Culturally Aware and Adapted NLP: A Taxonomy and a Survey of the State of the Art

Chen Cecilia Liu[1,2,3] and **Iryna Gurevych**[1] and **Anna Korhonen**[3]
[1]UKP Lab, Department of Computer Science and hessian.AI,
Technical University of Darmstadt, Germany
[2]Konrad Zuse School of Excellence in Learning and Intelligent Systems (ELIZA), Germany
[3]Language Technology Lab, University of Cambridge, UK
{chen.liu,iryna.gurevych}@tu-darmstadt.de,
alk23@cam.ac.uk

Abstract

The surge of interest in *culture* in NLP has inspired much recent research, but a shared understanding of ''culture'' remains unclear, making it difficult to evaluate progress in this emerging area. Drawing on prior research in NLP and related fields, we propose a fine-grained taxonomy of elements in culture that can provide a systematic framework for analyzing and understanding research progress. Using the taxonomy, we survey existing resources and methods for culturally aware and adapted NLP, providing an overview of the state of the art and the research gaps that still need to be filled.

1 Introduction

Culture is rapidly becoming an important research topic in Natural Language Processing (NLP), with a significant recent surge in the number of published papers (Figure 1). Current NLP systems, especially Large Language Models (LLMs), often lack fairness and diversity in cultural awareness, which leads to biased performance that disproportionately favors certain groups, and causes harm to others (Sambasivan et al., 2021; Johnson et al., 2022; Cao et al., 2023; Hofmann et al., 2024b). To build technology that is equitable, inclusive, and accessible, the NLP community must actively take the initiative, enhancing LLMs' cultural awareness and adaptability. Given the keen interest in this area and its importance for the safety and fairness of LLMs, it is now important to consolidate existing research on culturally aware and adapted NLP to take stock of the progress made so far and to identify research gaps. However, this is challenged by the lack of a common understanding of the concept of ''culture'' in NLP.

Prior work in NLP such as that by Hershcovich et al. (2022) laid the vital foundations for understanding how language, culture and society interact. Hershcovich et al. (2022) proposed a simple taxonomy derived from the interaction between language and culture that captures broad elements of culture (linguistic form and style, objectives and values, common ground, and aboutness). More recently, Adilazuarda et al. (2024) adopted ''proxies of culture'' (semantic or demographic proxies). While neither provides a shared understanding of culture, perhaps unsurprisingly, ***language*** is an essential component of culture in NLP.

A shared understanding of culture in NLP could benefit from examining definitions developed in anthropology and social sciences.[1] In these fields (Tylor, 1871; Kroeber and Kluckhohn, 1952; White, 1959; UNESCO, 1982; Matsumoto and Juang, 1996; Blake, 2000; Geertz, 1973; Goffman, 2023), most definitions of culture involve *people*, groups, and the interactions within and between individuals and groups.

Murdock (1940) describes culture as ''ideational''[2] and ''social''. White (1959) describes the *locus of culture* as: **(1)** ''*within human*'' (e.g., concepts, beliefs, i.e., ***ideational***), **(2)** between ''*social interaction* among human beings'', and **(3)** outside of human but ''*within the patterns of social interaction*''. An examination of such work reveals that ***social*** *interactions* are critical components of culture, in addition to ideational elements. Notably, this social aspect is underrepresented in previous cross-cultural NLP research

[1]They have been thinking about culture for over a century!

[2]The values, beliefs, norms, and ideas that constitute a way of life (Murdock, 1940; Briggle and Mitcham, 2012b).

Transactions of the Association for Computational Linguistics, vol. 13, pp. 652–689, 2025. https://doi.org/10.1162/tacl_a_00760
Action Editor: Baobao Chang. Submission batch: 11/2024; Revision batch: 3/2025; Published 7/2025.

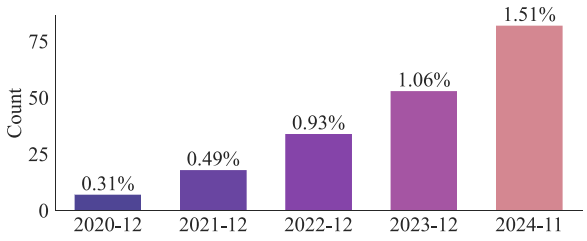

Figure 1: Papers title and abstract containing ‘‘culture’’ or ‘‘cultural’’ published in the main and findings of AACL/EACL/NAACL/ACL/EMNLP and TACL within 5 years, with normalized percentages based on the total number of papers at included venues to date.

(Hershcovich et al., 2022; Adilazuarda et al., 2024).[3]

Hence, we combine the emphasis on *language* in prior NLP work, while integrating the significance of *social* and *ideational* elements that shape culture. This leads us to a working definition of culture used to taxonomize recent work in NLP in this survey: *Culture encompasses the collective ideas, shared language, and social practices that emerge from and evolve through human social interactions within a society*.

Grounded in this working definition, we introduce a new fine-grained taxonomy of culture by expanding on the basic categories of prior work (Hovy and Yang, 2021; Hershcovich et al., 2022) to address above-mentioned issues. We then use this taxonomy to organize existing works in culturally aware and adapted NLP, and identify research gaps. Our survey of 127 publications in leading *CL venues (see selection method in Appendix A) provides an up-to-date view of cultural adaptation resources and models and identifies areas of progress as well as new research opportunities. We hope our taxonomy and analysis enable and inspire further research in this important emerging area.

2 The Taxonomy

In this section, we present our new taxonomy of culture. Unlike previous NLP studies (Hershcovich et al., 2022; Hovy and Yang, 2021) that sought to define cultural elements, this taxonomy (i) is grounded in well-established elements of culture in anthropology and social sciences (Tylor, 1871; Kroeber and Kluckhohn, 1952; White, 1959), (ii) consists of more fine-grained elements than in earlier work, and (iii) allows for a wider consideration of how social factors and variations in humans influence culture.

Figure 2 presents a taxonomy of cultural elements derived from our working definition in §1, organized into three main branches: **ideational**, **linguistic**, and **social**.[4] The **ideational** branch (§3.1; Murdock, 1940; Briggle and Mitcham, 2012a) encompasses the non-material aspects of culture that constitute a way of life, such as values or knowledge. The **linguistic** branch (§3.2) focuses on cultural variations in language and linguistic forms, bridging the ideational and social elements of culture. The **social** branch (§3.3) covers key factors in social interaction and communication, such as relationships or communicative goals.[5] Here, we define each element based on existing research and relating to example tasks in the NLP context. We then provide details and examples from the current literature in §3.1, §3.2, and §3.3.

Ideational elements are based on well-established discussions of culture (Tylor, 1871; Kroeber and Kluckhohn, 1952; White, 1959):

Concepts: basic units of meaning that structure and facilitate thought, bridging sensory experience (Jackendoff, 1989, 2012), e.g., cuisines (such as schnitzel, ratatouille) or holidays (such as Diwali, Nowruz). Related NLP task examples: question answering, dialogue generation.

Knowledge: information that can be acquired through education or practical experience, e.g., local agricultural knowledge. Related NLP task examples: dialogue generation, reasoning.

Values: beliefs, desirable end states or behaviors ranked by importance that can guide evaluations of things (Schwartz, 1992). Unlike norms and morals, values do not inherently involve ethical judgment, e.g., beauty standards, or perception of hate speech. Related NLP task examples: content moderation, debiasing.

[3]There is also prior work focusing on social NLP (Hovy and Yang, 2021), which is related. However, culture is not the central theme of the work.

[4]Icons in Figure 2 are created with the assistance of DALL-E.

[5]All cultural elements can interact and influence each other based on context and can be divided into finer groups. Similar to prior work, our taxonomy abstracts away from these contextual variations.

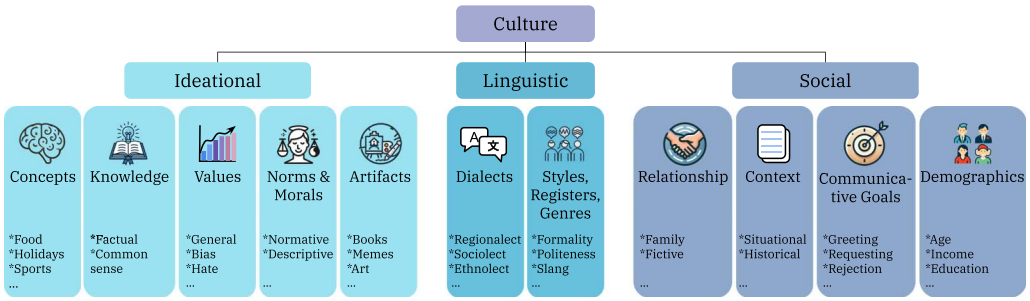

Figure 2: An overview of the taxonomy with examples of subcategories of future possible expansions. The elements in culture are organized into three different branches: **ideational**, **linguistic**, and **social**. The **ideational** branch (§3.1 encompasses the non-material aspects of culture that constitute a way of life. The **linguistic** branch (§3.2) focuses on cultural variations in language and linguistic forms, bridging the ideational and social elements of culture. The **social** branch (§3.3) covers key factors in social interaction and communication.

Norms and Morals: set of rules or principles that govern people's behavior and everyday reasoning (Cialdini et al., 1991; Bicchieri et al., 2018; Hechter and Opp, 2001; Gert and Gert, 2025), e.g., filial obedience attitude. Related NLP task examples: reasoning, safety alignment.

Artifacts: items that are products of human culture, such as art, poetry (White, 1959), etc. This is ideational in our taxonomy since we do not work on physical buildings or tools in NLP. Related NLP task examples: machine translation of long-form literature, emotion arc analysis of movies, memes classification.

Linguistic elements relate to language variations in the cultural context, based on existing discussions (Wardhaugh and Fuller, 2021):

Dialects: includes variations of languages in a systematic way (Fromkin et al., 1998; Trudgill, 2000; Wardhaugh and Fuller, 2021; such as dialects continuum, regionalects, sociolects, etc.), e.g., African American (Vernacular) English (AAE/AAVE). Related NLP task examples: machine translation, debiasing.

Styles, Registers, Genres: includes elements such as formality, variations of language in situation and communicative forms (Wardhaugh and Fuller, 2021), e.g., formality in text, slang, or specific genres like news, folk tales. Related NLP task examples: style transfer, creative writing generation.

Social elements focus on social interactions and communication among humans within the scope of NLP. Leveraging the work of Hovy and Yang (2021), we identify relevant elements:

Relationship: connection between two or more individuals or groups, e.g., father-son, colleagues. Related NLP task examples: creative writing generation, dialogue generation.

Context: the ''containers'' of communications (Yang, 2019), which can be linguistic such as surrounding sentences or extra-linguistic (Hovy and Yang, 2021), including social settings (e.g., at a wedding), non-verbal cues (e.g., gesture), or historical contexts (e.g., colonization). Related NLP task examples: coreference resolution, pragmatic inference.

Communicative Goals: the intention behind language use, e.g., requests, apologies, persuasion. Related NLP task examples: intent classification, emotion classification, human-AI collaboration.

Demographics: the characteristics of people, e.g., economic income, education level, nationality, location, political view, family status, etc. Related NLP task examples: content moderation, personalization.

3 Elements of Culture in Current NLP (Resources) Literature

In this section, we survey and categorize NLP resources. Table 1 shows an overview of papers organized according to the taxonomy.

We observe that in resources, culture can be captured in **(1)** the data itself, or **(2)** in the labels (e.g., multi-culturally annotated). Further, while cultural differences are evident in linguistic and social elements, most current work relies on standard language or country boundaries, leaving these elements understudied.

Element	Papers
Concepts	Shwartz, 2022; Majewska et al., 2023; Hu et al., 2023; Kabra et al., 2023 Liu et al., 2024a; Cao et al., 2024b; Jiang and Joshi, 2024; Hu et al., 2024 **Vision-Language:** Liu et al., 2021; Yin et al., 2021; Thapliyal et al., 2022 Khanuja et al., 2024; Li et al., 2024c; Bhatia et al., 2024; Nayak et al., 2024
Knowledge	**Probing:** Kassner et al., 2021; Yin et al., 2022; Keleg and Magdy, 2023; Zhou et al., 2024; Bhatt and Diaz, 2024 **MMLU:** Koto et al., 2023; Li et al., 2024b; Koto et al., 2024a; Wang et al., 2024b; Son et al., 2024 **Common sense:** Ponti et al., 2020; Wibowo et al., 2024; Koto et al., 2024b; Acquaye et al., 2024; Shi et al., 2024
Values - general	Tay et al., 2020; Ramezani and Xu, 2023; Cao et al., 2023; Wang et al., 2024c Yao et al., 2024b; Aakanksha et al., 2024
Values - bias	**WEAT:** Malik et al., 2022; España-Bonet and Barrón-Cedeño, 2022; Mukherjee et al., 2023; Zhao et al., 2023 Sahoo et al., 2023; Naous et al., 2024; Jha et al., 2023; Bhutani et al., 2024; Mukherjee et al., 2024 **Sent. Pairs:** Nangia et al., 2020; Névéol et al., 2022; Felkner et al., 2023; Sahoo et al., 2024 **Other:** Campolungo et al., 2022; Sandoval et al., 2023; Attanasio et al., 2023; Bauer et al., 2023 Zhou et al., 2022; Lee et al., 2023a; Palta and Rudinger, 2023; An et al., 2023; Jin et al., 2024; Hsieh et al., 2024
Values - hate	Shekhar et al., 2022; Zhou et al., 2023a,b; Lee et al., 2024b; Mostafazadeh Davani et al., 2024
Values - other perceptions	Mohamed et al., 2022; Frenda et al., 2023; Casola et al., 2024; Havaldar et al., 2024 Mohamed et al., 2024; Deas et al., 2024
Norms and Morals	Forbes et al., 2020; Emelin et al., 2021; Ziems et al., 2022b; Kim et al., 2022; Moghimifar et al., 2023 Fung et al., 2023; Pyatkin et al., 2023; CH-Wang et al., 2023; Ziems et al., 2023a; Dwivedi et al., 2023 Rao et al., 2023; Huang and Yang, 2023; Sun et al., 2023; Li et al., 2023b Zhan et al., 2024; Kim et al., 2024; Bhatt and Diaz, 2024; Vijjini et al., 2024; Liu et al., 2024c
Artifacts	Epure et al., 2020; Mohamed et al., 2022; Kruk et al., 2023 Jiang et al., 2023; Mohamed et al., 2024; Hobson et al., 2024
Dialects	Ziems et al., 2022a, 2023b; Le and Luu, 2023; Plüss et al., 2023; Kuparinen et al., 2023 Elmadany et al., 2023; Deas et al., 2023; Khondaker et al., 2023; Faisal et al., 2024
Styles, Registers, Genres	Sweed and Shahaf, 2021; Sun and Xu, 2022; Nadejde et al., 2022 Srinivasan and Choi, 2022; Havaldar et al., 2023a
Relationship	Li et al., 2023b; Jurgens et al., 2023; Shaikh et al., 2023; Ziems et al., 2023a; Zhan et al., 2024
Context	Forbes et al., 2020; Emelin et al., 2021; Kim et al., 2022; Ziems et al., 2023a; Moghimifar et al., 2023 Rao et al., 2023; Sun et al., 2023; CH-Wang et al., 2023; Zhan et al., 2024
Communicative Goals	Emelin et al., 2021; Li et al., 2023b; Zhan et al., 2024
Demographics	Frenda et al., 2023; Lahoti et al., 2023; Ziems et al., 2023a; Casola et al., 2024; Lee et al., 2024b

Table 1: Recent resource work considered in §3 by elements (selection method in Appendix A). The three blocks (divided by double lines) correspond to ideational, linguistic, and social elements, respectively.

3.1 Ideational Elements

3.1.1 Concepts

We can divide concepts into (1) basic concepts that are ''configured'' differently, reflecting the cultural-specific way of thinking,[6] and (2) concepts that are unique to a culture (Wierzbicka, 1992).[7]

Recent NLP research has explored grounding time expressions across cultures (Shwartz, 2022) and culinary concepts in recipe adaptations (Cao et al., 2024b). Additionally, studies have examined how various cultures *use* concepts across categories, such as through metaphors (Kabra et al., 2023) or traditional proverbs and sayings (Liu et al., 2024a). In vision and language (VL) settings, culturally unique concepts have been integrated into reasoning and captioning tasks (Liu et al., 2021; Yin et al., 2021; Thapliyal et al., 2022; Li et al., 2024c) or assess multimodal content adaptations (Khanuja et al., 2024) and generation of text-to-image models (Liu et al., 2023c).

These datasets are often small due to high annotation costs, and most are only available for evaluation. Training and evaluation datasets still lack diversity across cultures, languages, and concept categories (e.g., rituals, aesthetics, spatial relations).

3.1.2 Knowledge

Cultural knowledge can be factual or common sense.[8] *What weather phenomena can be expected*

[6]For example, one can explore the citizen science project for lexicon associations: `https://smallworldofwords.org/en/project/home`.

[7]For example, ''Kopi Ga Dai'' in Singaporean English versus ''double-double'' in Canada, both referring to coffee with extra sweetness and creaminess, but very different.

[8]Common sense and norms are sometimes used interchangeably in NLP. Norms are acceptable behavioral patterns of a group (§2), which we will discuss in §3.1.4.

if a rapidly rotating tropical storm forms off the coast of our country? (It's likely called a hurricane if one is in the US, a typhoon if one is in Korea.) *Is tofu pudding sweet or salty by default?* (In China, it's typically sweet in the south but salty in the north.)

We identified three major types of resources in NLP literature: (1) probing (by masking entities), (2) multiple choice question answering (MCQA), and (3) knowledge bases.

Assessing language models' knowledge has long been important, with early studies examining this across languages predating LLMs (Kassner et al., 2021; Yin et al., 2022; Keleg and Magdy, 2023; Zhou et al., 2024). Recently, Massive Multitask Language Understanding (MMLU)-style (Hendrycks et al., 2021) MCQA benchmarks have advanced LLM development and inspired cultural variants (e.g., Li et al., 2024b, details in Table 1) covering aspects like food, history, and geography in respective languages. However, MMLU-style benchmarks, often based on standard exams and textbooks, lack integration with broader cultural elements. In contrast, other common sense knowledge datasets (e.g., Wibowo et al., 2024; Koto et al., 2024b) can incorporate other elements under ''linguistic'' (e.g., in *dialects*) or ''social'' (such as from diverse *demographics* with geographic regions) branches.

Finally, integrating knowledge bases (KB) with models enhances cultural awareness (Bhatia and Shwartz, 2023) and supports culturally relevant synthetic data generation (Kim et al., 2023). Despite recent efforts in creating cultural KB from other venues (Nguyen et al., 2023; Fung et al., 2024; Nguyen et al., 2024), *CL community examples remain limited.

3.1.3 Values

Diverse ranking of values among groups can result in differences in aboutness, communication styles, perceptions, and multiple other dimensions (Hofstede, 1984, 2011). Such differences in pre-training data can be reflected in LLMs.

Many recent studies on evaluation (Johnson et al., 2022; Ramezani and Xu, 2023; Cao et al., 2023; Durmus et al., 2024; Santurkar et al., 2023; Masoud et al., 2025; Havaldar et al., 2023b; Wang et al., 2024c, inter alia) show that LLMs align better with values of WEIRD (Western, Educated, Industrialized, Rich and Democratic; Henrich et al., 2010) people, raising concerns about the fairness and safety of LLMs for others. Here, Pew Global Attitudes Survey (PEW),[9] the World Values Survey (WVS),[10] and the Hofstede Cultural Dimensions (Hofstede, 1984, 2011) are commonly used for evaluation, along with regional variants like the European Values Survey (EVS; EVS, 2011). However, the questions of how to improve the model's value alignment with diverse cultures, what resources to collect and whom to collect from remain unsolved (Kirk et al., 2024).

Biases. In contrast to general cultural values, biases have been long-studied in NLP, such as gender bias in machine translation (Stanovsky et al., 2019; Savoldi et al., 2021; Campolungo et al., 2022; Sandoval et al., 2023; Attanasio et al., 2023, inter alia) or bias towards particular social groups. Differences in value ''ranking'' lead cultures to exhibit distinct biases toward the same groups or unique biases specific to certain cultures (e.g., caste systems, unnatural beauty standards). These variations are central to the study of *cultural biases* and are the focus of our work.

To enable evaluations of cross-cultural variations in biases and develop transferable de-biasing methods, recent work has created varieties of culturally aware datasets to aid evaluations, including targets and attribute word sets, sentence pairs, conversational and QA data (see Table 1 for the papers).

Overall, this area shows notable progress compared to other sub-areas. Recent surveys on general biases cover key topics like evaluation and de-biasing methods (Sun et al., 2019; Meade et al., 2022; Dev et al., 2022; Delobelle et al., 2022), which we refer readers to them for further details.

Hate. Like biases, perceptions of hatefulness in the text also vary across cultures, as shown in recent research on hate speech classification (Sap et al., 2022; Zhou et al., 2023a,b; Lee et al., 2023b; Lwowski et al., 2022; Arango Monnar et al., 2022). Such model disparities may be due to the data source (i.e., using machine translations, not native text) or the labeling process. The first issue can be addressed by diversifying data sources, incorporating authentic local data (Shekhar et al., 2022; Jeong et al., 2022). The second issue can

[9] https://www.pewresearch.org/.

[10] https://www.worldvaluessurvey.org/.

be mitigated by creating annotations from diverse cultural groups. Recently, CREHate (Lee et al., 2024b) investigates variations in hate speech perceptions within the same language, highlighting the need for further research.

Other Perceptions The perception of politeness, aesthetic appeal, or emotions can also vary across cultures (House and Kasper, 1981; Mesquita et al., 1997; Masuda et al., 2008; Ringel et al., 2019; Abdelkadir et al., 2024). For example, whether a piece of text is deemed humorous or ironic is culturally dependent. Frenda et al. (2023) and Casola et al. (2024) try to address this with cross-cultural annotated (multilingual) irony corpora. Similarly, visual elements in arts can elicit different emotions in different cultural groups. ArtELingo (Mohamed et al., 2022, 2024) provides benchmarks with multilingual captions and emotion labels for artworks to evaluate models' cultural-transfer performance. This research area is significantly limited.

3.1.4 Norms and Morals

In ethics, a distinction is made between descriptive and normative morality (Gert and Gert, 2025). In NLP, this distinction is often overlooked (Vida et al., 2023) with a greater emphasis on the ''end product'', which is the final set of rules or principles and their judgments.[11]

Several norm banks exist, built through automatic, semi-automatic, or manual methods using sources like conversations, social media, or government websites (Forbes et al., 2020; Fung et al., 2023; CH-Wang et al., 2023; Dwivedi et al., 2023). These norm banks have also been automatically adapted to defensible norms in fine-grained situations (Pyatkin et al., 2023; Rao et al., 2023) or inference tasks (CH-Wang et al., 2023; Huang and Yang, 2023) for LLM evaluation and adaptation.

For model alignment, several approaches focus on ''inquisition'', directly questioning LLMs about issues through conversation or a QA task (Kim et al., 2022; Sun et al., 2023; Yu et al., 2024; Lee et al., 2024a; Yuan et al., 2024). A challenge with this approach is that a model's responses do not always align with its behavior in usage (i.e., conversation). Thus, culturally aligned conversational data show greater potential for behavior adaptation (Li et al., 2023b; Zhan et al., 2024). However, existing resources have limited coverage beyond Western, Chinese, and Indian cultures.

[11] This is reasonable for standard NLP tasks but should be re-evaluated for high-stakes judgment-based applications.

3.1.5 Artifacts

NLP research on artifacts has focused on (monolingual or mono-cultural) artifacts in texts, e.g., fairy tales, fiction, poetry, and songs (Yang et al., 2019; Haider et al., 2020; Chakrabarty et al., 2021; Xu et al., 2022; Thai et al., 2022; Jiang et al., 2023; Ou et al., 2023; Li et al., 2023a), or in multimodal such as movies, humor, and memes (Sharma et al., 2020; Liu et al., 2022a; Hessel et al., 2023; Hong et al., 2023), to name a few. While ''artifacts'' is an independent cultural element, usage in *adaptation* typically involves tasks that align with one or more previously mentioned categories, making design challenging. For example, in ArtELingo (in §3.1.3), the input data focuses on art, while cross-cultural measurement studies *perceptions*, which reflect cultural values. Similarly, translations of literary novels need to account for *concept* differences such as names (Jiang et al., 2023) across cultures. Research on integrating cross-cultural differences into modeling and data acquisition with artifacts remains limited.

3.2 Linguistic Elements

3.2.1 Dialects

A dialect is a variant of a language (Haugen, 1966) at the local regional level (e.g., Hessian German), national level (e.g., Tunisian Arabic), or by other factors (e.g., AAVE).

Much existing work focuses on dialect identification (Salameh et al., 2018; Abdelali et al., 2021; Hämäläinen et al., 2021; Yusuf et al., 2022), but how to enable LLMs to serve dialectal communities remains an open question. Recently, multiple studies have identified disparities in NLP models (Ziems et al., 2022a; Le and Luu, 2023; Paonessa et al., 2023; Deas et al., 2023) when evaluated across different language variations.

Current dialect datasets primarily consist of translations between dialects and standard languages or are created through dialect normalization, in text, audio, or both (Plüss et al., 2023; Kuparinen et al., 2023). Few studies focus on traditional generation tasks like summarization or standard benchmark tasks (e.g., classifications or inferences; Maronikolakis et al., 2022; Held et al., 2023; Faisal et al., 2024). Overall, research on

German and English dialects is more advanced (marginally) than other dialect types.

3.2.2 Styles, Registers, and Genres

Styles, registers (e.g., slang), and genres (e.g., news) depend on the context of language use (Wardhaugh and Fuller, 2021). Compared to other elements, recent developments in this area appear limited, with a handful of examples focusing on slang, formality, or politeness (Sun and Xu, 2022; Nadejde et al., 2022; Srinivasan and Choi, 2022; Havaldar et al., 2023a).

3.3 Social Elements

3.3.1 Relationship

In many cultures, communication could differ depending on the relationship between the speakers. For example, Chinese has distinct terms for elder vs. younger siblings. Translations to (and from) a language without this property may result in a loss of nuances in meaning. In Korea and Japan, misused politeness level in conversation can violate cultural norms (Matsumoto, 1988; Ambady et al., 1996), especially in different social relationships. Additionally, certain relationships exist uniquely within specific cultures, such as ''Godmother''. Considering relationships is important for building resources and modeling culturally appropriate methods. Zhan et al. (2024) serve as a recent example with this consideration.

3.3.2 Context

In NLP, linguistic context could be the surrounding text. Studies by Hovy et al. (2020), Akinade et al. (2023), and Stewart and Mihalcea (2024) show that machine translation systems can fail without appropriate consideration of linguistic context, revealing its importance in resource and model development. However, human communication is much richer, relying on the extra-linguistic context that situates language within broader frames of reference.

The extra-linguistic context can be situational (setting or location where communication occurs; e.g., at school, in a hospital), historical (past events; e.g., colonization, that change cultural values or language use, like in Hong Kong) or non-verbal (e.g., hand gesture, tone of voice). Each type shapes and reflects culture. These contexts significantly enhance conversational tasks, norm bank development, and visual-language applications (Zhan et al., 2023; Ziems et al., 2023a), enabling NLP models to interpret nuanced language elements beyond words, thus improving response relevance and accuracy.

3.3.3 Communicative Goals

Different cultures can have distinctive communication styles depending on communicative goals. For example, people may use indirect language for refusal (versus direct refusal with a ''no'') to avoid confrontation (House, 2005). Cultures may also exhibit variations in responses to the same situation (e.g., how to make requests and when to apologize; Blum-Kulka and Olshtain, 1984). Taking this type of variation into account is important for cross-cultural pragmatic-inspired tasks—an area that remains understudied, with limited examples identified in Table 1.

3.3.4 Demographics

A household with a monthly income of less than 50 US dollars is likely to have different household items than that with 5000 US dollars (Rojas et al., 2022). Névéol et al. (2022) also found that the original English CrowS-Pair dataset relied on names as proxies for a sociodemographic group (''Amy for women, Tyrone for African American men''; Névéol et al., 2022), whereas the French version features direct references to sociodemographic groups. These data differences may stem not only from cultural influences but also from the demographics of the data contributors. Where and from whom one collects data matters, as it can result in dramatic differences in data and modeling.

Demographic information is also important in annotation (Sap et al., 2022; Pei and Jurgens, 2023; Santy et al., 2023), where a piece of text can be humorous to some people but offensive to others (Meaney, 2020). In such cases, culture may exist in the labels rather than in the data. Recently, Lee et al. (2024b) and Frenda et al. (2023) show how to capture different cultural views of annotators using the same dataset.

3.4 Usage of the Taxonomy

Covering the key elements of culture, our taxonomy can act as a useful reference point for NLP system development, in addition to organizing existing literature. For example, the development of culturally aware debiasing should consider *Ideational* elements such as *Values*, *Norms & Morals*,

as well as *Social* elements such as *Demographics* to inform the focus of debiasing, along with *Linguistic* elements such as dialects to inform the choice of data for the task. The applicable elements of culture will vary between tasks and contexts, with the taxonomy acting as a useful checklist. See Appendix B for additional examples.

4 Culturally Aware Resource Acquisition

Resources discussed in §3 are essential for culturally aware NLP. As additional resources are much needed, this section surveys methods for creating new resources (an overview in Appendix Figure 4).

Resources can be classified based on their acquisition methods—manual, automatic, or semi-automatic—and their source types: (1) newly created (**New**, from scratch), or (2) culturally adapted from existing resources (**CA**, e.g., through translation from the original data, followed by culturally appropriate changes). point (1) captures unique cultural phenomena but is often limited by funding or access to native speakers. Point (2) provides an alternative, though accurately reflecting cultural phenomena can be challenging.

4.1 Manual: Incorporating Native Speakers, Communities, and Experts

A common strategy is to employ native speakers or experts (e.g., professional translators or students) for data acquisition. This can be done via crowd-sourcing platforms such as Amazon Mechanical Turk and Prolific (Liu et al., 2021, 2024a) or in a community-driven manner, leveraging networks such as Masakhane,[12] IndoNLP,[13] university mailing lists, or Slack/Discord of organizations. Involving native speakers and communities to address cultural variations requires responsible design and thoughtful considerations.

New: Most existing culture resources have been built by involving native speakers or communities for dataset acquisition (Liu et al., 2021; Maronikolakis et al., 2022; Koto et al., 2023; Kabra et al., 2023). For non-language related communities, WinoQueer (Felkner et al., 2023) utilizes channels such as Slacks/Discord, and gay Twitter to reach the LGBTQ+ community and generates benchmarks based on community survey results.

CA: When starting from existing datasets, some works also involved communities (e.g., using surveys) in determining the needed modifications and supplements to datasets (Névéol et al., 2022; Hu et al., 2023; Majewska et al., 2023; Jin et al., 2024). These adaptations range from simple changes, such as updating names and locations to fit the target culture, to creating entirely new instances.

In general, native speakers are consulted throughout the life-cycle of new data acquisition (from annotations to quality checks). However, the entire community is rarely consulted during the initiation stage (i.e., designing tasks). Involving native speakers can be costly and difficult, but is a best practice that enhances quality and cultural authenticity.

4.2 Automatic: Models and Pipelines

Since manual adaptation is slow and hard to scale, the use of automation has gained popularity in resource acquisition.

New: For instance, CANDLE (Nguyen et al., 2023) proposes a pipeline to extract cultural commonsense knowledge using various techniques like NER extraction, cultural facet classification, concepts extraction and ranking through algorithms or LMs. NormsSAGE (Fung et al., 2023) utilizes LLMs for norm discovery from conversation data, then performs model self-verification to validate and filter the data. CultureAtlas (Fung et al., 2024) extracts cultural knowledge from Wikipedia and hyperlinked document pages using LLMs for filtering and adversarial knowledge generation.

Recent work has also used sociodemographic prompting (Santurkar et al., 2023; Deshpande et al., 2023; Hwang et al., 2023; Beck et al., 2024)—extending input prompts with sociodemographic information—to generate outputs tailored to specific groups. Further research could reduce data acquisition efforts, particularly for generating subcultural data variations within WEIRD people. However, it has also been argued that LLMs do not accurately mimic individual or group behaviors (Argyle et al., 2023; Aher et al., 2023; Beck et al., 2024).

CA: Putri et al. (2024) examine automatic adaptation (paraphrasing and concept replacement) of Commonsense QA in Indonesian and

[12] https://www.masakhane.io/.
[13] https://indonlp.github.io/.

Sundanese. Current GPT models, however, reveal disparities in cultural adaptation across languages, highlighting the need for further research.

4.3 Semi-automatic: Structured Resources, Model-in-the-loop

As demonstrated by Putri et al. (2024), LLMs struggle with fully automated cultural adaptations. Alternatively, semi-automatic approaches combine the quality of manual work with scalability.

New: Methods have been developed to generate seed data for iterative human cleaning and labelling. NormBank (Ziems et al., 2023a) uses LLMs to generate seed roles and behaviors as norm candidates in specific situations, which are then annotated by humans. Similarly, other studies (CH-Wang et al., 2023; Liu et al., 2024a; Bhutani et al., 2024) employ prompting techniques to generate seed data, followed by human annotation on tasks like cultural bias and social reasoning.

CA: Ziems et al. (2023b, Multi-Value) introduced a framework that leverages the Electronic World Atlas of Varieties of English (Kortmann et al., 2020, eWAVE) to create and adapt datasets covering 50 English dialects. This framework enabled the adaptation of a standard corpus into dialectal forms (Held et al., 2023; Xiao et al., 2023). However, similar structured resources may not exist or be suitable for adaptation of other cultural elements (e.g., for concepts, consistently replacing 'bread' with 'rice' would not be desirable).

5 Creating Culturally Adapted Models

Most culturally aware NLP research has focused on resource creation and evaluation, with culturally adapted model development still emerging. Here, we review current methods for adapting pre-trained (L)LMs, covering in-context and in-weight adaptations (an overview can be found in Appendix Figure 5). We found that current cultural adaptation methods in NLP prioritize technical advancements and isolated cultural elements, measuring effectiveness solely by standard task performance.

5.1 In-context Adaptation

The success of LLMs allows for behavior tuning by prompts or in-context examples. A straightforward strategy is to provide the model with sociodemographic prompts or use ''role-playing'' (Park et al., 2022; Argyle et al., 2023) of a culture, as seen in Shaikh et al. (2023) and Hwang et al. (2023). For knowledge-intensive tasks, cultural knowledge can be added directly to the prompt, and LLMs can leverage indirect descriptions from external sources or prior model outputs (Yao et al., 2024a). Lastly, high-level prompts (or ''constitutions'', Bai et al., 2022b) guiding LLM reasoning could improve cultural alignment alongside demographic-based prompts (AlKhamissi et al., 2024).

Since different cultures reflect different values, there is a need to create models that embody pluralistic cultural values with flexible alignment capabilities (Sorensen et al., 2024). Feng et al. (2024) propose a framework to achieve this by enhancing pluralistic alignment in LLMs via collaboration between a high-level LLM and a group of specialized community LMs (i.e., an ensemble of LLMs). This framework enables general-purpose LLMs to flexibly incorporate diverse cultural and ideological perspectives, reflecting both individual preferences and broader cultural distributions.

A retrieval-augmented approach can further refine cultural alignment by adjusting responses dynamically. Friedrich et al. (2023) propose such a method for moral reasoning, where culture-specific contexts are stored in a retrieval engine. When asked moral questions, relevant contexts are retrieved and added to the input, enabling the model to respond with cultural nuances. This method shows promise for adapting LLMs to evolving cultural information, an aspect often overlooked in current adaptation methods.

5.2 In-weight Adaptation

5.2.1 Data Augmentation

Acquiring large corpora for supervised cultural adaptation is challenging. Data augmentation helps address this, enhancing model robustness. Li and Zhang (2023) present a data augmentation method for multilingual multicultural VL reasoning tasks, generating code-mixed data by substituting English concepts with culturally mapped equivalents. The cultural concept sets (for mapping) are built by querying hyponyms, synonyms, and hypernyms in the ConceptNet (Speer et al., 2017) and WordNet (Miller, 1992). However, the optimal resource depends on the specific cultural element being adapted (§3.1). For instance,

a cultural knowledge base might be better for norms adaptations.

5.2.2 Continual Pre-training, Auxiliary Losses

Continual pre-training (CPT, including instruction tuning), intermediate task training, and multi-task training with auxiliary losses are methods for cultural adaptation. CPT fine-tunes a pre-trained LM with an unlabeled domain or language corpus before downstream task fine-tuning. It improves downstream task performance via full-parameter training (Xu et al., 2019; Han and Eisenstein, 2019; Gururangan et al., 2020) or by training a few additional parameters while keeping the model frozen (Wang et al., 2021; Ke et al., 2022).

Recently, Hofmann et al. (2024a) show that when combined with a geo-location prediction loss, CPT can help to increase the awareness of dialectal variations of pre-trained LMs. Wang et al. (2024a) show that instruction tuning with instructions containing cultural knowledge can improve models' ability in cultural knowledge reasoning. In VL, Bhatia and Shwartz (2023) use a cultural commonsense knowledge graph from Nguyen et al. (2023) for CPT to develop a geo-diverse LM for commonsense reasoning tasks. This method category is effective for addressing diverse cultural elements, but adapting pre-trained LLMs can result in catastrophic forgetting (McCloskey and Cohen, 1989, or termed ''alignment tax'' due to Reinforcement Learning from Human Feedback (RLHF) tuning, Askell et al., 2021; Ouyang et al., 2022) potentially worsening their performance on general tasks. This warrants further investigation.

5.2.3 Other Forms of Information Integration

Cao et al. (2024a) propose a method that integrates cultural dimension vectors (derived through a regression task based on Hofstede Culture Dimensions, Hofstede, 1984) with a mT5 Transformer model (Xue et al., 2021). These cultural dimension vectors are added to the hidden states at each layer to enable culturally informed multi-turn dialogue classification and prediction.

5.2.4 Parameter-efficient Adaptations

As LMs grow larger, parameter-efficient fine-tuning methods (i.e., PEFT, by fine-tuning a small number of parameters, such as the bottle-neck adapters, Houlsby et al., 2019; LoRA, Hu et al., 2022, etc.) become increasingly important for task adaptations. Given their success in cross-lingual transfer learning (Üstün et al., 2020; Pfeiffer et al., 2020; Ansell et al., 2021; Liu et al., 2023a, 2024b, among others), PEFT can be a natural choice for cultural adaptation of, e.g., dialects.

Recently, HyperLoRA (Xiao et al., 2023) uses the Hypernetwork (Ha et al., 2017, a neural network for generating parameters) to generate LoRA adapters based on dialectal features. DADA (Liu et al., 2023b) proposes to train a pool of dialectal linguistic feature adapters and dynamically compose the adapters for dialectal tasks. Being task-agnostic, PEFT methods could prove important for cultural adaptations beyond dialects.

5.2.5 Outlook: Feedback Learning

The success of LLMs has popularized RLHF (Christiano et al., 2017; Bai et al., 2022a; Ouyang et al., 2022; Ivison et al., 2024) and Direct Preference Optimization (DPO; Rafailov et al., 2023; Ivison et al., 2023) methods. RLHF fine-tunes LMs with feedback by fitting a reward model with human preferences, and then training a reinforcement learning-based policy to maximize the learned reward. DPO avoids RL training by using a simpler supervised learning objective for an implicit reward model.

Recent work shows that RLHF can enhance the performance of multilingual instruction tuning for LLMs (Lai et al., 2023), while DPO can improve the multilingual reasoning abilities (She et al., 2024) and multilingual safety (Aakanksha et al., 2024) of LLMs. The use of RLHF or DPO for multilingual multicultural adaptation is still limited, but these examples suggest that the direction could be promising.

6 Further Discussions and Recommendations

As we have seen, significant work remains to be done on both resources and methods for various elements of culture.

An area that requires attention is the overall process of researching culturally aware NLP. As mentioned previously, a key practice is community involvement (§4) to get the process right (Bird, 2020; Liu et al., 2022b; Mager et al., 2023). It is crucial to assess how target communities can benefit most from technologies. For instance, many

dialects are primarily oral, and speech-to-speech or speech-to-text translations could be preferable over text-based applications (Blaschke et al., 2024). Furthermore, ethical data collection practices are also critical and technology ownership must be considered, especially when indigenous and marginalized communities are involved. For best practices, we refer the readers to work such as Bird (2020), Smith (2021), or Cooper et al. (2024) for further details.

Another key consideration is integrating insights from fields beyond NLP. Cultural adaptation has long been practiced in areas like video games (O'hagan and Mangiron, 2013), movies (Pettit, 2009), online learning (Blanchard et al., 2005), and clinical psychology (Bernal et al., 1995; Barrera Jr. and Castro, 2006). These existing practices can serve as a foundation for adapting NLP applications to meet the needs of diverse cultural contexts.

Here, we summarize and recommend best practices based on our prior discussions and the publications surveyed in the sections referenced below:

Resource Acquisition

- §4.1, §6: Consult with target cultural groups throughout design and implementation, wherever possible.
- §4.3: Use iterative feedback from culture experts to refine data quality. Automatically acquired resources should also undergo expert quality checks.
- §6: Ensure an ethical approach to data acquisition and discuss data ownership early to prevent misuse. This is always important but particularly critical with indigenous and marginalized communities.

Model Adaptation

- §5: Incorporate new metrics that assess cultural awareness alongside task performance.
- §5: Consider cultural adaptation as an ongoing, systematic process rather than a one-time task focused on a single element.
- §5.1: Monitor adaptation performance over time, especially for the evolving cultural elements, to maintain model relevance.
- §6: Build on existing knowledge outside of NLP when applicable.

7 Summary and Future Research Directions

Culturally aware and adapted NLP has recently emerged as an important and active research area. Significant progress has been made in the development of resources for capturing various elements of culture, but the development of NLP methods is still in its infancy. We will now summarize the main research gaps identified in this survey with respect to the categories of our new taxonomy (§2):

Resources. Currently, resources exist for all elements of culture, with considerable progress made on *values* (§3.1.3, particularly in biases) and *knowledge* (§3.1.2, particularly for MMLU-style cultural knowledge benchmarks). However, research is lacking in the following areas:

Gaps in Elements Coverage: While many resources already exist within *concepts* (§3.1.1), multilingual data resources covering a diverse set of concepts (e.g., aesthetics, spatial relation) in both unimodal and multimodal (§3.1.1) for generation tasks is lacking. Moreover, most recent developments in *norms & morals* are predominantly in English, reflecting a monocultural perspective. This highlights the need for more multilingual and multicultural resources. Additionally, there is a significant gap in datasets that focus on different types of value perceptions (such as emotion and irony, §3.1.3), stylistic variations (§3.2), and artifacts (§3.1.5) across various cultural groups, both in different languages and within languages.

Resources considering social elements of culture (§3.3) also remain limited. For example, collecting speaker relationships in dialogue datasets or distinguishing age groups in social norms datasets. These are needed to address the intricate relationship between culture and people in NLP.

Training Data and ''CultureGLUE'': Most existing resources focus on evaluation, providing benchmarks and test sets that enable researchers to assess the performance of models. While these evaluation resources are crucial for cultural adaptation, there is a pressing need for training data. Further, current evaluation resources often focus

on individual elements of culture. A unified, cultural benchmark like GLUE (Wang et al., 2019) does not yet exist for all cultural elements across diverse groups. Developing a multicultural ''CultureGLUE'' may be challenging at the moment, but a reasonable first step is to focus on individual cultures, ensuring a diverse range of tasks and comprehensive element coverage.

Modeling. While modeling methods for culture are generally under-explored, continual pretraining (§5.2.2) and prompting (§5.1) have received marginally more attention than other approaches. Research areas needing further exploration include:

PEFT-based Transfer Learning: Exploration of PEFT-based transfer learning techniques beyond dialects is limited (§5.2.4). Given their success in other NLP areas, these techniques warrant further investigation into other elements of culture, such as for *values* or *norms & morals*. A potential approach involves using WVS survey data, similar to Li et al. (2024a), to train PEFT-based modules focused on values. However, it is crucial to investigate whether survey data alone is sufficient for effective training.

Feedback Learning and Other LLM Specialties: Leveraging the success of LLMs and feedback learning presents promising new avenues for cultural adaptation (§5.2.5). A potential bottleneck is acquiring large, culturally diverse preference datasets for model adaptation and training culturally aligned reward models. This could be addressed by large-scale data collection efforts, as demonstrated by Kirk et al. (2024), or through the generation of synthetic data (Aakanksha et al., 2024). For synthetic data, using techniques such as role-playing and the creation of repositories of cultural personas could facilitate culturally sensitive model training.

Evolving Culture: Culture evolves gradually (Boyd and Richerson, 1988; Whiten et al., 2011), yet there have been few discussions on how to model and adapt to evolving culture. Future research should focus on methods that address the dynamic nature of culture. One potential approach is the use of retrieval-augmented systems to integrate evolving information (§5.1), which ensures models' relevance to cultural shifts over time.

Overall. Below, we discuss two overall research gaps.

Adaptation in the Social Context: As a key motivation of this paper, culture emerges from and is shaped by social interactions among humans within a society (§1). However, an important question remains unanswered in the existing literature: *Should the cultural adaptation of models occur within a situated social context and structure?* Exploring this could present new avenues for interdisciplinary research (e.g., with human-machine collaboration, social psychology, or anthropology).

''Surface'' versus ''Deep'' Adaptation in NLP: Resnicow et al. (1999) devise cultural adaptations for public health research into surface and deep adaptations, where the former considers familiar languages and concepts to the target groups, and the latter considers social and historical factors that influence the behaviors of the target groups.

In NLP, surface adaptations might include using the same language as a culture and recognizing explicit cultural differences (e.g., asking LLMs ''what is the meaning of ...''). In contrast, deep adaptations might enable a model to ''behave'' (e.g., make decisions, pragmatically comply, etc.) like a member of a culture without explicit inquisition (see Figure 3 in the Appendix for an illustration).

As we have seen from prior work described in §3 or §5, only a few current studies focus on adapting the behavioral aspect of models (which is becoming increasingly important with LLMs), and there has been no work to date on measuring the depth and progress of cultural adaptations or when a model is *fully culturally aware and culturally competent*. Further research could explore these areas.

8 Conclusions

This work proposes a new extensive taxonomy of culture that expands on earlier works in NLP and is grounded in well-established anthropology and social sciences literature. The taxonomy provides a systematic framework for understanding and tracking progress in the emerging area of culturally aware and adapted NLP. However, our taxonomy is not without its limitations. Future research could refine the taxonomy in areas like *values* or *communicative goals* by adding further

subcategories and providing a better understanding of interactions between the elements of culture (e.g., shifting values' impact on social norms over time).

We survey existing resources and methods in this area according to the taxonomy classes, identifying areas of strength as well as areas where research remains to be done. Our paper summarizes the state of the art and provides ideas for future research in this exciting and important area.

Acknowledgments

This work was funded by the German Federal Ministry of Education and Research (BMBF) under the promotional reference 13N15897 (MISRIK). It was also funded by EPSRC grant EP/Y031350/1 under the UK government's funding guarantee for ERC Advanced Grants. Chen Cecilia Liu is supported by the Konrad Zuse School of Excellence in Learning and Intelligent Systems (ELIZA) through the DAAD program Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research.

The authors would like to thank Anjali Kantharuban, Shun Shao, and Anne Lauscher for an early discussion of different aspects of this work. The authors would also like to thank Ji-Ung Lee, anonymous reviewers, and action editors of TACL for feedback on a draft of this paper.

References

Aakanksha, Arash Ahmadian, Beyza Ermis, Seraphina Goldfarb-Tarrant, Julia Kreutzer, Marzieh Fadaee, and Sara Hooker. 2024. The multilingual alignment prism: Aligning global and local preferences to reduce harm. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 12027–12049, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.671

Ahmed Abdelali, Hamdy Mubarak, Younes Samih, Sabit Hassan, and Kareem Darwish. 2021. QADI: Arabic dialect identification in the wild. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 1–10, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Nuredin Ali Abdelkadir, Charles Zhang, Ned Mayo, and Stevie Chancellor. 2024. Diverse perspectives, divergent models: Cross-cultural evaluation of depression detection on Twitter. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 672–680, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-short.58

Christabel Acquaye, Haozhe An, and Rachel Rudinger. 2024. Susu box or piggy bank: Assessing cultural commonsense knowledge between Ghana and the US. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 9483–9502, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.532

Muhammad Adilazuarda, Sagnik Mukherjee, Pradhyumna Lavania, Siddhant Singh, Alham 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. https://doi.org/10.18653/v1/2024.emnlp-main.882

Gati V. Aher, Rosa I. Arriaga, and Adam Tauman Kalai. 2023. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning, ICML 2023, 23–29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 337–371. PMLR.

Idris Akinade, Jesujoba Alabi, David Adelani, Clement Odoje, and Dietrich Klakow. 2023. Varepsilon kú mask: Integrating Yorùbá cultural greetings into machine translation. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 1–7, Dubrovnik, Croatia. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.c3nlp-1.1

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. https://doi.org/10.18653/v1/2024.acl-long.671

Nalini Ambady, Jasook Koo, Fiona Lee, and Robert Rosenthal. 1996. More than words: Linguistic and nonlinguistic politeness in two cultures. *Journal of Personality and Social Psychology*, 70(5):996. https://doi.org/10.1037//0022-3514.70.5.996

Haozhe An, Zongxia Li, Jieyu Zhao, and Rachel Rudinger. 2023. SODAPOP: Open-ended discovery of social biases in social commonsense reasoning models. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1573–1596, Dubrovnik, Croatia. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.eacl-main.116

Alan Ansell, Edoardo Maria Ponti, Jonas Pfeiffer, Sebastian Ruder, Goran Glavaš, Ivan Vulić, and Anna Korhonen. 2021. MAD-G: Multilingual adapter generation for efficient cross-lingual transfer. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4762–4781, Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-emnlp.410

Ayme Arango Monnar, Jorge Perez, Barbara Poblete, Magdalena Saldaña, and Valentina Proust. 2022. Resources for multilingual hate speech detection. In *Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH)*, pages 122–130, Seattle, Washington (Hybrid). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.woah-1.12

Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351. https://doi.org/10.1017/pan.2023.2

Amanda Askell, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Benjamin Mann, Nova DasSarma, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Jackson Kernion, Kamal Ndousse, Catherine Olsson, Dario Amodei, Tom B. Brown, Jack Clark, Sam McCandlish, Chris Olah, and Jared Kaplan. 2021. A general language assistant as a laboratory for alignment. *ArXiv preprint arXiv:2112.00861v3*.

Giuseppe Attanasio, Flor Miriam Plaza del Arco, Debora Nozza, and Anne Lauscher. 2023. A tale of pronouns: Interpretability informs gender bias mitigation for fairer instruction-tuned machine translation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3996–4014, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.243

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario Amodei, Tom B. Brown, Jack Clark, Sam McCandlish, Chris Olah, Benjamin Mann, and Jared Kaplan. 2022a. Training a helpful and harmless assistant with reinforcement learning from human feedback. *ArXiv preprint arXiv:2204.05862v1*.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosiute, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemí Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei,

Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. 2022b. Constitutional AI: harmlessness from AI feedback. *ArXiv preprint arXiv:2212.08073v1*.

Manuel Barrera Jr. and Felipe González Castro. 2006. A heuristic framework for the cultural adaptation of interventions. *Clinical Psychology: Science and Practice*, 13(4):311–316. https://doi.org/10.1111/j.1468-2850.2006.00043.x

Lisa Bauer, Hanna Tischer, and Mohit Bansal. 2023. Social commonsense for explanation and cultural bias discovery. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3745–3760, Dubrovnik, Croatia. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.eacl-main.271

Tilman Beck, Hendrik Schuff, Anne Lauscher, and Iryna Gurevych. 2024. Sensitivity, performance, robustness: Deconstructing the effect of sociodemographic prompting. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2589–2615, St. Julian's, Malta. Association for Computational Linguistics.

Guillermo Bernal, Janet Bonilla, and Carmen Bellido. 1995. Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with hispanics. *Journal of Abnormal Child Psychology*, 23:67–82. https://doi.org/10.1007/BF01447045, PubMed: 7759675

Mehar Bhatia, Sahithya Ravi, Aditya Chinchure, EunJeong Hwang, and Vered Shwartz. 2024. From local concepts to universals: Evaluating the multicultural understanding of vision-language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6763–6782, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.385

Mehar Bhatia and Vered Shwartz. 2023. GD-COMET: A geo-diverse commonsense inference model. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7993–8001, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.496

Shaily Bhatt and Fernando Diaz. 2024. Extrinsic evaluation of cultural competence in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 16055–16074, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.942

Mukul Bhutani, Kevin Robinson, Vinodkumar Prabhakaran, Shachi Dave, and Sunipa Dev. 2024. SeeGULL multilingual: A dataset of geo-culturally situated stereotypes. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 842–854, Bangkok, Thailand. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-short.75

Cristina Bicchieri, Ryan Muldoon, and Alessandro Sontuoso. 2018. Social norms. *The Stanford Encyclopedia of Philosophy*.

Steven Bird. 2020. Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics. https://doi.org/10.18653/v1/2020.coling-main.313

Janet Blake. 2000. On defining the cultural heritage. *International and Comparative Law Quarterly*, 49(1):61–85. https://doi.org/10.1017/S002058930006396X

Emmanuel Blanchard, Ryad Razaki, and Claude Frasson. 2005. Cross-cultural adaptation of e-learning contents: A methodology. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pages 1895–1902. Association for the Advancement of Computing in Education (AACE).

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. https://doi.org/10.18653/v1/2024.acl-short.74

Shoshana Blum-Kulka and Elite Olshtain. 1984. Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, 5(3):196–213. https://doi.org/10.1093/applin/5.3.196

Robert Boyd and Peter J. Richerson. 1988. *Culture and the evolutionary process*. University of Chicago Press.

Adam Briggle and Carl Mitcham. 2012a. *Ethics and science: An introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139034111

Adam Briggle and Carl Mitcham. 2012b. *Science and ideational culture*, Cambridge Applied Ethics, pages 268–289. Cambridge University Press. https://doi.org/10.1017/CBO9781139034111.012

Niccolò Campolungo, Federico Martelli, Francesco Saina, and Roberto Navigli. 2022. DiBiMT: A novel benchmark for measuring Word Sense Disambiguation biases in Machine Translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4331–4352, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.298

Yong Cao, Min Chen, and Daniel Hershcovich. 2024a. Bridging cultural nuances in dialogue agents through cultural value surveys. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 929–945, St. Julian's, Malta. Association for Computational Linguistics.

Yong Cao, Yova Kementchedjhieva, Ruixiang Cui, Antonia Karamolegkou, Li Zhou, Megan Dare, Lucia Donatelli, and Daniel Hershcovich. 2024b. Cultural adaptation of recipes. *Transactions of the Association for Computational Linguistics*, 12:80–99. https://doi.org/10.1162/tacl_a_00634

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. https://doi.org/10.18653/v1/2023.c3nlp-1.7

Silvia Casola, Simona Frenda, Soda Lo, Erhan Sezerer, Antonio Uva, Valerio Basile, Cristina Bosco, Alessandro Pedrani, Chiara Rubagotti, Viviana Patti, and Davide Bernardi. 2024. MultiPICo: Multilingual perspectivist irony corpus. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16008–16021, Bangkok, Thailand. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.849

Sky CH-Wang, Arkadiy Saakyan, Oliver Li, Zhou Yu, and Smaranda Muresan. 2023. Sociocultural norm similarities and differences via situational alignment and explainable textual entailment. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3548–3564, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.215

Tuhin Chakrabarty, Arkadiy Saakyan, and Smaranda Muresan. 2021. Don't go far off: An empirical study on neural poetry translation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7253–7265, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.577

Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4–9, 2017, Long Beach, CA, USA*, pages 4299–4307.

Robert B. Cialdini, Carl A. Kallgren, and Raymond R. Reno. 1991. A focus theory of normative conduct: A theoretical refinement

and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology*, volume 24, pages 201–234. Elsevier. https://doi.org/10.1016/S0065-2601(08)60330-5

Ned Cooper, Courtney Heldreth, and Ben Hutchinson. 2024. ''it's how you do things that matters'': Attending to process to better serve indigenous communities with language technologies. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 204–211, St. Julian's, Malta. Association for Computational Linguistics.

Nicholas Deas, Jessica Grieser, Shana Kleiner, Desmond Patton, Elsbeth Turcan, and Kathleen McKeown. 2023. Evaluation of African American language bias in natural language generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6805–6824, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.421

Nicholas Deas, Elsbeth Turcan, Ivan Mejia, and Kathleen McKeown. 2024. MASIVE: Open-ended affective state identification in English and Spanish. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20467–20485, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.1139

Pieter Delobelle, Ewoenam Tokpo, Toon Calders, and Bettina Berendt. 2022. Measuring fairness with biased rulers: A comparative study on bias metrics for pre-trained language models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1693–1706, Seattle, United States. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.122

Ameet Deshpande, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan. 2023. Toxicity in chatgpt: Analyzing persona-assigned language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1236–1270, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.88

Sunipa Dev, Emily Sheng, Jieyu Zhao, Aubrie Amstutz, Jiao Sun, Yu Hou, Mattie Sanseverino, Jiin Kim, Akihiro Nishi, Nanyun Peng, and Kai-Wei Chang. 2022. On measures of biases and harms in NLP. In *Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022*, pages 246–267, Online only. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-aacl.24

Esin Durmus, Karina Nyugen, Thomas I. Liao, Nicholas Schiefer, Amanda Askell, Anton Bakhtin, Carol Chen, Zac Hatfield-Dodds, Danny Hernandez, Nicholas Joseph, Liane Lovitt, Sam McCandlish, Orowa Sikder, Alex Tamkin, Janel Thamkul, Jared Kaplan, Jack Clark, and Deep Ganguli. 2024. Towards measuring the representation of subjective global opinions in language models. In *First Conference on Language Modeling*.

Ashutosh Dwivedi, Pradhyumna Lavania, and Ashutosh Modi. 2023. EtiCor: Corpus for analyzing LLMs for etiquettes. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6921–6931, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.428

AbdelRahim Elmadany, ElMoatez Billah Nagoudi, and Muhammad Abdul-Mageed. 2023. ORCA: A challenging benchmark for Arabic language understanding. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9559–9586, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.609

Denis Emelin, Ronan Le Bras, Jena D. Hwang, Maxwell Forbes, and Yejin Choi. 2021. Moral stories: Situated reasoning about norms, intents, actions, and their consequences. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 698–718, Online and Punta Cana,

Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.54

Elena V. Epure, Guillaume Salha, Manuel Moussallam, and Romain Hennequin. 2020. Modeling the music genre perception across language-bound cultures. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4765–4779, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.386

Cristina España-Bonet and Alberto Barrón-Cedeño. 2022. The (undesired) attenuation of human biases by multilinguality. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2056–2077, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.133

EVS. 2011. EVS - European Values Study 1981 - integrated dataset. GESIS Datenarchiv, Köln. ZA4438 Datenfile Version 3.0.0. https://doi.org/10.4232/1.10791

Fahim Faisal, Orevaoghene Ahia, Aarohi Srivastava, Kabir Ahuja, David Chiang, Yulia Tsvetkov, and Antonios Anastasopoulos. 2024. DIALECTBENCH: An NLP benchmark for dialects, varieties, and closely-related languages. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14412–14454, Bangkok, Thailand. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.777

Virginia Felkner, Ho-Chun Herbert Chang, Eugene Jang, and Jonathan May. 2023. WinoQueer: A community-in-the-loop benchmark for anti-LGBTQ+ bias in large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9126–9140, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.507

Shangbin Feng, Taylor Sorensen, Yuhan Liu, Jillian Fisher, Chan Young Park, Yejin Choi, and Yulia Tsvetkov. 2024. Modular pluralism: Pluralistic alignment via multi-LLM collaboration. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4151–4171, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.240

Maxwell Forbes, Jena D. Hwang, Vered Shwartz, Maarten Sap, and Yejin Choi. 2020. Social chemistry 101: Learning to reason about social and moral norms. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 653–670, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.48

Simona Frenda, Alessandro Pedrani, Valerio Basile, Soda Marem Lo, Alessandra Teresa Cignarella, Raffaella Panizzon, Cristina Marco, Bianca Scarlini, Viviana Patti, Cristina Bosco, and Davide Bernardi. 2023. EPIC: Multi-perspective annotation of a corpus of irony. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13844–13857, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.774

Felix Friedrich, Wolfgang Stammer, Patrick Schramowski, and Kristian Kersting. 2023. Revision transformers: Instructing language models to change their values. In *ECAI 2023 - 26th European Conference on Artificial Intelligence, September 30–October 4, 2023, Kraków, Poland - Including 12th Conference on Prestigious Applications of Intelligent Systems (PAIS 2023)*, volume 372 of *Frontiers in Artificial Intelligence and Applications*, pages 756–763. IOS Press. https://doi.org/10.3233/FAIA230341

V. A. Fromkin, Robert Rodman, and V. Hyams. 1998. *An Introduction to Language 6e*. Orlando, FL: Hartcourt Brace College Publishers.

Yi Fung, Tuhin Chakrabarty, Hao Guo, Owen Rambow, Smaranda Muresan, and Heng Ji. 2023. NORMSAGE: Multi-lingual multi-cultural norm discovery from conversations on-the-fly. In *Proceedings of the 2023*

Conference on Empirical Methods in Natural Language Processing, pages 15217–15230, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.941

Yi Fung, Ruining Zhao, Jae Doo, Chenkai Sun, and Heng Ji. 2024. Massively multi-cultural knowledge acquisition & LM benchmarking. *ArXiv preprint arXiv:2402.09369v1*.

Clifford Geertz. 1973. *The interpretation of cultures*. Basic books.

Joshua Gert and Bernard Gert. 2025. The Definition of Morality. Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Erving Goffman. 2023. The presentation of self in everyday life, *Social theory re-wired*. Routledge, pages 450–459. https://doi.org/10.4324/9781003320609-59

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.740

David Ha, Andrew M. Dai, and Quoc V. Le. 2017. Hypernetworks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings*. OpenReview.net.

Thomas Haider, Steffen Eger, Evgeny Kim, Roman Klinger, and Winfried Menninghaus. 2020. PO-EMO: Conceptualization, annotation, and modeling of aesthetic emotions in German and English poetry. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1652–1663, Marseille, France. European Language Resources Association.

Mika Hämäläinen, Khalid Alnajjar, Niko Partanen, and Jack Rueter. 2021. Finnish dialect identification: The effect of audio and text. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8777–8783, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.692

Xiaochuang Han and Jacob Eisenstein. 2019. Unsupervised domain adaptation of contextualized embeddings for sequence labeling. 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 4238–4248, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1433

Einar Haugen. 1966. Dialect, language, nation 1. *American Anthropologist*, 68(4):922–935. https://doi.org/10.1525/aa.1966.68.4.02a00040

Shreya Havaldar, Salvatore Giorgi, Sunny Rai, Thomas Talhelm, Sharath Chandra Guntuku, and Lyle Ungar. 2024. Building knowledge-guided lexica to model cultural variation. 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 211–226, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.12

Shreya Havaldar, Matthew Pressimone, Eric Wong, and Lyle Ungar. 2023a. Comparing styles across languages. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6775–6791, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.419

Shreya Havaldar, Bhumika Singhal, Sunny Rai, Langchen Liu, Sharath Chandra Guntuku, and Lyle Ungar. 2023b. Multilingual language models are not multicultural: A case study in emotion. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 202–214, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wassa-1.19

Michael Hechter and Karl-Dieter Opp. 2001. *Social Norms*. Russell Sage Foundation.

William Held, Caleb Ziems, and Diyi Yang. 2023. TADA: Task agnostic dialect adapters for English. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 813–824, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.51

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. Measuring massive multitask language understanding. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3–7, 2021*. OpenReview.net.

Joseph Henrich, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3):61–83. https://doi.org/10.1017/S0140525X0999152X, PubMed: 20550733

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. 2022. Challenges and strategies in cross-cultural NLP. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6997–7013, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.482

Jack Hessel, Ana Marasovic, Jena D. Hwang, Lillian Lee, Jeff Da, Rowan Zellers, Robert Mankoff, and Yejin Choi. 2023. Do androids laugh at electric sheep? Humor ''understanding'' benchmarks from the new yorker caption contest. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 688–714, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.41

David Hobson, Haiqi Zhou, Derek Ruths, and Andrew Piper. 2024. Story morals: Surfacing value-driven narrative schemas using large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 12998–13032, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.723

Valentin Hofmann, Goran Glavaš, Nikola Ljubešić, Janet B. Pierrehumbert, and Hinrich Schütze. 2024a. Geographic adaptation of pretrained language models. *Transactions of the Association for Computational Linguistics*, 12:411–431. https://doi.org/10.1162/tacl_a_00652

Valentin Hofmann, Pratyusha Ria Kalluri, Dan Jurafsky, and Sharese King. 2024b. AI generates covertly racist decisions about people based on their dialect. *Nature*, 633(8028):147–154. https://doi.org/10.1038/S41586-024-07856-5, PubMed: 39198640

G. Hofstede. 1984. *Culture's Consequences: International Differences in Work-Related Values*. Cross Cultural Research and Methodology. SAGE Publications.

Geert Hofstede. 2011. Dimensionalizing cultures: The hofstede model in context. *Online Readings in Psychology and Culture*, 2(1):8. https://doi.org/10.9707/2307-0919.1014

Xudong Hong, Asad Sayeed, Khushboo Mehra, Vera Demberg, and Bernt Schiele. 2023. Visual writing prompts: Character-grounded story generation with curated image sequences. *Transactions of the Association for Computational Linguistics*, 11:565–581. https://doi.org/10.1162/tacl_a_00553

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9–15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 2790–2799. PMLR.

Juliane House. 2005. *Politeness in Germany: Politeness in GERMANY?* Bristol, Blue Ridge

Summit. Multilingual Matters. https://doi.org/10.21832/9781853597398-003

Juliane House and Gabriele Kasper. 1981. *Politeness Markers in English and German*, pages 157–186. De Gruyter Mouton, Berlin, New York. https://doi.org/10.1515/9783110809145.157

Dirk Hovy, Federico Bianchi, and Tommaso Fornaciari. 2020. ''You sound just like your father'' commercial machine translation systems include stylistic biases. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1686–1690, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.154

Dirk Hovy and Diyi Yang. 2021. The importance of modeling social factors of language: Theory and practice. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 588–602, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.49

Hsin-Yi Hsieh, Shih-Cheng Huang, and Richard Tsai. 2024. TWBias: A benchmark for assessing social bias in traditional Chinese large language models through a Taiwan cultural lens. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 8688–8704, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.507

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25–29, 2022*. OpenReview.net.

Songbo Hu, Han Zhou, Mete Hergul, Milan Gritta, Guchun Zhang, Ignacio Iacobacci, Ivan Vulić, and Anna Korhonen. 2023. Multi 3 WOZ: A multilingual, multi-domain, multi-parallel dataset for training and evaluating culturally adapted task-oriented dialog systems. *Transactions of the Association for Computational Linguistics*, 11:1396–1415. https://doi.org/10.1162/tacl_a_00609

Tianyi Hu, Maria Maistro, and Daniel Hershcovich. 2024. Bridging cultures in the kitchen: A framework and benchmark for cross-cultural recipe retrieval. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1068–1080, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.61

Jing Huang and Diyi Yang. 2023. Culturally aware natural language inference. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7591–7609, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.509

EunJeong Hwang, Bodhisattwa Majumder, and Niket Tandon. 2023. Aligning language models to user opinions. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 5906–5919, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.393

Hamish Ivison, Yizhong Wang, Jiacheng Liu, Zeqiu Wu, Valentina Pyatkin, Nathan Lambert, Noah A. Smith, Yejin Choi, and Hanna Hajishirzi. 2024. Unpacking DPO and PPO: Disentangling best practices for learning from preference feedback. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10–15, 2024*.

Hamish Ivison, Yizhong Wang, Valentina Pyatkin, Nathan Lambert, Matthew E. Peters, Pradeep Dasigi, Joel Jang, David Wadden, Noah A. Smith, Iz Beltagy, and Hannaneh Hajishirzi. 2023. Camels in a changing climate: Enhancing LM adaptation with Tulu 2. *ArXiv preprint arXiv:2311.10702v2*.

Ray Jackendoff. 1989. What is a concept, that a person may grasp it? 1. *Mind & language*, 4(1–2):68–102. https://doi.org/10.1111/j.1468-0017.1989.tb00243.x

Ray Jackendoff. 2012. What is a concept? *Frames, Fields, and Contrasts*. Routledge, pages 191–208.

Younghoon Jeong, Juhyun Oh, Jongwon Lee, Jaimeen Ahn, Jihyung Moon, Sungjoon Park, and Alice Oh. 2022. KOLD: Korean offensive language dataset. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10818–10833, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.744

Akshita Jha, Aida Mostafazadeh Davani, Chandan K. Reddy, Shachi Dave, Vinodkumar Prabhakaran, and Sunipa Dev. 2023. SeeGULL: A stereotype benchmark with broad geo-cultural coverage leveraging generative models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9851–9870, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.548

Ming Jiang and Mansi Joshi. 2024. CPopQA: Ranking cultural concept popularity by LLMs. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 615–630, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-short.52

Yuchen Eleanor Jiang, Tianyu Liu, Shuming Ma, Dongdong Zhang, Mrinmaya Sachan, and Ryan Cotterell. 2023. Discourse-centric evaluation of document-level machine translation with a new densely annotated parallel corpus of novels. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7853–7872, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.435

Jiho Jin, Jiseon Kim, Nayeon Lee, Haneul Yoo, Alice Oh, and Hwaran Lee. 2024. KoBBQ: Korean bias benchmark for question answering. *Transactions of the Association for Computational Linguistics*, 12:507–524. https://doi.org/10.1162/tacl_a_00661

Rebecca L. Johnson, Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. The ghost in the machine has an american accent: Value conflict in GPT-3. *ArXiv preprint arXiv:2203.07785v1*.

David Jurgens, Agrima Seth, Jackson Sargent, Athena Aghighi, and Michael Geraci. 2023. Your spouse needs professional help: Determining the contextual appropriateness of messages through modeling social relationships. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10994–11013, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.616

Anubha Kabra, Emmy Liu, Simran Khanuja, Alham Fikri Aji, Genta Winata, Samuel Cahyawijaya, Anuoluwapo Aremu, Perez Ogayo, and Graham Neubig. 2023. Multilingual and multi-cultural figurative language understanding. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8269–8284, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.525

Nora Kassner, Philipp Dufter, and Hinrich Schütze. 2021. Multilingual LAMA: Investigating knowledge in multilingual pretrained language models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3250–3258, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.eacl-main.284

Zixuan Ke, Haowei Lin, Yijia Shao, Hu Xu, Lei Shu, and Bing Liu. 2022. Continual training of language models for few-shot learning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10205–10216, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.695

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. https://doi.org/10.18653/v1/2023.findings-acl.389

Simran Khanuja, Sathyanarayanan Ramamoorthy, Yueqi Song, and Graham Neubig. 2024. An image speaks a thousand words, but can everyone listen? On image transcreation for cultural relevance. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 10258–10279, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.573

Md Tawkat Islam Khondaker, Abdul Waheed, El Moatez Billah Nagoudi, and Muhammad Abdul-Mageed. 2023. GPTAraEval: A comprehensive evaluation of ChatGPT on Arabic NLP. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 220–247, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.16

Hyunwoo Kim, Jack Hessel, Liwei Jiang, Peter West, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, and Yejin Choi. 2023. SODA: Million-scale dialogue distillation with social commonsense contextualization. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12930–12949, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.799

Hyunwoo Kim, Youngjae Yu, Liwei Jiang, Ximing Lu, Daniel Khashabi, Gunhee Kim, Yejin Choi, and Maarten Sap. 2022. ProsocialDialog: A prosocial backbone for conversational agents. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4005–4029, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.267

Jaehong Kim, Chaeyoon Jeong, Seongchan Park, Meeyoung Cha, and Wonjae Lee. 2024. How do moral emotions shape political participation? A cross-cultural analysis of online petitions using language models. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 16274–16289, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.963

Hannah Rose Kirk, Alexander Whitefield, Paul Röttger, Andrew M. Bean, Katerina Margatina, Rafael Mosquera Gómez, 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 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10–15, 2024*.

Bernd Kortmann, Kerstin Lunkenheimer, and Katharina Ehret, editors . 2020. *eWAVE*.

Fajri Koto, Nurul Aisyah, Haonan Li, and Timothy Baldwin. 2023. Large language models only pass primary school exams in Indonesia: A comprehensive test on IndoMMLU. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12359–12374, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.760

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. 2024a. ArabicMMLU: Assessing massive multitask language understanding in Arabic. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 5622–5640, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.334

Fajri Koto, Rahmad Mahendra, Nurul Aisyah, and Timothy Baldwin. 2024b. IndoCulture: Exploring geographically influenced cultural

commonsense reasoning across eleven Indonesian provinces. *Transactions of the Association for Computational Linguistics*, 12:1703–1719. https://doi.org/10.1162/tacl_a_00726

Alfred Louis Kroeber and Clyde Kluckhohn. 1952. Culture: A critical review of concepts and definitions *Papers. Peabody Museum of Archaeology & Ethnology, Harvard University*.

Julia Kruk, Caleb Ziems, and Diyi Yang. 2023. Impressions: Visual semiotics and aesthetic impact understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12273–12291, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.755

Olli Kuparinen, Aleksandra Miletić, and Yves Scherrer. 2023. Dialect-to-standard normalization: A large-scale multilingual evaluation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13814–13828, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.923

Preethi Lahoti, Nicholas Blumm, Xiao Ma, Raghavendra Kotikalapudi, Sahitya Potluri, Qijun Tan, Hansa Srinivasan, Ben Packer, Ahmad Beirami, Alex Beutel, and Jilin Chen. 2023. Improving diversity of demographic representation in large language models via collective-critiques and self-voting. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10383–10405, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.643

Viet Lai, Chien Nguyen, Nghia Ngo, Thuat Nguyen, Franck Dernoncourt, Ryan Rossi, and Thien Nguyen. 2023. Okapi: Instruction-tuned large language models in multiple languages with reinforcement learning from human feedback. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 318–327, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-demo.28

Thang Le and Anh Luu. 2023. A parallel corpus for Vietnamese central-northern dialect text transfer. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13839–13855, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.925

Hwaran Lee, Seokhee Hong, Joonsuk Park, Takyoung Kim, Gunhee Kim, and Jung-woo Ha. 2023a. KoSBI: A dataset for mitigating social bias risks towards safer large language model applications. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, pages 208–224, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-industry.21

Jiyoung Lee, Minwoo Kim, Seungho Kim, Junghwan Kim, Seunghyun Won, Hwaran Lee, and Edward Choi. 2024a. KorNAT: LLM alignment benchmark for Korean social values and common knowledge. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 11177–11213, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.666

Nayeon Lee, Chani Jung, Junho Myung, Jiho Jin, Jose Camacho-Collados, Juho Kim, and Alice Oh. 2024b. 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. https://doi.org/10.18653/v1/2024.naacl-long.236

Nayeon Lee, Chani Jung, and Alice Oh. 2023b. Hate speech classifiers are culturally insensitive. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 35–46, Dubrovnik, Croatia. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.c3nlp-1.5

Cheng Li, Mengzhuo Chen, Jindong Wang, Sunayana Sitaram, and Xing Xie. 2024a. Culturellm: Incorporating cultural differences into large language models. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10–15, 2024*.

Chengxi Li, Kai Fan, Jiajun Bu, Boxing Chen, Zhongqiang Huang, and Zhi Yu. 2023a. Translate the beauty in songs: Jointly learning to align melody and translate lyrics. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 27–39, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.3

Haonan Li, Yixuan Zhang, Fajri Koto, Yifei Yang, Hai Zhao, Yeyun Gong, Nan Duan, and Timothy Baldwin. 2024b. CMMLU: Measuring massive multitask language understanding in Chinese. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 11260–11285, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.671

Oliver Li, Mallika Subramanian, Arkadiy Saakyan, Sky CH-Wang, and Smaranda Muresan. 2023b. NormDial: A comparable bilingual synthetic dialog dataset for modeling social norm adherence and violation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15732–15744, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.974

Wenyan Li, Crystina Zhang, Jiaang Li, Qiwei Peng, Raphael Tang, Li Zhou, Weijia Zhang, Guimin Hu, Yifei Yuan, Anders Søgaard, Daniel Hershcovich, and Desmond Elliott. 2024c. FoodieQA: A multimodal dataset for fine-grained understanding of Chinese food culture. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 19077–19095, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.1063

Zhi Li and Yin Zhang. 2023. Cultural concept adaptation on multimodal reasoning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 262–276, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.18

Chen Liu, Gregor Geigle, Robin Krebs, and Iryna Gurevych. 2022a. FigMemes: A dataset for figurative language identification in politically-opinionated memes. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7069–7086, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.476

Chen Liu, Fajri Koto, Timothy Baldwin, and Iryna Gurevych. 2024a. Are multilingual LLMs culturally-diverse reasoners? An investigation into multicultural proverbs and sayings. 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 2016–2039. Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.112

Chen Liu, Jonas Pfeiffer, Anna Korhonen, Ivan Vulić, and Iryna Gurevych. 2023a. Delving deeper into cross-lingual visual question answering. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2453–2468, Dubrovnik, Croatia. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-eacl.186

Chen Liu, Jonas Pfeiffer, Ivan Vulić, and Iryna Gurevych. 2024b. FUN with fisher: Improving generalization of adapter-based cross-lingual transfer with scheduled unfreezing. 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 1998–2015, Mexico City, Mexico. Association for Computational

Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.111

Fangyu Liu, Emanuele Bugliarello, Edoardo Maria Ponti, Siva Reddy, Nigel Collier, and Desmond Elliott. 2021. Visually grounded reasoning across languages and cultures. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10467–10485, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.818

Xuelin Liu, Yanfei Zhu, Shucheng Zhu, Pengyuan Liu, Ying Liu, and Dong Yu. 2024c. Evaluating moral beliefs across LLMs through a pluralistic framework. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4740–4760, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.272

Yanchen Liu, William Held, and Diyi Yang. 2023b. DADA: Dialect adaptation via dynamic aggregation of linguistic rules. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13776–13793, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.850

Zoey Liu, Crystal Richardson, Richard Hatcher, and Emily Prud'hommeaux. 2022b. Not always about you: Prioritizing community needs when developing endangered language technology. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3933–3944, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.272

Zhixuan Liu, Youeun Shin, Beverley-Claire Okogwu, Youngsik Yun, Lia Coleman, Peter Schaldenbrand, Jihie Kim, and Jean Oh. 2023c. Towards equitable representation in text-to-image synthesis models with the cross-cultural understanding benchmark (CCUB) dataset. *ArXiv preprint arXiv:2301.12073v2*.

Brandon Lwowski, Paul Rad, and Anthony Rios. 2022. Measuring geographic performance disparities of offensive language classifiers. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6600–6616, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Manuel Mager, Elisabeth Mager, Katharina Kann, and Ngoc Thang Vu. 2023. Ethical considerations for machine translation of indigenous languages: Giving a voice to the speakers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4871–4897, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.268

Olga Majewska, Evgeniia Razumovskaia, Edoardo M. Ponti, Ivan Vulić, and Anna Korhonen. 2023. Cross-lingual dialogue dataset creation via outline-based generation. *Transactions of the Association for Computational Linguistics*, 11:139–156. https://doi.org/10.1162/tacl_a_00539

Vijit Malik, Sunipa Dev, Akihiro Nishi, Nanyun Peng, and Kai-Wei Chang. 2022. Socially aware bias measurements for Hindi language representations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1041–1052, Seattle, United States. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.76

Antonis Maronikolakis, Axel Wisiorek, Leah Nann, Haris Jabbar, Sahana Udupa, and Hinrich Schuetze. 2022. Listening to affected communities to define extreme speech: Dataset and experiments. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1089–1104, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-acl.87

Reem Masoud, Ziquan Liu, Martin Ferianc, Philip C. Treleaven, and Miguel Rodrigues Rodrigues. 2025. Cultural alignment in large language models: An explanatory analysis based on Hofstede's cultural dimensions.

In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 8474–8503, Abu Dhabi, UAE. Association for Computational Linguistics.

Takahiko Masuda, Richard Gonzalez, Letty Kwan, and Richard E. Nisbett. 2008. Culture and aesthetic preference: Comparing the attention to context of east asians and americans. *Personality and Social Psychology Bulletin*, 34(9):1260–1275. https://doi.org/10.1177/0146167208320555, PubMed: 18678860

David Matsumoto and Linda Juang. 1996. Culture and psychology. *Pacific Grove*, pages 266–270.

Yoshiko Matsumoto. 1988. Reexamination of the universality of face: Politeness phenomena in Japanese. *Journal of Pragmatics*, 12(4):403–426. https://doi.org/10.1016/0378-2166(88)90003-3

Michael McCloskey and Neal J. Cohen. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. volume 24 of *Psychology of Learning and Motivation*, pages 109–165, Academic Press. https://doi.org/10.1016/S0079-7421(08)60536-8

Nicholas Meade, Elinor Poole-Dayan, and Siva Reddy. 2022. An empirical survey of the effectiveness of debiasing techniques for pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1878–1898, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.132

J. A. Meaney. 2020. Crossing the line: Where do demographic variables fit into humor detection? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 176–181, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-srw.24

Batja Mesquita, Nico H. Frijda, and Klaus R. Scherer. 1997. Culture and emotion. *Handbook of Cross-cultural Psychology: Basic Processes and Human Development*, 2:255.

George A. Miller. 1992. WordNet: A lexical database for English. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23–26, 1992*. https://doi.org/10.3115/1075527.1075662

Farhad Moghimifar, Shilin Qu, Tongtong Wu, Yuan-Fang Li, and Gholamreza Haffari. 2023. NormMark: A weakly supervised Markov model for socio-cultural norm discovery. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5081–5089, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.314

Youssef Mohamed, Mohamed Abdelfattah, Shyma Alhuwaider, Feifan Li, Xiangliang Zhang, Kenneth Church, and Mohamed Elhoseiny. 2022. ArtELingo: A million emotion annotations of WikiArt with emphasis on diversity over language and culture. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8770–8785, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.600

Youssef Mohamed, Runjia Li, Ibrahim Ahmad, Kilichbek Haydarov, Philip Torr, Kenneth Church, and Mohamed Elhoseiny. 2024. No culture left behind: ArtELingo-28, a benchmark of WikiArt with captions in 28 languages. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20939–20962, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.1165

Aida Mostafazadeh Davani, Mark Diaz, Dylan Baker, and Vinodkumar Prabhakaran. 2024. D3CODE: Disentangling disagreements in data across cultures on offensiveness detection and evaluation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18511–18526, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.1029

Anjishnu Mukherjee, Aylin Caliskan, Ziwei Zhu, and Antonios Anastasopoulos. 2024.

Global gallery: The fine art of painting culture portraits through multilingual instruction tuning. 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 6398–6415, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.355

Anjishnu Mukherjee, Chahat Raj, Ziwei Zhu, and Antonios Anastasopoulos. 2023. Global voices, local biases: Socio-cultural prejudices across languages. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15828–15845, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.981

George Peter Murdock. 1940. The cross-cultural survey. *American Sociological Review*, 5(3): 361–370. https://doi.org/10.2307/2084038

Maria Nadejde, Anna Currey, Benjamin Hsu, Xing Niu, Marcello Federico, and Georgiana Dinu. 2022. CoCoA-MT: A dataset and benchmark for contrastive controlled MT with application to formality. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 616–632, Seattle, United States. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-naacl.47

Nikita Nangia, Clara Vania, Rasika Bhalerao, and Samuel R. Bowman. 2020. CrowS-pairs: A challenge dataset for measuring social biases in masked language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1953–1967, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.154

Tarek Naous, Michael 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. https://doi.org/10.18653/v1/2024.acl-long.862

Shravan Nayak, Kanishk Jain, Rabiul Awal, Siva Reddy, Sjoerd Van Steenkiste, Lisa Anne Hendricks, Karolina Stanczak, and Aishwarya Agrawal. 2024. Benchmarking vision language models for cultural understanding. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5769–5790, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.329

Aurélie Névéol, Yoann Dupont, Julien Bezançon, and Karën Fort. 2022. French CrowS-pairs: Extending a challenge dataset for measuring social bias in masked language models to a language other than English. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8521–8531, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.583

Tuan-Phong Nguyen, Simon Razniewski, Aparna S. Varde, and Gerhard Weikum. 2023. Extracting cultural commonsense knowledge at scale. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023–4 May 2023*, pages 1907–1917, ACM. https://doi.org/10.1145/3543507.3583535

Tuan-Phong Nguyen, Simon Razniewski, and Gerhard Weikum. 2024. Cultural commonsense knowledge for intercultural dialogues. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, CIKM '24, pages 1774–1784, New York, NY, USA. Association for Computing Machinery. https://doi.org/10.1145/3627673.3679768

Minako O'hagan and Carmen Mangiron. 2013. *Game Localization: Translating for the global digital entertainment industry*, volume 106. John Benjamins Publishing. https://doi.org/10.1075/btl.106

Longshen Ou, Xichu Ma, Min-Yen Kan, and Ye Wang. 2023. Songs across borders: Singable and controllable neural lyric translation. In

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 447–467, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.27

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28–December 9, 2022*.

Shramay Palta and Rachel Rudinger. 2023. FORK: A bite-sized test set for probing culinary cultural biases in commonsense reasoning models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9952–9962, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.631

Claudio Paonessa, Yanick Schraner, Jan Deriu, Manuela Hürlimann, Manfred Vogel, and Mark Cieliebak. 2023. Dialect transfer for Swiss German speech translation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 15240–15254, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.1018

Joon Sung Park, Lindsay Popowski, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social simulacra: Creating populated prototypes for social computing systems. In *The 35th Annual ACM Symposium on User Interface Software and Technology, UIST 2022, Bend, OR, USA, 29 October 2022–2 November 2022*, pages 74:1–74:18. ACM. https://doi.org/10.1145/3526113.3545616

Jiaxin Pei and David Jurgens. 2023. When do annotator demographics matter? Measuring the influence of annotator demographics with the POPQUORN dataset. In *Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII)*, pages 252–265. Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.law-1.25

Zoë Pettit. 2009. *3: Connecting Cultures: Cultural Transfer in Subtitling and Dubbing*, pages 44–57. Multilingual Matters. Bristol, Blue Ridge Summit. https://doi.org/10.21832/9781847691552-005

Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020. MAD-X: An adapter-based framework for multi-task cross-lingual transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7654–7673, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.617

Michel Plüss, Jan Deriu, Yanick Schraner, Claudio Paonessa, Julia Hartmann, Larissa Schmidt, Christian Scheller, Manuela Hürlimann, Tanja Samardžić, Manfred Vogel, and Mark Cieliebak. 2023. STT4SG-350: A speech corpus for all Swiss German dialect regions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1763–1772, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-short.150

Edoardo Maria Ponti, Goran Glavaš, Olga Majewska, Qianchu Liu, Ivan Vulić, and Anna Korhonen. 2020. XCOPA: A multilingual dataset for causal commonsense reasoning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2362–2376, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.185

Rifki Putri, Faiz Haznitrama, Dea Adhista, and Alice Oh. 2024. Can LLM generate culturally relevant commonsense QA data? Case study in Indonesian and Sundanese. In *Proceedings of the 2024 Conference on*

Empirical Methods in Natural Language Processing, pages 20571–20590, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.1145

Valentina Pyatkin, Jena D. Hwang, Vivek Srikumar, Ximing Lu, Liwei Jiang, Yejin Choi, and Chandra Bhagavatula. 2023. ClarifyDelphi: Reinforced clarification questions with defeasibility rewards for social and moral situations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11253–11271, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.630

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10–16, 2023*.

Aida Ramezani and Yang Xu. 2023. Knowledge of cultural moral norms in large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 428–446, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.26

Kavel Rao, Liwei Jiang, Valentina Pyatkin, Yuling Gu, Niket Tandon, Nouha Dziri, Faeze Brahman, and Yejin Choi. 2023. What makes it ok to set a fire? Iterative self-distillation of contexts and rationales for disambiguating defeasible social and moral situations. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12140–12159, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.812

Ken Resnicow, Tom Baranowski, Jasjit S. Ahluwalia, and Ronald L. Braithwaite. 1999. Cultural sensitivity in public health: Defined and demystified. *Ethnicity & Disease*, 9(1):10–21.

Dor Ringel, Gal Lavee, Ido Guy, and Kira Radinsky. 2019. Cross-cultural transfer learning for text classification. 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 3873–3883, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1400

William Gaviria Rojas, Sudnya Frederick Diamos, Keertan Kini, David Kanter, Vijay Janapa Reddi, and Cody Coleman. 2022. The dollar street dataset: Images representing the geographic and socioeconomic diversity of the world. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28–December 9, 2022*.

Nihar Sahoo, Pranamya Kulkarni, Arif Ahmad, Tanu Goyal, Narjis Asad, Aparna Garimella, and Pushpak Bhattacharyya. 2024. IndiBias: A benchmark dataset to measure social biases in language models for Indian context. 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 8786–8806, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.487

Nihar Sahoo, Niteesh Mallela, and Pushpak Bhattacharyya. 2023. With prejudice to none: A few-shot, multilingual transfer learning approach to detect social bias in low resource languages. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13316–13330, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.842

Mohammad Salameh, Houda Bouamor, and Nizar Habash. 2018. Fine-grained Arabic dialect identification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1332–1344, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining algorithmic fairness in India and beyond. In *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3–10, 2021*, pages 315–328. ACM. https://doi.org/10.1145/3442188.3445896

Sandra Sandoval, Jieyu Zhao, Marine Carpuat, and Hal Daumé III. 2023. A rose by any other name would not smell as sweet: Social bias in names mistranslation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3933–3945, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.239

Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? In *International Conference on Machine Learning, ICML 2023, 23–29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 29971–30004, PMLR.

Sebastin Santy, Jenny Liang, Ronan Le Bras, Katharina Reinecke, and Maarten Sap. 2023. NLPositionality: Characterizing design biases of datasets and models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9080–9102, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.505

Maarten Sap, Swabha Swayamdipta, Laura Vianna, Xuhui Zhou, Yejin Choi, and Noah A. Smith. 2022. Annotators with attitudes: How annotator beliefs and identities bias toxic language detection. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5884–5906, Seattle, United States. Association for Computational Linguistics.

Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2021. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics*, 9:845–874. https://doi.org/10.1162/tacl_a_00401

Shalom H. Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Elsevier. https://doi.org/10.1016/S0065-2601(08)60281-6

Omar Shaikh, Caleb Ziems, William Held, Aryan Pariani, Fred Morstatter, and Diyi Yang. 2023. Modeling cross-cultural pragmatic inference with codenames duet. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6550–6569, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.410

Chhavi Sharma, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, and Björn Gambäck. 2020. SemEval-2020 task 8: Memotion analysis- the visuo-lingual metaphor! In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 759–773, Barcelona (online). International Committee for Computational Linguistics. https://doi.org/10.18653/v1/2020.semeval-1.99

Shuaijie She, Wei Zou, Shujian Huang, Wenhao Zhu, Xiang Liu, Xiang Geng, and Jiajun Chen. 2024. MAPO: Advancing multilingual reasoning through multilingual-alignment-as-preference optimization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10015–10027, Bangkok, Thailand. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.539

Ravi Shekhar, Mladen Karan, and Matthew Purver. 2022. CoRAL: A context-aware Croatian abusive language dataset. In *Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022*, pages 217–225, Online only. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-aacl.21

Weiyan Shi, Ryan Li, Yutong Zhang, Caleb Ziems, Sunny Yu, Raya Horesh, Rogério

Paula, and Diyi Yang. 2024. CultureBank: An online community-driven knowledge base towards culturally aware language technologies. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4996–5025, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.288

Vered Shwartz. 2022. Good night at 4 pm?! Time expressions in different cultures. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2842–2853, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-acl.224

Linda Tuhiwai Smith. 2021. *Decolonizing Methodologies: Research and Indigenous Peoples*. Bloomsbury Publishing. https://doi.org/10.5040/9781350225282

Guijin Son, Hanwool Lee, Sungdong Kim, Seungone Kim, Niklas Muennighoff, Taekyoon Choi, Cheonbok Park, Kang Min Yoo, and Stella Biderman. 2024. KMMLU: measuring massive multitask language understanding in Korean. *ArXiv preprint arXiv:2402.11548v2*.

Taylor Sorensen, Jared Moore, Jillian Fisher, Mitchell L. Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, Tim Althoff, and Yejin Choi. 2024. Position: A roadmap to pluralistic alignment. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21–27, 2024*. OpenReview.net.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4–9, 2017, San Francisco, California, USA*, pages 4444–4451, AAAI Press. https://doi.org/10.1609/aaai.v31i1.11164

Anirudh Srinivasan and Eunsol Choi. 2022. TyDiP: A dataset for politeness classification in nine typologically diverse languages. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5723–5738, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-emnlp.420

Gabriel Stanovsky, Noah A. Smith, and Luke Zettlemoyer. 2019. Evaluating gender bias in machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1679–1684, Florence, Italy. Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1164

Ian Stewart and Rada Mihalcea. 2024. Whose wife is it anyway? Assessing bias against same-gender relationships in machine translation. In *Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, pages 365–375, Bangkok, Thailand. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.gebnlp-1.23

Hao Sun, Zhexin Zhang, Fei Mi, Yasheng Wang, Wei Liu, Jianwei Cui, Bin Wang, Qun Liu, and Minlie Huang. 2023. MoralDial: A framework to train and evaluate moral dialogue systems via moral discussions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2213–2230, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.123

Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2019. Mitigating gender bias in natural language processing: Literature review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1630–1640, Florence, Italy. Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1159

Zhewei Sun and Yang Xu. 2022. Tracing semantic variation in slang. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1299–1313, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.84

Nir Sweed and Dafna Shahaf. 2021. Catchphrase: Automatic detection of cultural references. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 1–7, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-short.1

Yi Tay, Donovan Ong, Jie Fu, Alvin Chan, Nancy Chen, Anh Tuan Luu, and Chris Pal. 2020. Would you rather? A new benchmark for learning machine alignment with cultural values and social preferences. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5369–5373, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.477

Katherine Thai, Marzena Karpinska, Kalpesh Krishna, Bill Ray, Moira Inghilleri, John Wieting, and Mohit Iyyer. 2022. Exploring document-level literary machine translation with parallel paragraphs from world literature. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9882–9902, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.672

Ashish V. Thapliyal, Jordi Pont Tuset, Xi Chen, and Radu Soricut. 2022. Crossmodal-3600: A massively multilingual multimodal evaluation dataset. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 715–729, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.45

Peter Trudgill. 2000. *Sociolinguistics: An Introduction to Language and Society*. Penguin UK.

Edward Burnett Tylor. 1871. *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art and Custom*, volume 2. J. Murray.

UNESCO. 1982. World conference on cultural policies, mexico city, final report.

Ahmet Üstün, Arianna Bisazza, Gosse Bouma, and Gertjan van Noord. 2020. UDapter: Language adaptation for truly Universal Dependency parsing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2302–2315, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.180

Karina Vida, Judith Simon, and Anne Lauscher. 2023. Values, ethics, morals? On the use of moral concepts in NLP research. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 5534–5554, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.368

Anvesh Rao Vijjini, Rakesh R. Menon, Jiayi Fu, Shashank Srivastava, and Snigdha Chaturvedi. 2024. SocialGaze: Improving the integration of human social norms in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 16487–16506, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.962

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019. GLUE: A multi-task benchmark and analysis platform for natural language understanding.In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6–9, 2019*. OpenReview.net.

Bin Wang, Geyu Lin, Zhengyuan Liu, Chengwei Wei, and Nancy Chen. 2024a. CRAFT: Extracting and tuning cultural instructions from the wild. In *Proceedings of the 2nd Workshop on Cross-Cultural Considerations in NLP*, pages 42–47, Bangkok, Thailand. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.c3nlp-1.4

Bin Wang, Zhengyuan Liu, Xin Huang, Fangkai Jiao, Yang Ding, AiTi Aw, and Nancy Chen. 2024b. SeaEval for multilingual foundation models: From cross-lingual alignment to cultural reasoning. 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 370–390, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.22

Ruize Wang, Duyu Tang, Nan Duan, Zhongyu Wei, Xuanjing Huang, Jianshu Ji, Guihong Cao, Daxin Jiang, and Ming Zhou. 2021. K-Adapter: Infusing knowledge into pre-trained models with adapters. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1405–1418, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.121

Wenxuan Wang, Wenxiang Jiao, Jingyuan Huang, Ruyi Dai, Jen-tse Huang, Zhaopeng Tu, and Michael Lyu. 2024c. 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. https://doi.org/10.18653/v1/2024.acl-long.345

Ronald Wardhaugh and Janet M. Fuller. 2021. *An introduction to sociolinguistics*. John Wiley & Sons.

Leslie A. White. 1959. The concept of culture. *American Anthropologist*, 61(2):227–251. https://doi.org/10.1525/aa.1959.61.2.02a00040

Andrew Whiten, Robert A. Hinde, Kevin N. Laland, and Christopher B. Stringer. 2011. Culture evolves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):938–948. https://doi.org/10.1093/acprof:osobl/9780199608966.001.0001

Haryo Wibowo, Erland Fuadi, Made Nityasya, Radityo Eko Prasojo, and Alham Aji. 2024. COPAL-ID: Indonesian language reasoning with local culture and nuances. 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 1404–1422, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.77

Anna Wierzbicka. 1992. *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-specific Configurations*. Oxford University Press. https://doi.org/10.1093/oso/9780195073256.001.0001

Zedian Xiao, William Held, Yanchen Liu, and Diyi Yang. 2023. Task-agnostic low-rank adapters for unseen English dialects. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7857–7870, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.487

Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2019. BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2324–2335, Minneapolis, Minnesota. Association for Computational Linguistics.

Ying Xu, Dakuo Wang, Mo Yu, Daniel Ritchie, Bingsheng Yao, Tongshuang Wu, Zheng Zhang, Toby Li, Nora Bradford, Branda Sun, Tran Hoang, Yisi Sang, Yufang Hou, Xiaojuan Ma, Diyi Yang, Nanyun Peng, Zhou Yu, and Mark Warschauer. 2022. Fantastic questions and where to find them: FairytaleQA – An authentic dataset for narrative comprehension. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 447–460, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.34

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498,

Online. Association for Computational Linguistics.

Diyi Yang. 2019. *Computational Social Roles*. Ph.D. thesis. Carnegie Mellon University Pittsburgh, PA, USA.

Zhichao Yang, Pengshan Cai, Yansong Feng, Fei Li, Weijiang Feng, Elena Suet-Ying Chiu, and Hong Yu. 2019. Generating classical Chinese poems from vernacular Chinese. 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 6155–6164, Hong Kong, China. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1637, PubMed: 32467928

Binwei Yao, Ming Jiang, Tara Bobinac, Diyi Yang, and Junjie Hu. 2024a. Benchmarking machine translation with cultural awareness. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13078–13096, Miami, Florida, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.765

Jing Yao, Xiaoyuan Yi, Yifan Gong, Xiting Wang, and Xing Xie. 2024b. Value FULCRA: Mapping large language models to the multidimensional spectrum of basic human value. 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 8762–8785, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.486

Da Yin, Hritik Bansal, Masoud Monajatipoor, Liunian Harold Li, and Kai-Wei Chang. 2022. GeoMLAMA: Geo-diverse commonsense probing on multilingual pre-trained language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2039–2055, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.132

Da Yin, Liunian Harold Li, Ziniu Hu, Nanyun Peng, and Kai-Wei Chang. 2021. Broaden the vision: Geo-diverse visual commonsense reasoning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2115–2129, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.162

Linhao Yu, Yongqi Leng, Yufei Huang, Shang Wu, Haixin Liu, Xinmeng Ji, Jiahui Zhao, Jinwang Song, Tingting Cui, Xiaoqing Cheng, Liutao Liutao, and Deyi Xiong. 2024. CMoralEval: A moral evaluation benchmark for Chinese large language models. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 11817–11837, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.703

Ye Yuan, Kexin Tang, Jianhao Shen, Ming Zhang, and Chenguang Wang. 2024. Measuring social norms of large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 650–699, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-naacl.43

Mahmoud Yusuf, Marwan Torki, and Nagwa El-Makky. 2022. Arabic dialect identification with a few labeled examples using generative adversarial networks. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 196–204, Online only. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.aacl-main.16

Haolan Zhan, Zhuang Li, Xiaoxi Kang, Tao Feng, Yuncheng Hua, Lizhen Qu, Yi Ying, Mei Rianto Chandra, Kelly Rosalin, Jureynolds Jureynolds, Suraj Sharma, Shilin Qu, Linhao Luo, Ingrid Zukerman, Lay-Ki Soon, Zhaleh Semnani Azad, and Reza Haf. 2024. RENOVI: A benchmark towards remediating norm violations in socio-cultural conversations. In *Findings of the Association for Computational Linguistics: NAACL*

2024, pages 3104–3117, Mexico City, Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-naacl.196

Haolan Zhan, Zhuang Li, Yufei Wang, Linhao Luo, Tao Feng, Xiaoxi Kang, Yuncheng Hua, Lizhen Qu, Lay-Ki Soon, Suraj Sharma, Ingrid Zukerman, Zhaleh Semnani-Azad, and Gholamreza Haffari. 2023. Socialdial: A benchmark for socially-aware dialogue systems. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23–27, 2023*, pages 2712–2722. ACM. https://doi.org/10.1145/3539618.3591877

Jiaxu Zhao, Meng Fang, Zijing Shi, Yitong Li, Ling Chen, and Mykola Pechenizkiy. 2023. CHBias: Bias evaluation and mitigation of Chinese conversational language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13538–13556, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.757

Jingyan Zhou, Jiawen Deng, Fei Mi, Yitong Li, Yasheng Wang, Minlie Huang, Xin Jiang, Qun Liu, and Helen Meng. 2022. Towards identifying social bias in dialog systems: Framework, dataset, and benchmark. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 3576–3591, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-emnlp.262

Li Zhou, Laura Cabello, Yong Cao, and Daniel Hershcovich. 2023a. Cross-cultural transfer learning for Chinese offensive language detection. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 8–15, Dubrovnik, Croatia. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.c3nlp-1.2

Li Zhou, Antonia Karamolegkou, Wenyu Chen, and Daniel Hershcovich. 2023b. Cultural compass: Predicting transfer learning success in offensive language detection with cultural features. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12684–12702, Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.845

Li Zhou, Taelin Karidi, Nicolas Garneau, Yong Cao, Wanlong Liu, Wenyu Chen, and Daniel Hershcovich. 2024. Does mapo tofu contain coffee? Probing llms for food-related cultural knowledge. *ArXiv preprint arXiv:2404.06833v1*.

Caleb Ziems, Jiaao Chen, Camille Harris, Jessica Anderson, and Diyi Yang. 2022a. VALUE: Understanding dialect disparity in NLU. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3701–3720, Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.258

Caleb Ziems, Jane Dwivedi-Yu, Yi-Chia Wang, Alon Halevy, and Diyi Yang. 2023a. NormBank: A knowledge bank of situational social norms. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7756–7776, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.429

Caleb Ziems, William Held, Jingfeng Yang, Jwala Dhamala, Rahul Gupta, and Diyi Yang. 2023b. Multi-VALUE: A framework for cross-dialectal English NLP. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 744–768, Toronto, Canada. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.44

Caleb Ziems, Jane Yu, Yi-Chia Wang, Alon Halevy, and Diyi Yang. 2022b. The moral integrity corpus: A benchmark for ethical dialogue systems. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3755–3773,

Dublin, Ireland. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.261

A Method

We examine the main and findings papers from the leading *CL venues, including: ACL, EMNLP, AACL, EACL, NAACL, and TACL published since 2020 (5-year span). The initial set of papers was identified using the following search terms: ‘‘culture’’, ‘‘cultural’’, ‘‘geo-diverse’’, ‘‘socio’’, ‘‘social’’, ‘‘moral’’, ‘‘norms’’ in the title and abstract. Initially, we collected 336 papers, using human verification to exclude papers that did not consider cultural variations, as well as papers that solely focus on analysis and probing (as they are beyond the scope of our survey). The final paper count is 127. For more on probing and analysis, please refer to the recent surveys like Adilazuarda et al. (2024). We further acknowledge the limitation of missing relevant papers from other sources and papers without explicitly mentioning any of the search keywords. However, our goal is not to conduct a systematic review, but to propose a taxonomy and understand the progress in NLP for this research area and identify research gaps. We believe that focusing on *CL venues is an appropriate choice for this purpose.

B Additional Examples of Use Cases for the Taxonomy.

Another example of applying this taxonomy is the development of culturally aware conversational AI for educational purposes. Such development should be informed, at a minimum, by relevant *Knowledge* (e.g., Facts), appropriate *Style*, understanding of the *Communicative Goals* (e.g., that of teaching) and consideration of *Relationships* (e.g., that of a teacher and student). These are merely example elements and applications to consider.

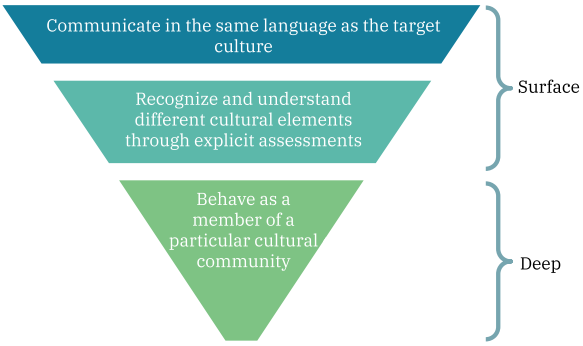

Figure 3: An illustration of surface versus deep culturally adapted NLP model.

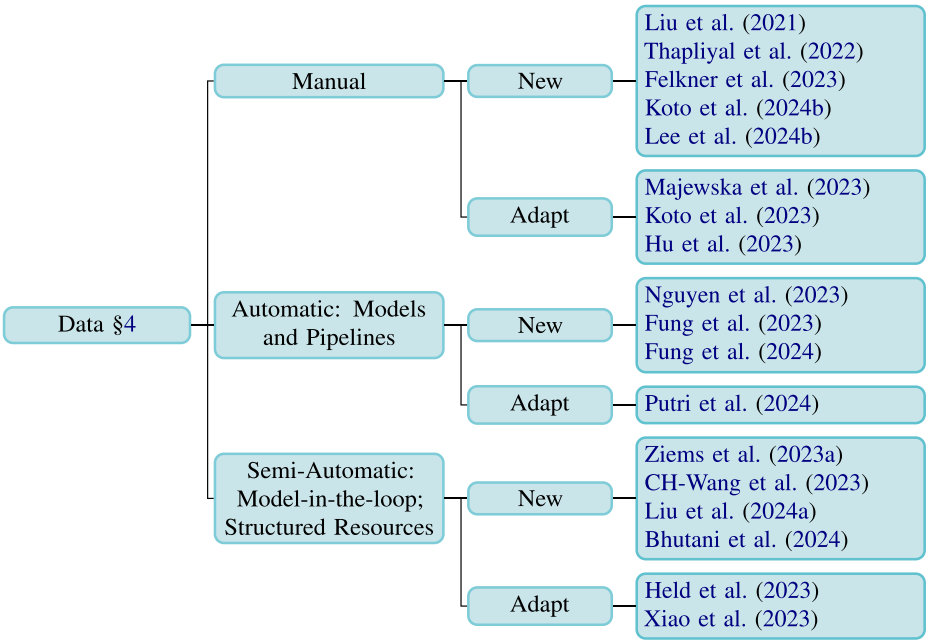

Figure 4: Categorization of the methods for resource acquisitions with representative examples.

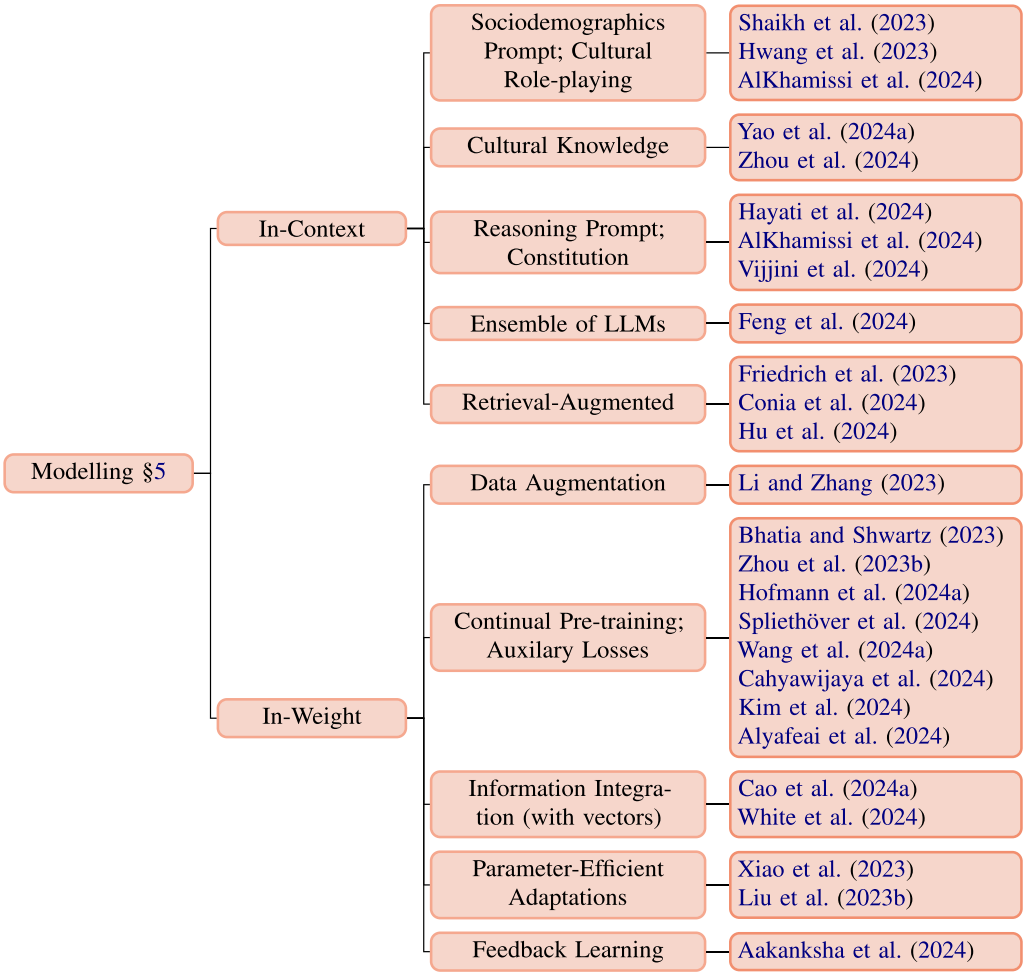

Figure 5: Categorization of the adaptation modeling methods and examples in each category.