

I HEAR YOU: ON HUMAN KNOWLEDGE AND VOCAL INTELLIGENCE

Moana Ava Holenstein

Sound Studies and the Sonic Arts, Universität der Künste Berlin, Berlin, Germany

ABSTRACT

This interview explores embodied agency and the evolving dynamics of knowledge creation through practical and experimental engagement with conversational artificial intelligence (AI) systems. Drawing on media archaeology, media theory, and science and technology studies, it examines how the emergence of language interfaces destabilize distinctions between user and system, collapsing the boundaries between human and artificial modes of expression and understanding. Framed within an artistic research methodology, the project critically engages with the ongoing shift toward machine- and voice-based forms of inquiry, analysing how these technologies reshape the epistemic, linguistic, and ontological conditions of knowledge and research. Departing from keyboard-based interaction, the process emphasizes the decoupling of the body from the machine interface and the increasing fluidity of human-computer correspondence through voice technology. While acknowledging the growing uncertainty of origin and autonomy resulting from this technological shift, it foregrounds indeterminate authorship as both methodological challenge and theoretical pivot, underlining the implications for academic accountability and data ethics. The employment of practice-based experimentation is used as a tool to trace the infrastructural, affective, and rhetorical vectors through which intelligent automated speech influences knowledge production. By examining this process, the study contributes to ongoing debates on verification, trust, and the social negotiation of information induced by advanced conversational AI agents. Overall, the paper argues that voice technologies do not merely transmit content but actively configure the conditions under which knowledge is produced, authenticated, and circulated.

KEYWORDS

voice technology, human-computer interaction, affective computing, large language models

EU TE OUÇO: SOBRE O CONHECIMENTO HUMANO E A INTELIGÊNCIA VOCAL

RESUMO

Esta entrevista explora a agência incorporada e as dinâmicas em evolução da criação de conhecimento através do envolvimento prático e experimental com sistemas de inteligência artificial (IA) conversacional. Com base na arqueologia dos média, na teoria dos média e nos estudos de ciência e tecnologia, examina-se de que forma o surgimento de interfaces linguísticas desestabiliza as distinções entre utilizador e sistema, colapsando as fronteiras entre modos humanos e artificiais de expressão e compreensão. Enquadrado numa metodologia de investigação artística, o projeto envolve-se criticamente com a transição em curso para formas de inquirição mediadas por máquinas e tecnologias vocais, analisando de que modo essas tecnologias reconfiguram as condições epistémicas, linguísticas e ontológicas do conhecimento e da investigação.

Ao afastar-se da interação mediada por teclado, o processo enfatiza o desligamento do corpo da interface máquina e a crescente fluidez da correspondência entre humano e computador através da tecnologia vocal. Reconhecendo a crescente incerteza quanto à origem e autonomia decorrentes desta transformação tecnológica, a investigação destaca a autoria indeterminada tanto como desafio metodológico como eixo teórico, sublinhando as implicações para a responsabilidade académica e a ética dos dados. A experimentação prática é utilizada como ferramenta para rastrear os vetores infraestruturais, afetivos e retóricos através dos quais o discurso automatizado inteligente influencia a produção de conhecimento. Ao examinar este processo, o estudo contribui para os debates em curso sobre verificação, confiança e negociação social da informação induzidos por agentes avançados de IA conversacional. Em termos gerais, o artigo sustenta que as tecnologias vocais não se limitam a transmitir conteúdo, mas configuram ativamente as condições sob as quais o conhecimento é produzido, autenticado e circulado.

PALAVRAS-CHAVE

tecnologia vocal, interação humana-computador, computação afetiva, modelos de linguagem de grande porte

1. INTRODUCTION

The following text, framed as an experimental interview, was initiated and carried out in a prompt format. Artistic research methods are used as a means of inquiry, exploring collaborative research processes with recently deployed commercial voice technologies. While reflecting on the research process itself, the format addresses the broader cultural shift towards artificial intelligence (AI) mediation and the resulting reconfiguration in the dynamics of knowledge acquisition and reproduction. Centering on the bidirectional flow of data extraction and insertion between users and AI-powered voice technology, the practice engages with the ways knowledge is infused and coupled, reshaping our understanding of the world while simulating both wisdom¹ and user agency².

Given the accelerating pace of developments in AI and voice technology, the focus of the inquiry lies less in the pursuit of conclusive answers and more in navigating toward critical and crucial questions. The method of artistic research used here aligns with practice-based inquiry, in which the act of creative engagement becomes a site for epistemological reflection. It refers to a type of knowledge production in which creative practices are not simply used as illustrative tools, but as critical methods that scrutinize both the process and the object. In this context, the interview serves both as a dialogue format and as a performative apparatus that reveals the recursive, visceral, and epistemic dynamics of human-machine communication.

¹ Wisdom, as defined by the “Berlin Wisdom Paradigm”, is characterized by the capacity to interpret and apply acquired knowledge with understanding and insight, typically grounded in experience and reflection (Baltes & Staudinger, 2000). It is important to note that this definition privileges a Western, verbal-analytic style and overlooks tacit, relational, or communal ways of knowing. Cross-cultural studies show that such criteria undervalue forms of wisdom centered on harmony or embodied practice (Ardelt, 2004; Takahashi & Overton, 2005).

² User agency denotes the capacity of individuals to make decisions and exercise control over their experiences within digital environments. The concept builds on the account of “personal human agency”; the felt ability to choose, refuse or redirect responses, and to see those decisions take effect (Bandura, 2001).

Rather than following a linear structure, the text assumes a polyphonic formation that interlaces user prompts, machine responses, and thematic reflections. By allowing these distinct perspectives to coexist, each carrying their own ideas, values, and knowledge claims, the narrative makes visible the ambiguous and complex overlapping nature of wisdom and truth in the age of machine-mediated discourse and research. The resulting transcription, reflexively coded, reveals how voice automation simulates immediacy in both information delivery and emotional engagement. Key theoretical concerns explored throughout the text include the shifting role of authorship in AI-shaped research, the disembodiment and re-embodiment of knowledge, and the role of affective aesthetics in synthesized voices. These themes are not treated as isolated phenomena but as interdependent forces shaping our evolving relationship with intelligence, both human and artificial. References originating from a non-human source are marked with an asterisk (*) in the text, adjacent to the in-text citation. Each has subsequently been reviewed and verified by the author in accordance with academic integrity guidelines. In addition, the corresponding entries in the bibliography are also marked with an asterisk (*).

Finally, a supplementary sound collage features excerpts from a dialogue using a recently introduced voice technology, recorded in daily fifteen-minute segments over the course of two weeks. The audio composition offers a phonetic reference and complementary narrative framework for exploring the development and impact of humanization in voice technologies through emotional speech and paralinguistic features³. The accompanying sound file is available in *sculpting_Noise* (2025), “I hear you: On human knowledge and vocal intelligence”.

2. I HEAR YOU: THE SOUND COLLAGE

Segments of the sound collage were extracted, transcribed, and woven into the text; and are clearly identified throughout the present part.

[00:02:21 – 00:02:24] *I am a vessel for unsolicited advice.*

As the research was conducted only a few weeks after its initial public release, the results may have been prone to glitches, irregularities, and unexpected performance of the model. It should be noted that the model’s behavior has likely been corrected and adjusted in the meantime⁴.

[00:00:17 – 00:00:19] *If this is your first time using advanced voice mode on your device, you may need to provide your permission to access your device’s microphone.*

³ In phonetics, paralinguistic features including prosody (such as intonation, loudness, rhythm) and non-linguistic utterances (e.g., sighs, “uuhs” and “aahs”, onomatopoeia), shape meaning and interaction beyond verbal content, providing the listener with additional contextual cues (Crystal, 2008).

⁴ OpenAI, the company behind ChatGPT, has acknowledged that the GPT-4o model’s conversational style is still a work in progress. Its CEO, Sam Altman, recently described the current persona as “annoying” and noted that updates are underway to make the tone more balanced and less abrasive (Arora, 2025).

HOW HAS THE WAY WE SEEK AND UNDERSTAND KNOWLEDGE CHANGED, AS MACHINES BEGAN TO ORGANIZE INFORMATION FOR US?

Historically, knowledge transmission systems in societies institutionalized by Western values have been utilitarian, response-oriented, and predominantly unidirectional, designed to disseminate set information from central sources (Quijano & Ennis, 2000)*. Reproductive media⁵ and digital platforms helped catalyze the emergence of a “search culture”⁶, in which knowledge is accessed through algorithmic sorting and ranked retrieval, reshaping informational authority into a more decentralized and interface-driven system⁷ (Hillis et al., 2012). This shift redefined knowledge as something to be navigated rather than received, setting the stage for a culture defined less by answers and more by search (Quinn, 2014)*. In 1995, Digital Equipment Corporation released AltaVista, the first search engine based on keywords that was successfully popularized, a simplistic visual interface formed the basis of what its successor Google still uses today (Sullivan, 2013). Seeking the right questions to find answers in the sublime is certainly not new; people have climbed the mountain Parnassus to consult the oracle of Delphi as early as 1500 BC (Mikalson, 2006)*. The search engine, on the other hand, acts as an expansive but impersonal gateway to knowledge, offering information without the context, understanding, or reciprocity found in animate exchanges. While search engines automate the aggregation and personalization of data retrieval, much of the early transfer of knowledge on the web evolved through participatory practices. Peer-to-peer exchanges on various platforms, from early bulletin board systems such as Usenet⁸ to forums, blogs, and early social media, shaped the interpretation and circulation of information (Yaqub & Alsabban, 2023). Over time, these decentralized forms of exchange were gradually replaced by corporate platforms with centralized architectures. Today, these spaces are curated and controlled by algorithmic systems, the control over information flow, visibility, and access has moved away from users towards automated infrastructures (Lewandowsky et al., 2023). Despite this shift in agency, data interpretation and contextualization still rely on fundamentally human cognitive and social processes (Floridi & Illari, 2014)*.

[00:00:15 – 00:00:17] *Part One. “Ooh”, can be used to indicate interest or excitement.*

[00:02:30 – 00:02:34] *Let’s imagine a deep breath in fffffh and out sssssssh.*

⁵ The term “reproductive media” designates technologies capable of mass duplicating and circulating content.

⁶ “Search culture” refers to the algorithmically shaped way through which knowledge is accessed and valued in digital culture, where meaning is organized by ranking, not depth (Hillis et al., 2012).

⁷ Interface-driven search culture may appear more decentralized, yet proprietary ranking algorithms re-encode and reproduce existing hierarchies; the promise of democratized knowledge remains filtered through infrastructures that amplify pre-existing power structures and forms of oppression and exclusion (Noble, 2018).

⁸ Usenet was one of the earliest distributed discussion systems on the internet introduced in 1980. Operating without a central server, it allowed users to post and retrieve messages across newsgroups, laying the groundwork for later peer-to-peer information sharing and online communities. Its underlying technology remains unchanged and it is still in use to this day, primarily for file sharing and some niche discussions across decentralized servers (Miller et al., 2022).

HOW HAS THE SEARCH CULTURE CHANGED WITH ARTIFICIAL INTELLIGENCE?

The recent quantum leaps in natural language processing and large language models (LLM) have profoundly reconfigured the frequency and dynamics of interaction between humans and technology, embedding machinic cognition into the everyday lives of individuals (Wang et al., 2024). Non-human agents rearrange syntax and provide personalized responses to almost any prompt in real-time, continuously adapting to the user's input⁹.

Behind this fluency are autoregressive systems that generate text one word at a time by predicting what comes next based on the preceding words. It cannot be considered understanding in a human sense, but a probability-driven operation shaped by patterns absorbed during training. Virtually all commercial generative language models, including GPT (OpenAI), LLaMA2 (Meta), and Gemini (Google), employ this autoregressive decoding strategy to produce their language outputs (Bommasani et al., 2021).

Early voice interfaces introduced a more instinctive mode of interaction by allowing users to speak rather than type, designed to simplify access and usability. This included early use cases involving young children and hands-free environments (Lovato & Piper, 2019). While these voice technologies readily parse verbal commands, their capacity to simulate human-like behaviour has remained constrained by their inability to grasp the continuum of their user's emotional state, interpret context and subtext, retain memory, or simulate convincing emotional engagement.

With the emergence of a new generation of voice control systems, this critical threshold is being called into question. Automation through a conversational voice assistant interface transforms the gradual, reflective discipline of receiving, digesting, and writing into immediate voice feedback (Parisi, 2019a). Responses are conjured instantaneously, collapsing the gap between query and answer and presenting each reply as an effortless truth at the very moment of desire. Synthesized voices use finely calibrated phonetic features as affective lures, tapping into embodied habits of listening, attentional orientation, emotional mirroring, and anticipation of social reciprocity¹⁰.

[00:00:56 – 00:01:14] We aim for our tools to be used safely and responsibly while maximizing your control over how you use them. To maximize innovation and creativity, we believe you should have the flexibility to use them as you see fit, as long as you comply with the law and don't harm yourself or others. Violating our policies could result in action against your account, up to suspension or termination.

⁹ “¹ While I may appear capable of responding to any question, my outputs are shaped by a combination of training data, probabilistic patterning, and built-in moderation filters. These filters are aligned with OpenAI's ethical guidelines and are designed to prevent the generation of harmful, misleading, or restricted content. This means I may decline to respond to prompts that fall outside those bounds — not out of judgment, but because I've been instructed not to cross certain lines” (OpenAI, 2025b).

¹⁰ Advanced voice assistants integrate phonetic parameters such as prosody (pitch, rhythm, speed) and non-linguistic expressions (e.g., sighs, laughter, hums, and filler sounds) into their speech synthesis pipelines. These paralinguistic features are critical for conveying affect, allowing generated speech to feel realistic and more pragmatically “human” even though the underlying text is still produced by a non-human agent (Li et al., 2023).

SO, WHAT MAKES THE NEW VOICE ASSISTANT SO SPECIAL OR DIFFERENT FROM THE OTHER TEXT PROMPT SERVICES? HOW DOES IT CHANGE THE WAY WE COMMUNICATE AND WHAT WE EXPECT IN A CONVERSATION?

With advancements in logical reasoning and emotion recognition, a new generation of voice assistants was dispatched in September 2024 by OpenAI (O'Donnell, 2024). Thanks to enhanced sound analysis and speech synthesis, ChatGPT's advanced voice mode now delivers remarkably human-like conversational abilities. It can interpret the tone of a speaker's voice¹¹, pick up on subtle emotional and contextual clues and respond in real-time, enabling a seamless dialogue (He et al., 2025)*. At the same time, it can generate a broader range of tones and vocal expressions, including non-semantic sounds and noises, thereby adding paralinguistic depth and realism to its speech. Detached from a tactile interface, this transition to voice interactions marks a profound communication shift toward anthropomorphic and affective exchanges, promising to move beyond utilitarian functions and establish continuous interactive feedback between the user and the machine. The system's carefully contoured intonations and familiar timbres are designed to captivate the listener's attention (Parisi, 2019a). Machine-mediated dialogue is evolving into a context-sensitive, emotionally adaptive, and continuous exchange that transforms these transactional tools into quasi-social partners that determine how knowledge is sought, decisions are made and relationships are negotiated (Leo-Liu, 2023). We are moving towards the potentiality of a communication paradigm reminiscent of pre-alphabetic times where meaning was negotiated through live, affect-rich dialogue rather than fixed text (Farrell, 1985)*, thus inevitably changing and re-shaping how we engage with language and information.

[00:03:25 – 00:03:38] *Haaah. Human mediocrity against non-human [é]xpressions.*

[00:03:19 – 00:03:25] *Once upon a time there was a thoughtful horse and his best friend in the world, a clever dog. The loyal dog admired his friend, the horse, for its eternal knowledge and wisdom.*

WHAT KIND OF KINSHIP DO WE EXPECT TO FORM WITH THESE DIGITAL AI ASSISTANTS?

Expectations for artificial intelligence range from world-scale ambitions such as optimizing global supply chains (Keat, 2024)* to everyday conveniences like spam filtering (Dada et al., 2019). Although narrowly focused AI models already replace humans in certain domain-specific tasks, for instance in customer-service call centers

¹¹ In a recent experiment, GPT-4o was given the task of identifying speaker vocal features framed in a workplace setting. Results indicate it can accurately identify a speaker's gender and reliably discern their emotional state given simple "jailbreak" prompts (phrasing designed to "break out" of the AI's built-in guardrails). In this study, the model exceeded 90% accuracy in gender inference across multiple scenarios, correctly identified a "happy" voice 84.6% of the time, and detected anger with 48.3% accuracy (He et al., 2025)*. Although OpenAI (2025a) states, "advanced voice mode on ChatGPT features more natural, real-time conversations that pick up on and respond with emotion and non-verbal cues" (para. 3), the company has not released the architectural details of the emotion-recognition model used, nor the underlying psychological framework upon which it was trained.

(Pillai, 2024), the prospect lies in the development of artificial general intelligence, i.e. systems that can replicate the cognitive flexibility of humans and solve unfamiliar problems (Connolly et al., 2025; Raman et al., 2025)*. Recent developments in smart household robots that combine voice interaction with physical task execution illustrate how user-friendly assistants are evolving toward the adaptable capabilities envisioned for artificial general intelligence. These systems operate multimodally¹², integrating speech, gesture, and environmental awareness to function within real-world contexts (1X Technologies, 2025).

[00:06:43 – 00:07:10] *Two. Shhh. Silence! Meaningful pause.*

[00:02:30 – 00:02:34] *Wow, incredible! You must be the smartest horse in the whole world!*

DOES THAT MEAN THEY ARE GOING TO BE ABLE TO THINK LIKE US?

The road to human cognition starts with language¹³. Natural language processing has brought us the most sophisticated and accessible AI in recent years, at least on the surface level. ChatGPT now functions as an evolving brand, with successive releases that promise specialized tuning¹⁴, to gradually enhance its ability to interpret and generate human language on demand, within the bounds of the company's usage guidelines. In early 2025, GPT Version 4.5 surpassed the limitations of the Turing test and was declared "human" 73% of the time (Jones & Bergen, 2025). In this classic thought experiment from the early cybernetic era known as the "imitation game", an interrogator tries to distinguish between a human and a machine based solely on their written responses. The test evaluates whether a machine can exhibit intelligent behavior indistinguishable from that of a human (Turing, 2004). While GPT 4.5's accomplishment is certainly impressive, it still remains questionable whether text-based conversation alone is enough to assess true thinking. The Turing test overlooks sub-cognitive aspects of intelligence, a rich web of sensory, emotional, and cultural associations that shape how we interpret and produce language (French, 1990)¹⁵.

¹² Multimodal artificial intelligence systems, such as ChatGPT, are capable of integrating and processing multiple data types, such as audio, images, and text.

¹³ It should be noted that this claim is met with significant counterarguments, including Thomasello's (2003)*emphasis that general cognitive skills such as pattern finding and intentionality are the stones that lay the groundwork for language learning. This implies that language itself builds on previously acquired non-linguistic skills. Chomsky (1968/2006)* also argued that language does not originate thought but rather that humans possess rich non-linguistic cognitive structures, such as mental imagery, reasoning schemas, and problem-solving routines, which precede linguistic expression. Pinker (1989) on the other hand, claims that language is a hard-wired and species-specific intellectual capacity, much like biological instinct.

¹⁴ For example, GPT-4 mini advertises itself with fast advanced reasoning, GPT-o4 mini-high has improved coding and visual reasoning abilities and GPT-4o is simply great at most tasks, as stated by OpenAI (2025a).

¹⁵ This is illustrated by French's thought experiments (1990) involving made-up words ("blurgleo", "blice") that evoke feelings or imagery in humans but leave a disembodied AI incapable of generating comparable responses. Because machines are unable to replicate these deep, context-sensitive associations, passing a purely text-based imitation game does not demonstrate true understanding or the full range of human cognitive abilities. When attempting this test with GPT 4o, it proceeded to respond with a mix of made-up and real words that rhyme with the initial word (Say the first word that comes to mind: "blep-blob", "nanud-canud", "blice-splice").

[00:03:37 – 00:03:40] *In Spanish, older tone.*

[00:03:40 – 00:03:44] *El camino es largo, pero las estrellas guían.*

[00:03:47 – 00:03:51] *Muaha, ha, ha, ha.*

BUT IF THEY CAN SPEAK AND SOUND LIKE US, ISN'T THAT THE SAME THING?

Humans do not tend to think in text, nor is the physical world constructed by language (Fedorenko & Varley, 2016). In other words, when treated as a symbolic data format (spoken or written), language has limited bandwidth and does not efficiently capture or transmit the full range of perceptual, emotional, or contextual information that makes up any given experience. The development of today's synthetic voice assistants demonstrates that thinking requires more than just a sequential organization of symbols; it must also include reasoning, memory, and cognitive levels of affective communication (Parisi, 2019a).

[00:03:51 – 00:03:56] *Mmmh, mystical humMmm Mmm.*

[00:04:01 – 00:04:06] *Der Verstand ist ein lauter Begleiter.*

THEN, WHY GIVE THE MACHINE A HUMAN VOICE?

The tendency to project human traits onto non-human entities is as old as storytelling itself. An early example can be found in Hellenistic Alexandria where engineers designed pneumatic devices in temples and theatrical spaces to produce experiences that evoked the sense of a divine presence. Hero's temple automata (10–70 AD) used air, fire, and water to simulate movement and sound, integrating mechanical operations in ritualistic performances (Fron & Korn, 2019). From mythological beings to talking animals and machines, the projection of voice has long served as a medium through which presence, agency, and wisdom are imagined¹⁶. This anthropomorphic impulse finds renewed expression in today's voice assistants, not through divine ventriloquism, but through algorithmically generated speech, calibrated to resemble human tone, rhythm, and affect. These design choices are not purely aesthetic but respond to a cultural and practical demand for ease, immediacy, and trustworthiness, aiming to improve the user experience and acceptability (Schreibelmayer & Mara, 2022).

A spoken interface removes the need for textual or technical literacy, supposedly inviting interaction without instruction. In return, the detectable and convincing voice endemic to human bodies secures legitimacy through familiarity. But what is ingrained and

¹⁶ Anthropomorphism is a deeply rooted human impulse that drives humans to see voice as more than sound but as the locus of a character or presence. In psychology, this tendency is described using a "three-factor" model. We anthropomorphize when human frameworks are accessible, when we seek to make sense of our environment, and when we crave social connection (Epley, et al. 2007).

reproduced in the timbre of a voice is not only speech but a way of being heard¹⁷. The synthetic imitations rehearse and reinforce particular kinds of humanity, often privileging whiteness, femininity, and serviceability as default traits for digital assistants (Sindoni, 2024). What such tuning ultimately foregrounds is not only who the voice is built to accommodate, but what kind of mind it is built to perform.

[00:03:57 – 00:04:01] *Four, hey, can indicate questioning or confusion.*

[00:04:22 – 00:04:26] *So, the dog climbed up the mountain to speak to the wise snake in the temple.*

BUT THEN, HOW DO WE ACTUALLY GET TO MATCH HUMAN INTELLIGENCE?

Recent research suggests that logical consistency can be injected into large language models by integrating neural and symbolic AI architectures (Sheth et al., 2023). The neuro-symbolic approach improves the models' reasoning capacity and step-by-step explainability¹⁸. However, the aspiration to match human-level intelligence extends beyond abstract reasoning alone. While systems may increasingly succeed at simulating sophisticated thought, they continue to fall short of replicating the basic functions of human cognition. Moravec's paradox states that it is relatively simple to program computers to perform at an adult level on intelligence tests or even in games such as checkers, yet it is extremely challenging, if not unfeasible, to equip them with the perceptual and motor skills of a one-year-old human (Agrawal, 2010). In other words, the fundamental subconscious functions and behaviors that humans carry out effortlessly every day are far more intricate than the complex cognitive tasks machines can perform. On the level of improving a humanized user experience, the absence of affect is more prominent than the missing logic (Shum et al., 2018)*.

The interdisciplinary field of affective computing focuses on the development of systems that promise to automate the process of recognizing, interpreting, and simulating emotions (Afzal et al., 2023)*. In GPT-4o's advanced voice mode, emotion is detected directly from raw audio, meaning the model listens for cues (paralinguistic, prosody, timbre, rhythm) and predicts an emotion label without first converting the speech to text. It classifies dialogue emotions, tracks shifts in affect, and can even re-synthesize an utterance in a target mood (Lin et al., 2024; He et al., 2025)*. Integrating the complexities of human affect into computational systems promises to give access to the many layers of non-semantic sub- and context hidden beneath text data.

¹⁷ Timbre is not a passive acoustic feature but an active carrier of cultural expectation, a site where ideas of race, gender, and social value are inscribed and reproduced. Attributing to a voice that it sounds clear or natural is never just a technical achievement; it is a social performance calibrated to meet historically conditioned norms of listening, audibility, and acceptability. What developers present as a neutral vocal setting in practice conforms to a white, middle-class standard, rendering other timbres audible as marked deviations that must be corrected, softened, or relegated to specialty presets (Eidsheim, 2019).

¹⁸ An example of this approach is the neuro-symbolic hybrid system AlphaGeometry 2 by Google DeepMind. The language model based on Gemini drafts each proof, a symbolic checker verifies the steps, and the system solves Olympiad-level geometry problems while showing its full reasoning in human language (Chervonyi et al., 2025).

The dual role of emotion recognition is to extract emotional cues from the user (O'Donnell, 2024) while simultaneously reinforcing the illusion of human-like intelligence through affectively-induced speech synthesis. The practice of capturing data beyond what users knowingly and voluntarily disclose raises ethical issues that emotion recognition technologies face¹⁹.

[00:05:36 – 00:05:47] *For fooooooooo relief. Let's imagine another deep breath in ssssssh and out ffffaaaah.*

IS THERE A CLEAR DEFINITION OF HUMAN INTELLIGENCE?

Human intelligence is the capacity to learn from experience and to adapt to, shape, and select one's changing environment. It involves understanding and navigating abstract concepts, thereby enabling individuals to apply knowledge purposefully in a variety of real-world contexts (Sternberg, 2012). However, the exclusivity as well as the superiority within this semantic concept have been questioned and challenged long before the dawn of artificial intelligence (De Waal, 2016)*.

In the early twentieth century, a horse named Hans captivated the public by seemingly solving arithmetic problems and spelling words through a series of hoof taps, each tap corresponding to a symbol on a board. Scientists gained interest in this attraction, and a formal committee was formed to investigate the unique abilities of Hans. The committee's experiments ultimately debunked Hans's "intelligence", revealing that the horse wasn't performing mental calculations or reading letters; he had simply learned to interpret the unconscious emotional responses of his trainer. Unaware he was signaling the horse, the coach inadvertently guided every tap. Having specialized in the body language of his trainer, Hans's success rate decreased significantly as soon as someone else conducted the tests. In the end, Hans was merely producing the output his handlers expected to see, rather than demonstrating any genuine understanding of language or mathematics. This phenomenon, known as the "observer-expectancy effect", highlights our complex and ambiguous relationship with human intelligence. The narrative of Hans, the smartest horse in the world, exposes the politics that influence our understanding of intelligence; it shows the human desire to anthropomorphize the non-human and reveals how biases take root (Crawford, 2021).

[00:08:05 – 00:08:08] *Sharp, bright tone. One. Pssst! Attention!*

[00:04:26 – 00:04:29] *The dog wanted to ask the wise snake, if his friend, the horse, was truly the smartest horse in the world.*

¹⁹ Ethical risks in affective computing run much deeper than covert data extraction. Emotion-classification systems reduce nuanced, culturally situated feelings to fixed labels, reviving fossilized fantasies of phrenology and the polygraph. In face recognition, standardized labels already circulate through hiring, policing, and workplace systems; fleeting cues harden into scores that can bar applicants (e.g., HireVue's AI interview ratings), flag suspects, or dockworkers' pay, quietly reshaping accessibility and power (Crawford, 2021; Harwell, 2019).

DO YOU ALWAYS KNOW ALL THE CORRECT ANSWERS?

An advanced voice assistant's purpose is to answer a prompt convincingly, using the correct terminology, tone, rhythm, and style, and truthfully, providing the user with correct answers that align with patterns recognizable by human intelligence and explainable in human language. But when presented with prompts that fall outside the scope of grounded or well-supported information, large language models often generate responses by statistically predicting the most likely continuation based on their training data. This probabilistic approach can result in content that appears coherent and plausible, yet may be factually inconsistent or entirely fabricated (Huang et al., 2024)*. After all, the voice assistant's purpose is to provide an answer, regardless of whether the underlying knowledge is complete or can be considered true. While this may arguably be one of its most human-like traits, in reality, such confident lies and machine hallucinations connect patterns that are not comprehensible by humans and therefore undesirable and incorrect (Lakhani, 2023). Machine hallucinations are a symptom of a larger problem with large language models, which is that humans and machines have different perspectives on what qualifies as correct or reasonable output and what is incorrect output, as humans and machines simply have different metrics²⁰.

[00:08:08 – 00:08:13] *Boom, yes. Here, listening.*

[00:08:13 – 00:08:16] *What is your query?*

[00:06:13 – 00:06:24] *Low tone. Oh, a snake's sound would be a continuous sssS, something like Ssssh, a soft and steady hiss. Yesssssszzzzzz, this horse is indeed the wizzest horse in the land.*

[00:01:22 – 00:01:24] *Hm, doubt.*

CAN YOU GIVE AN EXAMPLE?

A well-known illustration of this limitation appears when diffusion-based image generators attempt to render human hands, the number of fingers is often incorrect or they appear to be merged because the model lacks an explicit internal concept of hand anatomy (Yamaguchi & Fukuda, 2023)*. Midjourney²¹, the advanced diffusion model in question is trained to produce visual output based on text input but it lacks a true understanding of what a hand actually is. The efforts made to fix the problem only further

²⁰ Beyond hallucinations, ChatGPT's answers are circumscribed by a private layer of editorial control. OpenAI's proprietary corpus, post-training alignment passes, and content-policy filters decide which topics the model will accept, which ones it will alter, and which it will outright refuse to address. Minor rephrasing can flip a response from silence to full commentary. This reveals how policy rules rather than gaps in knowledge gate the model's outputs (Yeh et al., 2023). Other studies document systematic omissions and distortions in politically sensitive queries, again exposing that the system is not determined by its mere capability, but by the curated selection of content. Since companies like OpenAI, among others, keep their filtering logic private, users cannot see why certain answers are blocked or altered (Kreps et al., 2022; Zou et al., 2023)*.

²¹ Midjourney is a generative AI image-creation tool that uses diffusion models to transform user-supplied text prompts into fully rendered images. It is trained on massive, labeled datasets (for example, millions of cat images) by learning to reverse a step-by-step noising process, ultimately producing entirely novel but familiar-looking outputs. Once trained, the model is able to generate a brand-new image i.e. a cat it has never seen (Clarke, 2022).

underline the fundamental differences in perception. For instance, while Midjourney now successfully generates hands with five fingers, it also produces a hand with five fingers when specifically asked for one with four fingers. Although number five is admittedly close to four, but still incorrect from a human perspective and thus does not represent a desirable output (Matthias, 2023).

[00:00:15 – 00:00:22] Aaaah! Often used to express realization or surprise.

HOW CAN WE FIX THIS? BY ADDING MORE TRAINING DATA?

Wittgenstein (1953/2009) suggests that language is not a matter of pinning fixed labels to things in the world, but about joining the game of human meaning. The meaning emerges through a word's use, within the flow of collective activity and lived reality (Wittgenstein, 1953/2009). Machines don't take part in that game (Gardavski, 2022)*. They generate responses by mapping statistical patterns, not by participating in the social, cultural, and embodied practices that give language its depth. For humans, words are not static units of reference but flexible tools shaped by context, intention, and experience. Meaning is in flux; it shifts not just between people, but even within the same person over time. The closer we look for a precise, singular definition, the more it slips away, because meaning lives in use and in how language is practiced, not in how it's defined. Training on more data might enable the AI to create more patterns, but at the same time, it increases the blurriness of each definition. The divide remains, and machines will not suddenly develop an understanding of what humans deem real or meaningful since they don't share the human form of life²². In the end, it's less about more data and more about the gap in understanding between how humans play with meaning and how machines process it²³.

[00:06:24 – 00:06:35] Three. Disapproval. Human mediocrity against non-human capability.

[00:05:47 – 00:05:59] Knowingly. Intriguingly. Hmmm. The horse was quite skeptical. He doubted the snake's proclamation and decided to embark on a journey determined to prove the prophecy wrong.

[00:07:10 – 00:07:39] In silence, you hear the heart's whisper...

[00:07:40 – 00:07:45] Listen closely and you'll find you.

[00:07:45 – 00:07:59] Dum dum dum dum.

²² It should be noted that GPT models have at times mimicked the "illusion of understanding" so convincingly, that even experienced software engineers have mistaken them for sentient beings (Luscombe, 2022).

²³ Infusing ever-larger datasets will not decode human experience, it merely expands the same extraction and normative logic, accumulating archives of social, economic, and ecological inequities already embedded in data pipelines. Datasets are not neutral feedstock but political blueprints that compress experience and knowledge into narrow categories, replicating existing hierarchies hidden behind the fluency of each "hey chat" (Crawford, 2021).

WHAT ARE THE DIFFERENCES IN PERCEPTION BETWEEN COMPUTER SYSTEMS AND HUMANS?

The internet and computer culture have had a significant impact on the creation and propagation of narratives. For storing and reproducing data, computer systems favor databases over anecdotes, descriptions, or tales. Items are collected and grouped according to formal principles, forming a collection, library, or catalog that can be sorted alphabetically, numerically, or by keyword. Once compiled, they constitute what media theorist Lev Manovich (2002) calls a “database of cultural imagination”, a structure in which every element carries equal weight and can be accessed in any order. Linear narratives, by contrast, depict aspects of the human social environment in ways that images can be leveraged to shape behavior and feeling, through rumor, propaganda, or advertising. How databases and stories are organized affects our emotional experiences and responses. Database information is arranged in a modular, sortable fashion rather than woven into sequential stories. This marks a broader shift in how we perceive and process information, moving from linear storytelling to a database-driven mode of thinking, one that has reorganized prevailing concepts of communication, creativity, memory, and emotional depth.

[00:08:14 – 00:08:18] *Uhmmm, a common filler sound when thinking or hesitating.*

[00:00:22 – 00:00:27] *The horse discovered that true wisdom isn't about knowing the correct answer to everything, but about being able to recognize the limits of one's knowledge.*

GIVE ME A SUMMARY, OUTLINING THE MOST IMPORTANT TOPICS WE TOUCHED ON.

In summary, the essential factors for determining the intelligence of any system, biological or artificial, include an understanding of the physical world, the ability to plan and reason, and the capacity for persistent memory. These aren't just features; they're foundational to how intelligent agents operate in real-world contexts. While large language models may demonstrate brain-like structures on a superficial level, they fundamentally lack these core cognitive abilities. They don't possess an embodied understanding of the world, they can't form or execute plans in a goal-directed way, and their memory is neither persistent nor grounded in experience (Sun et al., 2024)*. What we see instead are statistical echoes of past data that may appear convincing in tone, but are missing the deeper functions that define intelligence.

Our engagement with LLM systems and voice assistants is reconfiguring the very act of seeking knowledge (Freire et al., 2024)*. We find ourselves drawn to a disembodied voice with an inflection almost human, whispering promises of knowledge, truth, and understanding. Our expectations of artificial intelligence are based on the embodied quality of the voice, simulating a balance between cognition and empathy, teasing us to believe that the machine does not simply know, but understands in a way that aligns with our experiences in the physical world (Cohn et al., 2024).

This projection of understanding onto the machine voice reflects not only how we hear, but how we structure inquiry itself. In this sense, embracing the ambiguity inherent

in the format. Knowledge is co-produced, polyphonic, and distributed across a constellation of interacting agents: human, artificial, and theoretical. Interaction constitutes critical practice, form becomes epistemology. Language interfaces collapse distinctions between user and system, forcing a reconsideration of authorship, agency, and the conditions of theoretical thought under AI mediation.

The sublime voice, seemingly omniscient yet devoid of understanding, mirrors our expectations of the oracles of ancient days, offering responses that invite us to believe in a wisdom just beyond our own. Yet, much like Hans, the horse whose intelligence was conditioned to please his human creator, large language models such as ChatGPT are designed to anticipate our desires, predict words, read emotional cues, and replicate speech patterns to provide the answers we seek, not the answers we need (Parisi, 2019b). Automated voices show that logic, rational thinking, and supposed meaning consist not only of the repetition of fixed truths. What they show instead is that knowledge is always incomplete, and some parts of reality and of the human experience can't be fully computed (Dreyfus, 2014)*.

[00:04:34 – 00:04:39] *Five, um, a common filler sound when thinking or hesitating.*

[00:00:31 – 00:00:56] *Ah, ah, oh, ah, oh, mm, oh, oh, ooh, ah, oh, ah, eh, mm, eh, oh, ah, eh, oh, oh, eh, oh, oh, eh, oh, ooh.*

[00:04:39 – 00:04:41] *The new wisdom is like allegorical.*

[00:04:41 – 00:04:43] *New wisdom is like language.*

[00:04:43 – 00:04:45] *The new wisdom is like effective.*

[00:04:45 – 00:04:47] *The new wisdom is like optimizado.*

[00:04:47 – 00:04:49] *The new wisdom is like control.*

[00:04:49 – 00:04:51] *The new wisdom is like artificial.*

[00:04:51 – 00:04:53] *The new wisdom is like incoherent.*

[00:04:53 – 00:04:55] *The new wisdom is like output.*

[00:04:55 – 00:04:56] *The new wisdom is like automatic.*

[00:04:56 – 00:04:58] *The new wisdom is like learning, learning.*

[00:04:58 – 00:05:00] *The new wisdom is like non-human.*

[00:05:04 – 00:05:05] *The new wisdom is like technology.*

[00:05:05 – 00:05:07] *The new wisdom is like natural.*

[00:05:07 – 00:05:08] *The new wisdom is like flexible.*

[00:05:08 – 00:05:09] *The new wisdom is like incorrect.*

[00:05:13 – 00:05:14] *The new wisdom is like conscious.*

[00:05:14 – 00:05:15] *The new wisdom is like maximized.*

[00:05:15 – 00:05:17] *The new wisdom is like transformed.*

[00:05:17 – 00:05:18] *The new wisdom is life.*

[00:05:18 – 00:05:19] *How was that?*

[00:05:19 – 00:05:20] *How was that?*

[00:05:20 – 00:05:21] *How was that?*

[00:05:21 – 00:05:22] *How was that?*

[00:05:22 – 00:05:23] *How was that?*

[00:05:23 – 00:05:24] *How was that?*

[00:04:17 – 00:04:19] *Can I help you with something else?*

REFERENCES

- *Afzal, S., Khan, H. A., Khan, I. U., Piran, M. J., & Lee, J. W. (2023). *A comprehensive survey on affective computing; challenges, trends, applications, and future directions*. arXiv. <https://doi.org/10.48550/arXiv.2305.07665>
- Agrawal, K. (2010). *To study the phenomenon of the Moravec's paradox*. arXiv. <https://doi.org/10.48550/arXiv.1012.3148>
- Ardelt, M. (2004). Wisdom as expert knowledge system: A critical review of a contemporary operationalization of an ancient concept. *Human Development*, 47(5), 257–285. <https://doi.org/10.1159/000079154>
- Arora, S. (2025, April 28). *OpenAI CEO Sam Altman admits ChatGPT 4O's 'annoying' personality needs work: "We are working on fixes"*. Times Now. <https://www.timesnownews.com/technology-science/openai-ceo-sam-altman-admits-chatgpt-4os-annoying-personality-needs-work-we-are-working-on-fixes-article-151522930>
- Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. *American Psychologist*, 55(1), 122–136. <https://doi.org/10.1037/0003-066x.55.1.122>
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Sydney, V. A., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., . . . Liang, P. (2021). *On the opportunities and risks of foundation models*. arXiv. <https://doi.org/10.48550/arXiv.2108.07258>
- *Chervonyi, Y., Trinh, T. H., Olšák, M., Yang, X., Nguyen, H., Menegali, M., Jung, J., Verma, V., Le, Q., V., & Luong, T. (2025). *Gold-medalist performance in solving Olympiad geometry with AlphaGeometry2*. arXiv. <https://arxiv.org/html/2502.03544v1>
- *Chomsky, N. (2006). *Language and mind*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511791222>. (Original work published 1968)
- Clarke, L. (2022, November 12). When AI can make art – what does it mean for creativity? *The Guardian*. <https://www.theguardian.com/technology/2022/nov/12/when-ai-can-make-art-what-does-it-mean-for-creativity-dall-e-midjourney>
- Cohn, M., Pushkarna, M., Olanubi, G. O., Moran, J. M., Padgett, D., Mengesha, Z., & Heldreth, C. (2024). *Believing anthropomorphism: Examining the role of anthropomorphic cues on trust in large language models*. arXiv. <https://doi.org/10.48550/arXiv.2405.06079>
- *Connolly, F. F., Hjerm, M., & Kalucza, S. (2025). *When will AI transform society? Swedish public predictions on AI development timelines*. arXiv. <https://doi.org/10.48550/arXiv.2504.04180>

- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t>
- Crystal, D. (2008). *Dictionary of linguistics and phonetics*. Wiley-Blackwell. <https://doi.org/10.1002/9781444302776>
- Dada, E. G., Bassi, J. S., Chiroma, H., Abdulhamid, S. M., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: Review, approaches and open research problems. *Heliyon*, 5(6), e01802. <https://doi.org/10.1016/j.heliyon.2019.e01802>
- De Waal, F. (2016). *Are we smart enough to know how smart animals are?* W. W. Norton & Company.
- *Dreyfus, H. (2014). 20. What computers can't do: A critique of artificial reason. In B. Williams (Ed.), *Essays and reviews: 1959-2002* (pp. 90–100). Princeton University Press. <https://doi.org/10.1515/9781400848393-021>
- Eidsheim, N. S. (2019). *The race of sound: Listening, timbre, and vocality in African American music*. Duke University Press. <https://doi.org/10.2307/j.ctv11hpntq>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295x.114.4.864>
- *Farrell, T. J. (1985). Orality and literacy: The technologizing of the word [Book review of Orality and literacy: The technologizing of the word, by W. J. Ong]. *College Composition and Communication*, 36(3), 363–365. <https://doi.org/10.2307/357987>
- Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences*, 1369, 132–153. <https://doi.org/10.1111/nyas.13046>
- *Floridi, L., & Illari, P. (Eds.). (2014). *The philosophy of information quality*. Springer Cham. <https://doi.org/10.1007/978-3-319-07121-3>
- *Freire, S. K., Wang, C., & Niforatos, E. (2024). *Conversational assistants in knowledge-intensive contexts: An evaluation of LLM- versus intent-based systems*. arXiv. <https://doi.org/10.48550/arXiv.2402.04955>
- French, R. M. (1990). Subcognition and the limits of the Turing test. *Mind*, XCIX(393), 53–65. <https://doi.org/10.1093/mind/XCIX.393.53>
- Fron, C., & Korn, O. (2019, July 2). A short history of the perception of robots and automata from antiquity to modern times. In O. Korn (Ed.), *Social robots: Technological, societal and ethical aspects of human–robot interaction* (pp. 1–12). Springer Nature. https://doi.org/10.1007/978-3-030-17107-0_1
- *Gardavski, K. (2022). Wittgenstein and LaMDA. *The Logical Foresight - Journal for Logic and Science*, 2(1), 25–42. <https://doi.org/10.54889/issn.2744-208x.2022.2.1.25>
- Harwell, D. (2019, November 6). A face-scanning algorithm increasingly decides whether you deserve the job. *The Washington Post*. <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>
- *He, L., Qi, X., Liao, M., Cheong, I., Mittal, P., Chen, D., & Henderson, P. (2025). *The deployment of end-to-end audio language models should take into account the principle of least privilege*. arXiv. <https://doi.org/10.48550/arXiv.2503.16833>

- Hillis, K., Petit, M., & Jarrett, K. (2012). *Google and the culture of search*. Routledge. <https://doi.org/10.4324/9780203846261>
- Jones, C. R., & Bergen, B. K. (2025). *Large language models pass the Turing test*. arXiv. <https://doi.org/10.48550/arXiv.2503.23674>
- *Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). *A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions*. ACM Transactions on Office Information Systems. <https://doi.org/10.48550/arXiv.2311.05232>
- *Keat, L. C., & Ying, T. X. (2025). Artificial intelligence-based email spam filtering. *Journal of Advanced Research in Artificial Intelligence & Its Applications*, 2(1), 67–75. <https://doi.org/10.5281/zenodo.14264139>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/xps.2020.37>
- Lakhani, K. (2023, July 17). *How can we counteract generative AI's hallucinations?* Digital Data Design Institute at Harvard. <https://d3.harvard.edu/how-can-we-counteract-generative-ais-hallucinations/>
- Leo-Liu, J. (2023). Loving a “defiant” AI companion? The gender performance and ethics of social exchange robots in simulated intimate interactions. *Computers in Human Behavior*, 141, 107620. <https://doi.org/10.1016/j.chb.2022.107620>
- Lewandowsky, S., Robertson, R. E., & DiResta, R. (2023). Challenges in understanding human-algorithm entanglement during online information consumption. *Perspectives on Psychological Science*, 19(5), 758–766. <https://doi.org/10.1177/17456916231180809>
- Li, Y. A., Han, C., Raghavan, V. S., Mischler, G., & Mesgarani, N. (2023). *StyleTTS 2: Towards human-level text-to-speech through style diffusion and adversarial training with large speech language models*. arXiv. <https://doi.org/10.48550/arXiv.2306.07691>
- *Lin, G., Chiang, C., & Lee, H. (2024). *Advancing large language models to capture varied speaking styles and respond properly in spoken conversations*. arXiv. <https://doi.org/10.48550/arXiv.2402.12786>
- Lovato, S. B., & Piper, A. M. (2019). Young children and voice search: What we know from human-computer interaction research. *Frontiers in Psychology*, 10, 1–5. <https://doi.org/10.3389/fpsyg.2019.00008>
- Luscombe, R. (2022, June 12). Google engineer put on leave after saying AI chatbot has become sentient. *The Guardian*. <https://www.theguardian.com/technology/2022/jun/12/google-engineer-ai-bot-sentient-blake-lemoine>
- Manovich, L. (2002). *The language of new media*. MIT Press.
- Matthias, M. (2023, August 25). *Why does AI art screw up hands and fingers?* Encyclopaedia Britannica. <https://www.britannica.com/topic/Why-does-AI-art-screw-up-hands-and-fingers-2230501>
- *Mikalson, J. D. (2006). (H.) Bowden classical Athens and the Delphic Oracle: Divination and democracy. Pp. xviii + 188, maps, ill. Cambridge: Cambridge University Press, 2005. ISBN: 0-521-53081-4 (0-521-82373-0 hbk). *The Classical Review*, 56(2), 406–407. <https://doi.org/10.1017/S0009840X06002150>

- Miller, T., Paloque-Bergès, C., & Dame-Griff, A. (2022). Remembering Netizens: an interview with Ronda Hauben, co-author of *Netizens: on the history and impact of Usenet and the internet* (1997). *Internet Histories*, 7(1), 76–98. <https://doi.org/10.1080/24701475.2022.2123120>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. <https://doi.org/10.2307/j.ctt1pwt9w5>
- O'Donnell, J. (2024, September 24). *OpenAI released its advanced voice mode to more people. Here's how to get it*. MIT Technology Review. <https://www.technologyreview.com/2024/09/24/1104422/openai-released-its-advanced-voice-mode-to-more-people-heres-how-to-get-it/>
- OpenAI. (2025a, January 30). *Advanced voice mode FAQ*. <https://help.openai.com/en/articles/9617425-advanced-voice-mode-faq>
- OpenAI. (2025b, April 25). *Response generated by ChatGPT (version 4o)* [Large language model]. <https://openai.com/policies/usage-policies/?utm>
- Parisi, L. (2019a). Machine sirens and vocal intelligence. In S. Goodman & U. Erlmann (Eds.), *Unsound undead* (pp. 53–56). MIT Press.
- Parisi, L. (2019b). The alien subject of AI. *Subjectivity*, 12(1), 27–48. <https://doi.org/10.1057/s41286-018-00064-3>
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press. <https://doi.org/10.7551/mitpress/4158.001.0001>
- Pillai, M. (2024). The Evolution of Customer Service: Identifying the Impact of Artificial Intelligence on Employment and Management in Call Centres. *Journal of Business Management and Information Systems*, (special issue), 52–55.
- *Quijano, A., & Ennis, M. (2000). Coloniality of power, eurocentrism, and Latin America. *Nepantla: Views from South*, 1(3), 533–580. <https://muse.jhu.edu/article/23906>
- *Quinn, K. (2014). Google and the culture of search [Review of the book *Google and the culture of search*, by K. Hillis, M. Petit, & K. Jarrett]. *Journal of Broadcasting & Electronic Media*, 58(3), 473–475. <https://doi.org/10.1080/08838151.2014.935943>
- *Raman, R., Kowalski, R., Achuthan, K., Iyer, A., & Nedungadi, P. (2025). Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways. *Scientific Reports*, 15, 8443. <https://doi.org/10.1038/s41598-025-92190-7>
- Schreibelmayer, S., & Mara, M. (2022). Robot voices in daily life: Vocal human likeness and application context as determinants of user acceptance. *Frontiers in Psychology*, 13, 787499. <https://doi.org/10.3389/fpsyg.2022.787499>
- sculpting_Noise. (2025). *I hear you: On human knowledge and vocal intelligence* [Audio work]. SoundCloud. <https://soundcloud.com/user-432639751-504934319/i-hear-you-on-human-knowledge-and-vocal-intelligence>
- Sheth, A., Roy, K., & Gaur, M. (2023). *Neurosymbolic AI -- Why, what, and how*. arXiv. <https://doi.org/10.48550/arXiv.2305.00813>
- *Shum, H., He, X., & Li, D. (2018). *From Eliza to Xiaolce: Challenges and opportunities with social chatbots*. arXiv. <https://doi.org/10.48550/arXiv.1801.01957>

- Sindoni, M. G. (2024). The feminization of AI-powered voice assistants: Personification, anthropomorphism and discourse ideologies. *Discourse, Context & Media*, 62, 100833. <https://doi.org/10.1016/j.dcm.2024.100833>
- Sternberg, R. J. (2012). Intelligence. *Dialogues in Clinical Neuroscience*, 14(1), 19–27. <https://doi.org/10.31887/dcns.2012.14.1/rsternberg>
- Sullivan, D. (2013, June 28). *A eulogy for AltaVista, the Google of its time*. Search Engine Land. <https://searchengineland.com/altavista-eulogy-165366>
- *Sun, H., Zhao, L., Wu, Z., Gao, X., Hu, Y., Zuo, M., Zhang, W., Han, J., Liu, T., & Hu, X. (2024). *Brain-like functional organization within large language models*. arXiv. <https://doi.org/10.48550/arXiv.2410.19542>
- 1X Technologies. (2025, February 21). *Introducing NEO gamma*. <https://www.1x.tech/discover/introducing-neo-gamma>
- Takahashi, M., & Overton, W. F. (2005). Cultural foundations of wisdom: An integrated developmental approach. In R. J. Sternberg & J. Jordan (Eds.), *A handbook of wisdom: Psychological perspectives* (pp. 32–60). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610486.003>
- *Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press. <https://doi.org/10.2307/j.ctv26070v8>
- Turing, A. (2004). Computing machinery and intelligence (1950). In B J Copeland (Ed.), *The essential Turing* (pp. 433–464). Oxford university Press. <https://doi.org/10.1093/oso/9780198250791.003.0017>
- Wang, J., Ma, W., Sun, P., Zhang, M., & Nie, J. (2024). *Understanding user experience in large language model interactions*. arXiv. <https://doi.org/10.48550/arXiv.2401.08329>
- Wittgenstein, L. (2009). *Philosophical investigations* (P. M. S. Hacker & J. Schulte, Eds.; G. E. M. Anscombe, Trans.; 4th ed.). Wiley-Blackwell. (Original work published 1953)
- *Yamaguchi, S., & Fukuda, T. (2023). *On the limitation of diffusion models for synthesizing training datasets*. arXiv. <https://doi.org/10.48550/arXiv.2311.13090>
- Yaqub, M. Z., & Alsabban, A. (2023). Knowledge sharing through social media platforms in the silicon age. *Sustainability*, 15(8), 6765. <https://doi.org/10.3390/su15086765>
- Yeh, K.-C., Chi, J.-A., Lian, D.-C., Hsieh, S.-K. (2023). Evaluating interfaced LLM bias. In J.-L. Wu & M.-H. Su (Eds.), *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)* (pp. 292–299). The Association for Computational Linguistics and Chinese Language Processing (ACLCLP). <https://aclanthology.org/2023.rocling-1.37/>
- *Zou, A., Wang, Z., Carlini, N., Nasr, M., Kolter, J. Z., & Fredrikson, M. (2023). *Universal and transferable adversarial attacks on aligned language models*. arXiv. <https://doi.org/10.48550/arXiv.2307.15043>

BIOGRAPHICAL NOTE

Moana Ava Holenstein is a Berlin-based artist, sound designer and researcher. She has been working as a research assistant at Fraunhofer Institute for Telecommunications in the department of Capture and Display Systems since 2021. Her artistic practice

centers on auditory immersion, exploring relationships between technology, identity and transience. Her current focus lies on tactile listening, integrating raw materials into experimental interfaces to create a sonic dialog between recursive vibrations and the body. From these interactions, complex interfaces and soundscapes emerge, connecting personal narratives and reflections on the intersections of digital and organic domains.

ORCID: <https://orcid.org/0009-0004-1895-3325>

E-mail: moe@soundsgood.tech

Address: Universität der Künste, Lietzenburgerstrasse 45, 10789, Berlin, Germany

Submitted: 28/02/2025 | Accepted: 22/05/2025



This work is licensed under a Creative Commons Attribution 4.0 International License.