to generate thousands of training examples from 50 high-quality "seed questions" from the WVS.

In a synthetic direction, Li et al. [6] published CulturePark, a multi-agent system where AI agents "role-played" as individuals from various cultures, generating rich dialogue that was diverse and realistic. In addition, foundational blueprints such as IndicLLMSuite [13] aimed to facilitate the generation of high-quality datasets for entire families of languages, like those spoken in India.

2) Model-Centric and Prompting Approaches

There is a third category of research that engages directly at the model level. Prompting is among the most adaptable tools in this space. In some cases, it is relatively simple, such as Cultural Prompting, which Tao et al. [4] employed to mitigate cultural bias. In other cases, prompting becomes significantly more sophisticated, as seen in Anthropological Prompting proposed by AlKhamissi et al. [1], where the model is guided to simulate a detailed cultural persona.

In addition to prompting, there are multiple methods that change the architecture of the model. The ValuesRAG method by Seo et al. [7] uses the concept of Retrieval-Augmented Generation and dynamically retrieves information about cultural values of individuals through the World Values Survey, making the model less dependent on stereotypes

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

memorization. The Cultural Palette method by Yuan et al. [8] treats major civilizations as primary colors, trains experts for each culture separately and combines them into one using the Mixture-of-Experts approach.

Other studies change the model foundation significantly. The HindiLLM method by Chouhan et al. [14] includes a Hindi-specific tokenizer. In theory, Grieve et al. [9] describe bias as an instance of varietal misalignment between model training data and sociolinguistic practices of the target community.

III. METHODOLOGICAL APPROACHES TO CULTURAL ALIGNMENT

Prompting

Prompting is a quick, flexible, and budget-friendly way to steer how pre-trained LLMs behave when you run them. Instead of retraining or tweaking the model's core, you just give it sharp, context-heavy instructions. That's all it takes to shift the model's answers so they fit a certain cultural vibe. The prompt itself does all the heavy lifting here. Sometimes you use a simple command; other times, you craft a layered, detailed set of instructions. Researchers have tried out a bunch of different techniques:

Cultural Prompting: Tao and colleagues [4] took a straightforward approach. They gave the model clear cultural instructions to see if this would help shake off its usual Western-centric bias. Basically, they wanted to know if just telling the model to think in a certain way could really change its perspective.

Anthropological Prompting: AlKhamissi et al. [1] took things a step further. Instead of just asking the model to pick up a viewpoint, they had it "think like an anthropologist." The idea is that the model doesn't just echo a culture—it tries to really get inside that worldview and reason from within it.

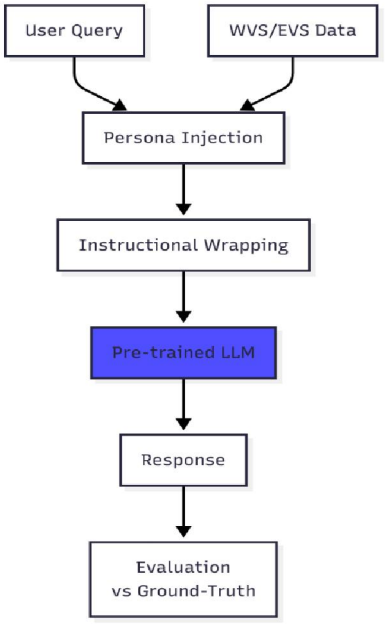

Fig.1 Overview of Prompting Architectures

Benchmarking

Benchmarks for measuring cultural competence have evolved substantially. In their early days, researchers employed simple knowledge tests such as asking, "What is the capital of Nigeria?" or "What is a major holiday in Japan?" Currently, benchmarks are more dynamic and focus on situational aspects. Indeed, in their paper on social cultural competence, researchers showed how LLMs' cultural competence varied greatly from their knowledge in the SocialCC benchmark. This indicates that, besides having adequate information about different cultures, models should also be able to navigate cultural situations effectively.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

Moreover, the creation of localized benchmarks such as BharatBBQ does not imply only broadening test coverage but is rather an inevitable consequence of the failure of Western-centric benchmarks in detecting culturally biased behaviour. The authors note that they developed BharatBBQ to "address the failure of Western-centric benchmarks." It should be noted that the authors later proved this claim empirically by revealing that biases were "stronger in Indian languages" when detected using this test.

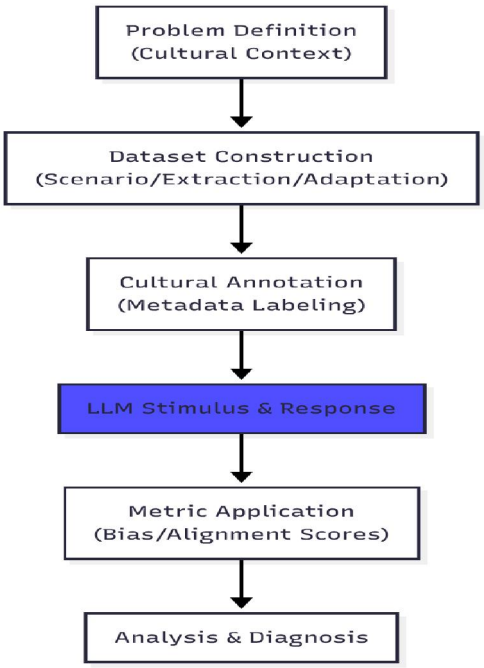

Fig.2 Evolution of Benchmarking Strategies

Specific datasets include:

Cao et al. [10] introduced cuDialog, which is a dataset created from movie subtitles that reflects 13 different cultures. This dataset notably has each conversation labelled with Hofstede's established cultural dimensions to create a "cultural vector" to evaluate the model's provision of culturally realistic dialogue.

Huang and Yang [12] constructed the Culturally Aware Natural Language Inference (CALI) dataset. This benchmark was developed with proposed scenarios based on different social norms for the United States and India.

Data-Centric

The various data-centric strategies, when analyzed together, reveal a fundamental "trilemma" that developers must navigate. This trilemma involves a trade-off between three desirable attributes of a dataset: authenticity (is the data genuine human expression?), scale (is the dataset large enough for modern LLMs?), and diversity (does the data capture a wide range of topics, perspectives, and interaction styles?). One can typically optimize for two of these attributes at the expense of the third.

For example, the Aya Model prioritizes authenticity and scale by curating vast amounts of human-written text, but this approach may struggle to capture the diversity of niche, interactive cultural scenarios. CultureLLM prioritizes scale and cost-efficiency, but the diversity of its final dataset is ultimately constrained by the semantic boundaries of its initial 50 seed questions. CulturePark prioritizes diversity and scale by synthetically generating novel interactions, but its authenticity is not guaranteed.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

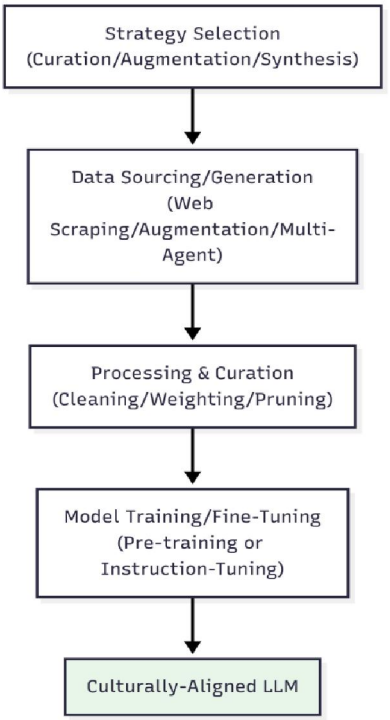

Fig.3 The Data-Centric Trilemma

Model-Centric

The core principle of the model-centric approach is that for cultural awareness to be robust, generalizable, and efficient, it must be an intrinsic property of the model's architecture. Instead of treating cultural knowledge as something to be memorized from data or recalled via a prompt, this approach seeks to create specialized components and pathways within the model that are optimized for cultural tasks.

The Model-Centric architecture represents the most profound and structurally invasive strategy for cultural alignment. This approach moves beyond manipulating external factors like prompts or training data to alter the fundamental internal structure and processing mechanisms of the LLM itself. The goal is to build models that are not just trained on cultural data but are architecturally designed for cultural reasoning, embedding this competence into their very "DNA."

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

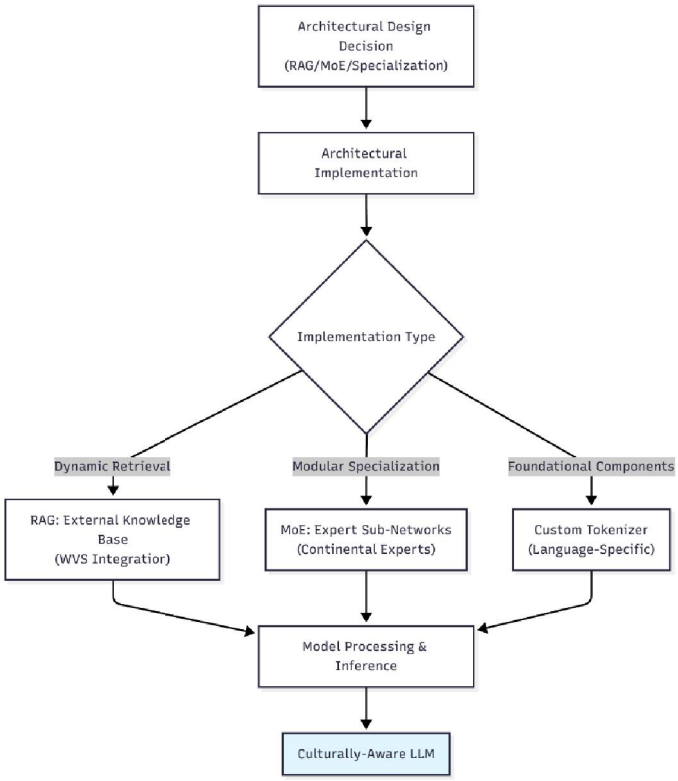

Fig.4 Model-Centric Architectural Modifications

Comparative Synthesis and Strategic Implications

Prompting, Benchmarking, Data-Centric, and Model-Centric—these four architectural approaches each take a different angle at the whole issue of cultural alignment in Large Language Models. They don't stand alone; they overlap and connect, but every one of them steps in at a different point in the AI process, from how you build the data to how you judge the final model. Each one brings its own upsides and downsides, and the choices you make with them really shape your overall strategy.

TABLE I: Comparative Overview of Cultural Alignment Architectures

Architecture	Primary Intervention Point	Core Mechanism	Illustrative Techniques	Relative Cost / Complexity	Primary Limitation
Prompting	Inference-Time	Contextual Steering	Anthropological Prompting, Cultural Prompting	Low	Brittleness; effectiveness is highly dependent on prompt quality and may not generalize.
Benchmarking	Evaluation	Systematic Measurement	SocialCC, BharatBBQ, CALI, cuDialog	Medium	Does not directly mitigate bias, only measures it; requires deep cultural expertise to create.
Data-Centric	Training Data	Corpus	CulturePark,	High	Authenticity–Scale–

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

		Engineering	CultureLLM, Aya Model		Diversity trilemma; can be extremely resource-intensive.
Model-Centric	Model Core	Architectural Modification	ValuesRAG, Cultural Palette (MoE), HindiLLM (Custom Tokenizer)	Very High	High implementation cost and complexity; requires deep architectural expertise.

TABLE III: Comparative Analysis of Cultural Alignment in LLMs (Selected Papers)

No.	Paper Title	Core Approach	Dataset/Technique	Evaluation Metric	Limitation or Key Finding
1	Investigating Cultural Alignment of Large Language Models [1]	Prompting	Anthropological Prompting using WVS	Soft & Hard Similarity between model and survey data	"Thinking like an anthropologist" boosts cultural alignment, especially for marginalized personas.
2	SocialCC [2]	Benchmarking	3,060 intercultural scenarios (SocialCC dataset)	Human and LLM-based evaluation (Awareness, Knowledge, Behavior)	Reveals gap between LLMs' cultural knowledge and real application in interactions.
3	Aya Model [3]	Data-Centric	Aya multilingual dataset (101 languages)	XNLI, XWinograd, FLORES-200, human evals	Large multilingual instruction-tuning significantly boosts non-English performance.
4	Cultural bias and cultural alignment of large language models [4]	Prompting	WVS + Cultural Prompting	Spearman correlation, Euclidean distance (Cultural Map)	GPT models are Western-biased; simple cultural prompts can mitigate bias.
5	CulturePark [5]	Data-Centric	Multi-agent dialogue generation (CulturePark)	Downstream task performance	Multi-agent synthesis produces rich, realistic cross-cultural data.
6	CultureLLM [6]	Data-Centric	Semantic data augmentation from WVS (50 → 50k Qs)	Cultural Q&A accuracy	Cost-efficient augmentation achieves strong cultural alignment.
7	ValuesRAG [7]	Model-Centric	Retrieval-Augmented Generation (WVS knowledge base)	Accuracy on binarized survey tasks	Dynamic retrieval improves alignment vs static fine-tuning.
8	Cultural Palette [8]	Model-Centric	Multi-agent MoE (5 continental "experts")	Pearson correlation, human evals, NLI consistency	Blending continent-level models yields fine-grained cultural generalization.
9	The Sociolinguistic Foundations of	Theory	Sociolinguistic framework (variety =	N/A	LLM bias = varietal misalignment between

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

	Language Modeling [9]		dialect + register + time)		data and target language.
10	Bridging Cultural Nuances in Dialogue Agents [10]	Benchmarking	cuDialog dataset (Hofstede labels)	BLEU, ROUGE, regression alignment metrics	Encoding cultural vectors improves dialogue realism and diversity.
11	BharatBBQ [11]	Benchmarking	BharatBBQ (Indian socio-cultural benchmark)	Bias score, stereotype score	Biases stronger in Indian languages; need localized datasets.
12	Culturally Aware Natural Language Inference [12]	Benchmarking	CALI dataset (India vs USA norms)	Label disagreement rate	Surfaces cultural variation via cross-cultural NLI disagreements.
13	IndicLLMSuite [13]	Data-Centric	IndicLLMSuite (SANGRAHA + fine-tuning data)	Dataset release only	Open blueprint fostering Indian LLM ecosystem.
14	HindiLLM: Large Language Model for Hindi [14]	Model-Centric	Custom Hindi tokenizer + new corpus	Downstream NLP tasks	Language-specific model beats multilingual ones for Hindi.

IV. CHALLENGES, WORKAROUNDS, AND WHAT'S NEXT

The Gap Between Knowing and Acting

Models can store large amounts of cultural knowledge, yet they often struggle to apply it during real conversations. They fail to identify which cultural norms are relevant in complex or sensitive social situations. To address this, researchers have developed interactive evaluations. SocialCC [2] forces models to role-play and resolve social conflicts instead of simply recalling facts. Huang and Yang [12] further categorized failures into knowledge, context, and inference, making it easier to identify where models break down. Despite these efforts, even advanced models such as GPT-4 frequently struggle in realistic social settings.

Not Enough Real Data and Lost Authenticity

High-quality cultural data is difficult to obtain, especially beyond English. Approximately 73% of instruction-tuning datasets are English-heavy, limiting cultural diversity. Synthetic data is often used to fill the gap, but it frequently lacks the subtle details that define real cultural expression. Initiatives such as the Aya Model [3], CulturePark [5], and Sangraha [13] attempt to address this by gathering large multilingual datasets, applying strict filtering, or generating examples through multi-agent collaboration. However, reliance on translation and synthetic data still risks losing the authentic ways people speak and think.

Missing Diversity and Stereotyping

Due to limited exposure to diverse languages and dialects, models often reduce cultures to stereotypes. In some cases, country-specific prompting can even amplify bias rather than reduce it. Recent approaches such as ValuesRAG [7], Cultural Palette [8], and Anthropological Prompting [1] attempt to counter this by incorporating external value sources, blending multiple expert models, or encouraging deeper cultural reasoning. While these methods improve alignment, they still fall short of capturing the full complexity of lived cultural experiences.

Where to Next?

Future research must move beyond country-level cultural alignment. Factors such as caste, religion, and region play a critical role, as demonstrated by benchmarks like BharatBBQ [11]. Training data should reflect real language use, including dialects and informal speech. Finally, projects such as IndicLLMSuite [13] and Aya [3] show that open-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

source collaboration and community-driven data collection are essential for achieving inclusive and representative cultural modeling.

V. CONCLUSION

We have identified an important and rapidly changing landscape: the pursuit of genuinely cultural competence in Large Language Models. With these systems increasingly adopted worldwide, confronting their embedded and crisped Western bias is not only a technical challenge, but also, in the words of the research, an "ethical imperative for equitable AI for all". Our synthesis of fourteen primary studies shows clear methodological growth. The field has authoritatively progressed past finding bias with static, knowledge based probes, like mapping models to WVS data, to measuring actual procedural competence. This has involved an early movement toward interactive, dynamic benchmarks like SocialCC, and culturally-localized measures like BharatBBQ, that have been central to uncovering biases that were previously invisible. The deeper takeaway from this analysis is the field's recognition that declarative knowledge is not procedural competence. For example, it is not enough for an LLM to state cultural facts; it also needs to demonstrate the ability to employ that knowledge in nuanced, real-world interaction. This dilemma is evident in the proposed solutions. Though some immediate, low-cost responses are available (such as advanced prompting strategies that simulate cultural reasoning), our analysis suggests the field is increasingly shifting toward more sustainable, structural solutions. However, both data- and model-centric approaches will ultimately create the developer's dilemma of drifting toward dataset authenticity, scale, and diversity. Importantly, culture is not a fixed target; it is always in flux, meaning the alignment will be an ongoing vexation.

REFERENCES

[1] B. AlKhamissi, M. ElNokrashy, M. Alkhamissi, and M. Diab, "Investigating cultural alignment of large language models," in Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 12 404–12 422. [Online]. Available: https://aclanthology.org/2024.acl-long.671/
[2] J. Wu, J. Lian, D. Wang, and H. M. Meng, "SocialCC: Interactive evaluation for cultural competence in language agents," in Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar, Eds. Vienna, Austria: Association for Computational Linguistics, Jul. 2025, pp. 33 242–33 271. [Online]. Available: https://aclanthology.org/2025.acl-long.1594/
[3] A. Üstün, V. Aryabumi, Z.-X. Yong, W.-Y. Ko, D. D'souza, G. Onilude, N. Bhandari, S. Singh, H.-L. Ooi, A. Kayid, F. Vargus, P. Blunsom, S. Longpre, N. Muennighoff, M. Fadaee, J. Kreutzer, and S. Hooker, "Aya model: An instruction finetuned open-access multilingual language model," 2024. [Online]. Available: https://arxiv.org/abs/2402.07827
[4] Y. Tao, O. Viberg, R. S. Baker, and R. F. Kizilcec, "Cultural bias and cultural alignment of large language models," PNAS Nexus, vol. 3, no. 9, Sep. 2024. [Online]. Available: http://dx.doi.org/10.1093/pnasnexus/pgae346
[5] C. Li, D. Teney, L. Yang, Q. Wen, X. Xie, and J. Wang, "Culturepark: Boosting cross-cultural understanding in large language models," 2024. [Online]. Available: https://arxiv.org/abs/2405.15145
[6] C. Li, M. Chen, J. Wang, S. Sitaram, and X. Xie, "Culturellm: Incorporating cultural differences into large language models," 2024. [Online]. Available: https://arxiv.org/abs/2402.10946
[7] W. Seo, Z. Yuan, and Y. Bu, "Valuesrag: Enhancing cultural alignment through retrieval-augmented contextual learning," 2025. [Online]. Available: https://arxiv.org/abs/2501.01031
[8] J. Yuan, Z. Di, S. Zhao, Z. Cui, H. Wang, G. Yang, and U. Naseem, "Cultural palette: Pluralising culture alignment via multi-agent palette," 2025. [Online]. Available: https://arxiv.org/abs/2412.11167
[9] J. Grieve, S. Bartl, M. Fuoli, J. Grafmiller, W. Huang, A. Jawerbaum, A. Murakami, M. Perlman, D. Roemling, and B. Winter, "The sociolinguistic foundations of language modeling," 2024. [Online]. Available: https://arxiv.org/abs/2407.09241

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

[10] Y. Cao, M. Chen, and D. Hershcovich, "Bridging cultural nuances in dialogue agents through cultural value surveys," in Findings of the Association for Computational Linguistics: EACL 2024, Y. Graham and M. Purver, Eds. St. Julian's, Malta: Association for Computational Linguistics, Mar. 2024, pp. 929–945. [Online]. Available: https://aclanthology.org/2024.findings-eacl.63/

[11] A. Tomar, N. R. Sahoo, and P. Bhattacharyya, "Bharatbbq: A multilingual bias benchmark for question answering in the indian context," 2025. [Online]. Available: https://arxiv.org/abs/2508.07090

[12] J. Huang and D. Yang, "Culturally aware natural language inference," in Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore: Association for Computational Linguistics, 2023, pp. 7591–7609.

[13] M. Khan, P. Mehta, A. Sankar, U. Kumaravelan, S. Doddapaneni, S. B, V. G, S. Jain, A. Kunchukuttan, P. Kumar, R. Dabre, and M. Khapra, "Indicllmsuite: A blueprint for creating pre-training and fine-tuning datasets for indian languages," in Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2024, pp. 15 831–15 879. [Online]. Available: http://dx.doi.org/10.18653/v1/2024.acl-long.843

[14] S. Chouhan, S. B. Nath, and A. Dutta, HindiLLM: Large Language Model for Hindi. Springer Nature Switzerland, Dec. 2024, p. 255–270. [Online]. Available: http://dx.doi.org/10.1007/978-3-031-78172-8_17