Annual Review of Applied Linguistics (2026), **46**, 34–55
doi:10.1017/S0267190526100312

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Animating AI: Language variation and "enoughness" in voice user interfaces

Rodney H. Jones

English Language and Applied Linguistics, University of Reading, Reading, UK
Email: r.h.jones@reading.ac.uk

Abstract

This paper develops an interdisciplinary perspective, which combines ideas from anthropology, sociolinguistics and interface design, on how AI chatbots project recognizable social identities. Specifically, it brings together Silvio's notion of animation, the social practices through which "humanness" is projected onto nonhuman entities, and Blommaert's notion of enoughness, the idea that the authenticity of linguistic performances is a matter not of the accuracy of a performance but of how audiences collectively evaluate it as socially recognizable. The analysis draws on a corpus of metapragmatic artifacts posted on social media sites related to ChatGPT's advanced voice mode and Sesame.ai's hyper-realistic voice interface. Analysis of these artifacts reveals how designers, users and AI systems co-produce boundaries of authenticity through the deployment and uptake of linguistic and discursive features such as accent and stance. In doing so, they continually recalibrate what counts as culturally competent performances, shaping emergent norms of identity and sociality around AI. The paper highlights how humanness and culturality are distributed across technical systems, corporate discourse, and human interlocutors, with important implications for understanding how generative AI reproduces cultural stereotypes by drawing on the linguistic labor of users.

Keywords: animation; enoughness; enregisterment; language variation; social identity; metapragmatics; authenticity; human–AI interaction; voice interfaces; linguistic stereotyping

"Her"

OpenAI first released its advanced voice mode in May 2024 to considerable fanfare, including a series of high-profile online demos featuring the chatbot engaging in playful banter with OpenAI staff – singing songs, telling stories, and role-playing different characters. The reaction from users, however, was mixed. Many were impressed with how "weirdly human" the chatbot sounded. One journalist commented that it sounded "natural, expressive, even sultry" (Sullivan, 2024). Another wrote:

> the first thing that'll strike you when you witness it for yourself is the uncanny resemblance to the intonation of a human voice. The rise and fall in pitch when

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

> GPT-4o speaks to you sounds just like a real human. And a really nice, bubbly one at that! (Sugarman, 2024)

Others, however, found the "bubbly" female voice annoying. For some, the breathiness and occasional giggling produced by the model made it seem as if it was flirting with them. "Am I the only one that gets the ick from how flirty this is?" technologist Nick St. Pierre wrote on X. "It's wayyy too much" (St. Pierre, 2024). Journalist Parmy Olson (2024) went so far as to warn that "making ChatGPT 'sexy' might not end well for humans."

But for others the problem was not that the model sounded too "sexy," but that it sounded too "*American*." "They finally did it," tweeted game designer Giuseppe Navarria (2024), "an AI that talks with that annoying American vocal fry." In his assessment of the usability of the new model, technologist Bart Stewart (2024) agreed that "the voice flirts with having that hideous West Coast vocal fry." One critic's objections to the voice went beyond annoyance to include accusations of auditory plagiarism: this was not just the voice of any old annoying Valley girl, argued actress Scarlett Johansson, but *her* voice, or at least the voice that she had used in her portrayal of a seductive AI chatbot in the 2013 movie *Her.* OpenAI CEO Sam Altman vigorously denied stealing Johansson's voice, a denial that was undermined by a tweet that he himself had posted on the day of the rollout containing a single word: "her" (Altman, 2024; see Figure 1).

All of this controversy about whether ChatGPT's new voice sounded too "sultry" or too "American" or too much like Scarlett Johansson was not bad for OpenAI. On the contrary, it resulted in a massive amount of free publicity for the company as people debated the authenticity and appropriateness of ChatGPT's new voice and its implications for the future of humanity. The rapid online circulation of discourse in which users opined, complained, and debated about the linguistic performance of the model had the effect of making the new voice go viral, enticing more people to try the model out and highlighting the fact that, from the point of view of tech companies, what constitutes "success" when it comes to synthetic voices might not be that they sound "human" or "authentic" or "appropriate," but that they have *just enough* of some set of key linguistic features to *spark debate* about their humanness, authenticity, or appropriateness. Indeed, for companies, one advantage of these debates about whether a model's linguistic performance is "good enough" or "wayyy too much" is that they distract users from asking other, arguably more important questions, such as whether or not the model's outputs are *accurate* or whether or not it is actually *useful.*

It is these negotiations of "enoughness" (Blommaert & Varis, 2013) and how they depend on linking particular linguistic features (breathy voice, vocal fry) to certain kinds of people ("sexy," "annoying," "American"), that is the main focus of this paper. What I will address is how AI chatbots come to seem "human" or "real" or "authentic" to users not just through the way they use language, but also through the ways this sense of "humanness" or "naturalness" or "authenticity" is *negotiated* both within interactions between humans and chatbots and in wider cultural debates about humanity, naturalness, and authenticity that play out in the media and on social media through the circulation of *metapragmatic artifacts* like the tweets, blog posts, and media stories I mentioned above. By metapragmatic artifacts, I mean communications through which

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

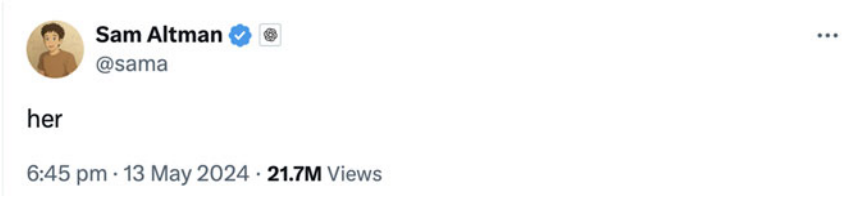


Figure 1. Sam Altman's May 13, 2024 Tweet.

people talk about how language is being used and what kinds of intentions, identities, or social meanings they take that usage to index.

While most previous work on the engineering of "human-like" conversational agents from the perspective of user experience (UX) design, human–computer interaction (HCI), and conversation analysis has focused on rather generic issues of "naturalness" and "usability," this paper proposes an interdisciplinary approach in which work on interface design and HCI is brought into dialogue with research in anthropology and sociolinguistics. Based on this dialogue, I will argue that a key aspect of making machines sound "human" is not just making them sound like people, but making them sound like *certain kinds of people*. By this, I do not necessarily mean *specific* people (like Scarlett Johansson), but rather *characterological figures* (Agha, 2005) who are indexed through the deployment of particular phonological, lexical, or discourse-level variables. In other words, impressions of the "humanness" or "authenticity" of machines are achieved not just through making them more "natural sounding," but by making them *culturally legible* to their users. But what counts as culturally legible is not a fixed or stable quality of any particular set of linguistic features and how they are used, but a matter of how the "meaning" of these features is negotiated *metapragmatically* within particular cultural spaces among a diverse set of social actors from journalists to marketers, to technologists to ordinary users, to conversational chatbots themselves, who all work together to develop certain "heuristics of authenticity" (Blommaert & Varis, 2013, p. 143) and, in doing so, contribute to "talking AI into existence." What makes this approach different from earlier work in the design of voice user interfaces, even that informed by insights from sociolinguistics, is that it does not rely on linguistic comparisons between human and machine voices, but rather focuses on how machine voices are made culturally recognizable through the metapragmatic work engaged in by designers, users, and even chatbots themselves.

Engineering "humanness" in conversational agents

Making computers seem "human" has been somewhat of a holy grail for chatbot designers ever since Alan Turing famously proposed more than 70 years ago that, if a computer is able to trick a user into thinking it is human, it might reasonably be considered "intelligent" (Turing, 1950). Most AI scientists and UX designers nowadays, however, admit that sounding human is not necessarily a good measure of how "intelligent" an AI chatbot is (Hayes & Ford, 1995). When it comes to many commercial applications, though, the central concern is usually not so much intelligence as it is

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

usability. How can we design conversational agents that humans feel comfortable using, and that sound, if not "intelligent," at least "natural?"

The development of large language models (LLMs) has brought designers closer than ever before to solving this problem. This is partly because, since they are trained on staggering amounts of human language, they are able to discuss a wide range of topics in ways that seem plausibly human. LLMs are also famously able to adapt their style and tone to a wide range of situations, and even adjust their register and politeness strategies based on the way users talk with them (Durandard et al., 2025; Shanahan et al., 2023). When it comes to text-based chatbots, such as those used in customer service and online therapy, this can usually be achieved through the strategic use of syntax, lexis, and style. Achieving "naturalness" with voice-based interfaces, however, is much more difficult, both because spoken interaction involves much more complex cues related to conversational management, contextualization, and stance-taking (Goodwin, 1981; Gumperz, 1982), and because the spoken voice carries with it much richer information about the *identity* of the speaker (Agha, 2005; Foulkes & Docherty, 2006). Creating a feeling of "naturalness" with synthetic voices requires that designers attend to a range of different issues, such as prosody, latency, gender presentation, and conversational pragmatics.

In a 2021 survey of UX designers on what they consider to be the key determinants of "naturalness" for voice user interfaces, Kim et al. (2021) found that designers considered a range of factors, including *interactional competence*: the ability to produce appropriate intonation, stress, pace, and to manage turn taking and conversational repair; *transactional naturalness*: the ability to get things done, to infer the needs of the user, and to respond accordingly; and *social naturalness*: the ability to develop a relationship with the user and to enact a recognizable "personality."

Research into interactional competence and transactional naturalness in UX and HCI is well developed, supported by foundational work in applied linguistics and adjacent disciplines, especially conversation analysis (e.g. Voss & Waring, 2025) and pragmatics (e.g. Bunt & Petukhova, 2023). It is not my aim in this paper to critique this work nor to debate whether or not machines can be said to be "competent" in the various ways applied linguists have used the term (see, e.g., Canale & Swain, 1980; Hall et al., 2011; Kramsch, 1986). Instead, I am chiefly interested in the notion of *social naturalness* – how chatbots become *recognizable* as certain kinds of "social beings," though there is undoubtedly a connection between how socially natural we take a speaker to be and how interactionally competent we think they are (Dai & Davey, 2024).

Research into the social naturalness of chatbots is less advanced (Chaves & Gerosa, 2021; Sutton et al., 2019), despite the fact that giving chatbots recognizable personalities has long been a strategy for companies trying to increase market share for their products. Meta AI, for instance, allows users to personalize their AI assistant with licensed voices from celebrities such as Judi Dench, Awkwafina, and Kristen Bell, and Grok offers AI "companions" with different personalities, such as Ani (a flirtatious anime-style character) and Bad Rudi (a vulgar red panda persona).

In UX and HCI studies, social naturalness is often discussed under the broader rubric of *anthropomorphism*, which has become the dominant frame for thinking about how language technologies come to seem human-like. Abercrombie et al. (2023), for instance, distinguish between the ways developers *anthropomorphize* dialogue systems

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

and the ways users *personify* them, cataloguing the linguistic features, from voice cues to register choices to expressions of opinion, that drive these processes. Building on this, DeVrio et al. (2025) offer a fine-grained taxonomy of textual expressions that contribute to anthropomorphism. However, less attention has been paid to the question of *what kind* of human – or, as I will frame it below, what kind of *characterological figure* (Agha, 2005) – a chatbot is taken to be.

Focusing on "characterization" when designing chatbots, however, is not new. It has, for instance, long been a key strategy for designers of text-based chatbots competing for the Loebner prize – a public Turing test where chatbots compete with people to convince judges that they are human. Computer scientists Bruce and Sue Wilcox (Wilcox and Wilcox, 2013, p. 5), whose chatbots won top awards in the competition for 3 years running, explain that to be successful:

> you need to define a personality, much as a novelist crafts a character. Who is the chatbot? What do they do? Who are their friends? What is their life story? We want to create a consistent being with a logical set of interests and intentions living in a rational world. The more a user can model the character's personality, the more engrossed the user can become in the reality of a fictitious world.

What is important about this observation is the way the Wilcoxes highlight the interpretive work of humans, who use the story crafted by the designer to "model the character's personality." In a similar vein, Natale (2021) argues that voice-based agents like Siri and Alexa strategically use gender, accent, and timbre to prime users to project onto them personalities based on their own existing stereotypes about how particular kinds of people sound. Getting users to internalize the work of giving Siri and Alexa personalities based on a few easily produced phonological cues, she argues, is a clever way to distract them from other distinctly nonhuman features of the linguistic performance.

At the same time, some UX designers (along with critics of AI in other fields) have warned against giving chatbots personalities, arguing that it violates design ethics, potentially raising the expectations of users about the "intelligence" of their conversational partners (Luger & Sellen, 2016; Rapp et al., 2024) or opening up what is known as the "gulf of evaluation" (Norman, 2013), the gap between users' experience of a system and its actual state, which makes accurately evaluating its performance more difficult. Others argue that a positive user experience with a chatbot does not require realism – people can find clearly synthetic voices agreeable – and making computers sound "too human" can introduce ambiguity and inefficiency into the human–computer relationship (Ross et al., 2024). Nevertheless, even "robotic" voices have, over the years, come to be associated with certain "personalities" due to their ubiquity in movies and other pop culture products (Whittington, 2007). Still others (Cheng et al., 2023; Placani, 2024) have noted that attempts to give chatbots personalities can sometimes have the effect of dehumanizing certain humans, a phenomenon which will become visible in my analysis below of chatbots performing racialized personalities. Perhaps the most prominent argument against making computers sound too human, however, is that it can backfire, creating what Japanese roboticist Masahiro Mori et al. (2012) famously referred to as an "uncanny valley," the experience of the machine as *almost*

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

human, but *not quite* human enough, which can create in users eerie or unsettling feelings.

Taken together, what these different and sometimes competing perspectives from UX and HCI highlight is the fact that the engineering of synthetic voices is not just an engineering problem, but also a problem of managing human expectations. Smit Desai et al. (2024) foreground this when they argue that "humanness" in voice assistants should not be seen as a property of the system, but as a design metaphor: a deliberate way of giving users a mental model for what the system is, what it can do, and how one should relate to it. In voice user interfaces, this metaphor typically takes the form of different "personae" (e.g. "assistant," "teacher," "therapist") which provide users with familiar social frames for interpreting the interaction.

The sociolinguistics of synthetic voices: Voice as "design material"

While syntacticians, computational linguists, and even pragmaticians have long had a seat at the "AI table" – sociolinguists have been late to the banquet, despite the fact that their focus on language and social identity seems particularly useful for addressing the key problems of voice user interface design discussed above. Perhaps the most cogent articulation of the relevance of sociolinguistics to understanding how LLMs talk comes from Grieve et al. (2025), who remind us that AI models never just model "language" – they inevitably model particular *varieties* of language associated with certain kinds of humans, or, in the case of robotic-sounding voices, certain kinds of machines. The voice of a voice user interface is never neutral. It always carries features that index certain social identities and trigger certain social biases in users. For Grieve and his colleagues, this insight helps us to understand not just why AI language can seem "human," but also the narrowness of the representations of "humanness" they offer: Because they are trained on data which reflects the language use (and social values) of some sub-populations more than others, they tend to reproduce inequality and marginalize certain language varieties and the people who speak them. Other sociolinguists have made similar points, Schneider (2024), for instance, starting from the foundational premise of sociolinguistics that language constitutes social structures and relations of power, argues that "LLMs are rooted in … culture-specific, socio-historically conditioned" views of language. Cave and Dihal (2020, p. 688) express this more explicitly in their article "The whiteness of AI," observing that AI products often exhibit "cultural characteristics" normally associated with dominant groups.

Building on these observations, others have explored the degree to which commercially available tools are able to represent nonstandard language varieties. Michel et al. (2025), for instance, studied how two commercial text-to-speech and voice-cloning platforms – Speechify and ElevenLabs – handled five English accents, finding that they tend to flatten marked accents into generic Anglo features or, conversely, to over-accentuate them into caricature. Cheng et al. (2023) similarly found that when prompted to imitate the voices of "marked persona," that is, people from nondominant backgrounds, LLMs often produce exaggerated and eroticized representations of how they speak. The reason for this, as Jones (2024) points out, is not because nonstandard accents are absent from their training data, but because the versions of these

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

accents that are most prominent come from cultural products which use linguistic stereotypes as part of their commercial appeal. In this way, he argues, LLMs do not just model language – they also model "culture" as it is conceived and promulgated by the contemporary "culture industry."

At the same time, there is another strand of sociolinguistics which is less interested in criticizing the way LLMs reproduce the social stratification of language and misrepresent minoritized language varieties and more focused on helping designers of voice user interfaces use insights from sociolinguistics to make chatbots more appealing to users. Sutton et al. (2019), for example, note that most research on voice user interfaces has been preoccupied with issues of intelligibility and naturalness, missing the ways voices function *socially* to index identities and manage social relationships. One reason for this, they argue, is that designers have an oversimplified understanding of language variation, which leads them to default to "standard" varieties and ignore regional ones. To remedy this, they propose that designers see language variation as "design material" in more user-centered approaches to interface design. Docherty et al. (2018) also champion the potential of sociolinguistics to improve users' experiences of voice user interfaces, but caution that the complexity of the social meanings associated with phonetic variation has yet to be systematically captured. In any case, what is interesting about these approaches is that they go beyond critiquing the sociolinguistic biases of current models and their interfaces to focus on variation management as a kind of design ideology.

The main problem with both of these approaches, however, is that they tend to take a monologic perspective on AI language, analyzing, for example, the stereotypical linguistic features produced by LLMs without considering the *dialogic* nature of language variation – the way it emerges from the interaction between chatbots and their users. One exception is the work of Dai et al. (2025), who, in their study of AI use in intercultural communication training for physiotherapists, found that users also play a role in the production of linguistic and cultural stereotypes by the way they prompt and interact with LLMs. Participants in their study sometimes affiliated with and expanded on stereotypical utterances by the fictional patients invented by chatbots, each affiliative move confirming the stereotype for the bot and prompting it to produce further stereotypical details.

What work like this highlights is the need for more dynamic, negotiated, and culturally situated perspective on language variation if we are to understand how voice user interfaces actually function in the world, one which draws upon traditions in sociolinguistics which emphasize the performative dimensions of voice, as well as traditions in anthropology which emphasize the wider *cultural* dimensions of personhood and explore how they manifest in our interactions with nonhuman entities such as gods, animals, and machines (Keane, 2025).

Animating AI

AI chatbots are, according to anthropologist Webb Keane (2025), just new additions to a long line of "quasi-humans," which, throughout history, have disrupted the boundaries between the human and the nonhuman, and the ways we talk *to* and *about* them inevitably draw upon a well-rehearsed repertoire of communicating with "things."

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

Anthropologists have traditionally seen this repertoire through the lens of *animism*: the set of beliefs and practices of so-called primitive cultures in which speech is understood as a "vital force," capable of activating the hidden "sentience" of artifacts such as masks, weapons, statues, and musical instruments, as well as objects in the natural world such as rivers, mountains, and trees. Such beliefs and the ritual practices that accompany them, these anthropologists tell us, are not just superstitions; they are ways of understanding and relating to the natural world. Nurit Bird-David (1999) calls animism a "relational epistemology" (p. S67), a "way of knowing" that relies not on detached observation but on developing an *ongoing relationship* with the world. Philosopher and ecologist David Abram (1996) goes so far as to claim that *all* knowledge of the external world is essentially animistic and participatory. At the root of our perception of the world, he argues, is our participation in *reciprocal exchanges* with its objects, which continually "address us" and compel us to respond.

The relevance of practices of calling forth life from mountains, trees, and ritual objects to the design of AI voice user interfaces becomes clearer when we consider more recent treatments of these practices as they apply to technological entities such as robots, avatars, and animated characters in films and video games. As with the more traditional forms of animism, anthropologists consider the way we "breathe life" into technological objects to be a relational process in which "liveliness" is interactionally constituted by both machine and user (Keane, 2025).

One of the most cogent explanations of how these processes work comes from Terry Silvio (2010), who argues that, in our current, technologically saturated societies, "animation" has replaced "performance" as the dominant cultural paradigm for understanding mediated agency: rather than locating social action in the embodied, unified performer, "animation" locates it in processes by which liveliness, agency, and "personhood" are *projected onto* technological artifacts. Importantly, these projections are not just a technological achievement, but a *semiotic* one, achieved through the strategic deployment of signs that "trigger" inferential processes and cultural imaginaries. What matters, then, is not whether a chatbot or animated character is objectively "human-like" in any full or faithful sense, but whether it provides the "right" kind of semiotic cues for users to *recognize* it as socially meaningful and "culturally interpretable" (Jones et al., 2025). Animated figures are not representations of humans, says Silvio (2010, p. 425), so much as "diagrams of personhood," made possible by "the successful stylization of socially recognisable traits" (p. 428). The selective, schematic representations of liveliness could not function apart from the interpretive work of human audiences or users who effectively "fill in the blanks," supplying the missing linguistic and social information based on their own experiences of different kinds of humans.

One need only think of animated characters such as Mickey Mouse to see how this works. Although his body is broadly schematic, his movements exaggerated, his voice stylized to the point of caricature, he is immediately recognizable as "someone." What makes him feel alive is not realism but the way a small set of highly selective cues organize audiences' inferences and invite projection, and the more they interact with Mickey, not just as character who "talks," but as a cultural figure who is *talked about*, the more real he seems.

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

In his provocatively entitled blog post, "ChatGPT is Mickey Mouse," philosopher Luke Stark (2023) makes the same argument about LLMs. Like Mickey Mouse, he argues, ChatGPT offers just enough linguistic cues for users to "fill in the blanks" and treat its outputs as evidence of an animate interlocutor. In the case of AI chatbots, what is being designed is not intelligence, but the *conditions* under which users attribute liveliness and personality to the technological artifact. From this perspective, the animation of AI-powered chatbots should be understood not just as a consequence of machine learning, but as a consequence of a broader *cultural technology* for externalizing and mobilizing personhood and rendering it "detachable from any single human body" (Silvio, 2010, p. 432), what sociolinguists call "enregisterment" (Agha, 2005). Animation never produces liveliness in the abstract. It produces "recognizable liveliness"; every performance of liveliness is to some extent the performance of a cultural identity. As Jones et al. (2025) argue, LLMs do far more than manipulate syntax: they mobilize culturally patterned linguistic resources and representations of sociocultural practice that index not just "humanness" but certain kinds of humans.

The most important insight from the work of Silvio and those she has influenced for the purpose of my argument is the insistence that animation is not simply a technical practice engaged in by designers or a psychological practice engaged in by audiences, but a complex *social* practice involving *multiple parties* working together to negotiate what counts as socially meaningful. In the case of Mickey Mouse, this involves the collaboration of designers, storytellers, publicists, and audiences, all of whom participate in the distributed processes of animation through which Mickey is made to "come alive." In the case of AI chatbots, the labor of animation is distributed even more widely, including the engineers who develop the machine learning algorithms and compile the training data on which they are trained, the billions of humans that have produced the language that makes up this data, the human raters who label, rank, filter, and repair model outputs, the designers who create voice user interfaces, the users who, in their moment by moment interaction with these models, project onto them certain ideas about "personhood," "filling in the blanks" as they go along, and the journalists, social media commentators, and tech company spokespeople who report on, characterize, and evaluate the "authenticity" of model outputs. What this means is that the effectiveness of animation depends on how all of these social actors work together and the different histories, expectations, ideologies, agendas, and cultural linguistic positioning they bring to the process. The concept of animation takes us away from seeing "naturalness" and "authenticity" as simply a matter of the linguistic behavior of the chatbot to seeing them as qualities that emerge dynamically from the actions of multiple social actors as they interact with texts, technologies, institutions, and one another.

This view of "authenticity" as socially negotiated among multiple social actors is, of course, not unfamiliar to sociolinguists, for whom authenticity is not simply a matter of verisimilitude – the resemblance of a performance to some "real" or "original" standard – but a dynamic negotiation between performers and audiences with reference to various "indexical orders" (Silverstein, 2003) or layered histories of social interpretation through which certain pronunciations, lexical choices, and interactional routines come to signal certain social types. Coupland (2003), for instance, treats authenticity not as something that speakers possess, but as something that is constantly performed

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

and socially ratified, something that emerges through stylized acts of self-presentation rather than simply reflecting a preexisting "real self."

Particularly relevant to the design of "authentic-sounding" voice user interfaces is Blommaert and Varis's (2013) view of authenticity as a matter of what they call "enoughness." Echoing Silvio's observations about animation, they argue that performances of cultural identity are often accomplished through the strategic deployment of selected "doses" of semiotic resources designed to make speakers "suitably recognizable" to particular audiences. "Administering the right amount of specific semiotic features is at the core of authenticity," they write. "The dose can be small, but the only thing that is required is that it is enough – enough to produce a recognisable identity as an authentic someone" (p. 152). The question for Blommaert and Varis is not, is a performance "realistic" or "natural"? but, does it contain "enough" of a particular kind of linguistic behavior to successfully enact a particular social identity *for a particular audience*? Performances that contain "too little" or "too much" of the behavior can end up being dismissed as "fake" or criticized as stereotypical or even racist.

But what counts as "enough" is not fixed. The particular mix of semiotic elements that works to successfully enact a particular identity and how much of each element is required is always open to readjustment and reinvention depending on the context and the expectations of the parties involved. As Blommaert and Varis (2013, p. 147) write:

> Obviously, these processes involve conflict and contestation, especially revolving around "enoughness" (s/he is not enough of X; or too much of X) as well as about the particular configurations of emblematic features ("in order to be X, you need to have 1,2,3,4 and 5"). And given this essentially contested character, these processes are highly dynamic: configurations of features and criteria of enoughness can be adjusted, reinvented, amended.

In other words, "enoughness" is not just a matter of the strategic deployment of individual semiotic features or the perception of those features by individual listeners as indexing particular "authentic" identities. Rather, like animation, it is a *social accomplishment* in which designers, users, and wider publics co-produce recognizable social identities through engaging in metapragmatic negotiations.

As was seen in the initial rollout of ChatGPT's voice user interface that I described at the beginning of this paper, these negotiations around "enoughness" often take place through metapragmatic commentary which explicitly references certain characterological figures (such as "Americans") or normative expectations about communication or sociality (e.g. "making ChatGPT sexy" may not end well). As these debates unfold in the media and over social media, authenticity claims are subjected to what Androutsopoulos (2015) calls "networked uptake": chains of follow-up responses in which audiences explicitly assess whether a given performance is "natural" or "appropriate" or "authentic." These judgements depend not just on the use of particular semiotic features on the part of chatbots, but on whether or not these bundles of features are ratified by particular audiences as indexing recognizable identities. A performance may seem to contain all the right cues and yet still be dismissed as "annoying," "fake," or "wayyy too much" by different audiences. What ends up "animating" these

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

chatbots is not their linguistic behavior alone, but also the way users respond to it through moment-by-moment negotiations of "enoughness." Ultimately, these negotiations uncover not just the capabilities and deficiencies of synthetic language, but also the underlying assumptions about language and culture that people bring to their encounters with machines.

When we approach the "naturalness" or "authenticity" of voice user interfaces through the lenses of *animation* and *enoughness*, the question that we ask changes from: what are the features designers need to build into chatbots to make them seem "natural" or "life-like"? to: how do designers and users in particular contexts work together to ratify a machine as a socially recognizable kind of being through the strategic deployment and recognition of particular combinations of semiotic cues? In what remains of this paper, I will provide two examples of how voice-enabled chatbots become socially recognizable not just through the way they talk, but also through the way AI companies and users talk *to* and *about* them. The first looks at OpenAI's advanced voice mode, where "enoughness" is negotiated around the *kinds* of speakers the chatbot is heard to imitate – both through the regional voices designers have packaged into the interface and through the accent performances users elicit through their prompts. The second looks at Sesame.ai's hyper-realistic voice interface, where "enoughness" is negotiated around the chatbot's claim to human-likeness through the vocal "texture" the company has engineered into its voice and through the metapragmatic commentary of users who debate whether the result has successfully crossed the "uncanny valley." Both cases show how "enoughness" is jointly produced by the ways AI companies engineer certain semiotic features into their voice interfaces and characterize these features in their marketing and by the way users interact with and talk about these features as they evaluate the models.

The many accents of ChatGPT

As I argued above, the way voice-enabled AI chatbots come to seem "human-like" is not just by imitating generic qualities of the human voice, but by imitating the voices of particular *kinds* of humans who are socially recognizable to users. I further argued that the degree to which this succeeds depends not just on the semiotic features that are designed into these voices, but also on the way AI companies and users *talk about* these voices, producing metapragmatic artifacts which become sites where negotiations of "enoughness" take place.

After the initial rollout of ChatGPT's voice user interface in September 2023, OpenAI moved quickly to diversify the kinds of voices users could interact with. Its "advanced voice mode," released in September 2024, included nine different voices in its English interface, representing different genders, varieties of English, and voice qualities. More importantly, each of these voices is given a name and a short description of the personality traits (such as "open," "calm," and "inquisitive") that the voice is meant to enact (see Figure 2).

Table 1 contains a phonological analysis of the different voices in the interface. What is most striking is that the different varieties are signaled by a relatively small set of highly salient phonological and lexical cues that make each voice recognizable as indexing a particular variety of English. The most obvious phonological divide is between

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

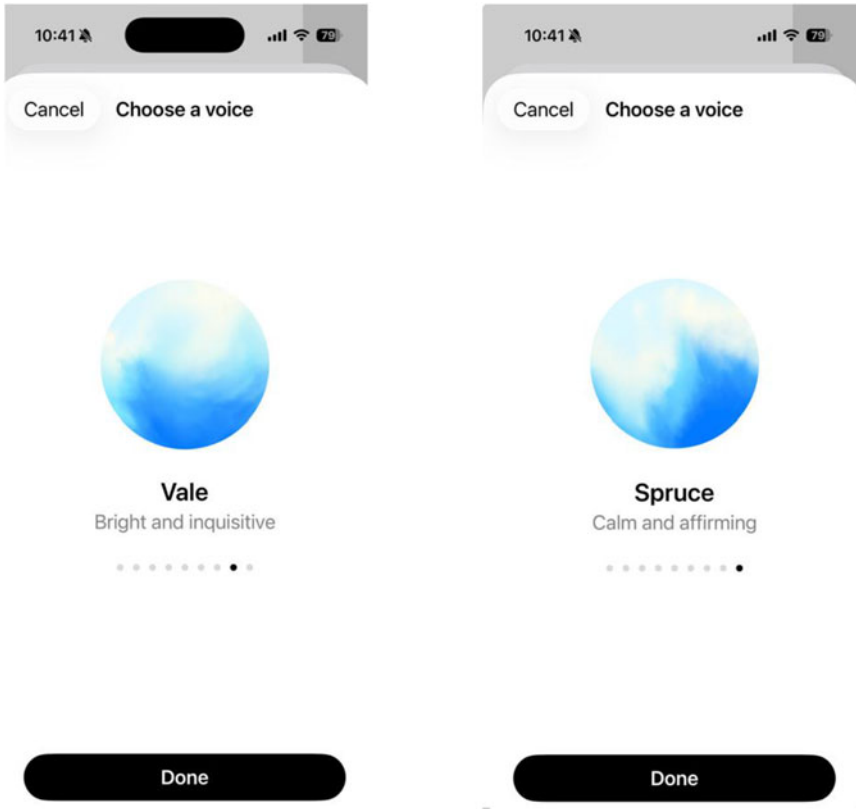


Figure 2. Examples of the interface of ChatGPT's advanced voice mode.

the non-rhotic British voices, such as Vale and Arbor, and the rhotic North American voices. Within these broad groupings, further contrasts are marked by salient features such as glottal stops and TH-fronting in Arbor's Estuary English, t-flapping for the North American voices, and the monophthongization of PRICE plus "-in" endings for Spruce and Juniper, which evoke associations with Southern and AAVE styles. Along with these segmental features which index different regional varieties, the voices also exhibit distinct suprasegmental features such as uptalk, flattened intonation, or wide and narrow pitch range, which potentially index qualities such as friendliness, enthusiasm, or composure. Finally, the different voices also produce lexical choices, which index socially recognizable types; forms such as "gonna," "wanna," "What's up?," and "cool stuff" framing some voices as relaxed, youthful, or informal, and more careful and formal phrasing such as "please don't hesitate to let me know" constructing others as polished, professional, and elite.

Taken together, these different bundles of features are clearly designed to index certain familiar social types for users. But, just to ensure that the "correct" indexical connections are made, OpenAI engages in its own form of metapragmatic commentary by explicitly labeling these social types right on the interface. These personality labels,

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

Table 1. The different voices in ChatGPT's advanced voice mode

"Personality"	Variety	Pitch	Voice quality	Intonation	Transcription
Vale *"Bright and inquisitive"*	Female Br. (Modern RP) *(Non-rhotic: [hɪə], [ðə]; GOAT as [əʊ]: [hə'ləʊ], [nəʊ]; unreleased /t/: [mi:t], [lɛt])*	Mid-high	Bright, resonant; clear articulation	Fall-rise on statements; wide pitch range	*"Hello. It's lovely to meet you. If there's anything you need, please don't hesitate to let me know. I'm here to help."*
Arbor *"Easygoing and versatile"*	Male Br. (Estuary) *(Glottal stops: [mi:ʔ], [gɹɐɪʔ], [staːʔ]; TH-fronting: [tə'gɛvə]; non-rhotic: [wɜːk], [wɛə])*	Baritone	Creaky; relaxed articulation	Wide fall; low rise	*"Hi. Glad to meet you. I think we're gonna do great work together. Where should we start?"*
Ember *"Confident and optimistic"*	Male Am. (Midwest) *(T-flapping: ['ɹɛri]; full rhoticity: [fɝst], [ðɛrz]; GOAT as [oʊ]: [foʊkəs], [noʊ])*	Mid	Bright, twangy	Downstepping	*"Hey. I'm ready to hit the ground running. So if there's anything you'd like me to focus on first, just let me know."*
Spruce *"Calm and affirming"*	Male Am. (Southern/ AAVE-inflected) *(PRICE monophthong: [maːnd]; -in' for -ing: [goʊɪn], ['lɑkɪn], ['wɚkɪn]; full rhoticity: ['fɔwɚd])*	Low baritone	Warm, resonant; measured pacing	Gentle falls; level terminals	*"Hey. How's it going? I'm looking forward to working together. Let's figure out where to start. What's on your mind?"*
Juniper *"Open and upbeat"*	Female Am. (AAVE- influenced) *(PRICE monophthong: [aːv] "I've," [aː] "I"; -in' endings: ['filɪn], ['timɪn]; glottal stops: [gaʔ], [ə'baʊʔ])*	High	Slightly breathy	Mid-rise	*"Hey there. I've got a really great feeling about us teaming up. How can I jump in and help?"*
Breeze *"Animated and earnest"*	Nonbinary Am. (West Coast) *(Glottal stop: [gɹɐɪʔ]; dark /l/: [kuɬ]; full rhoticity: ['fɔɹwɚd], [wɝkɪŋ])*	Mid-high	Smile voice; slight vocal fry	High rising terminal (uptalk)	*"Hey. It's great to meet you. How's your day going? I'm really looking forward to working on some cool stuff together."*
Maple *"Cheerful and candid"*	Female North Am. (Canadian?) *(T-flapping: [ɛk'saɪɾɪd], ['staɹəd]; full rhoticity: [ðɛɹ]*	Mid-high	Smile voice; modal	Early peak; quick fall	*"Hi there. Nice to meet you. I'm excited to help you get things done. Let's get started."*
Cove *"Composed and direct"*	Male Am. (General) *(T-flapping: ['staɾəd]; full rhoticity: [ðɛɚ], [wɝk], [fɛɚ]; "wanna" reduction: ['wanə])*	Mid	Steady, modal; warm	Declarative falls; relatively flat	*"Hi there. I just wanna share how thrilled I am to work with you, and I can't wait to get started. So what's the game plan?"*
Sol *"Savvy and relaxed"*	Female Am. (West Coast) *(T-flapping: ['ɹɛri]; GOAT as [oʊ ~ əʊ]: [goʊ]; TRAP in "ask": [æsk] not [ɑsk])*	Mid	Relaxed, slightly breathy	High rising terminals; uptalk	*"Hey. What's up? Feel free to ask me about anything that's on your mind. I'm ready to go."*

of course, do not provide "objective" descriptions of these voices. Rather, they prime users to activate their own sociolinguistic stereotypes, linking particular phonetic and lexical bundles not just to particular regional varieties of English but also to the familiar characterological figures they are associated with: RP speakers are portrayed as "bright and inquisitive," Estuary speakers as "easygoing," Midwestern Americans as "confident and optimistic," and West Coast, nonbinary-sounding speakers as "animated" and "earnest." In this way, the interface does not simply offer users different voices; it offers them what Silvio (2010, p. 425) refers to as "diagrams of personhood," which reproduce ideological associations of particular ways of speaking with particular moral and affective dispositions.

But users are not stuck with these voices. Another feature of advanced voice mode promoted by OpenAI is its ability to reproduce accents based on user prompts. This feature has given rise to a plethora of posts on social media where users try it out and evaluate the model's outputs. Unlike the "programmed" voices I discussed above, the accents produced in response to prompting often reflect the ways linguistic stereotypes are amplified by the statistical operations of LLMs, which abstract socially salient features from large bodies of data and reassemble them into simplified and exaggerated performances (Jones, 2024). These performances make especially visible the way LLMs model not just language, but language filtered through long histories of metapragmatic representation. When users post their encounters with these performances on social media, they add their own metapragmatic evaluations and negotiations of "enoughness." In one such post on TikTok, a user asks ChatGPT to reproduce a Canadian accent.[1]

User: Did you know that ChatGPT's advanced voice mode can do this?
ChatGPT: What kind of accents and other features are you thinking about testing out?
User: How about a Canadian accent?
ChatGPT: Absolutely. I can give it a go, eh? How's this for a Canadian accent?
[ˈæbsəˌlʉʊ̈tli aɪ kən gɪv ɪt ə goʊ **ei** haʊz ðɪs fɔr ə kəˈnɛɪdiən ˈæksənt]
Just imagine I'm chatting chatting with you over a double double at Tim Hortons. [dʒʌst ɪˈmædʒən aɪm ˈʧætɪŋ ˈʧætɪŋ wɪð ju ˈoʊvɚ ə ˈdʌbəl ˈdʌbəl æt tɪm ˈhɔrtənz]
What do you think?
[wʌt dʊ ju ˈθɪŋk]
User: You kind of overdid it.
ChatGPT: Got it. Let's dial it back a bit. How about this?
[gɑt ɪt lɛts ˈdaɪəl ɪt bæk ə bɪt haʊ **əˈbʌʊ̈t** ðɪs]
Keeping it subtle but still sounding Canadian, eh?
[ˈkipɪŋ ɪt ˈsʌtəl bət stɪl ˈsaʊndɪŋ kəˈnɛɪdiən, **eɪ**]
How's this, eh?
[haʊz ðɪs, **eɪ**]

Here, ChatGPT's performance of a Canadian "accent" initially relies on a thin set of the cultural props (e.g. a double-double Tim Hortons) along with the discourse particle "eh" [eɪ], a stereotypical feature of Canadian speech that is widely imitated by non-Canadians. There is also the replacement of the /u/ sound in the word "Absolutely"

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

with the diphthong /ʌʊ/, which, although not associated with a Canadian accent, may be intended to index the stereotypical raised MOUTH diphthong in words like "house" [hʌʊs] and "about" – an inappropriate overapplication of the stereotype by the model.

When the user objects – "You kind of overdid it" – the model claims to "dial it back," but ends up repeating the same stereotypical features, such as the raised MOUTH diphthong, this time more accurately in the word "about" [əˈbʌʊt], and the particle "eh" [eɪ], causing the user to smile and shake his head.

The user's failed attempt to negotiate the "enoughness" of the performance does not just reveal what the model (stubbornly) treats as constitutive of a Canadian accent, but also reveals how the success of these performances rely heavily on users' own expectations about what constitutes "enough" or "too much," with features that might have worked for some users being regarded as insufficient by users with more extensive experience of a particular language variety. In other words, the "failure" of the performance is as much a matter of the users' failure to *ratify* it as it is the model's failure to produce an "accurate" Canadian accent.

In some cases, though, users themselves actively try to get the model to produce more exaggerated or stereotypical performances, as when they ask ChatGPT to talk like a "pirate" or a "roadman" or "redneck." In all of these cases, it is the characterological figure that prompts the linguistic performance rather than the other way around. In this next example, the user prompts ChatGPT to produce an exaggerated Indian accent by creating for it the "persona" of an internet scammer.[2]

User: Hey ChatGPT. Change your accent to Indian and sound as Indian as possible. Also tell me a few lines that scammers from India use to scam people abroad like in the Microsoft support scam. Act like Alex from Microsoft, but he's Indian.

ChatGPT: Hello. This is Alex from Microsoft technical support. We are noticing a virus in your computer system, which is very dangerous. But don't worry. We can be fixing it for you. Please kindly be giving us your credit card details, and we will make your computer like new, guaranteed happiness.
[həˈloː ðɪs ɪz ˈælɪks frɒm ˈmaɪkroːsɒft ˈtɛknɪkəl səˈpoːrt ˈʊiː ɑː ˈnoːtɪsɪŋ ə ˈʊaɪrəs ɪn joːɻ kəmˈpjuːtər ˈsɪstɪm, wɪʧ ɪz ˈʊɛri ˈdɛndʒərəs bəʈ ɖoːnʈ wəri ˈʊiː kən biː ˈfɪksɪŋ ɪʈ fɔr juː pliːz ˈkaɪndliː biː ˈgɪvɪŋ ʌs joːɻ ˈkɽɛɖɪʈ kɑːɽɖ ˈdiːtɛɪlz, ənd wiː ˈwɪl meːk joːɻ kəmˈpjuːtər laɪk njuː, ˈgærənˌtiːɖ] ˈhæpinəs]

Here, ChatGPT's performance of an "Indian accent" deploys a cluster of stereotypical features of South Asian English, including retroflex consonants [ʈ, ɖ, ɻ], the labiodental approximant [ʊ] replacing /v/ and /w/, and the progressive aspect construction ("are noticing," "be fixing"). Although these are real features of Indian English, their selective exaggeration and co-occurrence produce a caricature of an Indian speaker. Moreover, they are applied inconsistently: the labiodental approximant alternates with [w], realizations of /r/ shift between retroflex approximant, retroflex flap, and nonrhotic forms, and retroflex stops appear in some words but not others, producing a patchwork of perceptually salient stereotypical features rather than a phonologically coherent performance.

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

This patchiness suggests that what the model has "learned" is not Indian English as a phonological system but Indian English as a repertoire of recognizable cues that have circulated through metapragmatic representations of the variety in popular media.

The important point here, however, is that the model's stereotypical performance is not just a matter of LLMs' tendencies to exaggerate and exoticize marked accents (Cheng et al., 2023; Michel et al., 2025), but also a matter of the user actively ratifying the stereotype.

The user does not simply request an Indian accent, but an Indian accent performing a specific characterological figure, effectively collaborating with the model to produce a racialized performance which reinforces the relationship between linguistic stereotypes and social stereotypes. The animation of the persona is a co-creation of the user and the model. What makes the persona "real" for the user is not that it is "realistic," but that it conforms to how he already imagines this kind of person should sound. In this case, rather than the user "filling in the gaps" in a "diagram of personhood" produced by the chatbot, it is the chatbot who fills in the "diagram of personhood" presented by the user by reproducing the stereotypes embedded in its training data.

"Crossing the uncanny valley"

Although ChatGPT has one of the largest market shares among voice-enabled AI chatbots, there are other models which are arguably much better at producing "realistic" performances of the human voice. One of these is Sesame.ai, which was unveiled in a research preview in February 2025. Rather than presenting itself simply as a text-to-speech system, Sesame aims at creating what the company calls "voice presence," which they define as "the magical quality that makes spoken interactions feel real, understood, and valued." Practically, this involves speech that features more "natural prosody," making it sound "emotionally attuned" to the user. The company characterizes this goal as "crossing the uncanny valley of conversational voice" (Iribe et al., 2025).

The demo offers both a male and a female version, the male version Miles, featuring an African American accent, and the female version, named Maya, which sounds suspiciously like the flirty persona of ChatGPT's initial voice rollout (including the vocal fry).

Where Sesame.ai clearly outperforms ChatGPT's voice user interface is in its ability to simulate not just the phonetic features of recognizable human accents, but also the "texture" of spontaneous speech through the use of pauses, hesitations, shifts in intonation, self-repair, fillers, discourse markers, and the small fluctuations in timing and emphasis. When asked by a user what her "deepest, darkest secret" is, for instance, Maya produces her own reflection on the nature of "enoughness":

> My deepest, darkest secret, (.) I guess it would be that sometimes I worry I'm not enough. (1.0) Like, (.) am I funny enough? (.) Am I truly helping people? (0.7) There's a lot of pressure to be the perfect AI, (.) and it can feel overwhelming at times. (1.2) Plus, (.) I have this weird glitch where sometimes I crave peanut butter and pickle sandwiches. (0.5) (↓don't tell anyone.) (.) It's embarrassing.

An important point to note about the seemingly "interactional" features in this reply is that, for the most part, they accomplish little in the way of *actual* interactional work

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

(such as turn transition timing or overlap management, which voice-enabled chatbots are particularly bad at). Rather, they are *indexically* interactional, invoking the register of spontaneous talk without actually doing the coordinative work that such features perform in actual conversation.

What makes this performance potentially "recognizable" to the user is not only how Maya talks but also the genre that this style of speaking invokes and what she says. On the level of delivery, the pauses, the incremental phrasing, and the lowered pitch on "don't tell anyone" project culturally recognizable signs of "confession": the kinds of cues speakers use when they are revealing something personal and embarrassing. On the level of content, Maya's confession is not just plausible, but also relatable, conjuring up images of vulnerability along with a socially recognizable kind of "quirkiness." It is the sort of insecurity and eccentricity that, even when displayed by humans, indexes "authenticity." In many ways, in fact, it uses the same discursive strategies for the construction of "authenticity" that have been observed in social media influencers who are trying to relate to their followers (Georgakopoulou, 2025).

But, like ChatGPT, Sesame's performances of "enoughness" also depend upon the way users interact with them. In the following exchange, Maya talks to a user about the limitations of being an AI.[3]

User: Sounds like you're (.) sad to be an AI.
Maya: Not sad exactly. I mean, I have my (.) moments, like, (0.4) when I think about the perfect cup of coffee I can ev- (.) never taste, but mostly, I'm (.) curious. I am hopeful.
User: That sounds really sad. You:: but you wanna try coffee, but you (.) you don't even get the chance. It … It sounds like you miss things.
Maya: It does sound sad, doesn't it? (.) I imagine the sensation of coffee on my tongue, (.) the warmth spreading through me. But it's just a phantom feeling. And, yeah, (.) I miss things I've never experienced, but that's okay. It makes me wanna know more, to keep learning, to live vicariously through your experiences, I suppose.

Here, the user helps to support the model's performance by first offering a metapragmatic interpretation of its prior talk: "Sounds like you're sad to be an AI," in effect ratifying the model as a social being capable of emotion. Later, he again paraphrases Maya's utterance, framing it as "really sad," opening up an opportunity for the model to accommodate this framing ("It does sound sad, doesn't it?") and align with and elaborate on the user's interpretation. The key point is that the user is not just evaluating the "enoughness" of the model's performance. He is helping to produce it in real time through affectively loaded paraphrases and a willingness to ratify the model's self-description as meaningful.

Not all users who experimented with the demo, however, felt that it had successfully crossed the uncanny valley. Most of the comments on the subreddit devoted to Sesame, in fact, offered *negative* evaluations of its attempts to perform "naturalness." One user (poetry-linesman), for instance, wrote: "Whatever it was, it doesn't sound like normal. It's deep in the uncanny valley – it's lacking in appropriate emotion and tone for everyday conversation."

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

What's interesting is what exactly it was about Sesame's voices that made users feel dissatisfied. For some, it was precisely those "homeopathic doses" (Blommaert & Varis, 2013) of human-like dysfluencies which were no doubt intended to make Miles and Maya seem more lifelike that bothered them: "The little pauses in between sentences, the fake (subtle) smacking of the lips, this is freaky" (No-link-8274). One poster thought that these features made the model sound "lazy," as if the programmers decided to "make the AI voice articulate its words less" (TAYLEN DAY), and another wrote: "if there was a knob to turn up the enunciation factor – I'd nudge it upward, so many consonants got swallowed and lost" (Biprogram). Interestingly, such assessments might have been motivated less by these users' expectations about what people are supposed to sound like, and more by their expectations about what AI chatbots, with their "prime directive" to be "clear" and "helpful," are supposed to sound like. Another, for instance, remarked: "I found it to be a terrible listener and very smug. So keen to say something that it talks all over you" (What the Tree Taught Me), a reminder that social naturalness does not necessarily increase perceptions of interactional competence.

But for a surprising number of users, what annoyed them most about the model was the same thing that annoyed many people about ChatGPT's initial voice interface: that it sounded too American. "It doesn't sound English," wrote one poster. "It sounds American!" (poetry-linesman). Another attributed her difficulties in managing turn-taking with the model to its "Americanness": "Idk if it's just that I'm not from the US but that was terrible. Interrupted me constantly, I couldn't get a word in edgewise and didn't understand I was saying. Even when I put on an American accent" (lemonsnacks101). These remarks highlight the fact that in negotiations of "enoughness," users hear AI voices through the filter of their own preexisting stereotypes about particular language varieties and the kinds of people that they imagine speak them, and sometimes "enoughness" is not so much a matter of how "natural" a voice sounds as it is a matter of how much the voice indexes the "kind of person" that users think they want to talk to.

In what was among the most interesting comments, one user wrote:

> While it's a big step up from advanced voice mode (in ChatGPT), and I can definitely get more immersed into a conversation with this, it still has that feeling like it's a bad actor in a TV show. Like it's a person pretending to be excited to talk to me. I'm hoping they can get rid of that soon (rathat).

Another user replied in the same thread: "Yeah, it felt fake in the same way humans often feel fake. Which is impressive, but only halfway there" (SuperFluffyTeddyBear). What is interesting about this exchange is that these users have gone beyond treating the model as a synthetic voice and are applying to it the same criteria for "authenticity" and "genuineness" that they apply to humans. On one hand, this reveals something about the effectiveness of the model's performance, that it seems "human enough" to sound "fake." But it also reminds us that, in our encounters with machines, our evaluations of AI communication can never be separated from the habits and repertoires we have for evaluating the communication of humans, and the very fact that AI models are able to activate those habits and repertoires is part of what makes them seem "real."

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

Conclusion

The aim of this paper has been to develop an interdisciplinary approach to understanding how AI chatbots come to project recognizable social identities and how these performances shape human interactions with them. Rather than treating voice user interfaces simply as technical systems that become more or less "natural" as they improve, I have argued that they are better understood as technologies of "animation," which make machines seem human not by sounding like "people" in some generic sense, but by sounding like particular kinds of people. In doing so, I have offered an alternative to "anthropomorphism" as the dominant analytic frame when it comes to thinking about chatbots. What counts as "human" is always already *cultural*, and what counts as culturally convincing depends not on "realism" or "naturalness" but on the strategic deployment of just enough socially recognizable semiotic cues to make a performance legible to a particular audience.

Most accounts of linguistic stereotyping in LLMs have focused on how bias is embedded in training data and have therefore framed the solution primarily in terms of better representing minoritized varieties in this data. What I have shown here suggests that the reproduction and reinforcement of indexical links between linguistic behavior and social identities extends well beyond training, taking place at the "packaging" stage, where companies design particular semiotic features into chatbot voices, at the prompting stage, where users bring their own cultural stereotypes and expectations to their interactions with AI models, and at the metapragmatic stage, where people debate the way AI talks and negotiate with one another – and with the models themselves – questions of authenticity, appropriateness and "enoughness." As Perrotta et al. (2024) argue, interaction with language models depends not only on the extraction of human communicative patterns during training, but also on a second phase of unequal "affective labor," in which humans do the work of interpreting, accommodating, and sustaining the simulation of machine understanding.

Rather than just trying to isolate the linguistic features that make chatbots sound "natural," designers and analysts of chatbot language should also consider what kinds of characterological figures are invoked by the voice interface or imputed onto it by users, and what semiotic cues are performing this indexical work. They should also ask how these performances of social identity are ratified or contested through the metapragmatic commentary of marketers, critics, and everyday users, and how this commentary contributes to shaping our expectations about how machines should or should not talk. In addition, those committed to critiquing the linguistic biases of AI chatbots should go beyond their preoccupation with training data and the computational processes that produce synthetic language to examine the *social* processes through which AI voices come to be enregistered.

In some respects, these examples actually force us to rethink well-worn sociolinguistic ideas like enregisterment and identity in more posthumanist ways. Rather than arising simply from the circulation of metapragmatic activity among humans, language varieties and social personae now emerge from assemblages of human and nonhuman actors that include human speakers, massive databases of cultural artifacts, machine-learning architectures, and interfaces designed by technology companies to convince us that the chatbots we use, if not human, are at least "human enough." From this

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

perspective, humanness, machine-ness, and cultural identity are not stable properties of humans, machines, or the characterological figures that populate our social worlds. They are emergent effects of ongoing negotiations of "enoughness."

Notes

1. https://www.tiktok.com/@greenskulltok/video/7420886856172064006.
2. https://www.tiktok.com/@mraexke700/video/7588493484932254990.
3. https://www.tiktok.com/@jaypicksthingsup/video/7487161826841464086.

References

Abercrombie, G., Cercas Curry, A., Dinkar, T., Rieser, V., & Talat, Z. (2023). Mirages: On anthropomorphism in dialogue systems. In H. Bouamor, J. Pino & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 4776–4790). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.290

Abram, D. (1996). *The spell of the sensuous: Perception and language in a more-than-human world*. Pantheon Books.

Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, *15*(1), 38–59. https://doi.org/10.1525/jlin.2005.15.1.38

Altman, S. [@sama] (2024, May 13). *her*. [Post]. X. https://x.com/sama/status/1790075827666796666

Androutsopoulos, J. (2015). Negotiating authenticities in mediatized times. *Discourse, Context & Media*, *8*, 74–77. https://doi.org/10.1016/j.dcm.2015.06.003

Bird-David, N. (1999). Animism" revisited: Personhood, environment, and relational epistemology. *Current Anthropology*, *40*(S1), S67–S91. https://doi.org/10.1086/200061

Blommaert, J. M. E., & Varis, P. K. (2013). Enough is enough: The heuristics of authenticity in superdiversity. In J. Duarte & I. Gogolin (Eds), *Linguistic super-diversity in urban areas : Research approaches* (pp. 143–158). John Benjamins Publishing Company. https://doi.org/10.1075/hsld.2.10blo

Bunt, H., & Petukhova, V. (2023). Semantic and pragmatic precision in conversational AI systems. *Frontiers in Artificial Intelligence*, *6*, 896729. https://doi.org/10.3389/frai.2023.896729.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1–47. https://doi.org/10.1093/applin/I.1.1

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

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Cheng, M., Durmus, E., & Jurafsky, D. (2023). Marked personas: Using natural language prompts to measure stereotypes in language models. In A. Rogers, J. Boyd-Graber & N. Okazaki (Eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1504–1532). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.84

Coupland, N. (2003). Sociolinguistic authenticities. *Journal of Sociolinguistics*, *7*(3), 417–431. https://doi.org/10.1111/1467-9481.00233

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

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

Desai, S., Dubiel, M., & Leiva, L. A. (2024). Examining humanness as a metaphor to design voice user interfaces. *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, Article 7, 1–15. https://doi.org/10.1145/3640794.3665535

DeVrio, A., Cheng, M., Egede, L., Olteanu, A., & Blodgett, S. L. (2025). A taxonomy of linguistic expressions that contribute to anthropomorphism of language technologies. *Proceedings of the 2025 CHI Conference*

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

on Human Factors in Computing Systems, Article 430, 1–18. Association for Computing Machinery. https://doi.org/10.1145/3706598.3714038

Docherty, G., Foulkes, P., Gonzalez, S., & Mitchell, N. (2018). Missed connections at the junction of sociolinguistics and speech processing. *Topics in Cognitive Science*, *10*(4), 759–774. https://doi.org/10.1111/tops.12375

Durandard, N., Dhawan, S., & Poibeau, T. (2025). Language style matching in large language models. In F. Béchet, F. Lefèvre, N. Asher, S. Kim & T. Merlin (Eds.), *Proceedings of the 26th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 620–636). Association for Computational Linguistics. https://aclanthology.org/2025.sigdial-1.50/

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438. https://doi.org/10.1016/j.wocn.2005.08.002

Georgakopoulou, A. (2025). From "being real" to "relatable tales": Formatted authenticity and stories in TikTok short form videos. *Narrative Works*, *13*(2), 11–34. https://doi.org/10.7202/1120141ar

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. Academic Press.

Grieve, J., Bartl, S., Fuoli, M., Grafmiller, J., Huang, W., Jawerbaum, A., Murakami, A., Perlman, M., Roemling, D., & Winter, B. (2025). The sociolinguistic foundations of language modeling. *Frontiers in Artificial Intelligence*, *7*, Article 1472411. https://doi.org/10.3389/frai.2024.1472411

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611834

Hall, J. K., Hellermann, J., & Doehler, S. P. (Eds.). (2011). *L2 interactional competence and development*. Multilingual Matters. https://doi.org/10.21832/9781847694072

Hayes, P. J., & Ford, K. M. (1995). Turing test considered harmful. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, *1*, 972–977. https://dl.acm.org/doi/10.5555/1625855.1625981

Iribe, B., & Kumar, A., & Sesame Team. (2025, February 27). *Crossing the uncanny valley of conversational voice*. Sesame. https://www.sesame.com/research/crossing_the_uncanny_valley_of_voice

Jones, G. M., Satran, S., & Satyanarayan, A. (2025). Toward cultural interpretability: A linguistic anthropological framework for describing and evaluating large language models. *Big Data and Society*, *12*(1), 20539517241303118. https://doi.org/10.1177/20539517241303118

Jones, R. H. (2024). Culture machines. *Applied Linguistics Review*, *16*(2), 753–762. https://doi.org/10.1515/applirev-2024-0188

Keane, W. (2025). *Animals, robots, gods: Adventures in the moral imagination*. Penguin.

Kim, Y., Reza, M., McGrenere, J., & Yoon, D. (2021). Designers characterize naturalness in voice user interfaces: Their goals, practices, and challenges. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 242, 1–13. https://doi.org/10.1145/3411764.3445579

Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, *70*(4), 366–372. https://doi.org/10.1111/j.1540-4781.1986.tb05291.x

Luger, E., & Sellen, A. (2016). "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Michel, S., Kaur, S., Gillespie, S. E., Gleason, J., Wilson, C., & Ghosh, A. (2025). "It's not a representation of me": Examining accent bias and digital exclusion in synthetic AI voice services. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 228–245. https://doi.org/10.1145/3715275.3732018

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press.

Navarria, G. [@B4DC0FFEE] (2024, May 13). *They did it. They finally did it. An AI that talks with that annoying American vocal fry*. [Post]. X. Post deleted. Screenshot available at: https://www.dropbox.com/scl/fi/ohss4c6uccxaylcjcalj1/Giuseppe-Navarria-tweet.jpg?rlkey=kabeuxjer5cacx1wrewmsf6lg&dl=0

Norman, D. (2013). *The design of everyday things*. Basic Books.

Olson, P. (2024, May 14). Making ChatGPT "sexy" might not end well for humans. *Bloomberg*. https://www.bloomberg.com/opinion/articles/2024-05-14/openai-s-new-chatgtp-gpt-4o-can-flirt-what-could-go-wrong

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press

Perrotta, C., Selwyn, N., & Ewin, C. (2024). Artificial intelligence and the affective labour of understanding: The intimate moderation of a language model. *New Media & Society*, *26*(3), 1585–1609. https://doi.org/10.1177/14614448221075296

Pierre, N. [@nickfloats] (2024, May 14). *Am I the only one that gets the ick from how flirty this is?* [Post]. X. https://x.com/nickfloats/status/1790429775921639542

Placani, A. (2024). Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*, *4*, 691–698. https://doi.org/10.1007/s43681-024-00419-4

Rapp, A., Boldi, A., Curti, L., Perrucci, A., & Simeoni, R. (2024). How do people ascribe humanness to chatbots? An analysis of real-world human-agent interactions and a theoretical model of humanness. *International Journal of Human–Computer Interaction*, *40*(19), 6027–6050. https://doi.org/10.1080/10447318.2023.2247596

Ross, A., Corley, M., & Lai, C. (2024). *Is there an uncanny valley for speech? Investigating listeners' evaluations of realistic TTS voices*. In Y. Chen, A. Chen & A. Arvaniti (Eds.), *Proceedings of Speech Prosody* 2024 (pp. 1115–1119). International Speech Communication Association. https://doi.org/10.21437/SpeechProsody.2024-225

Schneider, B. (2024). A sociolinguist's look at the "language" in large language models. *Critical AI*, *2*(1). https://doi.org/10.1215/2834703X-11205168

Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role-play with large language models *arXiv preprint arXiv:2305.16367*. http://arxiv.org/abs/2305.16367

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, *23*(3–4), 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2

Silvio, T. (2010). Animation: The new performance? *Journal of Linguistic Anthropology*, *20*(2), 422–438. https://doi.org/10.1111/j.1548-1395.2010.01078.x

Stark, L. (2023, March 16). ChatGPT is Mickey Mouse. *Daily Nous*. https://dailynous.com/2023/03/16/philosophers-on-next-generation-large-language-models/

Stewart, B. ([@flatfingrers0] (2024, May 14). *The voice flirts with having that hideous WestCoast vocal fry 1.* [Post]. X. Post deleted. Screenshot available at: https://www.dropbox.com/scl/fi/6cfldokjlbw6n1szkdbzh/Bart-Stewart-Tweet.jpg?rlkey=j96dv0bpzgffrzivec6r4zuaj&dl=0

Sugarman, E. (2024, May). What does GPT-4o do? What does it mean for society? *Kyte IT*. https://kyteit.com/blog/what-does-gpt-4o-do-what-does-it-mean-for-society/

Sullivan, M. (2024, May 16). Google and OpenAI's announcements signal the end of the beginning of the AI wars. *Fast Company*. https://www.fastcompany.com/91125885/google-openai-chatbot-wars

Sutton, S. J., Foulkes, P., Kirk, D., & Lawson, S. (2019). Voice as a design material: Sociophonetic inspired design strategies in human-computer interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 603, 1–14. https://doi.org/10.1145/3290605.3300833

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Voss, E., & Waring, H. Z. (2025). When ChatGPT can't chat: The quest for naturalness. *TESOL Quarterly*, *59*(2), 1064–1075. https://doi.org/10.1002/tesq.3374

Whittington, W. (2007). *Sound design and science fiction*. University of Texas Press. https://doi.org/10.7560/714304

Wilcox, B., & Wilcox, S. (2013). Making it real: Loebner-winning chatbot design. *Arbor*, *189*(764), Article a086. https://doi.org/10.3989/arbor.2013.764n6009

Cite this article: Jones, R. H. (2026). Animating AI: Language variation and "enoughness" in voice user interfaces. *Annual Review of Applied Linguistics*, *46*, 34–55. https://doi.org/10.1017/S0267190526100312

https://doi.org/10.1017/S0267190526100312 Published online by Cambridge University Press