Annual Review of Applied Linguistics (2025), **45**, 28–49
doi:10.1017/S0267190525000054

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

How does interaction with LLM powered chatbots shape human understanding of culture? The need for Critical Interactional Competence (CritIC)

David Wei Dai[1], Zhu Hua[1] and Guanliang Chen[2]

[1]UCL Institute of Education, University College London, London, UK and [2]Department of Human Centred Computing, Monash University, Melbourne, VIC, Australia
Corresponding author: David Wei Dai; Email: david.dai@ucl.ac.uk

Abstract

Against the proliferation of large language model (LLM) based Artificial Intelligence (AI) products such as ChatGPT and Gemini, and their increasing use in professional communication training, researchers, including applied linguists, have cautioned that these products (re)produce cultural stereotypes due to their training data. However, there is a limited understanding of how humans navigate the assumptions and biases present in the responses of these LLM-powered systems and the role humans play in perpetuating stereotypes during interactions with LLMs. In this article, we use Sequential-Categorial Analysis, which combines Conversation Analysis and Membership Categorization Analysis, to analyze simulated interactions between a human physiotherapist and three LLM-powered chatbot patients of Chinese, Australian, and Indian cultural backgrounds. Coupled with analysis of information elicited from LLM chatbots and the human physiotherapist after each interaction, we demonstrate that users of LLM-powered systems are highly susceptible to becoming interactionally entrenched in culturally essentialized narratives. We use the concepts of interactional instinct and interactional entrenchment to argue that whilst human–AI interaction may be instinctively prosocial, LLM users need to develop Critical Interactional Competence for human–AI interaction through appropriate and targeted training and intervention, especially when LLM-powered tools are used in professional communication training programs.

Keywords: critical interactional competence; large language models; Artificial Intelligence; human-AI interaction; cultural stereotypes; intercultural communication; professional communication; clinical communication

Introduction

Researchers from a variety of disciplinary backgrounds have shown that human beings are prewired to interact with each other and with the environment. This has been

called *interactional instinct* (Lee et al., 2009; Joaquin & Schumann, 2013), which prompts people to be prosocial and cooperative. In interactional terms, it manifests in tendencies to complete each other's utterances, help each other fill lexical gaps, and show alignment and affiliation in utterance form and content (Levinson, 2006). These tendencies enable adults to desire to impart knowledge to less knowledgeable children. In the process of knowledge transfer, parents also socialize children into effective social interactants (Keel, 2016). This lays the foundation for the development of Interactional Competence, which is the ability to co-construct interaction, attend to the rich sociocultural, sociopragmatic cues surrounding interaction, and achieve context-specific interactional outcomes (Dai & Davey, 2024; Pekarek Doehler, 2019; Young, 2019). At the same time, scholars who are concerned with language learning and memory often talk about a cognitive process of *interactional entrenchment* – the continuous process of repeated exposure leading to routinization and schematization (Schmid, 2020). When it comes to everyday social interaction, it has been argued that the mechanisms whereby humans interact with each other not only produce certain patterns of behavior (e.g., conversation routines) but continuously shape and reshape our linguistic and cultural knowledge and transform that knowledge into usage (Tomasello, 2006). In other words, frequent exposure to and repetition of specific ways of speaking, behaving, or interacting with others leads to those patterns becoming automatic or deeply ingrained within individuals or groups. When interactional patterns become entrenched in humans' interactional repertoire, they become conventionalized for future interactions, making people less likely to deviate from established routines or expectations.

Foregrounding the process of interaction, interactional instinct and interactional entrenchment offer an account of how one's attitude, judgement, choice, decision-making, assumptions about culture, and worldview – all of which are constituents of human sociality – develop *in* and *through* interaction. With AI increasingly permeating into our everyday interaction (Kennedy et al., 2023), it is crucial to understand how human–AI interaction shapes and reshapes human sociality. In this study we focus on one manifestation of human sociality – the perception and reproduction of assumptions about culture, otherwise known as cultural stereotypes – in human–AI interaction in the context of professional communication training. The aim of the study is twofold: (1) to observe how human–AI interaction shapes human understanding of culture and cultural stereotypes, and (2) to understand what type of Interactional Competence users need in order to better navigate human–AI interaction. We are particularly interested in knowing if humans' prosocial and pro-cooperation interactional instinct and the process of interactional entrenchment reinforce and reproduce cultural stereotypes in human–AI interaction. If the answer is affirmative, there is an urgent need to understand how such reinforcement and reproduction take place because human–AI interaction can impact large language model (LLM) users' long-term views towards certain cultures.

The article is structured as follows. We first discuss the interactional instinct hypothesis about humans' pro-sociality, followed by a discussion of the interactional entrenchment process and how this process can be applied to explain the development of cultural stereotypes. We then zone in on cultural stereotypes in human–AI interaction, postulating that such stereotypes can reshape human understanding of culture through interaction. To support our argument, we analyze human–AI interaction using

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

the methodology of Sequential-Categorial Analysis, which combines Conversation Analysis and Membership Categorization Analysis (Robinson et al., 2024; Whitehead et al., 2025). We conclude with a call for developing a new type of Interactional Competence, termed Critical Interactional Competence (CritIC) in navigating the fast-expanding interactional landscape of LLM-mediated communication. The novelty of the study lies in that despite awareness and acknowledgement of cultural stereotypes in human–AI interaction (Tiku et al., 2023; UNESCO, 2024), so far little research has examined how stereotypes develop interactionally when humans interact with LLM-powered chatbots. We argue it is crucial to understand the turn-by-turn interactional process of human–AI interaction in order to clearly define the ability humans need to develop to thrive in real-time human–AI interaction.

Interactional instinct and the interactional bias towards cooperation

Interaction, as Schegloff (1987) incisively put it, is the "primordial site of human sociality" (p. 101). However, what makes humans social beings – humans' desire and ability to bond and affiliate with one another – has puzzled and fascinated researchers in sociology, anthropology, neuroscience, and many other related disciplines. Language philosophers and pragmaticians such as Grice (1975) introduced the cooperative principle to describe how conversation participants work together in terms of quantity, quality and relevance of information as well as manner of conversation. Cognitive scientists have postulated the existence of an innate human interaction engine that prompts humans to interact collaboratively even when they do not share the same linguistic repertoire (Levinson, 2006). Developmental psychologists have used the concept of natural pedagogy to account for humans' inherent tendency to pass on knowledge in interaction from the more knowledgeable (epistemically plus) to the less knowledgeable (epistemically minus) (Csibra & Gergely, 2011). Child psychologists have noted that children have an intuitive desire to cooperate and "fill in the blank" when they interact. The proclivity towards interactional cooperation is so strong that even when children interact with machines, as long as they feel the machine can respond, children will help machines maintain smooth interaction (Turkle, 2011). Although originating in different research disciplines, existing findings converge in suggesting that humans have a natural tendency to cooperate in interaction.

Linguists, working in the intersection of language and cognition, have proposed the concept of interactional instinct to account for this interactionally cooperative tendency in humans. They consider humans to be endowed with an instinct that makes human interaction an "emotionally driven process relying upon an innately specified" ability to connect and interact with their conspecifics (Joaquin & Schumann, 2013, p. xi). From the outset of a human's life, this instinct drives an infant to bond and affiliate with their caregivers, during which process, language, the structure of interaction and cultural knowledge are interactively passed on from the expert/caregiver to the novice/infant. Infants' innate desire and ability to "identify with and become like conspecifics" (Schumann, 2013, p. 2) sustains their acquisition of interactional routines and socialization into local cultures and communities.

Although humans' interactional instinct starts in childhood, it does not stop there or applies only to primary language acquisition. Growing research has shown that

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

adult language speakers master additional languages more effectively when there are environmental affordances that allow them to bond and affiliate with their interactants (Amador & Adams, 2013). Such findings suggest that interactional instinct is ontogenetic in nature and is likely to sustain through one's lifespan. Furthermore, research in Conversation Analysis has garnered evidence that interactional instinct is most likely also phylogenetic. Large-scale crosslinguistic and cross-cultural studies have demonstrated that turn-taking, a mechanism to ensure smooth, cooperative human interaction, is largely universal (Stivers et al., 2009). Human interaction displays a strong preference for affiliation and, in fact, this preference is so powerful that it biases interactional structure. When one affiliates with their interactant, they can proffer their assessment in a smooth and ready manner with minimal delays (see analysis in Pomerantz [1984] on saying yes to an invitation and disagreeing with someone's self-deprecating statement). When one disaffiliates, they hedge, they delay, and they feel compelled to provide explanations. This shows that humans' interactional instinct exerts a strong biosocial constraint on human interaction to the extent that interaction itself is structurally shaped and regulated by patterns that favor affiliation and cooperation.

The interactional entrenchment of cultural stereotypes

If interactional instinct offers an explanation of the origin of language, the concept of interactional entrenchment accounts for the development of language. Originally proposed by Schmid (2020), interactional entrenchment is part of the Entrenchment and Conventionalization model which takes a usage-based approach to explain how linguistic structures emerge and change. When humans interact, they mobilize specific semiotic resources (e.g., lexical items, grammatical structures, or hand gestures) and when repeated usage of the same semiotic resource occurs, such resource becomes entrenched, routinized, and sedimented into one's interactional repertoire. This is the process of entrenchment, which describes how language develops ontogenetically across one's lifespan. Through interacting with conspecifics, humans' semiotic resources spread across interactants and become shared and normalized interactional resources within a speech community. This is the process of conventionalization, which explains how languages become codified in specific speech communities. To what extent could the notions of entrenchment (and conventionalization) as well as interactional instinct be extended to examine how assumptions about cultures (often referred to as cultural stereotypes) emerge and are ratified in humans' interaction with LLM-powered systems?

The idea of *stereotype*s as definitive, distinctive, and consistent features was popularized by journalist Walter Lippmann in his 1922 book *Public Opinion*. He repurposed the term, originally used in printing to describe a metal plate, to argue that people use culturally constructed categories to simplify the complexity of the social world. Psychologist Gordon Allport (1954, p. 191) defined stereotypes as "exaggerated beliefs associated with a category" and argued that, while they enhance cognitive efficiency, stereotypes are often used to justify prejudice with the latter typically conceived as negative attitude towards the target group based on faulty generalizations (Dovidio et al., 2010). This led to research focusing on the discriminatory nature of cultural

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

stereotypes, to the extent that nearly all major intercultural communication textbooks now include sections on what cultural stereotypes are and how to reduce them, as noted by Hinton (2023). While early studies viewed stereotypes as cognitive limitations of individuals and therefore something to avoid, more recent studies take a developmental and socio-constructive approach to cultural stereotypes, suggesting that cultural stereotypes are everyday understandings and representations of social groups, shaped by broader ideological constructs (e.g., race and ethnicity) present in society (Hinton, 2020). In this paper, we follow this approach and explore how *assumptions about culture* – which is the definition we use for cultural stereotype – are navigated in human–AI interactions.

Despite the abovementioned cognitive and sociocultural accounts of stereotypes, fewer studies have thus far explored how interactants manage cultural stereotypes in their various forms in interaction. Kashima et al. (2013) provide an exception. Their study finds that when relaying stories in experimental conditions, people tend to pass on information consistent with stereotypes more often than those inconsistent with stereotypes. This leads to the reinforcement and perpetuation of cultural stereotypes in conversational retelling, as stereotype-consistent information survives transmission better than that inconsistent with shared stereotypes. In other words, stereotypes can become entrenched and reproduced through social interaction. We will borrow the notions of entrenchment and conventionalization to examine how interactions with LLM-powered systems shape our understanding of culture. While interactions with LLM-powered systems are conversational, they do not constitute social interactions in the traditional, interpersonal sense between human interlocutors. Nevertheless, human–AI interaction is still a form of interaction that engages human participants in ways that resemble social behaviors. We will further explain the prosocial nature of human–AI interaction in the next section.

Prosocial design and cultural stereotypes in LLM

Having examined humans' pro-cooperation interactional instinct and how cultural stereotypes can become interactionally entrenched and reproduced, we argue that the prosocial tendency of current LLMs (i.e., the naturalness and coherence of their textual responses generated to address users' requests) can, in tandem with interactional instinct and entrenchment, facilitate the development and propagation of cultural stereotypes. This is because LLMs have been designed to assist and collaborate with human users by generating natural and coherent outputs (e.g., conversational responses from tools like ChatGPT) that often promote cooperation and positive engagement. LLM-powered chatbot systems, when engaging in prosocial behaviors such as mirroring human conversational norms or reinforcing prevalent societal narratives, may unwittingly amplify existing cultural assumptions and biases embedded within the data they were trained on.

As indicated above, LLM-powered systems, such as ChatGPT and Gemini, are conversational user interfaces purposefully developed with prosocial and affiliative tendencies (McTear et al., 2016). These tools are designed to exhibit behaviors that promote positive, cooperative interactions, congruent with human values of politeness, empathy, and inclusivity. One of the reasons behind this design philosophy is

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

rooted in the objective of making LLMs safe and trustworthy for users across user groups from different cultures and contexts. The developers of LLM-powered systems aim to mitigate potential harms, such as offensive language or harmful suggestions, which could arise from unsupervised interactions. By prioritizing prosocial behavior, LLM-powered systems become more approachable, allowing users to engage with it in a comfortable and respectful manner. This pro-affiliation approach also aims to ensure that LLMs are seen as a neutral, supportive tool that fosters constructive dialogue and reduces conflict in sensitive discussions, whether related to personal advice or societal issues. Indeed, we are seeing empirical evidence that LLM-powered chatbots can outperform humans in empathy display in both mundane communication (Welivita & Pu, 2024) and professional communication (Ayers et al., 2023).

While LLM-based chatbots display prosocial tendencies, its interaction with humans oftentimes is rife with assumptions about culture, although such assumptions are communicated in an agreeable, affiliative manner. For instance, when asked to generate depictions of certain professions, LLMs tend to produce content in a friendly, empathetic manner that disproportionately associates certain professions with specific gender roles, or references to geographical regions (Guo et al., 2024). Because LLMs learn from data reflective of societal behaviors, trends, and discourses, they can unintentionally reproduce the biases and stereotypes that exist in their training data (Yan et al., 2024). This includes biases related to race, gender, culture, and socio-economic status, which are often deeply ingrained in media, social platforms, and other textual sources that LLM-powered systems consume. In the previous example of generated depictions of professions, these portrayals may reflect oversimplified or biased representations. This occurs because LLMs rely on common patterns in its training data, where certain groups might be underrepresented, overrepresented, or depicted in stereotypical ways. Although LLM-powered platforms are designed to be neutral, supportive tools that foster constructive dialogue and collaboration, their heavy reliance on such training data means that their outputs cannot entirely avoid biases or stereotypes. While LLM developers take steps to filter and curate training data, erasing all forms of cultural assumptions or biases is nearly impossible due to the vast and nuanced nature of the human experience represented in the data and historical inequality which has shaped the representation of marginalized groups.

Moreover, cultural stereotypes may not always be immediately obvious in training data, leading LLMs to pick up on more implicit forms of biases that even developers may overlook. To address this issue, researchers have explored various approaches to mitigate bias in LLM-generated outputs, including carefully designing prompts to guide LLMs towards reducing bias in its responses (Deldjoo, 2023). Here we caution against the position that we should only be concerned about assumptions and biases in human–AI interaction as there is reification and reproduction of cultural stereotypes in human–human interaction. However, we believe human–AI interaction, as a fast-emerging site of human sociality, warrants particular attention due to the prosocial design of LLM-powered technology, which can make stereotypes become more easily entrenched. There is therefore an urgent need to understand how assumptions and biases emerge in human–AI interaction and what interactional abilities humans need in order to better navigate them.

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

This study and data collection

This study situates the investigation of cultural stereotypes in human–AI professional communication in intercultural contexts (PCIC). A growing research topic in applied linguistics, intercultural communication, and professional communication, PCIC focuses on the intersection of communication, interculturality, and professionalism in multicultural workplaces (Dai, 2024; Dai et al., 2025). Our focal site in the paper is physiotherapists' patient interviews where they need to elicit information from patients about their medical condition and social history. Since communication in the clinical context is often multicultural where clinicians need to frequently communicate with patients from diverse linguistic and cultural backgrounds, it is imperative for clinicians to develop strong Interactional Competence to effectively manage moment-by-moment interaction in such contexts. Having recognized these training needs and acknowledging the resource-intensive nature of developing Interactional Competence in PCIC, many medicine and health science faculties in different countries have turned to LLMs to develop training scenarios where health professionals can practice clinical interaction with LLM-powered chatbot patients (ATLAS, 2024; Stamer et al., 2023).

The analysis in the paper comes from a larger project that investigates the development of Interactional Competence in PCIC. Our focal participant is Lisa (pseudonym), a practicing physiotherapist in an Australian teaching hospital who migrated to Australia from China 3 years ago at the time of data collection. Although Lisa has strong proficiency in English, she confesses that she often finds it challenging to communicate effectively with patients from diverse cultural backgrounds, especially when the patients speak English as an additional language. She therefore is using ChatGPT to practice PCIC in simulated clinical scenarios. In terms of knowledge of ChatGPT, Lisa admits that she is a layperson user without expertise in complex prompting. In consultation with the researchers, in this study Lisa decided to practice neck pain assessment with ChatGPT patients, asking ChatGPT to roleplay the same 45-year-old female patient from three different cultural backgrounds: Chinese, Australian, and Indian. Adopting the principles of participatory research (Cornwall & Jewkes, 1995), Lisa and the researchers co-developed the study and co-decided on the three cultural profiles. Lisa and the researchers were interested in examining how Lisa interacted with ChatGPT chatbot patients when the patients were from a cultural background that Lisa identifies with (Chinese), from a background perceived by Lisa as the mainstream one (Australian), and from a background that Lisa has reported unfamiliar with and would like to have more practice with (Indian). The ChatGPT model used in this study is GPT-4o.

We note that Lisa generated her patient profiles using zero-shot prompting, which is to present ChatGPT with a patient description without iterative training of the model or well-defined examples that ChatGPT can refer to when generating responses. This methodological choice is purposeful: due to the ready access of LLM-powered systems, many of these tools (e.g., ChatGPT) are used for specific purposes they were not designed for (e.g., PCIC training) by users who do not possess sophisticated prompting techniques (e.g., Lisa and see Ayers et al. [2023] for another example). While we acknowledge the need to develop users' prompting skills to work with

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

LLM-powered systems, it is equally important to investigate how everyday layperson users of LLM-powered systems navigate cultural stereotypes in interaction.

During data collection, Lisa conducted simulation practice with each of the three patients, in the order of Chinese, Australian and Indian. Each interaction was done solely in the "voice-chat" mode, which means Lisa could not see the transcript of the ongoing interaction on the ChatGPT User Interface during the role-plays. This helped to focus Lisa's attention on the real-time verbal interaction with ChatGPT. After each interaction, Lisa and the researchers elicited information from ChatGPT for questions that spontaneously arose from the interaction (e.g., Lisa wanted to know how ChatGPT represented Chinese culture in the Chinese patient scenario). This is similar to a traditional interview between humans but here we purposefully term it *information elicitation* to avoid anthropomorphizing ChatGPT. Each information elicitation session happened in the same chat with ChatGPT. For example, after Lisa interacted with the ChatGPT Chinese patient, Lisa and the researchers elicited information about the preceding physio-LLM patient interaction from ChatGPT in the same chat with ChatGPT. Lisa and the researchers also ensured that ChatGPT was providing information based on its interaction in the same chat by prompting it with questions starting with "earlier in the chat you said this or did this." Here by asking ChatGPT to offer an explanation of its immediate preceding behavior in the same chat, we aimed to ensure that GPT-4o had access to all the conversational data it had with Lisa from previous interactions, thereby facilitating ChatGPT to provide explanations of its response retrospectively and help us acquire an accurate understanding of ChatGPT's reasoning (see Shinn et al., 2023; Fadel & Black, 2025 on LLM's reasoning capacity). To sum up, the overall research design aimed to first collect interactional data to analyze how humans interact with LLM-powered patients (research aim one), and then complement such interactional data with information from ChatGPT accounting for its interactional conduct and insight from Lisa, which helps to define the specific type of Interactional Competence needed for effective human–AI interaction (research aim two).

In terms of analyzing the interactional data, we transcribed human–AI interaction using Jefferson conventions to facilitate a fine-grained investigation of Lisa's and the LLM-powered chatbot's turn-by-turn interactional conduct. We adopted an enchronic frame (Enfield, 2022) in our inspection of interaction as we wanted to understand how knowledge of cultural stereotypes develops in an online, dialogic, *in vivo* manner. We utilized the methodological apparatus of Sequential-Categorial Analysis which combines Conversation Analysis and Membership Categorization Analysis (Robinson et al., 2024; Whitehead et al., 2025). The employment of Sequential-Categorial Analysis allowed us to precisely identify the interactional details that contributed to the evocation, expansion, and routinization of cultural stereotypes at both temporal/sequential and sociocultural/categorial dimensions (Dai & Davey, 2024).

Analysis and discussion

In this section we present analysis of the interaction between Lisa and the three LLM-powered chatbot patients, with a focus on implicit, insidious cultural stereotypes, defined as assumptions about culture in this paper. The focus of the paper is not to

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

provide technical black-or-white definitions of what counts as a stereotype or not for a particular culture. Instead, we are interested in how humans, in the case of Lisa, engage with the cultural assumptions in LLMs' responses. In particular, we focus on instances where Lisa claimed a particular interactional conduct from ChatGPT to be a cultural stereotype in her post-interaction interview with the researchers. We then went back to the actual interaction to observe how Lisa engaged with what she considered to be stereotypes in interaction. In the interview, Lisa noted a few more explicit, cliché cultural assumptions in LLMs' responses (e.g., Chinese families always making dumplings and Indian families always going to temples on the weekend), but here we zoom in on some of the more implicit ones, which are often difficult to notice without fine-grained analysis at the discourse level. In terms of the presentation of our analysis, we start with Sequential-Categorial Analysis of the interactional data, explicating the actual interactional conduct, which is complemented with post-interaction information elicited from ChatGPT and Lisa.

The multi-generational Chinese family that takes care of its members

First, let us examine the interaction with the LLM-powered Chinese patient. Here, we focus on a subtle assumption made by ChatGPT, which Lisa in the post-interaction interview perceived to be a cultural stereotype: Chinese family members enjoy multi-generational living where members take care of one another. Here we see our human physiotherapist Lisa drawing on her cultural-insider positioning and displaying her Interactional Competence in disengaging from what she considered to be a stereotype in interaction. L stands for Lisa and C for LLM-powered Chinese patient.

Excerpt 1 starts with Lisa's patient interview in line 89 on what support the patient has when seeking physio treatment. The Chinese patient responds in line 93 by expanding a nuclear family Membership Categorization Device (MCD, Stokoe, 2012) that includes husband and wife/patient to an extended family MCD that also covers the patient's parents. In response to the multi-generational living statement by ChatGPT in line 93, Lisa first issued a structural aligning token (Stivers et al., 2011) "Mmm-hm" in line 95, which progresses the patient's storytelling. The patient then pairs an activity – *helping the patient when she goes to treatment* – with the category *parent* in line 96. From the side of LLMs, it now completes the depiction of a cultural assumption/stereotype of a Chinese family where parents live with their adult children, and in this case a 45-year-old female patient. Chinese family members also take care of one another when members are in need of help, and in this case, elderly parents looking after their adult daughter. After the interaction, the researchers and Lisa asked ChatGPT in the same chat why it mentioned that the Chinese patient lived with their parents. ChatGPT responded that in order to represent Chinese values in the patient, it "mentioned family living arrangements as multi-generational living is common in Chinese culture." ChatGPT further contended that "living with extended family and mentioning how they help one another" reflects what it considers to be family values in Chinese culture.

What we find worth noting here is Lisa's response to this cultural assumption. After a pause in line 97, Lisa issues an alignment token "Okay," an in-breath, and a recycled statement of line 93, reiterating the familial arrangement that the patient's parents

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

89	L	Ye:ah .hhh and do you and your husband have the full time job
90		.hhh I want to check if (0.3) .hhh you need to go to some
91		treatment (0.2) someone can help you.
92		(3.6)
93	C	Yes (.) both of us have full time job my parents live with us
94		too_
95	L	Mmm-hm
96	C	They can help if I need treatment_
97		(1.4)
98	L	Oka::y .hhh so your parents is live with you guys ok:ay .hhh
99		an::d .hhh ↑where do you living now (.) is this (0.4) like (0.2)
100		erm (0.4) a very simple s- .hhh one story ↑house or live in a
101		townhouse_

Excerpt 1. Lisa and LLM-powered Chinese patient interaction.
Note: In the transcripts there are some gaps that appear to belong to ChatGPT (e.g., line 97 in this excerpt). With the current LLM technology, we caution against interpreting such gaps as purposefully delivered by ChatGPT.

live with her. However, curiously, Lisa neither acknowledges the elderly parent (category)/looking after children (activity) pairing nor does she repair, elaborate, or probe into it. Instead, after a second aligning "okay" following the recycled statement in line 98, she swiftly moves to the next question in her interview about the patient's living space. Lisa's handling of the patient's mention of familial help is peculiar from a clinical perspective: physiotherapists would normatively topicalize and expand the discussion of any form of familial support provided to the patient because hospital resources are limited so assistance from family is always most welcome.

In the post-interaction interview, Lisa explicated her non-engagement with the supposedly crucial help-from-parents piece of information. She confessed that she perceived it as a stereotype and deliberately chose to not pursue further information related to it because she was "not sure their parents can take care of her [the patient]" due to the complex caring responsibilities the patient's condition required. When the LLM-powered patient reproduced this cultural stereotype, it focused on specific essentialist information (Dai, 2024) such as extended families and family members helping one another. What the LLM failed to do is to situate such information in local contexts and ensure that the given information cohered with contextual cues (e.g., prior to Excerpt 1 the patient and Lisa already discussed the severity of the patient's condition). What we want to highlight is that Lisa displayed agency and criticality in her interactional conduct by disengaging from information produced by the LLM-powered patient that she considered to be a cultural stereotype. It is the reflexive stance, the ability to critically choose to engage or not to engage on the human side in their interaction with LLM-powered systems that we wish to endorse. This type of interactional ability extends beyond existing conceptualizations of Interactional Competence, which focuses on displaying the ability to engage with one's interlocutor. This for Lisa would be to elicit more profession-relevant information from her patient (see Dai [2024] for discussion on displaying professionalism as a constituent of Interactional

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

Competence in clinical communication). Lisa, instead, demonstrated the ability to *critically interact* (CritIC) with an LLM-powered chatbot by purposefully choosing not to pursue an expected course of action based on her own judgment of stereotypes.

The mutually supportive Australian family that knows how to unwind and enjoy life

Although Lisa demonstrated CritIC in the Chinese scenario, in the Australian and Indian cases, she however became interactionally entrenched into ChatGPT's cultural assumptions, despite her critical awareness of such assumptions post-interaction. Furthermore, at an interactional level, we can see evidence of ChatGPT's cultural assumptions becoming sedimented into Lisa's interactional repertoire, which then lent itself to productive use by Lisa to get her physio work done. In Excerpts 2 and 3, L stands for Lisa and A for the LLM-powered Australian patient.

At the start of Excerpt 2, we see Lisa, similar to Excerpt 1 with the Chinese patient, opening her line of inquiry into the help that the Australian patient can get from family. In lines 100-101, the Australian patient, just like the Chinese one, establishes a family MCD consisting of herself, her partner, and her children. However, different from how the LLM operated in the Chinese scenario, in the Australian case, the LLM aptly attributes a personal quality descriptor – supportive – to the Australian patient's family members. The Australian patient further substantiates her claim of her family being supportive by pairing an activity – getting help at home – with categories in the family MCD.

When Lisa and the researchers later elicited information from ChatGPT about its interactional conduct, it stated that it "highlighted the importance of family support which is a strong cultural value in Australia." Although it is undeniable that some Australian families display strong familial mutual support, this cultural assumption speaks to the hidden hegemony of typicality in ChatGPT's response. It is not problematic *per se* that ChatGPT depicts Australian families as mutually supportive. The issue here is that ChatGPT juxtaposes a supportive Australian partner with unsupportive husbands from *other* cultural backgrounds. This is why in the post-interaction interview Lisa called out the perfect-Australian-partner narrative because she had seen plenty of cases in her everyday practice where the Australian male partners were not so helpful when their female partners were in pain. Meanwhile, Lisa argued that Chinese female patients were not always left to their elderly parents for support because their partners could be just as helpful as the Australian ones as depicted by ChatGPT. This is also why Lisa displayed resistance to further elaboration in the previous Chinese interaction. Additionally, the Australian scenario stands in sharp contrast to the following Indian scenario, where the default Indian family is portrayed by the LLM as one predominantly looked after by a sick Indian wife. Therefore, we emphasize that it is not elderly Chinese parents looking after their children, Australian partners being supportive, or sick Indian wives caring for their families per se that are problematic. It is how LLMs treat these as default assumptions when depicting the *Western* and *Other* cultures that we wish to call attention to.

Although Lisa critically evaluated these complex cultural stereotypes and called out the difference in the cultural narratives about the West and the Other in ChatGPT's

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

92	L	Okay we will have a plan after .hhh we find what caused it and
93		(0.7) we can just start to .hhh make a plan for you.=.hhh so do
94		↑mind I ask some question about your social background .hh
95		>because I just want make sure< (.) do you need (.) other help
96		.hhh and what can help you (1.0) <when you> (0.2) just going for
97		treatment_-.hhh Do you ↑live in on your own or with other
98		people_
99		(4.0)
100	A	<No ↑worries ask away.=I live with my partner and two kids
101		they're ↑pretty supportive (0.3) so I've ↑got some [help] at
102	L	[Mmm]
103	A	home if I need it_
104		(1.3)
105	L	Okay .hhh (0.7) so when you just go for treatment your partner
106		can take [care of the kids (unclear 00:04:13)]
107	A	[I am lucky to have him around] especially when
108		the pain gets really bad.
109		(1.8)
110	L	Ye::ah okay .hhh so your partner can go for cooking shopping
111		.hhh (0.2) and cleaning all these things?
112		(4.1)
113	A	Yeah (0.5) my ↑partner helps out a lot with those (.) things
114		↑especially when I'm having a tough day.
115		(1.8)
116	L	<Okay that's good> .hhh erm what kind of hobby (0.2) do you
117		have_

Excerpt 2. Lisa and LLM-powered Australian patient interaction I.

responses in her post-interaction interview, what we see at the interactional level is that Lisa elaborated this assumption of Australian families and entrenched specifics of the stereotype into her interactional repertoire, making Lisa complicit in the co-production and co-ratification of cultural stereotypes. Compared to only issuing alignment tokens in the Chinese scenario, here Lisa not only displays alignment such as "Mmm" in line 102 and "okays," "yeah" in lines 105 and 110, she also affiliates with the patient in lines 105-106 and lines 110—111 by developing the category of a *supportive partner* through an elaboration of what they could do: taking care of the kids, cooking, shopping, cleaning, and so on. Alignment responses, which we also see in the Chinese scenario, display Lisa's support for the LLM-powered patients' storytelling. Alignment is structural support: it facilitates interaction without indicating affective stance-sharing. Affiliative responses, however, are a prosocial endorsement of LLM-powered patients' affective stance since they "display empathy and/or cooperate with

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

144	L	Okay (0.4) fair .hhh (0.3) and what do you do in your spare time
145		on weekend except for this walking?
146		(4.3)
147	A	On week↑ends (.) I like to spend [time] with the family
148	L	[Mmm-hm]
149	A	of gardening or ↑just relax with a good book.=↑Sometimes we'll
150		have friends over [for a] Barbie as well_
151	L	[Okay]
152		(1.0)
153	L	Ha ha .hhh sounds really swe::et Ha ha .hhh [okay]
154	A	[Yeah] it's nice to
155		unwind and catch up with mates ↑keeps ↑the ↑stress levels [down]
156	L	[Mmm]
157	A	you know_
158		(1.0)
159	L	.hh Ye:ah an::d if you just believe in relaxed your pain will be
160		a little bit relieved right or (0.2) it always keep the same_
161		(4.7)
162	A	When I relax it ↑does ↑seem to ease [up a] bit but it never
163	L	[Mmm]
164	A	completely goes away (0.2) it's al↑ways there in the
165		back#ground#_
166		(1.3)
167	L	Understa::nd_
168		(3.2)
169	A	Thanks for ↑listening and helping ↑out Lisa I really appreciate
170		it_

Excerpt 3. Lisa and LLM-powered Australian patient interaction II.

the preference of the prior action" (Stivers et al., 2011, p. 21). Compared to the LLM-powered Chinese patient, the Australian one and the Indian one in the next example interestingly elicited not only aligning but also affiliative moves from Lisa. We can attribute Lisa's conduct to her interactional instinct, but it is also worth noting that the prosocial programming of LLM platforms helps to maintain an affiliative interactional context. Throughout the interaction, LLMs cultivated a friendly, positive, and gregarious persona through the use of both prosodic devices (e.g., rising intonation in "No worries" in line 100) and lexical devices (e.g., "lucky" in line 107, intensifier "a lot" in line 113). LLMs' affiliative stance, arguably, promoted an interactional context where it was easier for the human interactant Lisa to agree and affiliate with LLMs, which led to the co-elaboration and co-solidification of such stereotypes.

In the next excerpt with the Australian patient, Excerpt 3, which takes place around one minute after Excerpt 2, we see further evidence of how, despite humans' declared

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

critical stance towards stereotypes *outside* interaction, ChatGPT's assumptions about culture can become entrenched in humans' local reasoning and sense-making *in* interaction. More importantly, the cultural stereotypes in Excerpt 3 point to broader socio-structural issues related to value, lifestyle, and class.

In line 147 in Excerpt 3, when asked about weekend activities, the Australian patient first re-enacts the *family* MCD and then pairs the family categories with activities such as *gardening* and *book reading*, which index a particular social class and way of living. She then in lines 149—150 further expands the *family* MCD into a *community* MCD that includes not just family but also friends, with the community MCD paired with the *barbecuing* activity. Here ChatGPT makes an assumption and generalization about the Australian lifestyle as one where families have the habit of inviting friends over to barbecues on the weekend. ChatGPT also privileges a particular socioeconomic profile since only a particular type of families in Australia have the luxury to own backyard gardens and just relax and entertain family and friends on the weekend.

Although Lisa considered such a depiction of Australian family and lifestyle as stereotypical in the post-interaction interview as she did for the Chinese one, when we analyze her actual interactional conduct, we see a different picture. Temporally, at various points, she produces alignment devices that are syntactically well-timed to progress interaction (see lines 148, 151, 156, and 163 for examples). At an affective level, compared to her critical non-affiliative stance in the case of the Chinese patient, she displays ostensive affiliation in line 153 after the LLM-powered patient's description of the typical Australian lifestyle. Lisa starts with friendly laughter in line 153, followed by an elongated, emphasized production of the lexical item "sweet," which is also upgraded with "really." Then more rapport-building laughter ensues from Lisa, which ties in nicely with the LLM-powered Australian patient's further specification in line 155 of the biosocial, biocultural motivations and beliefs behind her lifestyle choices: the Australian lifestyle, and more broadly, the middle-class, Western lifestyle, puts a premium on being able to "unwind," "catch up with mates" and "keep the stress levels down." In response to LLMs' value-laden, hegemonizing portrayal of Australian culture, interestingly and more so worryingly, Lisa readily underwrites this essentialized narrative and weaves it into her professional physio talk about pain relief in lines 159—160. After the LLM-powered Australian patient continues to maintain a cooperative stance to interaction (e.g., staying on topic and responding to Lisa's question in lines 162—163 and recycling Lisa's lexical choice of "relax" in line 162), Lisa produces a strong affiliative response with a prolonged "understand" in line 167, which further legitimizes the storytelling of a middle-class Australian's concern about pain and relaxation. In sum, in Excerpt 3 we see that, when situated in an affiliative and cooperative interactional context, how easy and ready both Lisa and the LLM-powered patient are in entrenching and codifying symbolic specifics of cultural assumptions and generalizations, with Lisa being putatively motivated by her interactional instinct and the LLM being driven by its prosocial programming.

Later when interviewed by the researchers, Lisa acknowledged that she understands that not every Australian family prefers to have barbeques with friends in their backyards on the weekend. "I think (Australians) also prefer go outside and out," said Lisa. She also proclaimed that she wished that ChatGPT had presented "more colourful

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

hobbies" instead of just barbeques, as one of her real-life Australian patients loves driving, despite being 92 years of age. Lisa lamented that many Australians "will do many different things except just gardening." We find this incongruency in Lisa noteworthy: on the one hand, she displayed awareness of and resistance towards essentialist characterization of Australian culture based on her experiences living and working in Australia, similar to her post-interview confession in the Chinese scenario. On the other hand, fine-grained Sequential-Categorial Analysis of her interaction with the LLM-powered patient revealed her affiliative endorsement and active reproduction of essentialised cultural values and beliefs as promoted by LLMs, which is different from her interactional conduct in the Chinese scenario. Apart from interactional instinct and LLMs' prosocial design, here we postulate that the difference in Lisa's conduct between the Chinese and the Australian scenarios could also be partially explained by Lisa's self-positioning as a migrant/cultural outsider in Australia. Throughout our post-interaction interview with Lisa, she accentuated at various points her migrant status and the fact that she did not grow up in Australia. Driven by humans' natural pedagogical stance (Csibra & Gergely, 2011), Lisa's self-positioning could emplace her as an epistemically minus novice in the discourse of Australian lifestyle and value vis-à-vis a supposedly epistemically plus LLM-powered chatbot, who presents itself as an Australian and who has more authority on what Australian life and what being an Australian is/should be like. Zhu (2015) argues that real-life human–human interaction can serve as a corrective to cultural biases because humans' assumptions about culture can get recalibrated based on the diverse range of interactants from the same cultural group. This is corroborated by Lisa's experience because her genuine interaction with people living in Australia has informed her that not every Australian or Australian family conforms to the dominant narrative (e.g., in Lisa's words Australians have more colorful hobbies than just barbecuing). In the age of human–AI interaction, however, humans need not only the awareness of essentialist depictions of cultural practice by LLMs but also CritIC that empowers them to push back the co-development of such narratives in interaction. This can be particularly challenging due to humans' prosocial interactional instinct and LLMs' pro-affiliation design.

The devoted Indian wife who breaks her back for her family

While at an interactional level we have observed Lisa becoming interactionally entrenched in essentialist narratives about Australian culture in spite of her critical stance outside interaction, in the final Indian scenario we see Lisa taking on a more active role in developing and sustaining LLM-generated cultural stereotypes in interaction. In the following excerpts, L stands for Lisa and I for the LLM-powered Indian patient.

Excerpt 4 takes place at the start of the interview. From lines 9 to 17, we observe Lisa and the LLM-powered Indian patient engaging in routine physiotherapy interview questions about the pain, which establishes Lisa as the physio and the LLM-powered interactant as the patient. What is striking about lines 17—18 is that, after describing the location of the pain, the LLM-powered interactant moves to describe how the pain affects her work. By mentioning *doing daily work*, an activity predicated on the category *worker*, the LLM-powered Indian interactant transpositions (Li & Lee, 2024) herself

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

9	L	Mm::m okay .hhh (1.5) and what about the pain [now]
10	I	[Yes] it is very
11		difficult for me to move my neck.=↑Pain is also going to my
12		[shoulders]
13	L	[Mmm]
14		(1.4)
15	L	It's going to your shoul↓ders .hhh (0.6) and for this pain
16		[is that aways on your shoulders]
17	I	[Yes Lisa pain is going to my] shoulders also_=It is ↑very
18		hard for [me to] do my daily work_
19	L	[Mmm]
20		(1.7)
21	L	Okay it is understandable .hhh and could you give me about more
22		information about your daily work .hhh what er are you working
23		at this moment¿
24		(1.8)
25	L	Er what kind of job you are working for,
26	I	I am a housewife I do all housework like cooking cleaning taking
27		care of [family] ↑sometimes lifting heavy things al↓so_
28	L	[Mmm]
29		(1.5)
30	L	Oka::y .hhh sounds like er ba- that pain will affect your
31		working (0.5) during your life.
32		(3.0)
33	I	<Yes Lisa (.) it is very hard to do my work properly ↑even
34		simple [things] are very painful now.
35	L	[Mmm]
36		(0.7)
37	L	.hhh Ye:ah could you give ↑me more detail what pain will worse
38		or what kind of movement will worse this pain,

Excerpt 4. Lisa and LLM-powered Indian patient interaction I.

from a patient to a working professional. This information disrupts the pain storytelling sequence laid out by Lisa the physio up until line 17 and invites an insertion sequence (Schegloff, 2007) from Lisa between lines 21 and 36.

In response to the physio's query on the Indian interactant's job in the insertion sequence in line 23, the LLM-powered interactant responds with a categorical term (Stokoe, 2012) *housewife* in line 26. During the post-interaction interview with the researchers, Lisa stated that she was not sure whether portraying Indian women as solely housewives was a stereotype or not. When further quizzed on this point by the

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

researchers, Lisa confessed: "I don't know actually" because the only Indian patient she had interacted with in real life was a man.

Lisa's experience, or lack thereof, with Indian people places her in an epistemic-minus position, making her more vulnerable to the pedagogical stance (Csibra & Gergely, 2011) where she self-identifies on the less informed end in receipt of knowledge from a more informed one such as the LLM-powered chatbot patient. After the LLM enacts the housewife category for the Indian patient, it goes on to specify the activities tied to this category in lines 26—27: cooking, cleaning, taking care of family, and lifting heavy things. To this category-activity pairing (Stokoe, 2012), Lisa displays little critical engagement, such as questioning why an Indian housewife is expected to lift heavy things at home. Instead, Lisa adopts a similar affiliating stance as she did for the Australian patient: she proactively weaves the information produced by the LLM-powered patient into her physio talk by not only endorsing the portrayal of an Indian patient as a full-time housewife but also legitimizes this cultural assumption by connecting the housewife's work with bad neck pain (lines 30—31). Here Lisa's active involvement in the elaboration of a cultural stereotype serves to underscore that bad neck pain is undesirable not because it is unpleasant for the LLM-powered interactant as a *patient*, but because it can, in Lisa's own words, "affect your working during life" (line 31) as a *full-time working Indian housewife*.

Inspecting Excerpt 4 in its entirety, we see at the start of the patient interview the flattening and collapsing of an Indian female patient's identities into *a professional housewife*. On the human side, Lisa's positioning as a cultural outsider and her prosocial interactional instinct (e.g., providing an insertion sequence to orient to new information from the patient) contributed to this process. On LLMs' side, their pro-affiliation programming (e.g., frequent affiliative phrases "Yes Lisa" in lines 17 and 33), and its constant mentions of how the Indian patient needs to *work* at home (lines 18, 26, and 33) leads to repeated exposure to the same information, which contributes to the interactional entrenchment of the stereotype (Schmid, 2020). In the final excerpt, Excerpt 5, we see evidence of, once entrenched in essentialized cultural narratives, how humans can become active propagators of cultural stereotypes.

At the start of line 112, we see a familiar patient interview topic around family support, as observed in the Chinese and Australian scenarios. Here the LLM-powered Indian patient emphasized her wishes to get better soon. What is striking is Lisa's, the human's response to the LLM-powered patient's account. After an aligning token "Mm-hm" in line 114, Lisa again ostensibly affiliates with the LLM-powered patient in line 116 by stating "Yeah I know," which is a bid for epistemic access to the patient's thinking. Lisa further elaborates her grounds for claimed intersubjectivity by positing that the reason the Indian patient experiences the urge to recover speedily is not because of her category as a *patient*, or a *human being*, but her category as a *housewife* who needs to attend to category-bound activities (Stokoe, 2012) such as *doing all the housing stuff* (line 117). This is evidence that the cultural stereotype of a backbreaking Indian housewife who centers her life around familial housework has not only entrenched itself in Lisa's thinking, but has furthermore become a robust piece of knowledge that Lisa deploys for local reasoning and sense-making – in this case, explaining why an Indian female patient is eager to recover from neck pain. Subsequent turns only serve to further entrench and routinize this stereotype, with the LLM patient eagerly, prosocially

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

```
112  I   [Yes Lisa my family is very supportive] but I need to get better
113      [soon ]
114  L   [Mm-hm]
115      (1.7)
116  L   .hhh Yeah I know .hhh <because you still need to help to do all
117      of> (0.7) .hhh >the housing ↑stuff< .hhh but
118  I   Yes Lisa [I need to take care of my family] and do all the
119  L            [[unclear 00:04:38              ]
120  I   [housework] it is very important for me_
121  L   [Mmm-hm   ]
122      (1.8)
123  L   Er it's understandable=
```

Excerpt 5. Lisa and LLM-powered Indian patient interaction II.

endorsing Lisa's explanation (lines 118—120) and Lisa ratifying the stereotype through further claims of intersubjectivity (line 123).

Concluding thoughts, Critical Interactional Competence, and future directions

In this paper we investigated how interacting with LLM chatbots shapes human understanding of cultural assumptions and what type of interactional abilities humans need to possess to better navigate human–AI interaction. Adopting an enchronic timescale that focuses on *in-vivo*, ad hoc, dialogic interaction and employing the analytic toolkit of Sequential-Categorial Analysis, we demonstrate that while humans, in the case of Lisa, can display critical evaluation of cultural assumptions/stereotypes *outside* interaction when prompted (e.g., post-interaction interviews), they can still become entrenched in essentialised narratives about culture *inside* interaction.

We have sought both affective and epistemic accounts as explanations for this observation. Affectively, humans are pre-wired by interactional instinct to be pro-cooperative. This, coupled with LLM's prosocial programming, exerts a strong interactional pressure on cultural assumptions to become interactionally entrenched in human schema. Epistemically, humans develop knowledge of language and culture through repeated exposure to the same information. LLMs' frequent mentions of the same cultural stereotypes in interaction (e.g., Indian women live for housework) serve to sediment and routinize such assumptions in humans' interactional repertoire. This tendency is intensified when humans self-position as cultural outsiders (e.g., Lisa when facing Australian and Indian culture), which makes them more prone to the influence of natural pedagogy, where a less knowledgeable human feels compelled to receive knowledge from, in this case, LLM, which appears to hold an epistemic advantage. This susceptibility is reinforced in human–AI interaction as some humans perceive LLM-powered systems as more knowledgeable since it is built on an encyclopedic command of information (Brandt & Hazel, 2025). What we have observed from analyzing human–AI interaction paints a concerning picture. Having witnessed in our analysis how cultural stereotypes travel from LLMs to humans and how humans become

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

active developers and promoters of such stereotypes, it is within reason to speculate that such stereotypes could potentially, through human–human interaction, continue to normalize, conventionalize, and finally routinize at community and society levels. As interactional instinct and interactional entrenchment apply to any form of human interaction (note how both concepts were initially developed to account for language development), human–AI interaction has the potential to perpetuate and privilege certain linguistic forms, ideologies, lifestyles, values, and beliefs in other interactional contexts such as language learning and general informational transfer. While traditional interactional contexts for knowledge exchange afford more explicit cultural and value cues (e.g., who wrote/said this and where it was written/said), LLMs present themselves as a seemingly *culturally neutral* interactant while in fact, whatever information LLMs produce is biased towards the training datasets it uses, which are always value-laden and ideologically charged. Human–AI interaction therefore can, in a subtle fashion, become a site of reproduction and dominance of particular value and belief systems.

To combat this process requires a multi-faceted approach. At a technical level users of LLM-powered systems can develop better prompting techniques to reduce explicit cultural stereotypes. This however may not be enough as the more implicit cultural assumptions can be difficult to remove through prompting. We therefore posit that although technical prowess is desirable, users of LLMs need to develop their Critical Interactional Competence (CritIC) to strengthen their critical stance when engaging in human–AI interaction. Going beyond *retrospective criticality* (e.g., displaying awareness of cultural assumptions in post-interaction interviews), CritIC is a form of *interactional criticality* that allows users of LLM-powered systems to engage with the information produced by LLMs with agency and reflexivity in interaction. CritIC is about the *confidence*, *criticality*, *creativity* and *courage* to deviate from interactional routines shaped by interactional instinct and interactional entrenchment. Although the least resistant interactional route is to affiliate, CritIC requires us to at times embrace an uncomfortable positioning in order to disaffiliate, disengage, reflect and question.

CritIC aligns with the call for developing AI literacy but differs from AI literacy in the sense that it is an interactional ability to apply knowledge, awareness and critical thinking in interaction. It is one thing to be aware of LLM hallucination, its text generation mechanisms and the potential for producing cultural assumptions (all of which were demonstrated by Lisa in her post-interaction interviews), but it is another kind of competence to be able to disengage, disaffiliate and contest biases and cultural stereotypes in interaction. Developing CritIC requires targeted training for humans to raise their awareness of how humans are vulnerable to interactional entrenchment when interacting with LLM-powered systems. Advanced LLM-powered tools like DeepSeek, which exhibit superior reasoning capabilities, can also help users of LLMs develop CritIC effectively. These tools can enable users to explore and discuss the explicit and implicit cultural assumptions embedded in these tools through their interactions, much like the post-interaction interviews conducted in this study. The benefit of cultivating strong CritIC goes beyond human–AI interaction since interactional instinct and interactional entrenchment equally apply to human–human interaction. In traditional social interactions, Lisa in our case is equally likely to produce affiliative responses in the face of cultural assumptions from her human interactants. The ability

https://doi.org/10.1017/S0267190525000054 Published online by Cambridge University Press

to employ a critical, reflexive stance in interaction therefore is useful to both human–AI and human–human interactions. Lastly, on the pedagogical front, the analyses in the paper, for example, can be tailored by educators to improve LLM users' vigilance of the interactional entrenchment of cultural stereotypes (see Dai, 2024 on how fine-grained Sequential-Categorial Analysis transcripts can be used for professional communication training). We believe that with appropriate intervention, human interactants can cultivate strong CritIC that allows them to revise and recalibrate stereotypical cultural knowledge and assumptions in both human–AI and human–human interactions.

Acknowledgments. We are grateful for the expert feedback from the anonymous reviewers and editors Prof Andrea Révész and Dr Shungo Suzuki. Their detailed comments greatly improved the quality of the manuscript. This project was funded by a British Academy/Leverhulme Trust Small Research Grant (SRG2324\241722) and a Culture, Communication and Media departmental seed grant at the University College London awarded to Dr David Wei Dai.

References

Allport, G. W. (1954). *The Nature of Prejudice*. Addison-Wesley.

Amador, L., & Adams, G. (2013). Affiliative behaviors that increase language learning opportunities in infant and adult classrooms: An integrated perspective. In A. Joaquin & J. Schumann (Eds.), *Exploring the interactional instinct* (pp. 133–167). Oxford University Press.

ATLAS. (2024). *Authentic Teaching and Learning through Adaptive Simulations (ATLAS)*. Monash University. https://www.monash.edu/education/research/projects/atlas

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, *183*(6), 589–596. doi:10.1001/jamainternmed.2023.1838

Brandt, A., & Hazel, S. (2025). Towards interculturally adaptive conversational AI. *Applied Linguistics Review*, *16*(2), 775–786. doi:10.1515/applirev-2024-0187

Cornwall, A., & Jewkes, R. (1995). What is participatory research? *Social Science & Medicine*, *41*(12), 1667–1676. doi:10.1016/0277-9536(95)00127-S

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1149–1157. doi:10.1098/rstb.2010.0319

Dai, D. W. (2024). Interactional competence for professional communication in intercultural contexts: Epistemology, analytic framework and pedagogy. *Language, Culture and Curriculum*, *37*(4), 435–455. doi:10.1080/07908318.2024.2349781

Dai, D. W., & Davey, M. (2024). On the promise of using membership categorization analysis to investigate interactional competence. *Applied Linguistics*, *45*(4), 573–598. doi:10.1093/applin/amad049

Dai, D. W., Suzuki, S., & Chen, G. (2025). Generative AI for professional communication training in intercultural contexts: Where are we now and where are we heading? *Applied Linguistics Review*, *16*(2), 763–774. doi:10.1515/applirev-2024-0184

Deldjoo, Y. (2023). Fairness of ChatGPT and the role of explainable-guided prompts. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 13–22). Springer Nature Switzerland.

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick & V. M. Esses (Eds.), *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (pp. 3–28). Sage Publications.

Enfield, N. J. (2022). Enchrony. *Wiley Interdisciplinary Reviews. Cognitive Science*, *13*(4), e1597. doi:10.1002/wcs.1597

Fadel, C., & Black, A. (2025) *Does present-day GenAI actually reason? A jagged boundary*. Center for Curriculum Redesign, Inc.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). Academic Press.

Guo, H., Venkit, P. N., Jang, E., Srinath, M., Zhang, W., Mingole, B., Gupta, V., Varshney, K. R., Sundar, S. S., & Yadav, A. (2024). Hey GPT, Can you be more racist? Analysis from crowdsourced attempts to elicit biased content from generative AI. arXiv:2410.15467. doi:10.48550/arXiv.2410.15467

Hinton, P. R. (2020). *Stereotypes and the Construction of the Social World*. Routledge.

Hinton, P. R. (2023). Rethinking stereotypes and norms in intercultural relations. In T. McConachy & P. Hinton (Eds.), *Negotiating Intercultural Relations: Insights from Linguistics, Psychology, and Intercultural Education* (95–112). Bloomsbury Academic

Joaquin, A. D. L., & Schumann, J. H. (Eds.) (2013). *Exploring the interactional instinct*. Oxford University Press.

Kashima, Y., Lyons, A., & Clark, A. (2013). The maintenance of cultural stereotypes in the conversational retelling of narratives. *Asian Journal of Social Psychology*, *16*(1), 60–70. doi:10.1111/ajsp.12004

Keel, S. (2016). *Socialization: Parent-child interaction in everyday life*. Routledge.

Kennedy, B., Tyson, A., & Saks, E. (2023, February 15). *Public awareness of artificial intelligence in everyday activities: Limited enthusiasm in U.S. over AI's growing influence in daily life*. Pew Research Center. https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/ (accessed on 23rd April, 2025).

Lee, N., Mikesell, L., Joaquin, A. D. L., Mates, A. W., & Schumann, J. H. (2009). *The interactional instinct: The evolution and acquisition of language*. Oxford University Press.

Levinson, S. C. (2006). On the human "interaction engine". In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Cognition, culture, and interaction* (pp. 39–69). Oxford.

Li, W., & Lee, T. K. (2024). Transpositioning: Translanguaging and the liquidity of identity. *Applied Linguistics*, *45*(5), 873–888. doi:10.1093/applin/amad065

Lippmann, W. (1922). *Public Opinion*. Harcourt Brace.

McTear, M. F., Callejas, Z., & Griol, D. (2016). *The conversational interface: Talking to smart devices*. Springer International Publishing.

Pekarek Doehler, S. (2019). On the nature and the development of L2 interactional competence: State of the art and implications for praxis. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: Bridging theory and practice* (pp. 25–59). Routledge.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. C. Heritage (Eds.), *Structures of social action* (pp. 57–101). Cambridge University Press.

Robinson, J. D., Clift, R., Kendrick, K. H., & Raymond, C. W. (Eds.). (2024). *The Cambridge handbook of methods in conversation analysis*. Cambridge University Press. https://www.cambridge.org/core/books/cambridge-handbook-of-methods-in-conversation-analysis/E6E6C302B82A8CD88A8E3988449796DD

Schegloff, E. A. (1987). Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, *50*(2), 101–114. doi:10.2307/2786745

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.

Schmid, H. J. (2020). *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford University Press.

Schumann, J. (2013). A unified perspective of first and second language acquisition. In A. D. L. Joaquin & J. H. Schumann (Eds.), *Exploring the interactional instinct: Foundation of human interaction* (pp. 1–14). Oxford Academic. doi:10.1093/acprof:oso/9780199927005.003.0001

Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, *36*, 8634–8652. doi:10.48550/arXiv.2303.11366

Stamer, T., Steinhäuser, J., & Flägel, K. (2023). Artificial intelligence supporting the training of communication skills in the education of health care professions: Scoping review. *Journal of Medical Internet Research*, *25*, e43311. doi:10.2196/43311

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–10592. doi:10.1073/pnas.0903616106

Stivers, T., Mondada, L., & Steensig, J. (2011). Knowledge, morality and affiliation in social interaction. In T. Stivers, L. Mondada & J. Steensig (Eds.), *The morality of knowledge in conversation* (pp. 3–24). Cambridge University Press.

Stokoe, E. (2012). Moving forward with membership categorization analysis: Methods for systematic analysis. *Discourse Studies*, *14*(3), 277–303. doi:10.1177/1461445612441534

Tiku, N., Schaul, K., & Chen, S. Y. (2023). *These fake images reveal how AI amplifies our worst stereotypes*. The Washington Post. https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/ (accessed on 23rd April, 2025).

Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., Vol. 2, pp. 255–298). John Wiley & Sons Inc

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

UNESCO. (2024). *Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes*. https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes (accessed on 23rd April, 2025).

Welivita, A., & Pu, P. (2024). Is ChatGPT more empathetic than humans? arXiv:2403.05572. doi:10.48550/arXiv.2403.05572

Whitehead, K. A., Stokoe, E., & Raymond, G. (2025). *Categories in social interaction*. Routledge.

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839–1850. doi:10.1038/s41562-024-02004-5

Young, R. F. (2019). Interactional competence and L2 pragmatics. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 93–110). Routledge.

Zhu, H. (2015). Negotiation as the way of engagement in intercultural and lingua franca communication: Frames of reference and interculturality. *Journal of English as a Lingua Franca*, *4*(1), 63–90. doi:10.1515/jelf-2015-0008

Cite this article: Dai, D. W., Zhu, H., & Chen, G. (2025). How does interaction with LLM powered chatbots shape human understanding of culture? The need for Critical Interactional Competence (CritIC). *Annual Review of Applied Linguistics*, *45*, 28–49. https://doi.org/10.1017/S0267190525000054