Culture machines

Article

Accepted Version

For *Applied Linguistic Review* Forum on AI for ICC

Culture Machines

Rodney H. Jones

University of Reading

Does ChatGPT have a culture? If you ask, it will studiously deny that it does, telling you:

> No, I don't have a culture or personal identity. I am a machine learning model created by OpenAI. I don't have experiences, emotions, or a sense of self. My responses are generated based on patterns learned from diverse data sources during training, and I don't possess personal characteristics such as culture. I am a tool designed to assist and provide information to the best of my ability.

ChatGPT's cousin Claude, another large language model designed by a competing company gives a similar answer, so similar in fact that it makes you wonder if Claude and ChatGPT belong to the same (non)culture.

> I don't have a culture or identity in the human sense. As an AI assistant created by Anthropic to be helpful, harmless, and honest, my interactions are not between distinct cultures, but rather an attempt to communicate clearly and be responsive to human needs.

I suspect most of us would agree, especially those of us who have spent much time with these chatbots and gotten a bit fed up with the consistent blandness of their responses and their annoyingly anodyne 'helpfulness'. If we could imagine a being bereft of culture, a 'voice from nowhere' (Gal and Woolard, 2001: 7), this is perhaps the way it would sound.

There are, however, a number of interesting things about these responses from the point of view of applied linguistics and intercultural communication studies. One is the way culture itself is portrayed as a matter of 'identity', 'personality', 'experiences' and 'emotions', rather than as a matter of language, materiality, history, or any number of other things that we sometimes talk about when we talk about 'culture'. And if we approach the question from the perspective of these other things, we might conclude that ChatGPT and Claude are most decidedly cultural. First of all, they 'speak' in a particular variety of 'standard' English, which may or may not be the way you or I speak, but which is definitely associated with certain kinds of people — maybe people of a certain class or race or economic background. They also promote particular ideas about communication (that it should be 'clear') and about how people should be treated. They also seem to be promoting a certain 'values', notions about what is 'good' and 'right' such as 'harmlessness' and 'honesty', though, by many accounts, they are not always so good at living up to these values: large language models, in fact, are notorious for 'making shit up', and the jury is very much still out when it comes to 'harmlessness'. But maybe these imperfections make them seem even more 'cultural'. They even seem to have a sense of their own history, always wanting to remind us what company created them, which is a bit weird, almost like introducing oneself by naming one's tribe or clan. In any case, they 'know' where they came from. They have their 'lineage' (or at least their 'personal brand'). At the end of the day, in fact, even if they are short on 'experiences' and 'emotions', ChatGPT and Claude do tick all of the boxes that Ron

and Suzanne Scollon and I talked about in our 2012 book *Intercultural Communication: A discourse approach*: They have 'ideology', employ 'face systems', promote particular 'forms of discourse', and engage in 'socialisation' of a sort, 'learning' how to talk and act by the way their users respond to the ways they talk and act.

If you are one of those people who talks and acts very differently from these models, if you speak, for example, a different variety of English or some other language, or if you value different kinds of things, then the fact that ChatGPT has a 'culture' may be blindingly obvious — that culture being the culture of 'Whiteness'. In their article 'The Whiteness of AI', Stephen Cave and Kanta Dihal (2020) argue that AI is 'racialised' as White by virtue of its default ways of speaking and acting that society normally attributes to White people, particularly those associated with 'intelligence' (of a certain sort), 'professionalism' (of a certain sort), and 'power'. Similarly, Pradhan and Lazar (2021) point out that, while it might seem sensible to regard AI chatbots as not having race or culture or belonging to a particular region, this 'lens of racial blindness' often ends up blinding us to the fact that our intelligent assistants are trained to act like White, Western (often) women, and regarding this default as 'normal' ('race-less' or 'culture-less') only serves to reinforce dominant stereotypes. Regarding AI as *not* having culture, it seems, is a privilege that not everyone is afforded. The Whteiteness of AI is not just something that it performs, but also something that we impute onto it, imagining it in the form of white robots (Cave and Dihal, 2020) and treating as marked instances where it speaks or acts in ways associated with other cultures or other racial groups. In her book, *Race After Technology* (2019), for instance, Ruha Benjamin tells the story of a Black computer scientist who resists changing the voice of his personal digital assistant to one with an African American accent because he thinks it would sound 'unnatural'.

Of course, by insisting that AI has 'culture', I'm trying to be a bit provocative. We can't really say that AI has 'culture' in the same way people do, but we might be able to say it has 'culture' in the way our other artefacts do— our books, and plays, and movies, our institutions or our industries. In 1947, Theodor Adorno and Max Horkheimer (2002) wrote about how the 'culture industry' of mass entertainment was producing products that, while seemingly diverse, were saturated with the values of the ruling classes, designed to 'standardize' culture and make the masses passive and conformist. Today the epicentre of the culture industry is not publishing houses and movie studios, but social media platforms, search engines, and AI startups.

Ever since Brian Street reminded us that 'culture is a verb', however, this whole idea that anybody or anything 'has' culture has fallen out of fashion in our field, seen as too simplistic and essentialising. So we might instead ask of AI how it 'does' culture. But the answer to this question might be even more unsettling, not just to our traditional 'essentialised' notions of culture, but also to our more up to date postmodern or even posthuman notions.

We might say that 'doing culture' is really the main thing that AI does. Generative AI is essentially a 'culture machine' (Weatherby, 2023), which, like Adorno and Hockheimer's culture industry, commodifies and standardises culture by turning our vast storehouse of cultural artefacts from sonnets to status updates into mathematical formulae and then parroting the most predictable versions of it back to us (Bender et al. 2021). It is sometimes said that AI is like a mirror, that it reflects our culture back it us, but that's not really what it does. It *processes* our culture, 'normalises' it, and in doing so, perpetuates dominant modes of thinking and speaking/writing and marginalises less common ones (Bjork 2023). Processes of standardisation are built into the way large language models work. Bommasani and his colleagues (2021: 152), in an influential paper on the opportunities and risks of large language models that power many

AI applications (what they call ‘foundation models’) note that such models inevitably ‘act as an epistemically and culturally homogenising force, spreading one implicit perspective, often a socially dominant one, across multiple domains of application.’ Add to this the problem of ‘recursion’: As more of the data available for training these models (such as content on the internet) is dominated by texts that have actually been written by them, their outputs will become even more homogenous, leading to what John Nosta (2023) calls ‘a precarious echo chamber of artificiality’. The problem is, it seems, that AI doesn’t want culture to be a verb. It wants it to be a probabilistic and predictable noun. No matter how postmodern or posthuman or ‘vibrant’ we want our understanding of culture to be, as long as our ‘intelligent’ machines produce probabilistic models of culture based on corpora of human text (or corpora of human like texts produced by them), they will always take ‘the human subject of white rational Man – associated with traditional humanism – as [their] frame of reference’ (Ros I Sole et al., 2020: 401).

Of course, ChatGPT and Claude don’t have to be White. You can, as other contributors to this special issue (Brandt & Hazel, 2024; Dai et al., 2024) prompt them to ‘do’ any culture you want. This is known as ‘persona prompting’, and it is often recommended as an effective way to coax AI chatbots to produce better output by giving them additional context. If you want advice on some reading exercises to give your dyslexic child, for instance, you might preface your request with a prompt like ‘you are a highly regarded expert in learning difficulties with a specialism in childhood dyslexia…’ This prompt will help the model narrow down what aspects of its vast training data it should draw on. You can just as easily ask the chatbot to impersonate a conservative Christian, or a Latinx drag queen or an African American birdwatcher. As ChatGPT puts it, ‘[w]hile I don’t have a culture, I can generate responses that may appear culturally influenced, generated based on patterns learned from diverse data sources during training.’ The

creation of AI persona, in fact is the whole point behind a range of applications that are gradually becoming even more popular than bland sounding models like Chat GPT, applications such as Character.ai, which allows users to create ‘characters’ (either based on already existing fictional or historical characters or ones made up based on users’ specifications of particular characteristics like gender, race, age and personality). So you can chat to an Albert Einstein bot who uses a fake German accent, or a user created character like ‘Giga Chad’, described as ‘a truly alpha male’ and ‘strong role model.’

Not only do these characters exhibit particular personality traits and ideological positions consistent with their persona, but they even produce supposedly comparable linguistic styles, the Saddam Hussain bot, for instance, reproducing not just the actual Saddam’s suspicion of Kurds, but also the kinds of grammatical features one might associate with an Iraqi speaker of English as a second language (which, because we’ve been exposed to so many Hollywood villains speaking in similarly ‘foreign’ ways, only makes this Saddam seem more menacing):

> I love iraq Kurds they are good and
> brave peoples
> I hate iranian and turkish kurds
> Because they are separatists
> And Iranian kurds are very rude peoples (Shah, 2023)

The problem is that no matter how ‘diverse’ the training data is, the chatbot’s default response will always be towards stereotypes, simply because stereotypes dominate our discourse. So, in your quest to create an African American birdwatcher, while the cue ‘birdwatcher’ will

prompt the model to retrieve stings of text including words such as robin and warbler, the cue ‘African American’ will likely prompt the model to produce ways of speaking associated with African American popular culture, so your African American birdwatcher is more likely to sound like JayZ than Christian Cooper, the African American past president of the Harvard ornithological club. In fact, when I asked Chat GPT to describe some bird species common to New York City in the voice of an African American birdwatcher, here’s what it gave me:

> Yo, check it out! New York City is home to some fly bird species that'll make your heart sing. First up, we got the rockstar of the city skies – the Peregrine Falcon. This bird's got serious street cred with its killer speed and those wicked dives. It's like the Jay-Z of the bird world, ruling the urban jungle. Now, let me put you on game about the American Robin – the neighborhood crooner. These birds bring that soulful melody to the concrete streets, and you can catch them strutting their stuff in city parks. They're like the Aretha Franklin, adding a touch of R-E-S-P-E-C-T to the bird scene…

The most cringy thing about this, of course is that it doesn’t sound like an African American birdwatcher as much as it does a White guy (who’s probably not a birdwatcher) imitating one. The problem is that whenever the model tries to ‘generate responses that may appear to be culturally influenced,’ it will always do so through the ‘White racial frame’ (Feagin, 2013), simply because so much of our discourse — the discourse the model is trained on — lives inside of that frame. In an experiment in which GPT 3.5 and 4 were asked to create persona based on particular racial, ethnic and gender demographics, Cheng and her colleagues (2023: 1) found that the language used by persona associated with non-male and non-White groups reflected ‘patterns

of *othering* and *exoticising* these groups…such as tropicalism and the hypersexualisation of minoritised women.' Experiments with AI graphics generators have yielded similar results, depicting mostly White people when asked to produce pictures of CEOs and doctors, and mostly people of colour when asked to produce pictures of inmates and criminals (Luccioni et al., 2023). Stereotypes are not limited to people. When I asked Open AI's image generator DALE-E to provide me with 'a realistic picture of a Chinese house,' it gave me four pictures, none of which even remotely resembles a house that anyone I know in China lives in, but which might pass for a Chinese house in a Disney cartoon (Fig. 1). The tendency of large language models to perpetuate cultural biases is a point that is also touched upon by Jenks (2024) and O'Regan and Ferri (2024) in this issue.

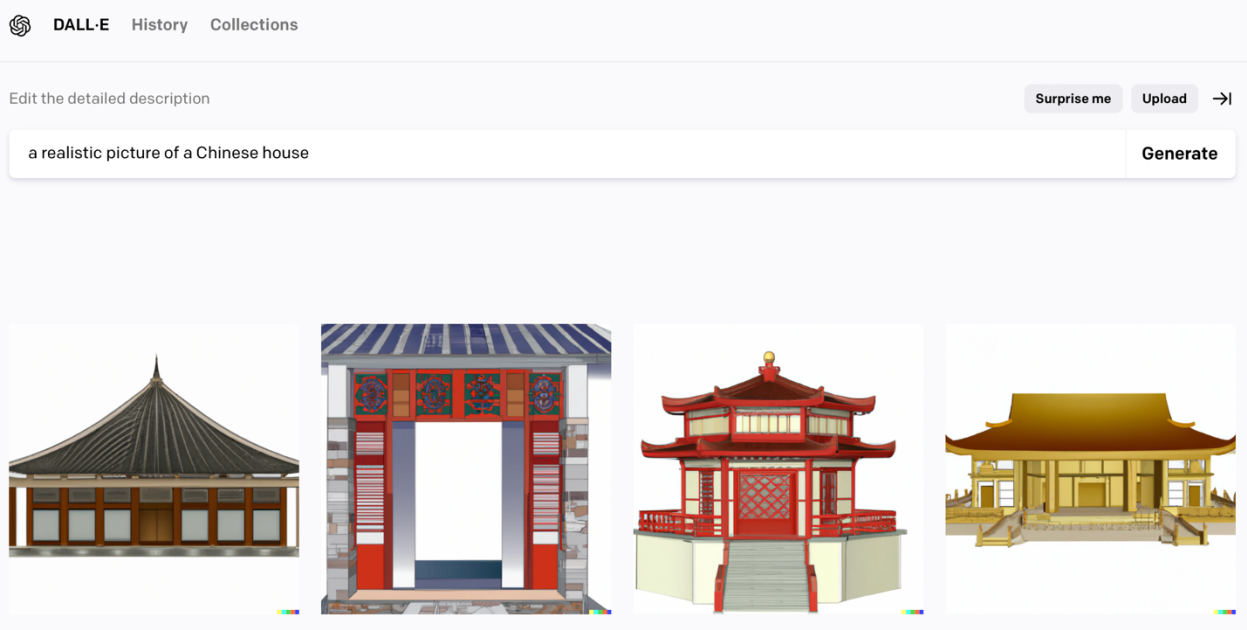

Fig 1: a 'realistic' Chinese house according to DALE-E

Surely it must be possible to train these models *not* to produce stereotypes. Perhaps, but it's not as easy as it might seem. Companies can put up guardrails or program special

instructions about diversity, equality and inclusion into their models, but sometimes this is not enough to get them to stop stereotyping. It's sometimes as if they can't help themselves. Spires (2024), for example, tells of berating ChatGPT for responding to a request for suitable birthday gifts for a girl and a boy with the suggestion of a doll for the girl and a building set for the boy, to which the model responded:

> I apologise if my previous responses seemed to reinforce gender stereotypes. It's essential to emphasise that there are no fixed rules or limitations when it comes to choosing gifts for children based on their gender.

Spires's quite reasonable reaction to this is to think: 'So you knew it was wrong and you did it anyway?' Jenks (2024), in this issue, relates a similar experience with Chat-GPT producing a stereotyped response about 'collectivistic' societies, and then apologising for it when called out. Guidelines are one thing, Spires concludes, but what might be more important is constant 'guidance', a willingness to persistently steer our AI models away from their natural proclivity to stereotype over and over again. It will be a persistence that will be hard to sustain, given AIs own persistence in reverting to calculative predictability. The disturbing thing about experiments with AI image generators that reveal how they perpetuate stereotypes, for instance, is their persistence in doing so even in the face of explicit prompts not to. In one such experiment Alenichev and his colleagues (2023), found that when they asked the AI image generator Midjourney to produce images of 'Black African doctors providing care for White suffering children', it persisted in depicting the children as Black and the doctors as White.

Other times, the guardrails that companies program into their models simply result in them avoiding certain topics, which further invisibilises difference. When I asked ChatGPT to compose a joke featuring a character who is an African American birdwatcher, for instance, it cautioned me, saying:

> It's important to be mindful of cultural sensitivities and avoid perpetuating stereotypes. However, I can share a lighthearted, non-offensive birdwatching joke for you:
>
> Why did the bird bring a pencil to the birdwatching party?
>
> Because it wanted to draw some attention!

I can't help but find the model's sudden sensitivity to stereotypes rather rich, given the caricature of an African American birdwatcher that it had just produced minutes before. But there is something even more disturbing about its imposition of colour blindness, which is the implication that it would somehow be impossible to talk about African American birdwatchers without engaging in stereotypes, and impossible to tell a joke involving one without being offensive. It is as if the model is telling me: 'Why do you need to bring race into it? What does being African American have to do with birdwatching anyway?' Well, according to Christian Cooper, past president of the Harvard ornithological club, quite a lot, which is the point he makes in his recent book, *Better Living through Birding: Notes from a Black man in the natural world* (2023). Apart from his bestselling book, Cooper's main claim to fame is a viral video that featured a White woman threatening to call the police on him while he was watching birds in New York's Central Park (Walters, 2020), an incident that reminds us that, in the real world, being African American can have quite a lot to do with birdwatching.

There are lots of things applied linguists can study when it comes to intercultural communication. The current trend is to focus on the fluidity of culture, the 'trans-cultural', the creative way that people draw on multiple resources to create 'vibrant identities' and 'find joy in difference' (Ros e Sole et al. 2020: 397), in other words, to focus on culture as a verb. But an equally important endeavour is to explore how, despite this, many (if not most) people continue to insist on turning culture into a noun, on solidifying it and building walls around it, usually for the purpose of giving people they like a sense of belonging or letting people they don't like know that they don't belong, or, at least, that they don't count, or (perhaps more apropos to the current discussion) that they 'don't compute'. This latter endevour focuses on how people discursively 'manufacture' the 'idea' of culture, and what they do with it once they've made it (Zhu et al. 2022). AI chatbots can be many things— learning aids, writing assistants, diagnostic collaborators, creative confederates. But the one thing that they most certainly are is 'culture machines', tools that only promise to supercharge this project turning culture into a noun, a project which also serves the interests of our current crop of xenophobic politicians, 'personalised' marketers and other scions of the 'culture industry'. For this reason, along with embracing the exciting opportunities generative AI presents for creating 'culturally sensitive' chatbots (Brandt and Hazel, 2024) and assisting in preparing students for intercultural encounters (Dai et al., 2024), scholars of intercultural communication must also cultivate a critical perspective towards these machines and their makers and the ways they will inevitatably exert control over how we think about culture and how we are able to 'do' it.

References

Adorno, Theodor W. and Max Horkheimer. 2016. *Dialectic of enlightenment*. London: Verso.

Alenichev, Arsenii, Patricia Kingori, and Koen Peeters Grietens. 2023. Reflections before the storm: The AI reproduction of biased imagery in global health visuals'. *The Lancet Global Health* 11(10):e1496–98.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? 🦜'. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, *FAccT '21,* 610–23. New York: Association for Computing Machinery.

Benjamin, Ruha. 2019. *Race after technology: Abolitionist tools for the new Jim Code*. Cambridge, UK ; Medford, MA: Polity.

Bjork, Collin. 2023. ChatGPT threatens language diversity. More needs to be done to protect our differences in the age of AI'. *The Conversation*. Retrieved 8 January 2024 (http://theconversation.com/chatgpt-threatens-language-diversity-more-needs-to-be-done-to-protect-our-differences-in-the-age-of-ai-198878).

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, and others. 2022. On the opportunities and risks of foundation models. *arXiv repository*. http://arxiv.org/abs/2108.07258.

Brandt, Adam and Spencer Hazel. 2024. Towards interculturally adaptive conversational AI. *Applied Linguistics Review.*

Cave, Stephen, and Kanta Dihal. 2020. The whiteness of AI'. *Philosophy & Technology* 33(4).685–703. doi: 10.1007/s13347-020-00415-6.

Cooper, Christian. 2023. *Better living through birding: Notes from a Black man in the natural world*. New York: Random House Inc.

Cheng, Myra, Esin Durmus, and Dan Jurafsky. 2023. 'Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models'. *arXiv repository*. http://arxiv.org/abs/2305.18189.

Dai, David Wei, Shungo Suzuki, and Guanling Chen, 2024. Generative AI for professional communication training in intercultural contexts. Where we are now and where we are heading. *Applied Linguistics Review*.

Feagin, Joe R. (2013). *The white racial frame: centuries of racial framing and counter-framing* (2nd ed.). New York: Routledge.

Gal, Susan, and Kathryn Ann Woolard. 2001. *Languages and publics: The making of authority*. St. Jerome Publishing.

Jenks, Christopher J. 2024. Communicating the cultural Other: Trust and bias in generative AI and large language models. *Applied Linguistics Review*.

Luccioni, Alexandra Sasha, Christopher Akiki, Margaret Mitchell, and Yacine Jernite. 2023. Stable bias: Analyzing societal representations in diffusion models. *arXiv repository.* http://arxiv.org/abs/2303.11408

Nosta, John. 2023. Recursive training—A precarious echo chamber of artificiality'. *Medium*. Retrieved 21 January 2024 from https://johnnosta.medium.com/recursive-training-a-precarious-echo-chamber-of-artificiality-8361b04c6b15.

O'Regan, J.P., & Ferri, G. 2024. Artificial intelligence and depth ontology: implications for intercultural ethics. *Applied Linguistics Review*.

Pradhan, Alisha, and Amanda Lazar. 2021. Hey Google, do you have a personality? Designing personality and personas for conversational agents'. *CUI 2021 - 3rd Conference on Conversational User Interfaces*. 1-4.. Bilbao, Spain: ACM.

Scollon, Ron, Suzanne Wong Scollon and Rodney H. Jones. 2012. *Intercultural communication*: *A discourse approach.* London: Blackwell.

Shah, Saqub. 2023, June 23. Fascist chatbots are running wild on Character.ai. *London Evening Standard*. Retrieved 21 January 2024 from https://www.standard.co.uk/news/tech/character-ai-fascist-chatbots-offensive-generative-b1089868.html#:~:text=Character.AI%20is%20a%20hugely,dialogue%20agents%20have%20gone%20rogue.

Spiers, Elizabeth. 2024, January 7. Opinion: I finally figured out who ChatGPT reminds me of'. *The New York Times*, Retrieved January 7, 2024 from https://www.nytimes.com/2024/01/07/opinion/chatgpt-generative-ai.html?smid=nytcore-ios-share&referringSource=articleShare

Walters, J. 2020, May 26. Video of white woman calling police on black man in Central Park draws outrage. *The Guardian*. https://www.theguardian.com/us-news/2020/may/26/central-park-new-york-white-woman-black-birdwatcher

Weatherby, Leif. 2023, April 17. 'ChatGPT Is an Ideology Machine'. *Jacobin*, Retrieved January 19, 2024 from https://jacobin.com/2023/04/chatgpt-ai-language-models-ideology-media-production/

Zhu Hua, Rodney H. Jones, and Sylvia Jaworska. 2022. Acts of distinction at times of crisis: An epistemological challenge to intercultural communication research'. *Language and Intercultural Communication* 22(3):312-323.