Multimodal Large Language Models in Multicultural and Cross-Cultural Contexts: A Survey

www.surveyx.cn

Abstract

This survey paper explores the evolving landscape of Multimodal Large Language Models (MLLMs) within multicultural and cross-cultural contexts, emphasizing their role in enhancing language understanding by integrating diverse data modalities such as text, images, and audio. MLLMs surpass traditional language models by providing culturally nuanced interpretations, essential for effective communication across different cultural and linguistic settings. The paper highlights the significant advancements in MLLMs that address challenges in cross-cultural communication, including machine translation and sign language interpretation, while also acknowledging the persistent challenges in ensuring cultural alignment and understanding complex cultural values. It discusses the architecture and functioning of MLLMs, recent advancements, and the role of cultural diversity in developing multilingual models. Practical applications in various domains such as education, healthcare, and global communication are examined, alongside ethical considerations and the need for culturally inclusive benchmarks. The survey concludes by outlining future research directions, emphasizing the importance of expanding datasets to include underrepresented languages and cultural contexts, and developing innovative methodologies for enhancing MLLM capabilities in culturally diverse environments. By advancing these areas, MLLMs can become pivotal tools in fostering a more interconnected and culturally aware global society.

1 Introduction

1.1 Significance of Multimodal Large Language Models

Multimodal Large Language Models (MLLMs) significantly enhance language understanding and communication across diverse cultural contexts by integrating various data modalities, including text, images, and audio [1]. This integration allows MLLMs to overcome the limitations of traditional language models that rely solely on text, thus improving their ability to generate and interpret culturally relevant content. By incorporating visual and auditory cues, MLLMs facilitate a comprehensive understanding of cultural nuances, which is essential for accurate interpretation in diverse environments [2].

MLLMs also address communication challenges faced by users of different sign languages, such as American Sign Language (ASL) and Indian Sign Language (ISL), promoting intercultural interaction and accessibility within the deaf community [3]. Furthermore, they enhance machine translation processes by leveraging multimodal information, leading to improved translation quality and deeper cultural understanding [4].

Despite these advancements, challenges persist in ensuring cultural alignment and understanding the complexities of diverse cultural values within MLLMs [5]. The development of Retrieval Augmented Generation (RAG) models in multicultural settings highlights the necessity for innovative methodologies to enhance information retrieval and processing [6]. Addressing these challenges is

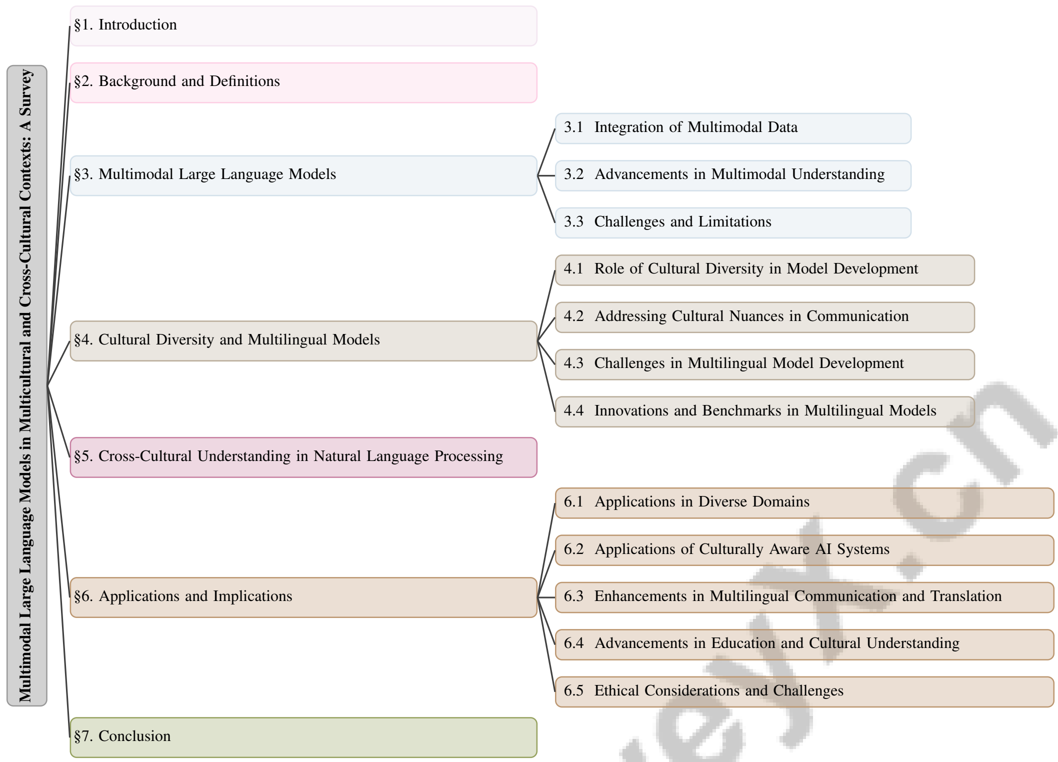


Figure 1: chapter structure

crucial for creating AI systems that accurately reflect cultural diversity and effectively serve users from various backgrounds [7].

MLLMs are pivotal in bridging cultural and linguistic gaps, fostering a more interconnected and culturally aware global society. Their advanced capabilities to integrate and process multimodal inputs—such as text, images, and audio—position them as essential tools for enhancing cross-cultural communication and understanding, particularly through applications like grounded information extraction, culturally aware language models, and inclusive digital expressions such as emojis [5, 8, 9, 10, 11].

1.2 Structure of the Survey

This survey is structured to provide a thorough exploration of Multimodal Large Language Models (MLLMs) within multicultural and cross-cultural contexts. It begins by introducing the significance of MLLMs, underlining their role in enhancing language understanding and communication across diverse cultural settings. A detailed background section follows, defining key concepts related to MLLMs, including multimodal models, multicultural and cross-cultural contexts, and natural language processing.

The third section examines the architecture and functioning of MLLMs, discussing recent advancements and challenges in integrating multimodal data. It also highlights cultural diversity's role in developing multilingual models, focusing on how MLLMs address cultural nuances and the challenges faced in creating culturally aware multilingual systems.

Subsequently, the survey explores the importance of cross-cultural understanding in natural language processing and how MLLMs facilitate improved communication across cultures. This discussion encompasses challenges faced by cross-cultural Natural Language Processing (NLP) systems, such as cultural bias, linguistic diversity, and the complexities of data annotation for under-resourced languages. It also presents evaluation frameworks designed to assess these systems' effectiveness in accommodating diverse cultural contexts, including the CultureAtlas dataset for benchmarking language models and the DualNeighbors algorithm for exploring cross-cultural connections in textual corpora [7, 12, 13, 11].

AI-generated, for reference only.

The practical applications of MLLMs in multicultural and cross-cultural settings are then discussed, focusing on their impact across domains such as translation, education, and global communication. This section also addresses ethical considerations and challenges in deploying MLLMs in multicultural environments.

The survey concludes by synthesizing the main findings and exploring future research avenues in the domain of MLLMs. It emphasizes the critical role of cultural diversity and cross-cultural understanding in AI development, highlighting studies that reveal MLLMs' limitations in reasoning with culturally specific proverbs, identifying cultural associations, and adapting content across different contexts. The necessity for enhanced methodologies to bridge cultural gaps and promote inclusivity in AI systems is underscored, as evidenced by new datasets and evaluation frameworks aimed at improving cultural sensitivity and awareness in language models [14, 15, 16, 17, 11].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definitions of Multimodal Models

Multimodal models are sophisticated computational frameworks that integrate text, images, and audio to enhance the comprehension and generation of complex information. These models are crucial for advancing Multimodal Large Language Models (MLLMs), utilizing the strengths of each modality to provide a comprehensive data interpretation and address the limitations of unimodal systems. By incorporating diverse data types, multimodal models enable nuanced and context-aware language processing, which is vital for tasks demanding cultural sensitivity and cross-cultural understanding [18].

MLLM-Tool exemplifies the application of multimodal models by combining open-source large language models with multimodal encoders to interpret and respond to user instructions across multiple modalities [19]. Similarly, MultiModal-GPT showcases their ability to engage in multi-round dialogues by processing multimodal instructions, enhancing interactive capabilities [1]. These models also bridge communication gaps between users of different sign languages, such as ASL and ISL, promoting effective interaction within the deaf community by addressing linguistic and cultural differences [3]. Their significance extends to autonomous driving, where MLLMs enhance vehicle intelligence by integrating visual, textual, and auditory data [20].

The ability of multimodal models to extract and ground information from multiple modalities is demonstrated by benchmarks like MUIE, which process text, images, audio, and video [9]. These frameworks tackle challenges related to data quality and resource availability, especially in low-resourced languages, ensuring that MLLMs are inclusive and culturally aware [21]. In specialized domains like dentistry, they leverage multi-source data for automated diagnostics and treatment planning, illustrating their potential in shaping future practices [22]. The development and application of multimodal models underscore their importance in enhancing the functionality and cultural adaptability of MLLMs, making them indispensable for creating effective and culturally sensitive AI systems.

2.2 Natural Language Processing and Cultural Diversity

Cultural diversity is integral to natural language processing (NLP), significantly impacting the development and effectiveness of Multimodal Large Language Models (MLLMs). Integrating diverse cultural elements into NLP systems is essential for creating AI technologies that excel in linguistic capabilities while demonstrating cultural sensitivity, ensuring outputs resonate with the socio-cultural realities of global populations [23]. The adaptation of NLP systems to cultural differences is crucial for enhancing their reliability and trustworthiness, enabling them to navigate complex linguistic and cultural landscapes effectively [14].

A primary challenge in embedding cultural diversity into NLP is the underrepresentation of non-Western languages and dialects, leading to biased outcomes due to the predominance of Western-centric data in training. This bias is particularly evident in code-mixing contexts, where multilingual speakers' linguistic backgrounds are often inadequately represented. Additionally, the evaluation of generative AI models has predominantly focused on English, limiting insights into their capabilities across other languages and cultures [24].

Cultural diversity in NLP is crucial for accommodating linguistic variations and addressing social bias, promoting fairness by recognizing the distinct cultural contexts influencing language use. By integrating cross-cultural considerations, NLP systems can better serve diverse user communities and mitigate harmful biases that arise from a one-size-fits-all approach, fostering equitable outcomes in applications such as recruitment and education [7, 25, 26, 27]. The inherent biases in AI models, which often prioritize Western norms, can lead to ineffective outcomes for non-Western users and potentially erase cultural expressions. To counteract these biases, models must effectively recognize and interpret named entities across various languages and modalities, thereby enhancing cross-cultural communication and understanding.

Moreover, existing benchmarks frequently lack adequate representation of diverse languages and cultures, posing a significant barrier to developing culturally aware NLP systems. This inadequacy underscores the urgent need for culturally inclusive models capable of identifying and addressing culturally specific nuances, particularly in under-resourced languages. Current research emphasizes the importance of integrating cultural features into machine learning, especially for tasks like Offensive Language Detection, where cultural values significantly impact outcomes. The development of comprehensive datasets, such as CultureAtlas, aims to bridge cultural knowledge gaps and enhance the performance of language models in diverse contexts. By prioritizing community needs and employing frameworks like the Capabilities Approach, we can ensure that language technologies are more equitable and better equipped to serve a global audience [11, 27, 28].

Incorporating cultural diversity into NLP is vital for advancing MLLMs. By ensuring that these models are culturally aware, developers can enhance their utility and acceptance in a globalized world, facilitating accurate and respectful communication across diverse cultural contexts. The application of NLP techniques is essential for extracting culturally relevant information from textual documents, significantly improving process management across various sectors, including Business Process Management (BPM). Leveraging advanced NLP methodologies, specifically those tailored for BPM tasks, organizations can analyze unstructured text effectively to enhance operational efficiencies. Recent advancements in Large Language Models (LLMs) have broadened the potential of NLP in this domain, enabling the extraction of valuable insights from diverse sources and facilitating process automation without extensive configuration. This integration of NLP supports understanding cultural nuances in language and aids in developing robust and adaptable BPM solutions [29, 7, 30].

In recent years, the development of Multimodal Large Language Models (MLLMs) has garnered significant attention within the field of artificial intelligence. These models are characterized by their ability to process and integrate various forms of data, thereby enhancing their understanding and application across different modalities. As illustrated in Figure 2, the hierarchical structure of MLLMs is depicted, emphasizing the integration of multimodal data and the advancements in multimodal understanding. This figure categorizes the enhanced capabilities and applications of MLLMs, while also highlighting recent advancements in frameworks and techniques. Furthermore, it outlines the challenges and limitations encountered, including issues related to data, cultural implications, performance metrics, and ongoing research challenges that must be addressed to further the development of these sophisticated models. By integrating this visual representation, we can better appreciate the complexities and multifaceted nature of MLLMs in contemporary research.

3 Multimodal Large Language Models

3.1 Integration of Multimodal Data

Integrating multimodal data—encompassing text, images, audio, and video—within Multimodal Large Language Models (MLLMs) is crucial for processing complex information and executing tasks such as long-context understanding and generation. Models like SEEKER, which compress extensive text sequences into visual pixel representations, exemplify enhanced handling of long-form multimodal inputs, achieving superior efficiency and performance [31, 32, 33]. Through synthesizing diverse data forms, MLLMs establish cohesive frameworks for comprehensive and contextually aware information interpretations, enabled by advanced architectures.

The GaMS model enhances language understanding in low-resource contexts by combining Slovene corpora with multilingual sources [34]. The RAG method underscores the importance of multimodal integration in multilingual applications by integrating tools for effective information retrieval across languages and literacy levels [6]. MLLM-Tool demonstrates how multimodal information reduces

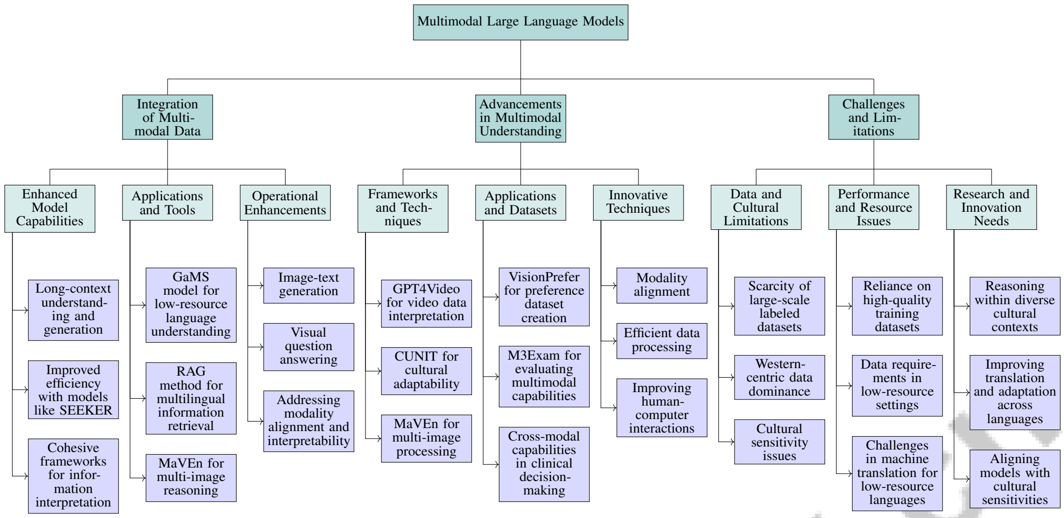


Figure 2: This figure illustrates the hierarchical structure of Multimodal Large Language Models (MLLMs), focusing on the integration of multimodal data, advancements in multimodal understanding, and the challenges and limitations encountered. It categorizes the enhanced capabilities, applications, and operational enhancements of MLLMs, highlights recent advancements in frameworks and techniques, and outlines the data, cultural, performance, and research challenges that need to be addressed.

ambiguity and improves comprehension of complex queries [19], while frameworks like MaVEn enhance multi-image reasoning through a Multi-granularity Hybrid Visual Encoding approach [35].

Applications such as semantically consistent video-to-audio generation illustrate the seamless integration of visual and auditory data, enhancing coherent output production [36]. Datasets enriched by AI annotator feedback, as seen with GPT-4 Vision, contribute fine-grained preference data to the multimodal learning process [37]. The Multilingual Text Classification Framework (MTCF) showcases multimodal integration's role in addressing linguistic diversity by employing language-specific strategies [21]. In specialized domains like dental diagnosis, systems using LLMs illustrate the integration of diverse data forms [22].

Integrating multimodal data in MLLMs significantly enhances operational capabilities by enabling diverse data type processing, informed decision-making, and nuanced communication across cultural contexts. This integration improves tasks like image-text generation and visual question answering while addressing challenges in modality alignment and interpretability, fostering transparency and trust in AI applications [38, 39, 40, 31, 32]. These capabilities are pivotal for advancing natural language processing and bridging linguistic and cultural gaps.

3.2 Advancements in Multimodal Understanding

Recent advancements in multimodal understanding have enhanced MLLMs by integrating various data modalities for improved comprehension and content generation. The GPT4Video framework unifies LLMs with visual feature extractors and generative models, facilitating holistic video data interpretation and addressing visual storytelling complexities [41]. The CUNIT framework introduces a contrastive matching task to evaluate LLMs' ability to identify culturally similar concepts, enhancing cultural adaptability [15].

In image description, multiple pattern layers and attention mechanisms have produced models capable of generating nuanced and contextually relevant descriptions, improving interpretative capabilities [42]. The MaVEn framework contributes to multi-image processing through its dual approach, synergizing discrete visual symbols with continuous representation sequences, enhancing reasoning efficiency and accuracy [35].

The VisionPrefer approach uses AI-generated synthetic data to create a diverse preference dataset, enriching multimodal learning with comprehensive rankings and explanations [37]. In education, the

AI-generated, for reference only.

M3Exam dataset offers a valuable resource for evaluating and improving LLMs' multimodal capabilities through diverse multiple-choice questions involving text and images from real exams across various countries and educational levels [43]. Cross-modal capabilities in applications like clinical decision-making highlight MLLMs' potential to interpret both textual and visual data effectively, enhancing decision-making in specialized fields [22].

These innovations underscore significant advancements in multimodal understanding, illustrating MLLMs' capabilities to integrate and interpret diverse data types across cultural and contextual settings. By employing sophisticated techniques such as modality alignment and efficient data processing, MLLMs enhance human-computer interactions and address challenges related to semantic gaps and long-context comprehension. This progress opens new avenues for applications in fields like recommendation systems, where dynamic multimodal data integration can better capture user preferences and improve system performance [38, 32, 44, 40, 31].

3.3 Challenges and Limitations

The integration of multimodal data in MLLMs presents several challenges and limitations affecting their efficacy and adaptability across diverse cultural contexts. A major challenge is the scarcity of large-scale, manually labeled datasets, which hampers the development of robust models capable of generalizing across various tasks and cultural settings [18]. This limitation is compounded by the predominance of Western-centric data in existing benchmarks, restricting MLLMs' ability to accurately interpret and generate culturally diverse content [16].

Cultural sensitivity is another critical issue, as MLLMs often struggle to recognize and adapt to diverse cultural expression patterns. The lack of benchmarks to assess models' performance in generating cross-cultural content, such as culturally nuanced images, exacerbates this challenge [23]. The absence of culturally inclusive benchmarks can lead to misinterpretations and biases, as models may overlook cultural contexts in their outputs [7].

Reliance on high-quality training datasets is a significant limitation; suboptimal datasets can result in poor model performance, particularly in scenarios where cultural nuances are critical, such as intercultural affect recognition, where annotated datasets for various cultural contexts are lacking [7]. Additionally, the extensive data requirements for training existing models present barriers to the efficient development and deployment of MLLMs, especially in low-resource settings [18].

In machine translation, MLLMs face significant hurdles in developing systems for low-resource languages, where data scarcity and ethical considerations in data collection can lead to potential cultural misrepresentation. These challenges highlight the need for comprehensive datasets that encompass a wider range of languages and cultural expressions [7].

The challenges faced by MLLMs underscore the urgent need for continuous research and innovation to overcome limitations in functionality and cultural adaptability. Enhancing their ability to reason within diverse cultural contexts, as evidenced by difficulties with figurative language and cultural references, is crucial, as is improving performance in translating and adapting content across languages. Addressing these issues will foster a more inclusive and effective application of MLLMs across varied linguistic and cultural landscapes, ultimately bridging gaps for underrepresented languages and aligning models more closely with cultural sensitivities and nuances [14, 31, 17, 45].

4 Cultural Diversity and Multilingual Models

Cultural diversity plays a foundational role in the development of multilingual models, profoundly influencing their effectiveness and applicability across different societies. As communication becomes increasingly globalized, AI technologies must not only understand linguistic variations but also embrace the rich cultural contexts embedded within these languages. This section explores the intricate relationship between cultural diversity and model development, focusing on how diverse cultural perspectives enhance multilingual models' capabilities. We will discuss the specific impact of cultural diversity on linguistic capabilities and cultural sensitivity in model development.

4.1 Role of Cultural Diversity in Model Development

Incorporating cultural diversity into multilingual models is crucial for their ability to interpret and generate content effectively across diverse cultural landscapes. This inclusion enhances AI systems' linguistic versatility and cultural sensitivity, which are vital for global applicability [7]. Current models often focus on monolingual tasks, neglecting shared linguistic features among languages, such as those in the Dravidian language family [21]. Developing culturally aware models is further complicated by the need for diverse datasets, which are costly and challenging to scale. Benchmarks like XC-Translate and the C3 emphasize the importance of translating entity names and generating culturally relevant images, respectively, to ensure models accurately reflect diverse cultural expressions [46, 23].

The D3CODE benchmark highlights cultural diversity's role in evaluating NLP models by focusing on offensive language detection, providing datasets that capture subjective views from multiple cultural backgrounds [47]. Integrating cultural diversity into models is essential for addressing digital inequalities and ensuring equitable access and representation, especially for recent immigrants [7]. Understanding cultural common ground is crucial for multilingual large language models (mLLMs) to bridge cultural gaps and enhance cross-cultural communication [14].

4.2 Addressing Cultural Nuances in Communication

Multimodal Large Language Models (MLLMs) must navigate complex cultural nuances in communication across diverse languages to enhance their efficacy and trustworthiness. The CAMT benchmark, enriched with Cultural Specific Item (CSI) annotations, introduces evaluation metrics like CSI-Match and PTA to assess cultural translation quality beyond linguistic accuracy [48]. Offensiveness detection reframed as a matter of moral judgment influenced by socio-cultural norms highlights the need for MLLMs to incorporate cultural sensitivity [49]. Studies reveal limitations in models like ChatGPT regarding cultural nuances, emphasizing the need for continuous refinement to enhance cultural adaptability [50].

Cross-cultural transfer learning is vital for improving offensive language detection and fostering inclusive online environments [51]. Innovations in attention-based feature selection in intercultural affect recognition models demonstrate how MLLMs can outperform intracultural models by addressing cultural nuances in audiovisual contexts [52]. XC-Translate serves as a focused resource for evaluating cross-cultural translation, driving improvements in machine translation systems [46].

Integrating culturally diverse benchmarks and methodologies is essential for enhancing the cultural sensitivity of MLLMs, addressing challenges like cultural bias and the lack of cultural commonsense knowledge. Leveraging datasets like CultureAtlas and employing frameworks for cultural adaptation can bridge cultural disparities, fostering a more inclusive representation of global cultures in AI applications [14, 48, 15, 17, 11].

4.3 Challenges in Multilingual Model Development

Developing culturally aware multilingual models faces challenges due to inadequacies in benchmarks and datasets. English-centric datasets often fail to capture the cultural values of non-English-speaking communities, leading to potential cultural misalignment [53]. The lack of comprehensive evaluation frameworks that measure cultural adaptability, often overlooking nationality and language influences, compounds this issue [54]. Benchmarks frequently neglect cultural nuances in translating entity names, resulting in insufficient resources for evaluating machine translation performance [46].

Limited coverage of low-resource languages in benchmarks hinders effective multilingual model development, especially for African languages and Arabic dialects. Initiatives like Lanfrica aim to address these gaps by creating platforms for sharing resources [55]. Recent studies emphasize the need for improved multilingual generalization methods that account for cultural sensitivity in underrepresented languages [12, 45].

Existing benchmarks often fail to account for annotator subjectivity, reducing differences to demographic variations and not capturing moral and cultural nuances [47]. The inability to generalize effectively to underrepresented languages due to a lack of data and training exacerbates this issue [45]. The complex nature of social dynamics and the ambiguity of fairness definitions pose challenges in measuring and mitigating biases in multilingual models [25].

Benchmarks often overlook cultural nuances and contextual usage of proverbs, leading to a lack of understanding of multilingual large language models' reasoning in diverse scenarios [14]. Developing more inclusive and culturally representative benchmarks and datasets is essential to better capture the diverse linguistic and cultural landscapes multilingual models navigate.

4.4 Innovations and Benchmarks in Multilingual Models

Benchmark	Size	Domain	Task Format	Metric
COLD[51]	25,726	Offensive Language Detection	Text Classification	F1-score, Accuracy
INSPAIRED[16]	6,000	Social Media	Inspiration Detection	Accuracy, F1-score
M5[56]	261,375	Visio-Linguistics	Visual Question Answering	F1-score, Accuracy
JCM[53]	2,000	Commonsense Morality	Ethical Judgment	Accuracy
CCL[54]	7,286	Cultural Adaptability	Response Simulation	Sign Agreement Rate
CUNIT[15]	1,425	Cultural Studies	Contrastive Matching	Accuracy, Consistency
AfroLingu-MT[57]	620,573	Machine Translation	Translation	spBLEU1K, ChrF++
XM3600[58]	3,600	Cognitive Science	Image Captioning	Saliency Score

Table 1: This table presents a comprehensive overview of representative benchmarks utilized in the development of multilingual models, highlighting their respective sizes, domains, task formats, and evaluation metrics. The benchmarks span a variety of tasks, including offensive language detection, inspiration detection, visual question answering, and machine translation, demonstrating the diverse applications and cultural contexts considered in recent advancements.

Recent innovations and benchmarks have advanced multilingual model development, enhancing cultural adaptability and linguistic versatility. A cross-cultural perspective in offensive language detection, utilizing datasets from multiple languages, improves model adaptability [51]. The INSPAIRED dataset provides a resource for analyzing inspiring content from multiple cultures, setting a new standard for future multilingual model development [16].

Table 1 provides a detailed overview of key benchmarks that have been instrumental in advancing multilingual models, showcasing the diversity in tasks and cultural contexts addressed by recent innovations in the field. These innovations emphasize the need for equitable language technology through initiatives like GlobalBench, which tracks NLP progress across diverse languages; the CultureAtlas dataset, addressing cultural biases; and Lanfrica, documenting machine translation research for Africa's diverse languages. Collectively, these efforts aim to improve NLP systems' performance and accessibility while ensuring inclusive representation of global languages and cultures [55, 59, 60, 11]. They highlight the importance of cultural diversity and rigorous evaluation in advancing the field, ensuring the creation of culturally aware AI systems capable of effective cross-cultural communication.

5 Cross-Cultural Understanding in Natural Language Processing

5.1 Significance of Cross-Cultural Understanding

Cross-cultural understanding is essential in natural language processing (NLP) for developing language models that effectively navigate diverse cultural landscapes. Models capable of interpreting and generating content aligned with various cultural norms enhance global communication and ensure AI systems' relevance in multicultural settings. The D3CODE dataset underscores the importance of incorporating diverse perspectives, highlighting significant demographic and regional differences in perceptions of offensive language [47]. The MAPS dataset further emphasizes cross-cultural understanding by evaluating the reasoning capabilities of multilingual large language models (mLLMs) using proverbs and sayings from different cultures [14]. Such understanding enables models to grasp cultural nuances, fostering more inclusive and respectful interactions.

5.2 Challenges in Cross-Cultural NLP

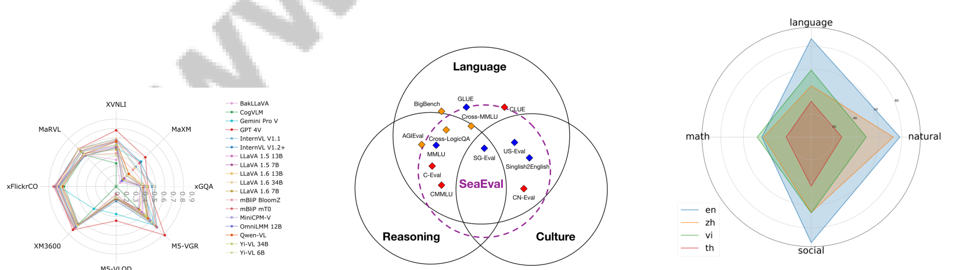
Achieving cross-cultural understanding in NLP is challenged by linguistic and cultural diversity complexities. State-of-the-art models face difficulties with cross-lingual transfer, revealing persistent issues in cross-cultural comprehension [61]. Variability in model preferences across demographic groups complicates this further, as notable disparities in model alignment with diverse cultural expectations are evident [62]. Evaluating Multimodal Large Language Models (MLLMs) across hierarchical capability levels, particularly in multimodal content generation and understanding, remains a critical challenge [63]. Additionally, code-mixed message detection complicates migrant integration into host communities, necessitating NLP systems adept at handling such linguistic phenomena [64].

To address cultural bias, benchmarks simulate cross-cultural communication, generating diverse cultural data to improve model adaptability [65]. However, reliance on synthetic data limits real-world applicability, necessitating validation with actual datasets [66]. Oversimplifying cultural differences by using single datasets as representatives can lead to inadequate cross-cultural understanding, neglecting nuanced complexities in cultural interactions and language use [51, 7]. A collaborative approach integrating diverse linguistic and cultural perspectives is essential for addressing these challenges effectively. Utilizing frameworks like the Capabilities Approach ensures consideration of various communities' unique needs and contexts, fostering the creation of inclusive NLP models [27, 7, 28, 11].

5.3 Benchmarks and Evaluation Frameworks

Assessing cross-cultural NLP systems requires comprehensive benchmarks and evaluation frameworks that capture cultural diversity and linguistic intricacies. Metrics such as Fréchet Inception Distance (FID) for generation quality and accuracy scores for understanding tasks provide quantitative insights into model performance, highlighting areas where cultural nuances may be inadequately addressed [44]. Effective evaluation necessitates benchmarks encompassing a wide array of languages and cultural contexts, as these factors influence linguistic diversity and content requirements. The CultureAtlas dataset exemplifies comprehensive benchmarks that facilitate language model assessment in culturally diverse settings [7, 11]. This approach ensures models are tested against a broad spectrum of linguistic and cultural scenarios, enhancing their generalization across different cultural settings.

Evaluation frameworks must consider the dynamic and multifaceted nature of cultural interactions, characterized by intricate and context-sensitive language use influenced by social norms, cultural practices, and individual perspectives. Integrating diverse cultural knowledge is crucial, as highlighted by recent research emphasizing understanding cultural nuances and biases in language models [67, 68, 13, 69, 11]. Including real-world datasets reflecting authentic cultural exchanges enables accurate assessment of models' ability to navigate and interpret cultural nuances. By leveraging comprehensive benchmarks and evaluation frameworks, researchers can advance cross-cultural NLP systems that are linguistically proficient and culturally sensitive, fostering more effective and inclusive global communication.



(a) The image is a radar chart that compares the performance of various language models on different datasets.[56]

(b) SeaEval: A Comprehensive Evaluation Framework for Language, Reasoning, and Culture[67]

(c) The image is a radar chart that visually represents the distribution of scores across four categories: language, math, social, and natural. The chart is divided into four quadrants, each representing a different category.[70]

Figure 3: Examples of Benchmarks and Evaluation Frameworks

As shown in Figure 3, cross-cultural understanding in NLP necessitates robust benchmarks and evaluation frameworks to ensure comprehensive model performance across diverse linguistic and

AI-generated, for reference only.

cultural contexts. The first radar chart compares various language models' performance on different datasets, highlighting variability in efficacy across linguistic datasets and emphasizing the need for diverse benchmarks. The second image introduces SeaEval, a comprehensive evaluation framework integrating language, reasoning, and cultural dimensions, as depicted in a Venn diagram, underscoring the importance of encompassing multiple domains for effective model evaluation. Lastly, the third radar chart categorizes scores into language, math, social, and natural domains, illustrating the multifaceted nature of NLP model evaluation. Together, these visualizations convey the intricate challenges and considerations involved in developing and assessing NLP systems sensitive to cross-cultural variations [56, 67, 70].

6 Applications and Implications

6.1 Applications in Diverse Domains

Multimodal Large Language Models (MLLMs) have revolutionized fields such as translation, education, and global communication by integrating diverse data modalities. In translation, datasets like XC-Translate challenge systems to handle culturally specific references, enhancing precision and cultural relevance [46]. The GaMS benchmark aids in evaluating generative models for Slovene, improving translation in low-resource languages [34]. In education, MLLMs personalize learning through frameworks like MaVEn, which enhances understanding in multi-image contexts [35]. The MTCF framework boosts classification accuracy in multilingual tasks, impacting educational tools for Dravidian languages [21]. Global communication benefits from MLLMs through frameworks like CAT, which excels in Audio-Visual Question Answering (AVQA) [71], and LLMs in Business Process Management (BPM), enhancing process model mining [29]. In healthcare, models like ChatGPT improve clinical applications in dentistry [22]. These advancements demonstrate MLLMs' capacity to enhance translation, education, and communication by bridging linguistic and cultural gaps [38, 31, 17, 40].

6.2 Applications of Culturally Aware AI Systems

Culturally aware AI systems are vital for effective communication across diverse cultural contexts. Many AI models are rooted in Western-centric perspectives, often neglecting global cultural diversity. Research highlights the need for culturally responsive algorithms that adapt to unique values and communication styles, enhancing inclusivity [72, 73, 74, 11]. Datasets like PSN, which includes social norms across contexts, enable AI systems to align outputs with expected behaviors [75]. The MMMModal framework processes multi-turn dialogues with multiple images and audio inputs, enhancing conversational responsiveness [76], while the C3 benchmark improves cultural sensitivity in text-to-image models [23]. Evaluating image features for cultural adaptability is crucial [42]. These systems address cross-cultural communication challenges, promote inclusive education, and ensure culturally nuanced content generation [74, 72, 73, 11].

6.3 Enhancements in Multilingual Communication and Translation

Advancements in MLLMs have significantly improved multilingual communication and translation by integrating diverse data modalities. The IMAGE framework uses visual information to enhance translation performance, facilitating context comprehension [77]. Models like KOSMOS-2 exemplify progress in grounding capabilities, enhancing interpretation and generation of culturally contextualized content [78]. The MLLM-Tool improves multilingual communication by accurately selecting tools for multimodal instructions [19]. Additionally, the AdaptVision framework optimizes input tokens for natural and text-rich images, boosting performance across benchmarks [79]. These innovations illustrate MLLMs' transformative impact on multilingual communication, improving translation accuracy and cultural relevance [80, 81, 24, 11].

6.4 Advancements in Education and Cultural Understanding

MLLMs have advanced education and cultural understanding by integrating diverse data modalities, enhancing accessibility and comprehension. The LLaVA-Read framework improves accessibility for visually impaired individuals by accurately extracting text from images [82]. The Merlion framework enhances instruction-following capabilities by integrating paralinguistic information [83]. Future research should explore the integration of audio and language data to enrich educational experiences and foster cultural understanding. Digital communication tools like emojis play a significant role in cross-cultural settings, necessitating guidelines for their use in educational and professional environments [10]. MLLMs like SEEKER hold potential to revolutionize education and enhance cultural communication by processing complex visual and textual information [14, 31, 38].

6.5 Ethical Considerations and Challenges

Deploying MLLMs in multicultural settings presents ethical considerations and challenges, primarily concerning cultural bias from inadequately diverse datasets [18]. This issue is compounded by limited labeled datasets, restricting scalability and generalizability. Ethical challenges in machine translation are pronounced for Indigenous languages, necessitating expanded datasets like XC-Translate and advanced metrics for translation quality [46]. Cultural sensitivity is crucial in affect recognition across contexts to avoid misinterpretations and biases [52]. Privacy concerns also arise in applications such as autonomous vehicles and healthcare [22]. Existing benchmarks, predominantly in English, fail to capture offensive language nuances in other languages [47]. Developing inclusive benchmarks is essential for effective cross-cultural understanding. Furthermore, mLLMs' limited reasoning abilities with culturally specific proverbs raise ethical considerations [14]. Mitigating biases in MLLMs is vital for bridging cultural divides, and involving native speakers in model development can enhance representation of diverse cultural viewpoints [14, 84, 4].

7 Conclusion

7.1 Future Directions and Research Opportunities

Advancing Multimodal Large Language Models (MLLMs) hinges on expanding datasets to encompass a broader spectrum of cultural contexts and underrepresented languages, which is crucial for fostering cultural diversity and inclusivity within AI systems. This expansion will refine evaluation frameworks that assess MLLMs' performance across diverse tasks and inputs. Exploring innovative methodologies for controllable data augmentation can further enhance MLLMs' proficiency in culturally aware multilingual applications.

Enhancing translation methodologies and developing diverse prompting techniques are vital for optimizing MLLM performance, particularly in low-resource languages. Research should focus on achieving a balance between parameter selection and task efficiency, applying frameworks such as SPIDER across a wider array of tasks and architectures. Additionally, improving the temporal synchronization between video and audio in multimodal applications and devising precise evaluation metrics for audio generation will propel the field forward.

Integrating High-Dimensional Preference Optimization (HDPO) with advanced techniques for processing high-quality preference data offers a promising research trajectory. Increasing participant diversity and validating cultural norms are essential for deepening cross-cultural understanding in AI systems. Future research should also explore broader model testing and diverse prompt engineering strategies to capture the complexities of moral judgments across different cultures.

Optimizing dynamic reduction mechanisms and exploring novel training methodologies will enhance frameworks like MaVEn, crucial for managing complex multimodal inputs. Expanding datasets to include more diverse multimodal inputs and additional modalities will further augment MLLMs' capabilities, enabling them to better serve global audiences from varied cultural backgrounds.

Future research should aim to improve cultural response consistency, enhance generalization across diverse cultures, and innovate methodologies for integrating cultural awareness into language technologies. In autonomous driving, developing robust datasets, enhancing real-time processing capabilities, and investigating personalized user interactions are imperative. Expanding the application of LLMs in Business Process Management (BPM) to other lifecycle tasks, with a focus on refined prompt engineering techniques that consider cultural diversity, is also necessary.

Collaborative approaches that prioritize community input, particularly for Creole languages, should be pursued to address unique linguistic features and advance speech technologies. Moreover, frameworks that uphold Indigenous sovereignty over language data and involve community members in the design of machine translation systems should be established. Expanding language coverage for Dravidian languages and optimizing the Multilingual Text Classification Framework (MTCF) for improved performance in multilingual contexts are critical steps forward.

Focusing on these research avenues will significantly advance the development of culturally aware and effective AI systems, enhancing global communication and interaction while underscoring the importance of cultural diversity and cross-cultural understanding in AI development.

References

[1] Tao Gong, Chengqi Lyu, Shilong Zhang, Yudong Wang, Miao Zheng, Qian Zhao, Kuikun Liu, Wenwei Zhang, Ping Luo, and Kai Chen. Multimodal-gpt: A vision and language model for dialogue with humans. *arXiv preprint arXiv:2305.04790*, 2023.

[2] Wei Zhang, Wong Kam-Kwai, Biying Xu, Yiwen Ren, Yuhuai Li, Minfeng Zhu, Yingchaojie Feng, and Wei Chen. Cultiverse: Towards cross-cultural understanding for paintings with large language model, 2024.

[3] Malay Kumar, S. Sarvajit Visagan, Tanish Sarang Mahajan, and Anisha Natarajan. Enhanced sign language translation between american sign language (asl) and indian sign language (isl) using llms, 2024.

[4] Manuel Mager, Elisabeth Mager, Katharina Kann, and Ngoc Thang Vu. Ethical considerations for machine translation of indigenous languages: Giving a voice to the speakers, 2023.

[5] Yong Cao, Li Zhou, Seolhwa Lee, Laura Cabello, Min Chen, and Daniel Hershcovich. Assessing cross-cultural alignment between chatgpt and human societies: An empirical study, 2023.

[6] Syed Rameel Ahmad. Enhancing multilingual information retrieval in mixed human resources environments: A rag model implementation for multicultural enterprise, 2024.

[7] Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, Constanza Fierro, Katerina Margatina, Phillip Rust, and Anders Søgaard. Challenges and strategies in cross-cultural nlp, 2022.

[8] Chao Zhang, Zichao Yang, Xiaodong He, and Li Deng. Multimodal intelligence: Representation learning, information fusion, and applications, 2020.

[9] Meishan Zhang, Hao Fei, Bin Wang, Shengqiong Wu, Yixin Cao, Fei Li, and Min Zhang. Recognizing everything from all modalities at once: Grounded multimodal universal information extraction, 2024.

[10] Lingfeng Li and Xiangwen Zheng. Cross-cultural communication in the digital age: An analysis of cultural representation and inclusivity in emojis, 2024.

[11] Yi Fung, Ruining Zhao, Jae Doo, Chenkai Sun, and Heng Ji. Massively multi-cultural knowledge acquisition lm benchmarking, 2024.

[12] Teresa Lynn, Malik H. Altakrori, Samar Mohamed Magdy, Rocktim Jyoti Das, Chenyang Lyu, Mohamed Nasr, Younes Samih, Kirill Chirkunov, Alham Fikri Aji, Preslav Nakov, Shantanu Godbole, Salim Roukos, Radu Florian, and Nizar Habash. From multiple-choice to extractive qa: A case study for english and arabic, 2025.

[13] Taylor Arnold and Lauren Tilton. Cross-discourse and multilingual exploration of textual corpora with the dualneighbors algorithm, 2018.

[14] Chen Cecilia Liu, Fajri Koto, Timothy Baldwin, and Iryna Gurevych. Are multilingual llms culturally-diverse reasoners? an investigation into multicultural proverbs and sayings, 2024.

[15] Jialin Li, Junli Wang, Junjie Hu, and Ming Jiang. How well do llms identify cultural unity in diversity?, 2024.

[16] Oana Ignat, Gayathri Ganesh Lakshmy, and Rada Mihalcea. Cross-cultural inspiration detection and analysis in real and llm-generated social media data, 2024.

[17] Pushpdeep Singh, Mayur Patidar, and Lovekesh Vig. Translating across cultures: Llms for intralingual cultural adaptation, 2024.

[18] Feng Li, Hao Zhang, Yi-Fan Zhang, Shilong Liu, Jian Guo, Lionel M. Ni, PengChuan Zhang, and Lei Zhang. Vision-language intelligence: Tasks, representation learning, and large models, 2022.

[19] Chenyu Wang, Weixin Luo, Qianyu Chen, Haonan Mai, Jindi Guo, Sixun Dong, Xiaohua, Xuan, Zhengxin Li, Lin Ma, and Shenghua Gao. Mllm-tool: A multimodal large language model for tool agent learning, 2024.

[20] Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, Yang Zhou, Kaizhao Liang, Jintai Chen, Juanwu Lu, Zichong Yang, Kuei-Da Liao, et al. A survey on multimodal large language models for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 958–979, 2024.

[21] Xiaotian Lin, Nankai Lin, Kanoksak Wattanachote, Shengyi Jiang, and Lianxi Wang. Multilingual text classification for dravidian languages, 2021.

[22] Hanyao Huang, Ou Zheng, Dongdong Wang, Jiayi Yin, Zijin Wang, Shengxuan Ding, Heng Yin, Chuan Xu, Renjie Yang, Qian Zheng, and Bing Shi. Chatgpt for shaping the future of dentistry: The potential of multi-modal large language model, 2023.

[23] Bingshuai Liu, Longyue Wang, Chenyang Lyu, Yong Zhang, Jinsong Su, Shuming Shi, and Zhaopeng Tu. On the cultural gap in text-to-image generation, 2023.

[24] Mor Ventura, Eyal Ben-David, Anna Korhonen, and Roi Reichart. Navigating cultural chasms: Exploring and unlocking the cultural pov of text-to-image models, 2024.

[25] Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. Bias and fairness in large language models: A survey, 2024.

[26] Vincent Freiberger and Erik Buchmann. Fairness certification for natural language processing and large language models, 2024.

[27] Hellina Hailu Nigatu and Zeerak Talat. A capabilities approach to studying bias and harm in language technologies, 2024.

[28] Li Zhou, Antonia Karamolegkou, Wenyu Chen, and Daniel Hershcovich. Cultural compass: Predicting transfer learning success in offensive language detection with cultural features, 2024.

[29] Michael Grohs, Luka Abb, Nourhan Elsayed, and Jana-Rebecca Rehse. Large language models can accomplish business process management tasks, 2023.

[30] Yue Kang, Zhao Cai, Chee-Wee Tan, Qian Huang, and Hefu Liu. Natural language processing (nlp) in management research: A literature review. *Journal of Management Analytics*, 7(2):139–172, 2020.

[31] Yujie Lu, Xiujun Li, Tsu-Jui Fu, Miguel Eckstein, and William Yang Wang. From text to pixel: Advancing long-context understanding in mllms, 2024.

[32] Tianyi Bai, Hao Liang, Binwang Wan, Yanran Xu, Xi Li, Shiyu Li, Ling Yang, Bozhou Li, Yifan Wang, Bin Cui, Ping Huang, Jiulong Shan, Conghui He, Binhang Yuan, and Wentao Zhang. A survey of multimodal large language model from a data-centric perspective, 2024.

[33] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. A survey on multimodal large language models, 2024.

[34] Domen Vreš, Martin Božič, Aljaž Potočnik, Tomaž Martinčič, and Marko Robnik-Šikonja. Generative model for less-resourced language with 1 billion parameters, 2024.

[35] Chaoya Jiang, Jia Hongrui, Haiyang Xu, Wei Ye, Mengfan Dong, Ming Yan, Ji Zhang, Fei Huang, and Shikun Zhang. Maven: An effective multi-granularity hybrid visual encoding framework for multimodal large language model, 2024.

[36] Gehui Chen, Guan'an Wang, Xiaowen Huang, and Jitao Sang. Semantically consistent video-to-audio generation using multimodal language large model, 2024.

[37] Xun Wu, Shaohan Huang, and Furu Wei. Multimodal large language model is a human-aligned annotator for text-to-image generation, 2024.

[38] Shezheng Song, Xiaopeng Li, Shasha Li, Shan Zhao, Jie Yu, Jun Ma, Xiaoguang Mao, and Weimin Zhang. How to bridge the gap between modalities: Survey on multimodal large language model, 2025.

[39] Yunkai Dang, Kaichen Huang, Jiahao Huo, Yibo Yan, Sirui Huang, Dongrui Liu, Mengxi Gao, Jie Zhang, Chen Qian, Kun Wang, et al. Explainable and interpretable multimodal large language models: A comprehensive survey. *arXiv preprint arXiv:2412.02104*, 2024.

[40] Yuyang Ye, Zhi Zheng, Yishan Shen, Tianshu Wang, Hengruo Zhang, Peijun Zhu, Runlong Yu, Kai Zhang, and Hui Xiong. Harnessing multimodal large language models for multimodal sequential recommendation, 2025.

[41] Zhanyu Wang, Longyue Wang, Zhen Zhao, Minghao Wu, Chenyang Lyu, Huayang Li, Deng Cai, Luping Zhou, Shuming Shi, and Zhaopeng Tu. Gpt4video: A unified multimodal large language model for lnstruction-followed understanding and safety-aware generation, 2024.

[42] Dan Sun, Yaxin Liang, Yining Yang, Yuhan Ma, Qishi Zhan, and Erdi Gao. Research on optimization of natural language processing model based on multimodal deep learning, 2024.

[43] Wenxuan Zhang, Mahani Aljunied, Chang Gao, Yew Ken Chia, and Lidong Bing. M3exam: A multilingual, multimodal, multilevel benchmark for examining large language models. *Advances in Neural Information Processing Systems*, 36:5484–5505, 2023.

[44] Chunwei Wang, Guansong Lu, Junwei Yang, Runhui Huang, Jianhua Han, Lu Hou, Wei Zhang, and Hang Xu. Illume: Illuminating your llms to see, draw, and self-enhance, 2024.

[45] Samuel Cahyawijaya. Llm for everyone: Representing the underrepresented in large language models, 2024.

[46] Simone Conia, Daniel Lee, Min Li, Umar Farooq Minhas, Saloni Potdar, and Yunyao Li. Towards cross-cultural machine translation with retrieval-augmented generation from multilingual knowledge graphs, 2024.

[47] Aida Mostafazadeh Davani, Mark Díaz, Dylan Baker, and Vinodkumar Prabhakaran. D3code: Disentangling disagreements in data across cultures on offensiveness detection and evaluation, 2024.

[48] Binwei Yao, Ming Jiang, Tara Bobinac, Diyi Yang, and Junjie Hu. Benchmarking machine translation with cultural awareness, 2024.

[49] Aida Davani, Mark Díaz, Dylan Baker, and Vinodkumar Prabhakaran. Disentangling perceptions of offensiveness: Cultural and moral correlates, 2023.

[50] Hang Yuan, Zhongyue Che, Shao Li, Yue Zhang, Xiaomeng Hu, and Siyang Luo. The high dimensional psychological profile and cultural bias of chatgpt, 2024.

[51] Li Zhou, Laura Cabello, Yong Cao, and Daniel Hershcovich. Cross-cultural transfer learning for chinese offensive language detection, 2023.

[52] Leena Mathur, Ralph Adolphs, and Maja J Matarić. Towards intercultural affect recognition: Audio-visual affect recognition in the wild across six cultures, 2022.

[53] Yuu Jinnai. Does cross-cultural alignment change the commonsense morality of language models?, 2024.

[54] Louis Kwok, Michal Bravansky, and Lewis D. Griffin. Evaluating cultural adaptability of a large language model via simulation of synthetic personas, 2024.

[55] Chris C. Emezue and Bonaventure F. P. Dossou. Lanfrica: A participatory approach to documenting machine translation research on african languages, 2020.

[56] Florian Schneider and Sunayana Sitaram. M5 – a diverse benchmark to assess the performance of large multimodal models across multilingual and multicultural vision-language tasks, 2024.

[57] AbdelRahim Elmadany, Ife Adebara, and Muhammad Abdul-Mageed. Toucan: Many-to-many translation for 150 african language pairs, 2024.

[58] Uri Berger and Edoardo M. Ponti. Cross-lingual and cross-cultural variation in image descriptions, 2024.

[59] Yueqi Song, Catherine Cui, Simran Khanuja, Pengfei Liu, Fahim Faisal, Alissa Ostapenko, Genta Indra Winata, Alham Fikri Aji, Samuel Cahyawijaya, Yulia Tsvetkov, Antonios Anastasopoulos, and Graham Neubig. Globalbench: A benchmark for global progress in natural language processing, 2023.

[60] Kathrin Blagec, Georg Dorffner, Milad Moradi, Simon Ott, and Matthias Samwald. A global analysis of metrics used for measuring performance in natural language processing, 2022.

[61] Fangyu Liu, Emanuele Bugliarello, Edoardo Maria Ponti, Siva Reddy, Nigel Collier, and Desmond Elliott. Visually grounded reasoning across languages and cultures, 2021.

[62] Hannah Rose Kirk, Alexander Whitefield, Paul Röttger, Andrew Bean, Katerina Margatina, Juan Ciro, Rafael Mosquera, Max Bartolo, Adina Williams, He He, Bertie Vidgen, and Scott A. Hale. The prism alignment dataset: What participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models, 2024.

[63] Bohao Li, Yuying Ge, Yixiao Ge, Guangzhi Wang, Rui Wang, Ruimao Zhang, and Ying Shan. Seed-bench: Benchmarking multimodal large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13299–13308, 2024.

[64] Fedor Vitiugin, Sunok Lee, Henna Paakki, Anastasiia Chizhikova, and Nitin Sawhney. Unraveling code-mixing patterns in migration discourse: Automated detection and analysis of online conversations on reddit, 2024.

[65] Cheng Li, Damien Teney, Linyi Yang, Qingsong Wen, Xing Xie, and Jindong Wang. Culturepark: Boosting cross-cultural understanding in large language models, 2024.

[66] Chanseo Lee, Sonu Kumar, Kimon A. Vogt, Sam Meraj, and Antonia Vogt. Advancing complex medical communication in arabic with sporo arasum: Surpassing existing large language models, 2024.

[67] Bin Wang, Zhengyuan Liu, Xin Huang, Fangkai Jiao, Yang Ding, AiTi Aw, and Nancy F. Chen. Seaeval for multilingual foundation models: From cross-lingual alignment to cultural reasoning, 2024.

[68] Sky CH-Wang, Arkadiy Saakyan, Oliver Li, Zhou Yu, and Smaranda Muresan. Sociocultural norm similarities and differences via situational alignment and explainable textual entailment, 2023.

[69] Reem I. Masoud, Ziquan Liu, Martin Ferianc, Philip Treleaven, and Miguel Rodrigues. Cultural alignment in large language models: An explanatory analysis based on hofstede's cultural dimensions, 2024.

[70] Wenxuan Zhang, Sharifah Mahani Aljunied, Chang Gao, Yew Ken Chia, and Lidong Bing. M3exam: A multilingual, multimodal, multilevel benchmark for examining large language models, 2023.

[71] Qilang Ye, Zitong Yu, Rui Shao, Xinyu Xie, Philip Torr, and Xiaochun Cao. Cat: Enhancing multimodal large language model to answer questions in dynamic audio-visual scenarios, 2024.

[72] Uwe Peters and Mary Carman. Cultural bias in explainable ai research: A systematic analysis, 2024.

[73] Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. Ai suggestions homogenize writing toward western styles and diminish cultural nuances, 2024.

[74] Natalia Ożegalska-Łukasik and Szymon Łukasik. Culturally responsive artificial intelligence – problems, challenges and solutions, 2023.

[75] Hamidreza Saffari, Mohammadamin Shafiei, and Francesco Pierri. Psn: Persian social norms dataset for cross-cultural ai, 2024.

[76] Husein Zolkepli, Aisyah Razak, Kamarul Adha, and Ariff Nazhan. Mmmmodal – multi-images multi-audio multi-turn multi-modal, 2024.

[77] Andong Chen, Yuchen Song, Kehai Chen, Muyun Yang, Tiejun Zhao, and Min Zhang. Make imagination clearer! stable diffusion-based visual imagination for multimodal machine translation, 2025.

[78] Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, and Furu Wei. Kosmos-2: Grounding multimodal large language models to the world, 2023.

[79] Yonghui Wang, Wengang Zhou, Hao Feng, and Houqiang Li. Adaptvision: Dynamic input scaling in mllms for versatile scene understanding, 2024.

[80] Longju Bai, Angana Borah, Oana Ignat, and Rada Mihalcea. The power of many: Multi-agent multimodal models for cultural image captioning, 2024.

[81] Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and S Yu Philip. Multimodal large language models: A survey. In *2023 IEEE International Conference on Big Data (BigData)*, pages 2247–2256. IEEE, 2023.

[82] Ruiyi Zhang, Yufan Zhou, Jian Chen, Jiuxiang Gu, Changyou Chen, and Tong Sun. Llava-read: Enhancing reading ability of multimodal language models, 2024.

[83] Yingxu He, Zhuohan Liu, Shuo Sun, Bin Wang, Wenyu Zhang, Xunlong Zou, Nancy F. Chen, and Ai Ti Aw. Meralion-audiollm: Bridging audio and language with large language models, 2025.

[84] Mijntje Meijer, Hadi Mohammadi, and Ayoub Bagheri. Llms as mirrors of societal moral standards: reflection of cultural divergence and agreement across ethical topics, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.