

Algorithmic Ventriloquism: The Contested State of Voice in AI Speech Generators

Ido Ramati 

Social Media + Society
January-March 2024: 1–10
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20563051231224401
journals.sagepub.com/home/sms



Abstract

This article explores the vocal human–machine relations embedded in text-to-speech (TTS) generators. Retracing the human sources behind the synthetic speech and tracking the remediation of the voice by the machine-learning algorithm, it argues that artificial intelligence (AI) speaking agents such as Siri and Alexa, as well as other TTS acts such as TikTok’s, are performing *algorithmic ventriloquism*. Speaking mechanically with the voices of professional voiceover artists, AI speech technologies algorithmically manipulate these voices, thus generating personas that hold an interconnected chain of tensions between the embodied and the virtual, the particular and the general, the human and the non-human, as well as between speech and writing. Algorithmic ventriloquism serves as an analytical framework to tie the techno-vocalic operation of the TTS system with its cultural, economic, philosophical, and sociolinguistic predicaments. The last section discusses the implications of algorithmic ventriloquism beyond the realm of the voice.

Keywords

text-to-speech, speech generators, human–machine relation, media ventriloquism, TikTok

In February 2023, Spotify introduced DJ, an artificial intelligence (AI) feature which, besides curating playlists for users on the basis of their previous listening, addresses them in a human voice as their “own private DJ” (SpotifyNews, 2023). This AI DJ names the tracks, describes their musical style, or makes comments like “Let’s keep this vibe going,” all in a voice that sounds as natural and flowing as if it came from a human DJ. Indeed, the vocal source for this algorithmic DJ is Xavier “X” Jernigan, who is “the head of Spotify’s cultural partnerships” and hosts one of Spotify’s daily podcasts about pop culture (Demopoulos, 2023). This feature joins existing text-to-speech (TTS) applications that are powered by AI neural network algorithms and are also based on specific individuals’ prerecorded speech: the voices of Siri, Alexa, and Google Assistant have been available since the beginning of the previous decade, and now we have TikTok, Spotify, and other online free-to-use TTS platforms. All of these algorithmically manipulate the speech voices of professional voice actors, and contemporary developments promise to apply this technology to the voice of almost any person.

This article analyzes the vocal human–machine collaboration manifested by such synthetic speech technologies. Contemporary TTS applications differ in their purpose, availability, use, and operational techno-vocal model, but they also have in common the blurring of the assumed boundaries

between human and mechanical speech; they voice a non-human speech that sounds human and originates in particular humans. Tracking the mechanical remediation of particular individuals’ voices, I examine AI-speaking agents as techno-vocal human–machine compounds which perform *algorithmic ventriloquism*. Ventriloquism is the body technique for sounding voices as if they are coming from somewhere else. Speaking mechanically with the voices of human individuals, AI speech technologies algorithmically manipulate these voices, thus generating personas that hold an interconnected chain of tensions between the embodied and the virtual, the particular and the general, the human and the non-human, as well as between speech and writing. These personas are products of the meticulous design of vocal-linguistic features: their names, the pre-scripted sarcastic answers they provide to tricky questions, as well as the direct suggestion for users to talk to them “as you would to a person” (Natale, 2021,

The Hebrew University of Jerusalem, Israel

Corresponding Author:

Ido Ramati, Department of Communication and Journalism and The Program in Cultural Studies, The Hebrew University of Jerusalem, Jerusalem, Israel.

Email: ido.ramati@mail.huji.ac.il

X: @IdoRamati



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

p. 107). Exploring them as agents of algorithmic ventriloquism foregrounds the vocal properties which, when remediated by the algorithm, become fundamental to the design of these personas. Algorithmic ventriloquism explains the person–persona complex embedded in TTS technologies in terms of the continuity and mutuality of human–machine relations.

The idea that each voice is distinct and analogous to the person voicing it has nourished research on, and the development of, voice identification technologies since at least the end of the nineteenth century (Eidsheim, 2019; Kang, 2022). However, perceiving the voice as an extension of the person, as a marker of an “intimate kernel of subjectivity” and as a defining trait of humans, may be traced back to Aristotle (Dolar, 2006, p. 14). Mechanical speech which sounds human challenges these perceptions of the human voice because it provides this voice with an external source. It points to the duality of the voice as both internal and external, singular and reproducible, and complicates the link between the voice and the person.

Mechanical speech, whether based on prerecorded speech or on the synthesis of human-like but robotic-sounding voices, has served in growing capacities since the 1930s, and its conceptual roots stretch even further back (Furui, 2010; Li & Mills, 2019; Napolitano, 2020). Nowadays, machines speaking in human voices are ubiquitous: from children’s toys and ATMs to call centers and alarm systems giving messages on public transportation. In most cases, these are not yet powered by machine-learning (ML) algorithmic networks, as, for example, voice assistants are. However, tech companies are already providing infrastructure for general TTS applications, such as Amazon Polly, and technological developments constantly decrease the number of computational resources needed for general TTS generators (Défossez et al., 2022). Recently, AI voices have become popular also in artistic expression: examples are Netflix’s *The Andy Warhol Diaries*, a documentary narrated by AI algorithmic network that was trained on recordings of the late artist’s voice, the *Holly+* project by Holly Herndon that allows anyone to upload audio files and download them sung back in Herndon’s voice, or the various AI generated Beatles, Oasis, or Drake songs which deploy the voices of these artists to sing lyrics they never sang.

Like the voice assistants, these TTS realizations manipulate the prerecorded human voices of particular individuals. They detach these voices from their original bodies, casting them as “acousmatic” (Chion, 1999), and providing them instead with a surrogate body which in some cases may “contradict, compete with, replace or even reshape” the original body of the speaker (Connor, 2000, pp. 35–36; Kane, 2014). Composed of both Natural Language Processing (NLP) models and voice analysis and generation models (Kang, 2022), TTS systems re-emphasize the features of the voice as a sound medium carrying lingual content. Detaching the voice from its dependency on the human body, they multiply, transform, and transfer it from one surrogate vessel to another, but with the price of subordinating its polyphony to

the mathematics of algorithms. This is, however, a two-way binding process: the multivocality of the datafied human voice spoken by the AI agent is dependent on the algorithmic operation; simultaneously, this operation depends on and is colored by the initial human vocal data on which it feeds. Algorithmic ventriloquism, I suggest, specifies this disembodiment and re-embodiment of the human voice as human–algorithm reciprocity which inherently involves sociocultural power plays.

Given that media technologies are always intermingling humans and machines, and that AI “thinking machines” have an agency that “only emerges in interaction and relationship with humans and their cultures” (Natale & Guzman, 2022, p. 628), it is beneficial to ask: who are the human individuals giving their voices to TTS algorithms? What happens to their voices inside the algorithmic model? And what kind of social and cultural presumptions are embedded in these techno-vocal human–machine alignments? As an organizing principle, algorithmic ventriloquism ties the techno-vocal operation of the TTS system with its cultural, economic, philosophical, and sociolinguistic predicaments. It enables reclaiming the human within the machine, demystifying the ideas of autonomy and independency attributed to AI algorithms and at the same time avoiding the anthropocentrism that dismisses the technology as merely human-made automata. Instead, by reverse-engineering the algorithmic process and unveiling its human vocal sources, algorithmic ventriloquism complicates our understanding of the relation between humans, their voices, and the algorithms that ventriloquistically manipulate these voices to speak back to humans. In this, algorithmic ventriloquism joins the contemporary sociocultural critique of ML networks which unveil the presumptions and “ground-truths” embedded in them (Burrell, 2016; Kang, 2023; Mackenzie, 2017).

The following sections explore first the material and phenomenological links between ventriloquism, media, voices, persons, and personas. They serve as theoretical baselines for the subsequent examination of the particularities of algorithmic ventriloquism, and for contextualizing the human–machine relations AI-speaking agents perform. Then, based on secondary technological literature and published interviews with professional voice actors, I analyze the operation of TTS algorithms and the voice work of the particular individuals granting their voices to them. Algorithmic ventriloquism describes the embroilment of the human with the technological in producing a voice-based persona, and tackles the social, economic, and linguistic aspects of this human–machine continuum by further problematizing seemingly simple questions, such as: who speaks? Who has the power of and over a voice? And what are the implications of casting an individual human’s voice into a machine? Although emerging from voice-body relations, algorithmic ventriloquism as an analytical category has implications stretching beyond the realm of the voice and can be used to study how various algorithmic technologies ventriloquize human actions. Seeing algorithms as ontological and epistemological apparatuses,

and perceiving humans and algorithms as interwoven, I propose algorithmic ventriloquism as a perspective for analyzing this human-algorithmic enmeshment. The final section discusses the broader implications of exploring contemporary AI/ML technologies in terms of algorithmic ventriloquism.

Ventriloquism and Media

As a practice for channeling voices that appear to come from another place, ventriloquism may cause curiousness or discomfort, but also amusement. It is ancient and has been associated with inexplicableness and madness, as well as necromancy and witchcraft, because of the gap it opens between a voice and its seemingly absent source. Historically, it has been connected to femininity—examples are the Oracle of Delphi or the Biblical Witch of Endor—and related to other performances of channeling voices through the body such as the psychic medium (Baron et al., 2021). Modern ventriloquism is commonly known as a form of entertainment, during which the ventriloquist casts her or his voice onto a dummy, drawing the audience's attention to the puppet as an alternative source for the voice that originates in the human body. Fundamentally, a multivocal performance, ventriloquism destabilizes the apparent oneness of the person: it confuses the assumed link between bodies and voices, undermining its attributed consistency and cohesiveness, and replaces them with ambiguity, multiplicity, and playfulness. Machines that re-sound human voices further tangle this ventriloquistic voice-body mixture by adding more optional sources to its polyphony.

Media theory has applied ventriloquism to explore various voice-related phenomena, from the metaphorical voice in written texts (Cooren, 2010) to the relation between media technologies and the human voice (Altman, 1980; Drenten & Psarras, 2023; Goldblatt, 2006; Riszko, 2017; Truax, 2001). Media can recast human voices and generate new relations between humans, their bodies, and their voices; ventriloquism describes the mechanism of this reorganization as an operative ontology that redefines “the relations between selves and their bodies” (Connor, 2000, p. 43). Assuming that each medium performs a different kind of ventriloquism (Ramati & Abeliovich, 2022), what are the particularities of algorithmic ventriloquism, as performed by AI agents that speak in human voices?

Algorithmic ventriloquism amplifies the dissociation of individuals from their voices which is inherent to media ventriloquism. Voice assistants such as Siri and Alexa, as well as other TTS acts such as those of TikTok and Spotify, algorithmically reassign the voices of particular humans, usually professional voice actors whose voices are submitted to personate the AI agent: that is, to give it vocal features that sound human. In doing so, these agents situate a person–persona complex at the heart of the human–machine vocal relations they manifest. Algorithmic ventriloquism is therefore a constitutive mechanism acting through recurring detachments

and relocations: by channeling and manipulating the dissociated voice of a person, a persona is created, which holds a flexible and not necessarily direct connection to its vocal source. Algorithmic ventriloquism depends on the datafication of the prerecorded human voice which opens new, beyond-human opportunities for manipulating this vocal data. The AI vocal persona celebrates these algorithmic potentials: it may play with the most basic features of human speech and voice—such as musicality, timbre, pitch, or accent—and even repackage the speech of a person to sound like someone else. These playful algorithmic capabilities exacerbate a gap that is already inherent in person–persona vocal relations.

Voices and Person(a)s

The person–persona link has etymological and material roots that are voice-related (Ihde, 2007). The Latin word *persona* denoted a “human being” but also “a part in a drama, assumed character,” because originally the persona was “a mask, a false face” (Harper, n.d.). Actors wore this mask to externally express traits of a character and at the same time to conceal their own face. The persona was “related to” the Latin verb *personare*, “to sound through”: the mask was a stage tool spoken through by the actor “and perhaps amplifying the voice” (Harper, n.d.). The persona–person link was much about the voice as it was about the assumed “false face”: the persona as a theatrical character and the person who was the individual behind the mask materially shared the same voice. Voicing through the persona was a technique for voicing out a personality and impersonating, that is, becoming another person.

This constitutive connection between the voice and the person is not just historical or limited to the theatrical stage. Human voices, especially in the context of speech, serve as markers of the person. Accents, inclinations, tonality, timbre, pitch, and many other vocal characters signify a particular individual's vocal signature. For this reason, they have served in forensic and biometric identification, although the level of accuracy of these forms of identification has been questioned (Eidsheim, 2019; Kang, 2022). This is also because the human voice is anything but stable: it changes through life and may sound different according to context and situation. Despite this instability, in everyday situations, we constantly rely on an assumed link between voice and person to identify speakers, whether others or ourselves, often finding ourselves mistaken.

Steven Connor (2000) described the phenomenological basis for this presumed connection between the voice and personhood: “nothing else about me defines me so intimately as my voice” because “there is no other feature of my self whose nature it is . . . to move from me to the world, and to move me into the world”; so “if my voice is mine because it comes from me, it can only be known as mine because it also goes from me” (Connor, 2000, p. 7). Connor depicts the voice as a transitive event that originates in bodies, but

moves between them, simultaneously internal and external: it attests to an inner self as its origin which projects a vocal extension of itself into space. This movement of the voice structures not just the self but also its relations to the world; voicing out is a technique both for self-constitution of a person and for externalizing a persona as part of social relations (LaBelle, 2019).

These are interconnected aspects of the voice: our voice allocates us roles as both speakers and its first listeners; when it projects ourselves out to the world, it also returns this vocal self to us. In this sense we reconstitute and reaffirm our selves to ourselves each time we speak: our vocal persona constantly reconstructs our personality. An example may be the ways in which we adjust our speech to our surroundings in everyday situations as a strategy for keeping a public “face”—to use the famous persona-related Goffmanian term. In addition, this speaker-listener duality holds a tension between the familiar and the strange, which is amplified in several instances, the most evident one being listening to a recording of our own voice, which can be an alienating experience because this is not how we imagine others hear us and because the idea of our own voice coming at us from an external origin is uncanny (Truax, 2001).

Technologies that record and replay voices defy the ephemerality of the voice but also its assumed singularity (Cavarero, 2005; Sterne, 2003). They challenge its strong association with the person in several ways: copies of this recorded voice may be manipulated, edited, travel through space and time, and be replayed over and again. We tend to think of our voices as ours, as property that is part of our identity and is subjected to our own exclusive control, but this stance is consistently challenged by the characteristics of the voice and the work of sound recording media. As much as the voice points to an identity and a self or personality, it also undermines conceptions about this self as fixed, coherent, continuous, and uniformed. TTS technologies further highlight these tensions in their ventriloquistic act: speaking in a particular person’s voice they “steal” and clone this person’s vocal markers of personhood; they undermine one’s assumed power, control, and exclusive ownership over one’s voice when they appropriate, manipulate, and revoice this voice from another source, external to this person. They synthesize a vocal mask which, by generating an algorithmic persona such as Siri, undermines the position of the person who initially gave it its voice.

Synthetic Speech, Mechanical Personas

The person–persona complex is well-researched in many fields, from theater and performance studies to psychology, anthropology, or the study of stardom. Similarly, the personification of non-humans is omnipresent and deeply rooted in culture: from personifying everyday items such as toys or cars to animating objects in ancient myths or contemporary movies (Humphry & Chesher, 2021). Popular culture has countless

representations of personified speaking computers: famous examples include the calm but ultimately murderous HAL 9000 of *2001: Space Odyssey*, the knowledgeable starship computer in *Star Trek*, or the sympathetic assistant that becomes a lover in *Her* (Faber, 2020). Historically, whenever actual speaking machines were presented to the public, they were also personified with human traits: from Joseph Faber’s 1845 female *Euphonia*, through Homer Dudley’s 1939 Voder, which was able to “do practically anything that can be done with the human voice” (MonoThyatron, 2011) to Apple’s 1984 Macintosh, which when asked by Steve Jobs to “speak for itself,” uttered “Hallo, I’m Macintosh. It sure is great to get out of this bag” (The Unofficial AppleKeynotes Channel, 2012).

In all these cases, the human presenters played with the machines’ vocal features to exhibit their attributed personalities: laughing, singing, imitating pet sounds, telling jokes, or answering silly questions with witty answers were vocal strategies of personification long before the Siri introductory event in 2011. In this event, software engineer Scott Forstall asked Siri “who are you?”; she replied “I am a humble personal assistant,” and he concluded with the same personating manner: “Siri is your humble, intelligent personal assistant that goes everywhere with you and can do things for you just by you asking” (The Unofficial AppleKeynotes Channel, 2013). Siri’s initial persona—“intelligent” but “humble”—was predesigned into the system and so are the personas of Alexa, Google Assistant, and other TTS agents such as TikTok’s Jessi or the aforementioned Spotify’s DJ.

Research on voice assistants pointed at several themes embedded in these personifications and critically analyzed the tendency to give them a feminine voice as default. Within this context, researchers described the history of the relation between the female secretary and technologies (Lingel & Crawford, 2020; Phan, 2017), the feminization of technology as a domestication strategy (Woods, 2018), or the surrendering of privacy to machines that deceive with personalized allure (Natale, 2021). The discussion suggested here targets the vocal materiality of the AI persona by following the remediation of the voices of specific persons through the machine. These AI personas—with their default feminine voices and deceiving surveillance strategies—rely on human voice work that may be understood in terms of digital labor (Fuchs & Seignani, 2013) and on algorithmic manipulation of extralinguistic individuating facets of the voice, such as accent or timbre, which are inherently culturally marked (Rangan et al., 2023). Eidsheim’s (2019) analysis of *Vocaloid*, a software which manipulates prerecorded singing voices, shows how listeners imagined and reflected ethnic presuppositions regarding the voice produced by the machine. Similarly, the AI personas of contemporary TTS acts are loaded with cultural, linguistic, economic, and gendered presumptions which are epitomized in their ventriloquistic performance. Algorithmic ventriloquism points to the continuous transitivity and circulation of these loads between humans and non-humans.

Technologies of synthetic speech have relied on human voices from their beginnings (Li & Mills, 2019). Archives of recorded speech served in studying vocalization and its physiology, and among other implications led to the design of machines for imitating human voices. Specifically, the study of phonetics has been a main contributor to the evolution of speech synthesizers, and vice versa (Mills, 2010). Human contribution has been and still is imperative to mechanical speech. From an operational perspective, this contribution in contemporary models may be divided between systems that voice prerecorded speech segments and systems that intervene in the deeper level of the phonemes.

The more basic and earlier techno-vocal algorithmic model sequences prerecorded meaning-bearing segments of speech, rearranging snippets of entire words or sentences. This is enough for generating messages in technologies that rely on the relatively low variability of what the machine needs to announce. For example, GPS navigators have a limited number of directionality-related utterances like “in + one + hundred + meters,” “turn + left,” and the like. Almost every word in these strings may be replaced by another prerecorded word, such as “right” instead of “left.” One weakness of this model is evident when these systems need to voice sentences with a higher level of phonetic variability, like street names. Another is the unwieldy musicality of the synthetic speech, derived from the connection of whole words without the ability to control their timbre, stress, tonality, or pitch.

Unlike these systems, AI voice applications do much better in pronouncing more complicated sentences with higher levels of phonetic diversity. Voice assistants like Siri or Google’s answer questions and provide information about almost anything. For this reason, most of the time they need to be able to vocalize and string almost any combination of phonemes. The vocal units that they manipulate are much smaller and the sound data that they rely on need to be richer with nuances. These systems decompose strings of phonemes of the recorded texts and recompose them into a new text, in a sense, just like humans do. Their speech sounds much more flowing and its musicality is more natural because they choose the best available utterance to the context of the entire word and sentence. To this end, the original recording must contain vocal data for various possible everyday combinations of phonemes so the synthesized utterances sound alive and human-like. In the last decade, these ML models have become increasingly sophisticated, and consequently, the sound of synthetic speech has become more natural.

There are several textual passages specifically fabricated to contain all phoneme combinations in the English language. The three most frequent are “The Caterpillar,” the “Rainbow Passage,” and the “Grandfather passage,” used regularly in speech therapy to test articulation capabilities, features of oral reading, and “speech motor functioning”

(Lammert et al., 2020; Reilly & Fisher, 2012, p. 84). Versions in other languages also exist (Bergerzon-Bitton & Ben-David, 2022). In these texts, the uttered words become functional lingual medium chosen not for their denotational sense but for their phonetic value, highlighting the vocal qualities of the person who reads them. In the voiceover industry, these passages are sometimes used to audit voice artists for a specific narration job; ML models may use recordings of people reading such passages to train the algorithm to disjoint and then reconnect phonemes into words and sentences. Susan Bennett, the voice actress behind the original Siri, gave an example for the type of sentences she was asked to record (not from the aforementioned passages): “Malitia oi hallucinate, buckry ockra ooze” and the like (Broussard, 2017). The phonetic units in these recordings were then concatenated into the various words, sentences, and paragraphs used in the Siri voice.

This algorithmic work of the ML models means that the “algorithmization” of the speech units—that is, devising them for the work of the algorithm—starts before the actual algorithm sets to work. The speech segments carried by voices of particular individuals become meaningful only in retrospect when they are remediated by the algorithm. In addition, this means that, although in their output, TTS systems are sound machines, in their inner automatic NLP operations, they are actually writing machines, subjecting speech sounds to written texts either in their capacity as speech recognition technologies (converting speech to text) or in their speech generation aspect (converting text to speech). Even current technological innovations aiming at speech-to-speech conversion translate the soundwaves into graphic representation and back (Lakhotia et al., 2021). In the datafication process of the human voice, it must become “machine readable” and statistical, and be converted into a graphic representation that the algorithm can read, pattern, manipulate, and pronounce. The vocal mask of AI speech generators depends therefore on written signifiers for its ventriloquistic operation. Algorithmic ventriloquism is a reading and writing process as much as it is listening and speaking operation. As such, its implications go beyond the realm of the voice and may describe other contemporary algorithmic operations, discussed in the final section of this article.

Humans of the AI Voices

Whose prerecorded voices do the AI agents ventriloquize? Some of the individuals who have given their voices, sometimes unknowingly, to an AI agent include professional voiceover artists such as Susan Bennett (Siri), Nina Rolle (Alexa), Kiki Baessell (Google), or Beverly Standing (TikTok). Bennett, whose voice was used for Siri from its initiation until the iOS 7 update of 2013, described how in July 2005 she recorded the aforementioned segments of speech for the voice database of the software company ScanSoft. ScanSoft eventually merged with Nuance, which

provided voices and speech recognition services to other companies as well (Parkinson, 2015).

According to interviews Bennett gave, she became aware that her voice was used for Siri only in 2011. Apple has never officially acknowledged nor confirmed the use of her voice. The same is true to John Briggs, who gave his voice to British accent Siri, and Karen Jacobsen, the voice actress behind the Australian accent Siri. Similarly, Amazon has neither confirmed nor denied that Nina Rolle was the voice of Alexa. In some of these cases, forensic comparisons between the human voices and the voice assistants indicated a match (AI-Heeti, 2021).

Algorithmic ventriloquism may explain the companies' lack of recognition of the human vocal contribution to the algorithm and draw attention to the politics of the AI voice industry. From the perspective of the companies, a faceless voice frees the company's brand from any association with a particular human; such ambiguity makes Siri, Alexa, or Google Assistant appear as if they have no human faces behind their vocal masks. This serves the companies in giving the AI agent an independent persona, not associated with any particular human individual. According to an interview with Briggs, who identified publicly as British Siri, Apple "wasn't pleased" with his "newfound fame" and "asked him not to talk publicly about Siri, saying the company isn't 'about one person'" (Colson, 2011). The idea of giving a persona to the voice assistant depends on the anonymity of the individual behind the vocal mask, so for keeping up the persona, the person behind the vocal mask must recede and stay in the shadows. Like in a live ventriloquist show, in which the audience is urged to sustain its disbelief and succumb to the illusion that the voice comes from the puppet rather than the puppeteer, so users of AI voice agents are presented with the agent's persona rather than the human who gave it its voice. The absence of the human body from which this voice originates and the presence of a surrogate container such as Amazon Echo or a smartphone support this algorithmic ventriloquism: the human must not be seen or located in order for the voice assistant persona to have its own voice.

It is exactly because the voice usually serves to personate individuals that these companies aim as much as possible to keep anonymous the human behind the assistant's persona: this persona gives voice to the company, both metaphorically and materially, so it better not be recognized as the person who professionally gives his or her voice to other companies. The attributed distinctiveness of the human voice becomes an obstacle to the individuation of the voice assistant, and by extension to the singularity of the tech company, and therefore must be repressed. The elasticity of the algorithmic voice makes it a perfect mask: once datafied, the voice may be manipulated in ways that make the traces of the original human voice behind it unhearable to the naked ear. This algorithmic intervention also supports the companies' claim for ownership of the voice, rather than that of the voice actor: if it no longer sounds like that human individual, then it can

be argued that the voice belongs to the company that regulates the algorithm. Algorithmic ventriloquism shifts the power of the voice and the power over the voice to the tech companies while silencing the humans whose original voices it channels.

This has ethical, legal, and financial implications: algorithmic voices may be reused and manipulated recurrently, without reimbursing their human sources. Indeed, several of the voice actors felt in retrospect that their payment did not reflect the recurring use of their voice, its traveling through applications, or the revenue that the tech giants made using their voices. For example, Beverly Standing, who did not know that TikTok used her voice for its algorithm, claimed damages in her lawsuit for "the emotional distress of having her likeness exploited without consent; loss of the ability to control the dissemination of her likeness; and loss of the ability to control the association of her likeness" (Smith, 2021). These arguments go beyond the financial aspects, highlighting the destabilizing impacts of algorithmic ventriloquism: when one algorithm uses a human voice, this particular individual cannot know where and to what ends his or her voice might eventually serve. Algorithmic ventriloquism is precarious both for the voice artist and for the company that uses this voice.

The play of power exhibited by the algorithmic container which hosts the human voice goes even further. The algorithmic persona sometimes informs the search for a human voice that best serves the imagined characteristics of that persona. In the case of Google Assistant, the initial voice actress, Kiki Baessell, was chosen to match a predesigned backstory described by James Giangola and then a "lead conversation and persona designer" at Google: "the Assistant comes from Colorado, which gives her a neutral accent. She comes from a well-read family and is the youngest daughter of a physics professor (who has a B.A. in art history from Northwestern University) . . . and a research librarian. She once worked for 'a very popular late-night-TV satirical pundit' as a personal assistant. She was always a smart kid, she won \$100,000 on the Kids Edition of 'Jeopardy.' Oh, and she also likes kayaking" (White, 2022). Whether Baessell liked kayaking or ever participated in Jeopardy was not relevant; the idea was to find someone whose voice sounded as if she did. Going into details such as the occupations of the voice assistant's parents or her traits as being "well read," "smart," having a "natural accent," and having an occupational history as an assistant, are all supposed to serve, eventually, the characteristics of the persona as "skillful," "professional," but also "energetic," like she's "up for kayaking" (White, 2022). The voice of the human actress was supposed to give the feeling that it echoed these human traits.

Following the live ventriloquist show script—in which for the purpose of comic reversal, the dummy takes over the performance and appears to be controlling the human puppeteer, and especially the voice and what is said on stage—so the logic of searching for a human voice actress who best fits

the predetermined characteristics of the AI persona exhibits a person–persona reversal. Algorithmic ventriloquism plays a recursive human–machine loop: the AI persona is described in human categories and molded according to detailed human clichés, so the speech actress may be vocally appropriated to fit these characteristics. The human voice, which is commonly understood to be unique and singular, becomes a type-cast, a script serving companies’ perception of how humans should appropriately sound and what they should voice. If “giving voice” is normally a marker of empowerment, this twist enhances corporate power in amplifying and enforcing normativity through the voice.

The accents in AI voices emblemize these relations between humans, algorithms, and the companies they speak for. If the human voice in general is perceived as an indicator of a particular individual, the accent is a vocal paralinguistic element associating one with a distinct social, ethnic, or geographic group. Google’s choice to search for a voice actress with a “natural accent” entailed looking for an accent that did not stick out and could not be pinned down. However, as voice assistants became more ubiquitous, more companies decided to localize their agents by recording different accents of English—that is, British, Australian, and so on—and later also in some other languages like French or Spanish. Currently, in most AI voice applications, users can choose between various languages and accents. This means that the companies, at least for the initial assistant’s voice, had to choose between dialects and regional vernaculars and usually decided to go with the one considered the most unmarked so it sounds more “natural.” For example, in the case of the British Siri, the accent may be described as a mix of “generic southern-English variety,” with lengthened /a/ sounds in words such as “‘ask’ and ‘answer’” (McCabe, 2013). Similarly, the initial voice actress behind the German Siri was Heike Hagen, speaking Hochdeutsch pronunciation (Stein, 2013). Google’s choice to describe their voice assistant as someone who was born in Colorado points to the wish for a non-specific American accent, or what is known as Standard American English. The issue of accent shows again that the synthetic voice holds a tension between the particular and what is considered general. Amplifying the understanding that accent is not an essential vocal trait but rather processual, context- and listener-dependent (Eidsheim, 2019)—and therefore may be subjected to algorithmic manipulation—the voice assistant persona is predesigned to sound both like someone *and* no-one-in-particular. In the case of Google’s voice assistant persona, to achieve a personal-impersonal voice, the Colorado (non)accent was picked as the zero-level of pronunciation. The element of accent shows how paralinguistic vocal qualities are key to understanding person–persona, human–AI agent relations. By voicing a particular voice and accent, they further entrench presumed sociocultural rankings for what is “normal” and “natural,” and what passes as unmarked.

Multivocality of AI Voices

However, over the last decade, TTS models have grown in sophistication, and the accent of the AI voice is no longer necessarily permanent. In addition to duplicating the voice of a particular person, they are able to transform it to sound like someone else’s voice. These changes may meddle with tonality, pitch, timbre, and other aspects of speech voices, as well as amalgamate the voices of different people into a new voice, or “coat” a speech voice to sound different—for example, to age a voice or make it speak in a different accent (Trueba & Klimkov, 2019). TTS algorithms are polyphonic, orchestrating a choral of human and machine-manipulated voices, actualizing them from a singular–plural repository: they contain multiple vocal potentials, but these are dependent on an initial prerecorded human voice. One human voice may serve different vocal personas; each AI persona is always already vocally abundant. In performing algorithmic ventriloquism, AI agents are inherently multivocal.

This quodlibet of voices predates the algorithmic operation and is rooted in the variability of the human voices initiating it. Through the last decade, the humans of the AI voices have been replaced several times. For example, Kiki Baessell’s voice for Google was replaced in 2016 with that of Antonia Flynn; similarly, the current Siri voice sources are not the ones who empowered it in previous years. From a technological perspective, AI voice algorithms are vessels or positions that may be filled with voices of changing voice actors. This perspective once again puts the power over the voice into the hands of the tech companies: repeatedly replacing the human voice artists supports the inclination of not identifying AI personas with particular individuals; it also points at the dispensability of the human vocal labor that is invested in initiating the AI voice. Humans are disposable providers of vocal raw data that feeds the algorithm. The AI voice needs a human source, but this source can change through time.

This changeability has several implications which reveal the complexity of AI person–persona relations. A few weeks after Standing’s lawsuit against TikTok was settled, the app presented a new algorithmic voice persona named “Jessi.” Several months later, Canadian voice actress and radio DJ Kat Callaghan revealed herself as the “TikTok TTS girl,” the person behind Jessi. Callaghan’s TikTok account celebrates her person–persona relation with Jessi: her videos creatively play with the differences between her voice and the Jessi voice, which is pitched higher to sound bouncier, or between her human linguistic capabilities and Jessi’s failures (like properly pronouncing the name Beyoncé). That TikTok’s moderators have not blocked her content might point toward a change of attitude on the part of the tech companies: instead of hiding the humans behind the algorithmic personas, they publicly foster these person–persona relations as playable material for creating more content and app-traffic. The popularity of Callaghan’s posts is beneficial both for her and for TikTok. Similarly,

Spotify revealed Xavier Jernigan's identity as part of the launch of the DJ feature, and the company's Twitter account highlighted that he is "The Voice of @Spotify." His noticeable urban accent contributes a "cool" vibe to the AI DJ persona, contrasting with the initial unmarked accents of Siri or Google Assistant. This also means that these human individuals have become assimilated into the brand of the company, exhibited as part of its product.

In some cases, voice professionals do not wish to be identified with a particular brand, because they do not want to limit their ability to get other voice acting work. Bennett, for example, explained, "she was initially hesitant to reveal herself as the voice of Siri because she was worried she'd be 'typecast and stereotyped, and that's something you don't want to be as a voice actor'" (Al-Heeti, 2021). Later she changed her mind and started celebrating being the first Siri voice. Similarly, Karen Jacobsen, the voice of Australian Siri, promotes herself professionally as the trademarked "The GPS Girl[®]" and her website announces that her voice is "heard in over a billion GPS and smartphone devices" (Jacobson, 2023). All these cases point at bidirectional relations between the AI persona and the human behind it, and that there are several ways that human actors can benefit from their digital vocal labor. That the AI persona is a position filled with changing voice actors may characterize it as a role, bringing to mind the theatrical qualities of voice acting. Voice professionals are essentially actors, moving from one role to another.

The "Alexa loses her voice" Amazon 2018 Super Bowl commercial played with this innate singular-plural multivocality and theatricality of algorithmic ventriloquism (TheAdsWorld, 2018). In the commercial, when Alexa suddenly loses her voice, Gordon Ramsay, Rebel Wilson, Cardi B, and Sir Anthony Hopkins step in, filling the AI position with their voices, consequently importing particular personas (the angry chef, the sassy rapper, the eerie psycho-killer, etc.). Eventually, Alexa's familiar voice returns, against a recording of Marvin Gaye and Tammi Terrell's duet "Ain't nothing like the real thing." This self-parodying commercial launched Alexa's feature of replacing the default voice with voices of celebrities such as Samuel L. Jackson, Shaquille O'Neal, and Melissa McCarthy. This option cost money—expanding the commodification of the voice to the users' end—and was limited in what the celebrities' voices could respond to; after 3 years, Amazon discontinued it (as Google did with their parallel service; Forristal, 2023). The temporary association of the AI persona with distinct voices of famous persons, and the ability to switch between voices and personas, eventually strengthened the default voice as "the real thing" which elastically shifts between vocal masks. The return of the default familiar voice reassures the status of this voice and persona as the original Alexa which may be masked with other personas. The AI agent, who is no-one-in-particular, becomes someone when it uses the voice(s) of anyone, whether or not they are famous.

This changeability suggests a possible future for the voice. Recent technological innovations promise to clone almost any voice with little recorded data. Apple's recent announcement of *personal voice*, a TTS feature that will allow users to type and voice texts in their own voice, followed other platforms that promise to "create a digital copy of your voice" (Apple, 2023; my-own-voice, 2023). This marks the diffusion of algorithmic ventriloquism and consequently the person-persona complex from the professional realm to that of everyone's life. This promises users who have lost their voice or experience speech impairments to be able to sound fluently in their own voice. However, once datafied, the duplicated voice may be used for a variety of purposes, from a parent's avatar reading a bedtime story to impersonating the person whose voice it cloned (Liszewski, 2022). These perils and potentials are intrinsic to the extension of any human voice beyond human vistas and into algorithmic terrains. Algorithmic ventriloquism becomes ubiquitous and any voice may continue to be remediated, migrate from one holder to another, transform, and eventually get a life of its own that depends only on the available technology. Algorithmic ventriloquism expands the understanding of the voice as an event that continues to roll, from one medium to another, from humans to non-humans, and back. This ventriloquistic process forefronts reciprocity and continuity as key principles of human-algorithm relations, which also obtains outside the realm of voices.

Coda: Algorithmic Ventriloquism Beyond Voices

In what ways can algorithmic ventriloquism be used to describe the operative mechanism of other AI systems? The big ML models, such as those empowering ChatGPT or DALL-E, similarly manipulate human-made resources—textual or graphic—to generate new, seemingly unconnected outputs. Like TTS systems, they hide the human investment that they depend on, thus engaging in the ventriloquist-dummy power play. Understanding them in terms of ventriloquism suggests that they are not simply "stochastic parrots" which imitate human activity, working as "systems for haphazardly stitching together sequences of linguistic forms . . . according to probabilistic information . . . but without any reference to meaning" (Bender et al., 2021, p. 617). Rather, it advocates analyzing them as continuing and depending on human processes, but also as operating in ways different from humans, thus muddling domains commonly perceived as exclusively and independently human. Ventriloquistic algorithms change what it means to be human because they extend human presence and capabilities beyond human territories.

As NLP engines, they work with linguistic units previously unavailable to humans: they analyze textual data in capacities greater than human comprehension or decompose and recompose micro-units of information which escape unaided human cognition. Since their operations deeply penetrate the fabric of

lingual creation (e.g., texts) and embodied experiences (e.g., speech), they may reveal structures and patterns of human behavior never previously noticed. These powerful ontological and epistemological features are also political tools, serving the power structures from which they emerge. Currently trusted in the hands of big-tech companies and subjected to their set of motivations, values, presumptions, and biases, they direct humanity toward a future shaped by these denominations. Algorithmic ventriloquism unveils the commitment of algorithms to the impetuses of their owners and unmasks the process of delegating the most basic human resources—such as the voice but also creativity and imagination—to the authority of companies motivated by profit. As such, algorithmic ventriloquism provides media research with an analytical perspective for examining human-algorithmic collaborations as always already committed to power relations and continuing intersections. Ventriloquizing algorithms that openly credit their human sources—such as Spotify’s use of Jernigan’s voice—set example for starting balancing these relations and acknowledging the human labor invested in the algorithmic operation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ido Ramati  <https://orcid.org/0000-0002-5364-2969>

References

- Al-Heiti, A. (2021, October 1). The original voice of Siri is now advocating for a more accessible web. *CNET*. <https://www.cnet.com/tech/mobile/the-original-voice-of-siri-is-now-advocating-for-a-more-accessible-web/>
- Altman, R. (1980). Moving lips: Cinema as ventriloquism. *Yale French Studies*, 60, 67–79.
- Apple. (2023, May 16). *Apple introduces new features for cognitive accessibility, along with live speech, personal voice, and point and speak in magnifier* [Press release]. <https://www.apple.com/newsroom/2023/05/apple-previews-live-speech-personal-voice-and-more-new-accessibility-features/>
- Baron, J., Fleeger, J., & Wong Lerner, S. (2021). *Media ventriloquism: How audiovisual technologies transform the voice-body relationship*. Oxford University Press.
- Bender, E. M., Gebu, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT'21)* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Bergerzon-Bitton, O., & Ben-David, B. M. (2022). Validation and cross-linguistic adaptation of the Frenchay Dysarthria Assessment (FDA-2) speech intelligibility tests: Hebrew version. *International Journal of Language & Communication Disorders*, 57, 1023–1049. <https://doi.org/10.1111/1460-6984.12737>
- Broussard, M. (2017, April 17). Siri voice actress Susan Bennett reveals more details about the origins of Apple’s virtual assistant. *Macrumors*. <https://www.macrumors.com/2017/04/14/siri-susan-bennett-details/>
- Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512. <https://doi.org/10.1177/2053951715622512>
- Cavarero, A. (2005). *For more than one voice: Toward a philosophy of vocal expression*. Stanford University Press.
- Chion, M. (1999). *The voice in cinema*. Columbia University Press.
- Colson, J. (2011, November 10). Voice behind UK Siri voice “Daniel” speaks out. *MacRumors*. <https://www.macrumors.com/2011/11/10/voice-behind-uk-siri-voice-daniel-speaks-out/>
- Connor, S. (2000). *Dumbstruck: A cultural history of ventriloquism*. Oxford University Press.
- Cooren, F. (2010). *Action and agency in dialogue: Passion, incarnation and ventriloquism*. John Benjamins Publishing Company.
- Défossez, A., Copet, J., Synnaeve, G., & Adi, Y. (2022). *High fidelity neural audio compression*. <https://doi.org/10.48550/arXiv.2210.13438>
- Demopoulos, A. (2023, March 2). Last night AI DJ saved my life? Testing Spotify’s virtual radio host. *The Guardian*. <https://www.theguardian.com/technology/2023/mar/01/spotify-ai-dj-service-testing#:~:text=While%20the%20honey%2Dvoiced%20guide,listeners%20a%20%E2%80%9Cpersonalized%20playlist%E2%80%9D>
- Dolar, M. (2006). *A voice and nothing more*. MIT Press.
- Drenten, J., & Psarras, E. (2023). Digital ventriloquism and celebrity access: Cameo and the emergence of paid puppeteering on digital platforms. *New Media & Society*, 25(12), 3350–3369. <https://doi.org/10.1177/14614448211041175>
- Eidsheim, N. S. (2019). *The race of sound: Listening, timbre, and vocalicity in African American music*. Duke University Press. <https://doi.org/10.2307/j.ctv11hpntq>
- Faber, L. W. (2020). *The computer’s voice: From Star Trek to Siri*. University of Minnesota Press.
- Forristal, L. (2023, May 30). Sorry, but Alexa can no longer talk like Samuel L. Jackson. *Techcrunch*. <https://tcrn.ch/45DmPzx>
- Fuchs, C., & Seignani, S. (2013). What is digital labour? What is digital work? What’s their difference? And why do these questions matter for understanding social media? *TripleC*, 11(2), 237–293.
- Furui, S. (2010). History and development of speech recognition. In F. Chen & K. Jokinen (Eds.), *Speech technology* (pp. 1–18). Springer. https://doi.org/10.1007/978-0-387-73819-2_1
- Goldblatt, D. (2006). *Art and ventriloquism*. Routledge.
- Harper, D. (n.d.). *Etymology of person*. Online Etymology Dictionary. <https://www.etymonline.com/word/person>
- Humphry, J., & Chesher, C. (2021). Preparing for smart voice assistants: Cultural histories and media innovations. *New Media & Society*, 23(7), 1971–1988.
- Ihde, D. (2007). *Listening and voice: Phenomenologies of sound* (2nd ed.). SUNY Press.
- Jacobson, K. (2023). Voice-over artist. *The GPS Girl*. <https://thegpsgirl.com/voice-over-artist/>

- Kane, B. (2014). *Sound unseen: Acousmatic sound in theory and practice*. Oxford University Press.
- Kang, E. B. (2022). Biometric imaginaries: Formatting voice, body, identity to data. *Social Studies of Science*, 54(4), 581–602.
- Kang, E. B. (2023). Ground truth tracings (GTT): On the epistemic limits of machine learning. *Big Data & Society*, 10(1), 20539517221146122. <https://doi.org/10.1177/20539517221146122>
- LaBelle, B. (2019). *Acoustic territories: Sound culture and everyday life* (2nd ed.). Bloomsbury.
- Lakhotia, K., Kharitonov, E., Wei-Ning Hsu, Adi, Y., Polyak, A., Bolte, B., Nguyen, T.-A., Copet, J., Baevski, J., Mohamed, A., & Dupoux, E. (2021). On generative spoken language modeling from raw audio. *Transactions of the Association for Computational Linguistics*, 9, 1336–1354.
- Lammert, A., Melot, J., Sturim, D., Hannon, D., Delaura, R., Williamson, J., Ciccirelli, G., & Quatieri, T. (2020). Analysis of phonetic balance in Standard English passages. *Journal of Speech, Language, and Hearing Research*, 63, 1–14.
- Li, X., & Mills, M. (2019). Vocal features: From voice identification to speech recognition by machine. *Technology and Culture*, 60(2), S129–S160. <https://doi.org/10.1353/tech.2019.0066>
- Lingel, J., & Crawford, K. (2020). “Alexa, tell me about your mother”: The history of the secretary and the end of secrecy. *Catalyst: Feminism, Theory, Technoscience*, 6(1), 1–25.
- Liszewski, A. (2022, May 31). This smart speaker will read stories to your kids with a deepfaked copy of your own voice. *Gizmodo*. <https://gizmodo.com/smart-speaker-takara-tomy-ai-deepfake-voice-parents-kid-1848997168>
- Mackenzie, A. (2017). *Machine learners: Archaeology of a data practice*. MIT Press.
- Mccabe, S. (2013, February 8). Investigating Siri’s English accents. *TidBits*. <https://tidbits.com/2013/02/08/investigating-siris-english-accents/#:~:text=Siri's%20British%20accent%2C%20incidentally%2C%20would,%E2%80%9D%20and%20%E2%80%9Canswer%E2%80%9D%20lengthened>
- Mills, M. (2010). Deaf jam: From inscription to reproduction to information. *Social Text*, 102(28–1), 35–58.
- MonoThyatron. (2011, July 10). *The Voder—Homer Dudley (Bell Labs) 1939* [Video]. YouTube. https://www.youtube.com/watch?v=5hyI_dM5cGo&t=181s&ab_channel=MonoThyatron
- my-own-voice (2023) <https://mov.acapela-group.com/>
- Napolitano, D. (2020). The cultural origins of voice cloning. *Computation, Communication, Aesthetics & X*. <https://2020.xcoax.org/pdf/xCoAx2020-Napolitano.pdf>
- Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press.
- Natale, S., & Guzman, A. L. (2022). Reclaiming the human in machine cultures: Introduction. *Media, Culture & Society*, 44(4), 627–637. <https://doi.org/10.1177/01634437221099614>
- Parkinson, H. J. (2015, 12 August). Hey, Siri! Meet the real people behind Apple’s voice-activated assistant. *The Guardian*. <https://www.theguardian.com/technology/2015/aug/12/siri-real-voices-apple-ios-assistant-jon-briggs-susan-bennett-karen-jacobsen#comments>
- Phan, T. (2017). The materiality of the digital and the gendered voice of Siri. *Transformations*, 29, 23–33.
- Ramati, I., & Abeliovich, R. (2022). Use this sound: Networked ventriloquism on Yiddish TikTok. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/14614448221135159>
- Rangan, P., Saxena, A., Tharoor Srinivasan, R., & Sundar, P. (2023). *Thinking with an accent: Toward a new object, method, and practice*. University of California Press.
- Reilly, J., & Fisher, J. L. (2012). Sherlock Holmes and the strange case of the missing attribution: A historical note on “The Grandfather Passage.” *Journal of Speech, Language, and Hearing Research*, 55(1), 84–88.
- Riszko, L. (2017). Breaching bodily boundaries: Posthuman (dis) embodiment and ecstatic speech in lip-sync performances by boychild. *International Journal of Performance Arts and Digital Media*, 13(2), 153–169.
- Smith, K. L. (2021, May 25). TikTok changes text-to-speech voice after original voice actor files lawsuit. *PopBuzz*. <https://www.popbuzz.com/internet/tiktok/text-to-speechvoice-change-law-suit-reason/>
- SpotifyNews. (2023, February 22). Spotify debuts a new AI DJ, right in your pocket. *For the Record* <https://newsroom.spotify.com/>
- Stein, S. (2013, October 14). Diese Frau sprach die Deutsche Siri. *Bild*. <https://www.bild.de/digital/smartphone-und-tablet/siri/diese-frau-sprach-die-deutsche-siri-32938552.bild.html>
- Sterne, J. (2003). *The audible past: Cultural origins of sound reproduction*. Duke University Press.
- TheAdsWorld. (2018, December 18). *Alexa loses her voice—Amazon super bowl LII commercial* [Video]. YouTube. https://www.youtube.com/watch?v=iNxxvsxU2rJE&ab_channel=TheAdsWorld
- Truax, B. (2001). *Acoustic communication* (2nd ed.). Greenwood Press.
- Trueba, J. L., & Klimkov, V. (2019, August 22). Neural text-to-speech makes speech synthesizers much more versatile. *Amazon Science*. <https://www.amazon.science/blog/neural-text-to-speech-makes-speech-synthesizers-much-more-versatile>
- The Unofficial AppleKeynotes Channel. (2012, December 7). *Apple annual shareholder event 1984—Macintosh introduction* [Video]. YouTube. https://www.youtube.com/watch?v=i3rJsy66bRw&ab_channel=theunofficialAppleKeynoteschannel
- The Unofficial AppleKeynotes Channel. (2013, March 23). *Apple special event 2011—Siri introduction* [Video]. YouTube. https://www.youtube.com/watch?v=agzItTz35QQ&list=PL0msAmi5SaOx5O2z8PVBmQjtMDN3_mi5C&index=4&t=813s&ab_channel=theunofficialAppleKeynoteschannel
- White, M. J. (2022, February 3). This is the real voice behind Google Assistant. *Slash Gear*. <https://www.slashgear.com/this-is-the-real-voice-behind-google-assistant-22708374/>
- Woods, H. S. (2018). Asking more of Siri and Alexa: Feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication*, 35(4), 334–349.

Author Biography

Ido Ramati (Ph.D. The Hebrew University of Jerusalem) is an assistant professor of communication and journalism and cultural studies at the Hebrew University of Jerusalem. His research interests include human–machine communication, history and philosophy of media, and the relation between media technologies, language, and society.