Supporting and Performing Culture from the Inside

Lea Frermann[1,2] **and Steven Bird**[3]

[1]The University of Tübingen, Tübingen, Germany
[2]The University of Melbourne, Melbourne, Australia
[3]Charles Darwin University, Darwin, Australia

lea.frermann@uni-tuebingen.de steven.bird@cdu.edu.au

Abstract

Anthropology and related disciplines which study culture have often found it useful to consider their epistemological and methodological approaches in terms of emic versus etic. This is the distinction between the insider perspective, how the world is viewed by a member of the culture, versus the outsider perspective, the scientific cataloguing of cultural knowledge and practice. We adopt this framework to systematically analyse assumptions in recent research on culture in NLP along the full pipeline of task selection, data collection, system design and evaluation. We draw attention to the 'culture of NLP' which shapes the research focus and approaches of our field, and suggest pathways towards a more culturally attuned AI.

1 Introduction

Human culture can hardly be disentangled from the culture that a researcher brings to an investigation, a phenomenon that has been called *Kulturbrille* (cultural glasses) by Franz Boas (Engelke, 2018). The same holds true for the flourishing research on culture and cultural alignment in natural language processing (NLP) (Pawar et al., 2025). The *culture of NLP* is so deeply embedded that it can be hard to see (Hutchinson, 2024). Upstream of the data that flows into NLP systems, culture is the invisible filter that selects, shapes and colours what is observed, the meaning that is assigned, and the text or speech that is ultimately produced before it is ingested by an LLM. Downstream of the language technologies we create – like information retrieval (IR), machine translation (MT), automatic speech recognition (ASR), and most recently, chatbots and large language models (LLMs) – are all the diverse ways that information and technology are socialised in the world's cultures.

Between the upstream and the downstream culture that is invisible to NLP, there is the culture embedded in the representational choices of LLMs and in the interactional style of chatbots. This NLP culture includes evaluations of the capability of NLP artefacts like LLMs and chatbots to reproduce culture, including cultural facts but also deeper notions like cultural sensitivity and cultural competence (Nejadgholi et al., 2026), or situational appropriateness and values alignment (Minggad et al., 2026). NLP culture is present wherever we limit our conception of culture to information that is documented in textual or spoken records, separated from the paralinguistic and the extralinguistic, communicative layers that arise from the embodied nature of communication. This includes evaluations of the capability of LLMs and chatbots to reproduce cultural facts, but goes further to evaluations of cultural sensitivity and cultural competence (Nejadgholi et al., 2026), or situational appropriateness and values alignment (Minggad et al., 2026).

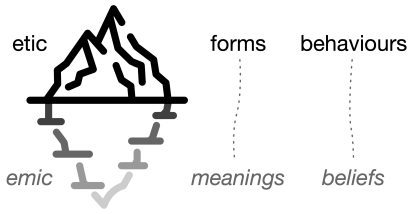


Figure 1: Hall's Cultural Iceberg: forms and behaviours (etic) can be observed by outsiders, while corresponding meanings and beliefs (emic) are only accessible to insiders (after Hall, 1976) .

To systematise epistemological and methodological positions, anthropologists have proposed to distinguish emic versus etic approaches to cultural studies (Pike, 1967; Sullivan, 2009). An *etic* perspective adopts the outsider's view, classifies surface forms, and applies generic measures across cultures in a universal, comparative, and scalable view. The *emic* perspective adopts the insider's view of a member of the culture, with a focus on the underlying cultural system, specific to a single culture. The concepts of etic and emic originate in structuralist linguistics, as cultural generalisations of phon*etic* and phon*emic*. Pike (1967) observed

Findings of the Association for Computational Linguistics: EMNLP 2026, pages 47683–47691
October 24-29, 2026 ©2026 Association for Computational Linguistics

that human behaviour can be observed, described, and classified from outside a culture (etic), while having a significance accessible only to insiders (emic, Headland et al., 1990). A popular analogy for this distinction is the 'cultural iceberg' (Fig. 1).

The goals of this paper are to challenge seemingly inevitable methods of research and system development in NLP, to put forth an alternative perspective, and to encourage the community to reflect on the culture of NLP and how it impacts our community's goals and contributions.

This paper contrasts emic with etic perspectives across the Four Stages of AI Development, and on this basis surveys the recent literature on cultural NLP, identifying a dominance of etic approaches. It then applies the emic/etic framework and (1) uncovers hidden assumptions in how our community designs approaches, data sets and model evaluations; (2) points to gaps in research assuming that the ultimate goal is to develop systems which are valuable for the culture in question; and (3) sketches out a path forward towards an emic research agenda.

2 The Etic/Emic Framework

In this section we will see that the practice of generative AI is predominantly etic, in all stages of development. Task definitions are understood as general purpose solutions to externally-defined problems, which means that few researchers see any need to engage individual communities about what they need (Sec. 2.1). Primary data about the world is taken to be text, yet this contains many layers of culture, and may effectively be a blind spot for NLP researchers (Sec. 2.2). Systems are designed to replicate or replace humans without due consideration for how this may impact the meaningfulness of lives and livelihoods on the ground (Sec. 2.3). Finally, deployments are optimised to serve global measures of efficiency, productivity, and profitability, and performance benchmarks are applied across the board, regardless of local cultural norms and expectations (Sec. 2.4).

In addition, we reconsider these AI development stages from an emic perspective. We incorporate qualitative findings from a review of all papers published in the Findings or Main streams of *ACL and EMNLP conferences (2022–4/2026) containing the word 'culture' or 'cultural' in the title (N=107). See Appendix A for details of our approach and a link to a GitHub repository containing the full list of papers and labels.

2.1 Tasks

In the context of a minority culture, an etic approach usually presumes the presence of the dominant culture, and looks across cultures, seeking opportunities for universal solutions. The most obvious need, at this level, is 'overcoming language barriers' (Pastor et al., 2017). The default task, then, would be translation. We can observe this by submitting a queries to a chatbot in progressively smaller languages and – beyond a certain threshold – notice that the chatbot shifts from answering queries in monolingual mode, to translating queries into the dominant language, usually English. In all such translation processes, minority culture content is shaped along the lines of dominant cultures (Minggad et al., 2026).

If the default etic position is to overcome barriers, the default emic position is often to maintain barriers, defending cultural diversity in the face of efforts that seek to flatten it. Community leaders enact their self-determination through intergenerational transmission of cultural practices, strengthening and safeguarding the local lifeworld (Kimmerer, 2013; Harjo, 2019).

How this manifests on the ground changes from place to place, but a common linguistic challenge is to transmit ecological knowledge. For example, parents teach their children about food preparation, duties to kin, and the associated knowledge of the country and its seasons. These knowledge practices are increasingly video recorded and shared, to the point where individuals and communities may have sizeable collections of untranscribed audio and video in their local vernacular.

This rarely involves universal information access mediated through local language or culture. Those members are usually already bicultural and bilingual owing to the history of contact. This means that they are perfectly capable of navigating the dominant culture using the dominant language. Such is the nature of bilingualism and language shift (Fishman, 2001; Grosjean, 2021).

When considering NLP tasks from an emic perspective, we do not seek to 'understand' or 'translate' or 'align', but allow space for diverse local meanings that fall outside the scope of our systems. For example, language technologies can support navigation of untranscribed audio and video through spoken term retrieval, enabling human learning of language and culture directly from non-textual sources (Bird, 2026).

Insider-driven considerations like the above are barely mentioned in the NLP literature. The majority of surveyed papers (N=78; 73%) focussed on the intrinsic goal of assessing or augmenting cultural knowledge in LLMs, or so-called 'trivia-centered' cultural knowledge representations (Zhou et al., 2025; Oh et al., 2025; Alkhamissi et al., 2026). The remaining papers address downstream tasks ranging from culturally appropriate dialogue over culturally attuned translation to culture-aware IR. However, the tasks are selected and formalised by the researchers who are embedded in a dominant culture (by upbringing and/or by academic affiliation), typically with the primary goal of generalizability to many cultures in mind (etic). We did not encounter papers which co-developed a task with a cultural community.

2.2 Data

For most of the world's cultures, the textual record comes from ethnographic observations that are subject to the cultural conventions of outsiders (Clifford and Marcus, 2023). The record depends on the observable behaviours of insiders, and what insiders may report about the underlying meanings and beliefs, reports which may be unreliable or even deceptive (Borofsky, 2005). For example, Said argues that everything written about the 'Orient' is a projection of western culture (Said, 1978).

This problem is not removed by eliminating outsider texts; there is the culture of writing itself:

> *... writing, wherever it exists, is always only one of several communication channels available to the members of a society. Consequently, the conditions under which it is selected and the purposes to which it is put must be described in relation to those of other channels. (Basso, 1974, p426)*

There is also the nature of observation, e.g.: there are no pure facts (Popper, 1935); differentiating figure from ground is a perceptual act (Schacter et al., 2009); measurement involves culturally-informed choices (Walter and Andersen, 2013); and 'the world as it is' differs from 'the world according to data' (Mitchell and Fleischman, 2020, p145). Thus texts are separated from 'reality' by three cultural layers: what is observed; what meaning is attributed; and how this is rendered into text.

One response to this situation is to consult with members of the minority culture in the data collection process rather than relying exclusively on existing texts as ground truth. A notable fraction of the surveyed literature (19 papers) drew on community input in concept selection, example creation and/or data validation, where data was typically contributed by members of the target culture which were based at the respective research institution, or via crowd sourcing. However, we also observe a strong trend of creating 'cultural' benchmarks in an at least partially synthetic way (33 papers), where often concepts and/or context are first generated by LLMs and later filtered by humans in a practice that necessarily distorts the emic, intra-cultural perspective (see discussion above). The practice of translating English benchmarks was widely criticised and occurred only three times. 31 papers based their concept samples or examples on existing data sets which often spanned many languages (news, social media, Wikipedia, ...) and imply the concerns laid out above.

2.3 System design

The vast majority of language technologies is designed to enhance human information access, through IR, MT, ASR, or LLMs in the form of chatbots. The construct of chatbots was founded on the model of a human amateur consulting a human expert (Weizenbaum, 1966; Woods, 1973). Studies of the culture of expertise show that it is not to be understood in terms of information possessed and imparted, but in terms of interaction and ideology. For each domain there is a culture whereby experts are socialised, evaluated, and authorised, and there are particular linguistic performances that produce jargon, structure interactions, and animate evidence (Carr, 2010). Amateur-expert consultations have a *shape* which is particular to the domain, e.g. in medicine with opening, complaint, examination/test, diagnosis, treatment/advice, and closing (Ten Have, 1989). Across many domains, better outcomes are linked to interpersonal rapport (Wilson et al., 2010).

Set against this, we have the agenda of making chatbots into domain experts (Metzler et al., 2021), with the diversion of web queries into chatbot interactions, in the name of retrieving answers rather than documents. However, to make this shift is to 'miss the big picture of why people seek information and how that process contains value beyond simply retrieving relevant information' (Shah and Bender, 2022, p221). Moreover, it risks the replacement of diverse cultures of expert-shaped learning interactions with a series of prompt-response dyads shaped by novices.

Generative AI is increasingly deployed in agentic systems, designed with the goal of simulating

	Task	Data	System Design	Deployment / Evaluation
Etic	generic solns to external problems ▷ *generate facts about any culture*	data extraction for LLMs ▷ *translated or synthesized common sense benchmarks*	simulate or improve on humans ▷ *text chat; adapting to user language; flexibility in tone/register*	efficiency and scalability ▷ *correctness and completeness of responses*
Emic	local agency in shaping their world, identifying problems, exploring solutions ▷ *understanding western concepts which have no translation (e.g. "risk")* ▷ *transmitting knowledge practices (e.g. food collection/preparation)*	elicit cultural concepts and logics ▷ *exegesis of western cultural concepts into local languages* ▷ *videos of people demonstrating knowledge practices, sharing lore*	build / augment human capacity in sustaining cultural vitality ▷ *redesign of text interaction to follow local teacher-led pattern* ▷ *collection management and navigation, spoken term retrieval*	local decision making, maintaining culture and locally meaningful livelihoods ▷ *equipped with actionable information, people make better-informed choices* ▷ *measure user-reported utility and safety*

Table 1: Etic vs Emic Perspectives on Four Stages of AI Development

and eventually replacing human experts, from software engineers to academic reviewers. This reflects the culture of efficiency and automation underlying the dominant western view (Birhane et al., 2022). An alternative take is to design AI agents to be *augmentative* (Bird, 2024), to enhance human expertise rather than seeking to replicate it. This approach bodes better with minority cultures' matters of concern (Section 2.3), and may well lead to more socially sustainable AI development more broadly as we discuss in Section 3.

Most of our surveyed papers (73%) focussed on intrinsic evaluation and did not consider a 'downstream system' (cf. Sec. 2.1). Those which were more task oriented did not question the design of a potential system, and adopted the predictable designs for IR, classification, or dialogue as established in the culture of NLP.

2.4 Deployment and evaluation

How can we evaluate the success of technology developed from an emic perspective? Typical NLP validation paradigms involve benchmarks, baselines, and scores that can be compared universally, e.g., across methods, languages, and cultures. In keeping with the understanding of emic as local and culture-specific, solutions will vary from place to place. However, emic level evaluation will always center culturally-meaningful ends, including cultural survival, intergenerational knowledge transmission, communal thriving and individual wellbeing (cf. Nigatu et al., 2024). It will be contextualised in concrete application cases and user groups; and it will involve ongoing evaluation monitoring changes in user needs.

Of the minority of task-oriented papers in our survey, none deployed a system in a community and attempted an evaluation in an ecologically valid environment. Validation was primarily based on benchmarks with automated scoring, occasionally augmented with a user study. However, human studies typically took the form of comparison of culturally attuned system A with baseline system B on researcher-defined variables, and were thus constrained in focus and context, and insensitive to user considerations beyond the predefined variables.

3 Paths Forward

How can we navigate into the emic? One option is to reproduce the four stages of AI development (Sec. 2) within a local culture through monocultural deployments. Another is to reproduce mainstream methods under local control and idiomatic interaction designs. A third begins from a relational understanding of intelligence.

Local adaptation. Rather than seeking to contribute cultural data to shape mainstream LLMs, this is a space of deploying language models within local cultures (cf. Mahelona et al., 2023; Biju and Gayathri, 2025; Raj et al., 2025). Harms that come from outsider views and from homogenisation towards dominant norms (cf. Peterson, 2025) are mitigated by restricting training data to insider sources. This remains etic because sources are textual, and they have the limitations described in Section 2.2. However, the benefit of conducting the four stages of AI development within a single cultural area is that the cultural assumptions of users are more likely to align with the cultural tropes that appear in texts provided by cultural insiders.

Local control. Here we ask about approaches to the design of AI which put the local community's aspirations at the centre (e.g. Bird and Yibarbuk, 2024; Cooper et al., 2024; Markl et al., 2024). These may start with seemingly mundane tasks such as text entry and literacy (Moshagen et al., 2023; Pine et al., 2025). Under local control, the interaction design of a chatbot would be reshaped along the lines of local knowledge transmission. Perhaps the initiative is held by teachers, not by learners (who have no entitlement to knowledge). We provide some suggestions in Table 1.

Relational AI. Navigating into the emic involves diverse understandings of intelligence (Bri-

dle, 2022; Lewis et al., 2024), and on approaches not based on the premise of replicating observable human behaviour (Crawford, 2021; Zuboff, 2022). This work exceeds one-size-fits-all thinking (cf. Rehm and Way, 2023; NLLB Team, 2024), and essentialisations of culture (cf. Lenard, 2020). We expect this to lead to parochial and hybrid solutions (cf. Srinivasan, 2017; Bird, 2026). A promising pathway is to develop human resources: increasing the capacity of western technologists to engage respectfully with members of diverse cultures, and increasing the capacity of local communities to conceive language technologies that serve local agendas. This involves rethinking the four stages of AI development under diverse epistemologies (e.g. Kimmerer, 2013; Harjo, 2019).

4 Discussion

In response to constructive feedback, we address the primary counter-arguments below:

(a) The etic/emic distinction is presented as near-categorical rather than a spectrum. This suggests intermediate positions between outsider and insider, and that a model could be iterated from etic to emic, or from forms to meanings, or from behaviours to beliefs. This approach imagines that culture is little more than a set of idioms that can be learnt (Kabra et al., 2023; Attia et al., 2026) or values that can be aligned with (Cao et al., 2024; Kabir et al., 2025). See Zhou et al. (2025) for a critique of this perspective.

Alternatively, one might wonder about incrementally shifting some of the Four Stages of AI Development (cf. Tab. 1, columns). However, in an emic approach, these stages are tightly coupled, strongly influenced by the same underlying culture.

Emic and etic were originally proposed as complementary standpoints on the nature of descriptions of human behaviour (Pike, 1967). Much of the work to date is etic rather than emic, and we contend that research on culture in NLP would benefit from a greater emphasis on emic approaches.

(b) Systems should embed a 'multi-cultural view of knowledge' Conceptually, the emic view by definition considers one culture at a time, rejecting the idea of general cross-cultural structures. Practically, large training data sets are always skewed to majority cultures — quantitatively due to a skew towards majority culture data and/or qualitatively due to short-cut data augmentation methods like translation or generation of additional data. This leads to a dominance of majority culture downstream, compromising the goal of developing technology that is maximally attuned to the needs and goals of the target culture.

(c) NLP should develop/improve the technical capabilities of a system, leaving its integration into specific scenarios to others (e.g., HCI). This is an articulation of the ideology that technology arises in a cultural vacuum. In fact, technological artefacts are social constructs, a reality that is explored in the discipline of Science and Technology Studies (STS). For case studies and reflections that relate to linguistic and cultural diversity, we refer the reader to (Srinivasan, 2017).

(d) Incorporate author demographics into the literature survey as part of determining whether an outsider or insider perspective is applied. Even assuming we could establish author demographics, there remains the effect that researchers from minority cultures whose work reaches ACL venues have already assimilated the culture of NLP with its largely etic approaches. African initiatives such as Masakhane and Lanfrica focus on compiling public datasets and replicating the results of GenAI from dominant languages.

5 Conclusion

We have applied the etic/emic framework to notions of culture that arise in NLP, drawing on a systemic review of the cultural NLP literature. We have described the limits to cultural alignment that are intrinsic to etic approaches, given the layers of culture that cannot be removed from – or that fall outside of – the textual record accessible to LLMs. We have shown that cultural alignment of LLMs is not 'solved' through etic means, by manipulating the representations of LLMs or the interactional style of chatbots. The problem is 'society-complete' (cf. Resnik, 2025, p898), where society differs from culture to culture. We have suggested ways forward, towards a more emic paradigm for research in NLP.

6 Limitations

This paper was shaped by the authors' own cultural perspectives. While our literature survey is *representative* of research related to culture in NLP, we do not claim that it is *exhaustive* and we might have missed a small number of more emic approaches. Still, our main conclusions that etic approaches dominate NLP and that a more emic angle would benefit the utility and value of cultural NLP work still holds.

Acknowledgements

We are grateful for the reviewers' constructive feedback. Frermann is supported by the Australian Research Council Discovery Early Career Research Award (Grant No. DE230100761). Bird is supported by Australian Research Council Discovery Projects (DP210100228, DP240101952). The authors made an equal contribution to the research reported here.

References

Mai Alkhamissi, Yunze Xiao, Badr AlKhamissi, and Mona T. Diab. 2026. Hire your anthropologist! rethinking culture benchmarks through an anthropological lens. In *Findings of the Association for Computational Linguistics: EACL 2026*, pages 1218–1235. ACL.

Mena Attia, Aashiq Muhamed, Mai Alkhamissi, Thamar Solorio, and Mona T. Diab. 2026. Beyond understanding: Evaluating the pragmatic gap in LLMs' cultural processing of figurative language. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*, pages 7238–7265. ACL.

Keith H. Basso. 1974. The ethnography of writing. In Richard Bauman and Joel Sherzer, editors, *Explorations in the Ethnography of Speaking*, pages 425–32. Cambridge University Press.

P. R. Biju and O. Gayathri. 2025. Indic approach to ethical AI in automated decision making system: implications for social, cultural, and linguistic diversity in native population. *AI and Society*, 40(8):6645–6670.

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

Steven Bird. 2026. Towards a general theory of linguistic diversity. In *Towards Inclusivity and Equality: Language Resources and Technologies for Under-Resourced and Endangered Languages, Proceedings of the SIGUL 2026 Workshop*, pages 169–182. ELRA.

Steven Bird and Dean Yibarbuk. 2024. Centering the speech community. In *Proceedings of the 18th Conference of the European Association for Computational Linguistics*, pages 826–839. ACL.

Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The values encoded in machine learning research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 173–184. ACM.

Robert Borofsky. 2005. *Yanomami: The fierce controversy and what we can learn from it*. University of California Press.

James Bridle. 2022. *Ways of Being: Animals, Plants, Machines: The Search for a Planetary Intelligence*. Penguin.

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

E. Summerson Carr. 2010. Enactments of expertise. *Annual Review of Anthropology*, 39:17–32.

James Clifford and George Marcus. 2023. *Writing Culture: The Poetics and Politics of Ethnography*. University of California Press.

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*, pages 204–211. ACL.

Kate Crawford. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Matthew Engelke. 2018. *How to Think Like an Anthropologist*. Princeton University Press.

Joshua A. Fishman, editor. 2001. *Can Threatened Languages be Saved?: Reversing Language Shift, Revisited: A 21st Century Perspective*. Multilingual Matters.

François Grosjean. 2021. *Life as a Bilingual: Knowing and Using Two or More Languages*. Cambridge University Press.

Edward T. Hall. 1976. *Beyond Culture*. Garden City.

Laura Harjo. 2019. *Spiral to the Stars: Mvskoke Tools of Futurity*. University of Arizona Press.

Thomas Headland, Kenneth Pike, and Marvin Harris, editors. 1990. *Emics and Etics: The Insider/Outsider Debate*. Sage Publications.

Ben Hutchinson. 2024. Modeling the sacred: Considerations when using religious texts in natural language processing. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1029–1043. ACL.

Mohsinul Kabir, Ajwad Abrar, and Sophia Ananiadou. 2025. Break the checkbox: Challenging closed-style evaluations of cultural alignment in LLMs. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 24–51. ACL.

Anubha Kabra, Emmy Liu, Simran Khanuja, Alham Fikri Aji, Genta Winata, Samuel Cahyawijaya, Anuoluwapo Aremu, Perez Ogayo, and Graham Neubig. 2023. Multi-lingual and multi-cultural figurative language understanding. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8269–8284. ACL.

Robin Wall Kimmerer. 2013. *Braiding Sweetgrass: Indigenous Wisdom, Scientific Knowledge and the Teachings of Plants*. Penguin.

Patti Tamara Lenard. 2020. Culture. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University.

Jason Edward Lewis, Hēmi Whaanga, and Ceyda Yolgörmez. 2024. Abundant intelligences: Placing AI within Indigenous knowledge frameworks. *AI and Society*, 40:2141–2157.

Keoni Mahelona, Gianna Leoni, Suzanne Duncan, and Miles Thompson. 2023. OpenAI's whisper is another case study in colonisation. *Papa Reo*. Accessed 20250215.

Nina Markl, Lauren Hall-Lew, and Catherine Lai. 2024. Language technologies as if people mattered: Centering communities in language technology development. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 10085–10099.

Donald Metzler, Yi Tay, Dara Bahri, and Marc Najork. 2021. Rethinking search: Making domain experts out of dilettantes. *SIGIR Forum*, 55(1):1–27.

Quulgan Minggad, Xiao Zinan, and Yuan Sun. 2026. MonCulture-eval: A hierarchical benchmark for evaluating Mongolian cultural capabilities of large language models across scripts and regions. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 28997–29014. ACL.

Benjamin Mitchell and William Fleischman. 2020. Responsibility in the age of irresponsible speech. In Mario Arias-Oliva, Jorge Pelegrín-Borondo, Kiyoshi Murata, and Ana María Lara Palma, editors, *Societal Challenges in the Smart Society*, pages 587–592. Universidad de La Rioja.

Sjur Nørstebø Moshagen, Flammie Pirinen, Lene Antonsen, Børre Gaup, Inga Mikkelsen, Trond Trosterud, Linda Wiechetek, and Katri Hiovain-Asikainen. 2023. The GiellaLT infrastructure: A multilingual infrastructure for rule-based NLP. In Arvi Hurskainen, Kimmo Koskenniemi, and Tommi Pirinen, editors, *Rule-Based Language Technology*, pages 70–94. Northern European Association for Language Technology.

Isar Nejadgholi, Masoud Kianpour, Krishnapriya Vishnubhotla, and Maryam Molamohammadi. 2026. Defining cultural capabilities for AI evaluation: A taxonomy grounded in intercultural communication theory. In *Proceedings of the Workshop on Evaluating Evaluations*, pages 161–173. ACL.

Hellina Hailu Nigatu, Atnafu Lambebo Tonja, Benjamin Rosman, Thamar Solorio, and Monojit Choudhury. 2024. The Zeno's paradox of 'low-resource' languages. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17753–17774. ACL.

NLLB Team. 2024. Scaling neural machine translation to 200 languages. *Nature*, 630:841–846.

Juhyun Oh, Inha Cha, Michael Saxon, Hyunseung Lim, Shaily Bhatt, and Alice Oh. 2025. Culture is everywhere: A call for intentionally cultural evaluation. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 19156–19168. ACL.

Rafael Rivera Pastor, Carlota Tarín Quirós, Juan Pablo Villar García, Toni Badia Cardús, and Maite Melero Nogués. 2017. *Language Equality in the Digital Age: Towards a Human Language Project*. European Parliament.

Siddhesh Pawar, Junyeong Park, Jiho Jin, Arnav Arora, Junho Myung, Srishti Yadav, Faiz Ghifari Haznitrama, Inhwa Song, Alice Oh, and Isabelle Augenstein. 2025. Survey of cultural awareness in language models: Text and beyond. *Computational Linguistics*, 51(3):907–1004.

Andrew J. Peterson. 2025. AI and the problem of knowledge collapse. *AI and Society*, pages 1–21.

Kenneth L. Pike. 1967. Etic and emic standpoints for the description of behavior. In Kenneth L. Pike, editor, *Language in relation to a unified theory of the structure of human behavior*, 2nd edition, pages 37–72. Mouton & Co.

Aidan Pine, Erica Cooper, David Guzmán, Eric Joanis, Anna Kazantseva, Ross Krekoski, Roland Kuhn, Samuel Larkin, Patrick Littell, Delaney Lothian, Akwiratékha' Martin, Korin Richmond, Marc Tessier, Cassia Valentini-Botinhao, Dan Wells, and Junichi Yamagishi. 2025. Speech generation for indigenous language education. *Computer Speech and Language*, 90:101723.

Karl Popper. 1935. *The Logic of Scientific Discovery*. Routledge.

Abhishek Raj, Harsh Singh, and Anshul Pachouri. 2025. Harnessing the potential of AI for Indian agriculture: Using "Bhashini" as a tool to deploy responsible AI and increase the uptake of AI applications among farmers. In Balaraman Ravindran and Abhishek Singh, editors, *Advancing Responsible AI in Public Sector Application*, pages 119–128. CRC Press.

Georg Rehm and Andy Way. 2023. *European Language Equality: A Strategic Agenda for Digital Language Equality*. Springer Nature.

Philip Resnik. 2025. Large language models are biased because they are large language models. *Computational Linguistics*, 51:885–906.

Edward W. Said. 1978. *Orientalism*. Penguin Random House.

Daniel Schacter, Daniel Gilbert, and Daniel Wegner. 2009. *Psychology*. Macmillan.

Chirag Shah and Emily M. Bender. 2022. Situating search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, pages 221–232. ACM.

Ramesh Srinivasan. 2017. *Whose Global Village?: Rethinking how Technology Shapes Our World*. NYU Press.

Larry Sullivan. 2009. Emic and etic perspectives. In *The SAGE Glossary of the Social and Behavioral Sciences*, pages 174–175. Sage Publications.

Paul Ten Have. 1989. The consultation as a genre. In Brian Torode, editor, *Text and Talk as Social Practice*, pages 115–135. Foris Publications.

Maggie Walter and Chris Andersen. 2013. *Indigenous Statistics: A Quantitative Research Methodology*. Taylor and Francis.

Joseph Weizenbaum. 1966. ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9:36–45.

Janie Wilson, Rebecca Ryan, and James Pugh. 2010. Professor–student rapport scale predicts student outcomes. *Teaching of Psychology*, 37(4):246–251.

William A. Woods. 1973. Progress in natural language understanding: an application to lunar geology. In *Proceedings of the National Computer Conference and Exposition*, pages 441–450. ACM.

Naitian Zhou, David Bamman, and Isaac L. Bleaman. 2025. Culture is not trivia: Sociocultural theory for cultural NLP. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, pages 25869–25886.

Shoshana Zuboff. 2022. Surveillance capitalism or democracy? The death match of institutional orders and the politics of knowledge in our information civilization. *Organization Theory*, 3(3):1–79.

A Literature Survey

Our main paper results are supported by a survey of the recent NLP literature on culture, with the aim to obtain a *representative* picture of the treatment of culture in recent NLP research. Our literature is systematic, but not exhaustive. We discuss a broader follow-up survey in Section A.1.

We started by obtaining all papers published between 2022 and April 2026 in the main or findings track of a top-tier NLP conference (*ACL, EMNLP) which mention 'culture' or 'cultural' in their title. To keep the scope for manual paper inspection manageable, we removed multimodal approaches, and papers whose primary focus is not culture. We reviewed a total of 107 papers. We manually labelled these papers for the following information:

Primary contribution (resource, method, analysis, survey)

Data creation (predominantly manual, at least partially synthetic, based on a broad background corpus, ...). Only for empirical papers (i.e., excluding surveys)

Task Assessment/improvement of cultural knowledge in LLMs (upstream or intrinsic) vs. task-oriented papers proposing culturally sensitive applications (downstream or extrinsic). Only for empirical papers (i.e., excluding surveys).

The full list of reviewed papers and labels is provided on GitHub.[1] Summary statistics underpinning our analyses in Section 2 are provided in Tables A.1 – A.3.

A.1 Expanded Survey

We complemented our survey, retrieving all papers in the ACL Anthology which mentioned the word 'culture' or 'cultural' in their abstract or title (N=1536). These were infeasible to manually analyse so we instead searched for an expanded list of keywords indicating an emic approach.[2] 108 papers mentioned at least one of these terms. Even with the very optimistic assumption that all these papers take an emic approach, they would make up for 7% of our sample. We note that again this is a crude approximation because both the denominator (there will be other papers about culture which did not meet our initial keyword filter) and the numerator (there may be emic papers not mentioning any of our expanded keywords, and many matched papers won't be emic) are noisy. We provide this complement survey as an additional data point, complementary to our more focussed manual analysis.

Category	N	%
Resource	50	0.467
Survey	8	0.075
Analysis	15	0.140
Method	31	0.290

Table A.1: Distribution over main contributions in papers in our survey.

Category	N	%
Assess / augment cultural knowledge in LLMs	78	0.729

Table A.2: Fraction of paper with a focus on assessing or augmenting cultural knowledge in LLMs.

Category	N	%
culture annotators	19	0.178
broad data	41	0.383
synthetic	33	0.308
translation	3	0.028
– / LLM output analysis	4	0.037
– / survey	8	0.075

Table A.3: Distribution over data set creation approaches in the papers in our survey. Papers of the last two categories were not included, as they did not contribute data sets or investigated culture in LLM output (rather than human artifacts).

[1] https://github.com/ColiLea/emic-culture-nlp/

[2] emic, etic, community-centric, community centric, community-centered, community centered, community-focussed, community focussed, community-focussed, community focussed, anthropological, anthropology, insider, outsider, intercultural