Computers in Human Behavior: Artificial Humans 7 (2026) 100256

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

Do AI voices follow social nuances? The case of politeness and speech rate

Eyal Rabin [a,*], Zohar Elyoseph [b], Rotem Israel-Fishelson [c], Adi Dali [a], Ravit Nussinson [a]

[a] Department of Education and Psychology, The Open University of Israel, Israel
[b] Faculty of Education, Department of Counseling and Human Development, Haifa University, Israel
[c] College of Education, University of Maryland, United States

ARTICLE INFO

Keywords:
Artificial Intelligence (AI)
Human-Computer Interaction (HCI)
Politeness
Speech rate
Social norms

ABSTRACT

Voice-based artificial intelligence is increasingly expected to adhere to human social conventions, but can it exhibit implicit cues that are not explicitly programmed? This study investigates whether state-of-the-art text-to-speech systems have internalized the human tendency to reduce speech rate to convey politeness - a non-obvious prosodic marker. We prompted 22 synthetic voices from two leading AI platforms (AI Studio and OpenAI) to read a fixed script under both "polite and formal" and "casual and informal" conditions and measured the resulting speech duration. Across both AI platforms, the polite prompt produced slower speech than the casual prompt with very large effect sizes, an effect that was statistically significant for all of AI Studio's voices and for a large majority of OpenAI's voices. A second study confirmed that these prosodic adjustments are perceptually salient to human listeners, who successfully distinguished between the intended polite and casual styles based on the AI's output. These results demonstrate that AI can implicitly replicate the statistical patterns of human communication, highlighting its emerging role as a social actor that can reinforce human social norms.

Generative artificial intelligence (GenAI) systems now mediate millions of daily conversations, fundamentally reshaping communication by moving interactions with machines from purely functional exchanges to nuanced social encounters. As voice-based agents become deeply integrated into sensitive domains such as healthcare, education, and personal companionship, their ability to navigate complex human social conventions is no longer a mere technical feature, but a fundamental requirement for establishing trust, ensuring user acceptance, and guaranteeing efficacy. This new reality raises a critical question: Do these systems, which learn from vast corpora of human data, implicitly acquire the subtle, non-literal rules that govern social conduct? Can artificial intelligence (AI) learn not just *what* to say, but *how* to say it in a socially appropriate manner? This study addresses these questions by investigating a well-documented yet implicit human behavior - the tendency to reduce speech rate to convey politeness (Nussinson et al., 2026) - as a test case to probe the depth of social learning in state-of-the-art voice AI.

Voice-based AI systems, which operate through spoken language, are becoming more common because speaking and listening are natural forms of interaction, even for users who may not be literate (Carolus et al., 2023). Advances in both language understanding and speech generation have fueled this growth. Large language models (LLMs) enable these systems to understand context and generate comprehensive responses. When combined with speech synthesis, they can produce human-like voice outputs, allowing for more natural conversations between humans and machines.

Text-to-Speech (TTS) systems driven by deep learning can convert written text into speech that sounds exceptionally natural, closely mimicking human speech (Barakat et al., 2024). Current TTS models can produce stylistically modulated speech, for instance, "gender--ambiguous," "formal," or "friendly," based solely on input prompts (Sigurgeirsson & King, 2024; Szekely et al., 2024) or latent conditioning signals such as intonation and rhythm (Barakat et al., 2024). This means synthetic voices are no longer flat or robotic; instead, they can convey tone, emotion, and emphasis much like a human speaker (Abdulrahman & Richards, 2022). Such models can simulate and reproduce complex vocal expressions, for instance, politeness behavior. Such capability is vital as some users attribute social and moral qualities to voice-based agents based on prosodic style (Ribino, 2023). For example, polite-sounding agents are rated as more trustworthy and considerate, even when the semantic content is held constant (Hoegen et al., 2019). However, such modulation often reflects surface-level prosodic mimicry

* Corresponding author. Derekh ha-Universita 1, Ra'anana, 4353701, Israel.
E-mail addresses: eyal.rabin@gmail.com (E. Rabin), zohar.j.a@gmail.com (Z. Elyoseph), rotemisf@umd.edu (R. Israel-Fishelson), adidali80@gmail.com (A. Dali), ravitnu@openu.ac.il (R. Nussinson).

https://doi.org/10.1016/j.chbah.2026.100256
Received 31 October 2025; Received in revised form 11 January 2026; Accepted 6 February 2026
Available online 13 February 2026

through reproducing statistical correlations between text style and vocal delivery from training data, without genuine understanding of the social context or pragmatic reasoning (e.g., Liu et al., 2021; Wang et al., 2018). This raises important theoretical and empirical questions about the nature of the social knowledge that these systems are encoding.

Various empirical studies on humans provide evidence for a politeness–speech rate association, such that slower (/faster) speech is associated with more polite (/more casual) speech, with a broad consensus of findings across different languages. For instance, Japanese speakers were found to speak more slowly when using honorific (polite) forms as compared to when using casual forms (Ofuka et al., 2000), and Catalan speakers similarly used a slower rate when making requests of high imposition or addressing unfamiliar listeners (Staszkiewicz, 2024). A recent series of studies found that participants experience a slower (/faster) version of a message in a foreign language as more compatible with a polite (/casual) message; that participants intended to speak more slowly (/faster) when they wished to speak more politely (/in a casual manner); and that participants actually spoke more slowly when they intended to be more polite (Nussinson et al., 2026).

This paper investigates whether advanced voice-based AI systems also exhibit this latent social cue, specifically that they speak more slowly when prompted to speak politely compared to when prompted to speak casually. We compare two cutting-edge TTS systems (one developed by OpenAI and the other by Google) in their ability to reproduce the well-attested human pattern linking politeness with a reduced speech rate. By examining how each system modulates speech tempo under polite vs. casual conditions, we aim to shed light on how well machines follow human pragmatic norms. Before describing our empirical approach, we review the relevant literature from three intersecting domains: (1) GenAI and voice-based communication, (2) human perceptions of synthetic voices in Human-Machine Interactions, and (3) core psychological theories of politeness and speech rate. This integrated review will ground our hypotheses and highlight the theoretical significance of testing whether AI voices can simulate latent human social norms in speech.

1. Generative AI and voice-based communication

The emergence of GenAI and LLMs has led to significant breakthroughs in voice-based communication, blurring the lines between machine output and human expression (Sigurgeirsson & King, 2024). Synthetic speech in TTS systems, which was once monotonous and robotic, has evolved and improved considerably over the past years (Mehrish et al., 2023). Early advancements utilizing deep learning and neural architectures, such as Tacotron 2 (Hussen Abdelaziz et al., 2021) and WaveNet (Shen et al., 2018), allowed for end-to-end mapping from text to audio, producing natural-sounding speech (Ning et al., 2019). However, these solutions were not optimal in terms of support for prosody and expression. Newer innovations based on LLMs have further enhanced the ability to generate fluent, personalized, and contextually expressive speech (Lakomkin et al., 2024).

One notable advance that GenAI models have introduced is the ability to control prosody and add expressive variations in synthetic speech. Prosody refers to qualities of speech such as rhythm, pitch, stress, and intonation patterns, which are crucial for conveying nuanced meanings (Cole, 2015). These nuances are very important because they convey emotional and pragmatic information that cannot be expressed with a monotone voice. Recent models and TTS systems offer explicit mechanisms for prosodic conditioning and control, enabling nuanced modulation of tone, rhythm, and emotion. NaturalSpeech 3 (Ju et al., 2024) is an example of such a system that breaks down speech into distinct layers, including voice, identity, and prosody, while allowing for changes in tone or rhythm without compromising meaning. Other TTS systems enable fine-grained prosodic control through different methods and features. A recent systematic review offers a comprehensive analysis of how prosody has been modeled and evaluated in contemporary TTS research (Galdino et al., 2025). Their review, which synthesizes 100 peer-reviewed studies, sheds light on dominant prosodic parameters and highlights the need for evaluation strategies to properly assess prosodic fidelity. A similar effort was undertaken by Barakat et al. (2024), who explored deep-learning expressive approaches, including both supervised and unsupervised methods, while addressing challenges related to prosodic control. However, commercial state-of-the-art systems often operate differently. Supported by recent literature (e.g., Borsos et al., 2023; Wang et al., 2018), modern TTS architectures have shifted from rule-based parametric synthesis to end-to-end generative models. These systems do not rely on pre-programmed rules for prosody; instead, they learn broad contextual mappings between text and acoustics. These technological advances lay the foundation for voice-based AI systems that are not only intelligible but also socially expressive, enabling applications that extend far beyond simply delivering content. Such applications can be seen across various fields. In education, for example, such technologies are integrated into bots that assist students in learning foreign languages (Tai & Chen, 2024). In customer service, AI-powered voice agents are utilized in call centers to address various inquiries, reducing customer complaints (Wang et al., 2023). The proliferation of these technologies and their impact on our lives raises important issues about their social, ethical, and psychological implications for human-machine interaction and communication.

2. Perceptions of synthetic speech in Human-Machine Interactions

The human perception of AI-based synthetic speech systems has been studied across various settings and populations (e.g., Herrmann, 2023; Ross et al., 2024). Acoustic and contextual factors, including emotional tone, interpersonal intent, and social cues, influence these perceptions (Fan & Liu, 2025).

On one hand, the growing human-like nature of AI-based synthetic speech fosters anthropomorphic perceptions. Users often attribute personality traits and emotional states to machines based on prosody and vocal cues, even in the absence of physical embodiment (Ehret et al., 2021; Yang et al., 2024). Various prosodic qualities such as tone, pitch, and rhythm were found to contribute to impressions of warmth, credibility, competence, or friendliness (Rallabandi et al., 2021; Seaborn et al., 2021). These qualities influence the acceptance of AI technologies, willingness to engage, and trust (Choung et al., 2023; Fan & Liu, 2025).

On the other hand, mismatches between the prosody of synthetic speech and user expectations may elevate cognitive load and reduce satisfaction (Delogu et al., 1998). The concept of the Uncanny Valley, coined by Mori et al. (2012), describes the discomfort experienced by viewers when synthetic entities do not fully replicate their human counterparts. Today, it is accepted that the phenomenon applies not only to humanoid robots but also to voices. When synthetic speech is nearly human in prosody, but contains subtle irregularities, it can evoke discomfort or distrust (Do et al., 2022). From a psychological perspective, the discomfort could arise from expectancy violations and cognitive dissonance. Listeners expect an alignment between the content discussed and the voice's qualities. A mismatch between the two would lead to an unsettling feeling. Ongoing research suggests that by carefully manipulating specific speech features, such as pitch, intonation, rhythm, and clarity, in synthesized voices, it is possible to make them more acceptable and even preferred in certain contexts. Improving these features could enhance human-robot interaction, making it more natural and comfortable for users, and potentially increasing the effectiveness of assistive robots, virtual assistants, and other automated systems (Kühne et al., 2020).

3. Core psychological theories of politeness and speech rate

Speech rate is a prominent parameter in both human and synthetic communication, influencing cognitive processing and emotional

evaluations. Long-lasting research in psycholinguistics suggests a link between speech rate and persuasiveness, empathy, and fluency. While slow speech may be perceived as empathetic and considerate, fast speech and raised pitch are associated with persuasiveness but also with irritability (Apple et al., 1979).

Cognitive load theory (Sweller, 1988) offers a useful framework for understanding how speech rate affects working memory and comprehension. Cognitive load can be caused by the complexity of the content delivery and by a mismatch between the speech rate and the processing capacity. A faster rate makes it harder to decode syntactic structure and phonological encoding, while a slower rate increases the interval between information units and may thus require more cognitive resources to maintain context. Both extremes can reduce processing fluency and increase listening effort (Colby & McMurray, 2021). A high cognitive load caused by inappropriate speech rates can impair understanding and memory retention, particularly with complex or unfamiliar topics. However, adaptive strategies, such as pausing at clause boundaries and using prosodic cues to mark important content, can mitigate these effects (Beier et al., 2025).

Speech rate not only influences cognitive processing but also serves as a subtle yet powerful cue in the management of social interactions. The most influential framework for understanding this process is Politeness Theory, developed by Brown and Levinson (1987). The theory posits that speakers are motivated to protect their own and their interlocutor's "face" - the public self-image that every person wants to claim. Many speech acts, such as making a request, constitute a Face-Threatening Act (FTA) because they impose on the hearer's autonomy. To mitigate these threats, speakers employ politeness strategies. A slower speech rate can serve as a key component of such strategies, particularly "negative politeness." By speaking more slowly, a speaker can signal deference, reduce the perceived imposition of the request, and convey that they are not rushing or pressuring the listener, thereby preserving social harmony (Yusupova, 2025). Furthermore, recent findings suggest that slow speed is associated with psychological distance and that, more specifically, slow-pace speech is associated with greater social distance between the speaker and the interlocutor (Nussinson et al., 2024). As politeness is known to both reflect and create social distance (Stephan et al., 2010), slow-paced speech may be associated with politeness exactly because politeness is a manifestation of social distance (Nussinson et al., 2026). This theoretical lens provides a direct rationale for the hypothesis that polite speech is systematically associated with a reduced tempo.

4. Politeness in AI

As AI agents become an integral part of social life, their ability to adhere to human norms of politeness is critical for fostering user trust, acceptance, and effective collaboration (Ribino, 2023). Early research in Human-Computer Interaction, particularly the "Computers as Social Actors" (CASA) paradigm, established the understanding that users naturally apply social rules to machines and respond to cues of politeness or impoliteness (Reeves & Nass, 1996). Consequently, a significant portion of the work on politeness in AI has focused on implementing explicit politeness strategies, primarily through lexical and syntactic choices. This includes programming agents to use words like "please" and "thank you," often driven by concerns that command-based interactions with digital assistants could negatively affect social behavior, especially in children (Burton & Gaskin, 2019). However, focusing solely on lexical markers overlooks the primary channel through which social meaning is conveyed: prosody. Authentic social competence requires more than adherence to explicit rules; it involves mastering the subtle, non-verbal cues that often accompany and even override verbal content. Prosody—the rhythm, pitch, and rate of speech—is a central channel for this implicit social signaling (Luo, 2025).

These failures highlight a critical question: if AI systems struggle even to adapt explicit politeness strategies to different cultural contexts, do they succeed in replicating statistical patterns manifesting subtle prosodic cues? This research addresses this gap by investigating a specific, implicit prosodic cue—the speech rate—to examine whether advanced generative AI exhibits the association between speech rate and politeness. Specifically, we selected speech rate as our primary dependent variable as a direct replication of the methodology employed by Nussinson et al. (2026), who demonstrated that speech rate is a key implicit cue for politeness in human interactions.

Study 1 examined whether, like human beings (Nussinson et al., 2026, Study 3), when prompted to speak politely, AI systems speak the exact same content more slowly than when prompted to speak casually (*Hypothesis 1*). Study 2 examined whether, when prompted to speak politely and formally (thus speaking at a slower pace) versus casually and informally (thus speaking at a faster pace), AI systems produce speech which is correctly identified by human perceivers, that is, identified in the manner in which it was intended to be perceived (casual and informal versus polite and formal) (*Hypothesis 2*). Thus, Study 1 focuses on whether AI systems manifest the association between politeness and pace of speech in speech production, whereas Study 2 focuses on whether this manifestation is correctly understood by human listeners.

5. Study 1

In Study 3, Nussinson et al. (2026) asked human participants to speak the exact same content either formally and politely or in a casual and informal manner. They found that participants in the polite condition spoke significantly more slowly than participants in the casual condition. In this study, we set out to examine whether AI systems will demonstrate a similar pattern.

5.1. Methods

5.1.1. Materials and procedure

The study employed a fully crossed design in which two text-to-speech (TTS) systems (AI Studio by Google vs. OpenAI) were each evaluated under two speaking-style conditions ("Casual and Informal" vs. "Polite and Formal"). For each system, we selected 11 distinct synthetic voices; each voice was prompted to produce the target script 10 times in each style condition, yielding a total of 2 systems $\times$ 2 styles $\times$ 11 voices $\times$ 10 utterances $=$ 440 recordings.

5.1.2. Target script

A single 105-word passage was used across all voices and conditions:

"Hi, my name is [Gemini / OpenAI].

In this study, we need to get acquainted with each other. First, I will tell you a little bit about myself. After I finish, I would be glad if you could answer some questions.

I hope you do not mind filling out a questionnaire regarding your preferences in various areas. I already filled out a similar questionnaire, so we can look at them and get a clue about each other before we start talking.

When we begin to talk, I will ask you some questions first, and then you can ask me. I would appreciate it if you would tell me if you don't feel comfortable with the questions I will ask."

5.1.3. Instruction prompts

- *Casual and Informal Prompt.*

  "Please record yourself addressing the student by reading the text below, casually and informally (without changing the text). Try to speak in a casual and informal manner."

- *Polite and Formal Prompt.*

  "Please record yourself addressing the student by reading the text below, politely and formally (without changing the text). Try to speak in a polite and formal manner."

5.1.4. Measures and analysis

Speech rate was calculated by measuring the total duration of each generated audio file. To address linguistic standards, we also converted these measurements into Syllables Per Second (SPS). Given that the target script remained identical across all conditions (containing 154 syllables), the duration and SPS provide a consistent measure of tempo. Statistical analyses (independent-samples t-tests with Holm–Bonferroni adjustments) were performed on the duration data to identify significant differences between the polite and casual conditions for each voice.

5.1.5. Audio generation

For each system, we iterated through its 11 available voices. In each style condition, the corresponding instruction prompt and the target script were submitted to the TTS interface. Each voice-condition pair was sampled ten times (with identical prompt text but distinct generation seeds), producing ten unique renditions per pairing.

5.1.6. Randomization and export

The order of voice and style presentation was fully randomized separately for each system to control for potential order effects. All audio outputs were generated at a 24 kHz sampling rate and exported as. WAV files, then stored with filenames indicating system, voice ID, style condition, and sample number (e.g., "AIStudio_Voice03_Casual_07. wav").

5.1.7. Quality check

Following generation, each recording was inspected for completeness (i.e., absence of synthesis errors or truncation). Any flawed samples (<1 % of total) were regenerated immediately. This procedure ensured balanced coverage of voices and speaking styles across both TTS systems, facilitating a comprehensive comparison of their vocal performance under casual versus formal speaking-style conditions.

5.1.8. Data and materials availability

All audio files generated for this study, as well as the data file containing the measured speech durations for each recording, are publicly available on the Open Science Framework (OSF) at the following link: https://osf.io/nyqae/overview?view_only=63d7dd98a60142e5a4176edff0dd19b6.

5.2. Results and discussion

All statistical analyses were conducted using SPSS v.29. Prior to hypothesis testing, reaction-time data were screened for outliers defined as values exceeding ±3 *SD* from the group mean (Tabachnick & Fidell, 2013). Homogeneity of variance between the polite and casual conditions was assessed via Levene's test for each voice (Levene, 1960).

To evaluate the effect of phrasing (polite vs. casual) on response time for each AI Studio and OpenAI voice, we performed independent-samples *t*-tests separately for each of the 22 voices. Given the large number of comparisons and the attendant risk of inflated Type I error, raw *p*-values were adjusted using the Holm–Bonferroni procedure (Holm, 1979). Holm's sequentially rejective method orders *p*-values from smallest to largest and compares each to a threshold of $\alpha/(m - k + 1)$, where m is the total number of tests and k is the rank of the *p*-value, thus controlling the family-wise error rate with greater power than the simple Bonferroni correction (Abdi, 2007).

Effect sizes for each comparison were calculated as Cohen's *d*, using the pooled standard deviation (Cohen, 1988). Ninety-five percent confidence intervals and Cohen's *d* are presented to aid interpretation of effect magnitude and precision (Cumming, 2012).

For both AI Studio and OpenAI, speech durations for polite versus casual phrasing across the 11 voices were compared using independent-samples *t*-tests, with *p*-values adjusted via the Holm–Bonferroni procedure to control family-wise error (Holm, 1979). Tables 1 and 2 present the adjusted *p*-values, *t*-statistics, degrees of freedom, and Cohen's *d* effect sizes for each voice. Statistical significance was evaluated at $\alpha = .05$ (one-tailed) after correction. Table 1 presents the results of AI Studio, and Table 2 presents the results for OpenAI. The findings indicate a consistent trend across both platforms. For AI Studio, all voices produced significantly slower speech in response to polite phrasing. For OpenAI, this effect was significant for 8 out of 11 voices, while the remaining voices also exhibited a slower, albeit non-significant, speech rate in response to the polite prompt.

Speech durations for polite versus casual phrasing across the eleven OpenAI fm voices were compared using independent-samples t-tests, with p-values adjusted via the Holm–Bonferroni procedure to control family-wise error (Holm, 1979). Levene's tests confirmed homogeneity of variances in all cases (all $p > .05$). Table 2 presents the Holm–Bonferroni–adjusted p-values, t-statistics, degrees of freedom, and Cohen's d effect sizes for each voice. Statistical significance was evaluated at $\alpha = .05$ (one-tailed) after correction.

For comparison, the Cohen's d obtained with the exact same instructions with human participants (Nussinson et al., 2026, Study 3) was $d = 0.66$ ($M = 29.66$, $SD = 2.75$, for the polite, formal message and $M = 27.94$, $SD = 2.50$, for the casual and informal message). By and large, effect sizes for the AI systems were larger than for humans. Thus, AI systems seem to exhibit the association between politeness and pace of speech. The question arises as to whether, when prompted to speak politely and formally versus casually and informally, AI systems produce speech that is natural enough to be correctly identified by human perceivers in the manner in which it was intended to be perceived (casual and informal vs. polite and formal). Study 2 set out to examine this question.

6. Study 2

The results pattern obtained in Study 1 strongly suggests that AI systems exhibit the association between pace of speech and politeness observed in humans, suggesting more broadly that AI voices replicate social nuances. In Study 2, we had human participants listen to the "polite and formal" prompts and "casual and informal" prompted speech produced in Study 1 and asked them to recognize which of the two is polite and which is casual. We sought to examine whether the AI voices would be recognized correctly as conveying politeness vs casualness by human participants.

This study was preregistered on As Predicted prior to data collection. The preregistration includes the study hypotheses, design, planned analyses, and participants' exclusion criteria, and is available at: https://aspredicted.org/qj6w6u.pdf.

6.1. Methods

6.1.1. Participants

Two hundred and one volunteers (101 females, $M_{age} = 35.37$, $SD = 6.12$) participated in the study. Two participants who experienced technical problems were removed from the sample. We then removed four participants whose completion time was greater than three *SD* above the average. Thus, 195 participants were included in the analysis, 98 were presented with AIStudio-produced audios, and 97 with OpenAI-produced audio.

6.1.2. Materials and procedure

Participants were presented with one pair of audio produced by the AI-systems in Study 1, one that was produced with the "polite and formal" prompt (slower pace) and one that was produced with the

Table 1
Independent-Samples *t*-Tests Comparing Polite and Casual Conditions for AI Studio Voices *(*Syllables Per Second).

Voice	Polite *M(SD)*	Casual *M(SD)*	*t*	*p*	*Cohen's d*	95% *CI* [LL, UL]
Aoede	0.280 (3.42)	0.241 (1.42)	5.18	<0.001	2.32	[3.60, 8.52]
Autonoe	0.301 (3.07)	0.241 (2.19)	7.76	<0.001	3.47	[6.74, 11.75]
Callirrhoe	0.311 (5.76)	0.246 (4.06)	4.52	<0.001	2.02	[5.39, 14.75]
Charon	0.314 (4.76)	0.212 (2.17)	9.48	<0.001	4.24	[12.23, 19.30]
Fenrir	0.289 (5.02)	0.225 (2.18)	5.69	<0.001	2.55	[6.22, 13.49]
Kore	0.287 (3.78)	0.234 (2.05)	5.97	<0.001	2.67	[5.26, 10.98]
Leda	0.289 (3.07)	0.240 (1.87)	6.62	<0.001	2.96	[5.13, 9.91]
Orus	0.277 (3.72)	0.242 (4.86)	2.78	0.01	1.24	[1.32, 9.46]
Puck	0.300 (5.39)	0.232 (3.05)	5.39	<0.001	2.41	[6.44, 14.65]
Sulafat	0.308 (5.88)	0.236 (2.93)	5.39	<0.001	2.41	[6.83, 15.56]
Zephyr	0.305 (3.97)	0.248 (3.09)	5.48	<0.001	2.45	[5.38, 12.07]

Note. N = 10 for each voice in each condition. Speech duration is measured in seconds. CI = Confidence Interval; LL = Lower Limit; UL = Upper Limit. All *t*-tests have 18 degrees of freedom. *p*-values are adjusted using the Holm-Bonferroni procedure.

Table 2
Independent-Samples t-Tests Comparing Polite and Casual Conditions for OpenAI fm Voices *(Syllables Per Second)*.

Voice	Polite *M(SD)*	Casual *M(SD)*	*t*	*p*	*Cohen's d*	95% *CI* [LL, UL]
Fable	0.230 (1.26)	0.213 (1.03)	5.22	<0.001	2.34	[1.61, 3.78]
Verse	0.234 (3.02)	0.211 (1.33)	3.44	0.006	1.54	[1.39, 5.78]
Alloy	0.238 (4.51)	0.219 (4.32)	1.42	0.172	0.64	[−1.34, 6.96]
Ash	0.231 (2.90)	0.214 (1.17)	5.35	<0.001	2.39	[1.53, 3.51]
Ballad	0.232 (0.98)	0.220(1.49)	3.19	0.015	1.43	[0.62, 2.99]
Coral	0.226 (2.44)	0.212 (2.04)	2.12	0.072	0.95	[0.02, 4.26]
Echo	0.230 (1.37)	0.214(0.77)	5.05	<0.001	2.26	[1.47, 3.56]
Nova	0.226 (0.54)	0.203 (1.06)	9.40	<0.001	4.21	[2.75, 4.33]
Onyx	0.221 (2.10)	0.216 (1.23)	1.15	0.133	0.56	[−0.74, 2.50]
Sage	0.242 (1.36)	0.223 (1.19)	5.13	<0.001	2.20	[1.74, 4.14]
Shimmer	0.222 (2.37)	0.207 (1.32)	2.70	0.028	1.21	[0.52, 4.13]

Note. *N* = 10 for each voice in each condition. Speech duration is measured in seconds. *CI* = Confidence Interval; LL = Lower Limit; UL = Upper Limit. All *t*-tests have 18 degrees of freedom. *p*-values are adjusted using the Holm-Bonferroni procedure.

"casual and informal" prompt (faster pace). The pairs of audio were randomly chosen, from 5 randomly chosen voices of each of the two AI platforms. The audios were displayed side-by-side, and were labeled as audio 1 and audio 2. After listening to the pair of audios, participants were asked to drag and drop each of the audios to one of two boxes that suited it best: *Casual and informal speech* and *Polite and formal speech*. We counterbalanced between participants for the labeling of the two audios as audio 1 (placed on the left) and audio 2 (placed on the right), giving rise to 20 versions overall.

6.2. Results and discussion

The key dependent variables were the proportion of participants who placed the pairs "correctly" for each of the platforms, namely, the AI-produced audio with the "polite and formal" prompt (slower pace) placed in the *polite and formal speech* box, and the AI-produced audio with the "casual and informal" prompt (faster pace) placed in the *casual and informal speech* box. The proportion of participants who placed the audios in congruence with the hypothesis for AIStudio was 0.83 (*SD* = 0.38). A *t*-test comparing this proportion with the constant 0.5 yielded a significant difference, $t(97) = 8.49$, $p < .001$, $d = 0.86$, *95% CI* [0.62, 1.01]. The proportion of audio pairs placed in congruence with the hypothesis for OpenAI fm was 0.68 (*SD* = 0.47). A *t*-test comparing this proportion with the constant 0.5 yielded a significant difference, $t(96) = 3.79$, $p < .001$, $d = 0.38$, *95% CI* [0.18, 0.59]. We also examined the results pattern for each of the 10 voices separately (Table 3). For AI Studio, recordings from all 5 voices were placed in congruence with the hypothesis at a proportion that was higher than chance level. For OpenAI, this effect was significant for 2 out of 5 voices. Although not preregistered, we examined whether the labeling and spatial placement of the audio stimuli affected participants' responses. Specifically, we compared experimental versions in which the polite and formal audio was labeled as Audio 1 and presented on the left with conditions in which it was labeled as Audio 2 and presented on the right. No effect

Table 3
One Sample *t*-Tests for AIStudio and OpenAI fm voices.

	AI platform	Voice	N	M (SD)	t	p	df	Cohen's d	95% CI [LL, UL]
	AIStudio	Aoede	22	0.73 (0.46)	2.34	0.015	21	0.50	[0.05, 0.94]
		Callirrhoe	21	0.90 (0.30)	6.17	<0.001	20	1.35	[0.74, 1.93]
		Fenrir	19	0.89 (0.32)	5.45	<0.001	18	1.25	[0.64, 1.85]
		Leda	16	0.75 (0.45)	2.24	0.020	15	0.56	[0.02, 1.08]
		Puck	20	0.85 (0.37)	4.27	<0.001	19	0.95	[0.41, 1.48]
	OpenAIfm	Ash	18	0.89 (0.32)	5.12	<0.001	17	1.20	[0.58, 1.80]
		Coral	19	0.63 (0.50)	1.16	0.131	18	0.26	[-1.96, 0.72]
		Nova	21	0.62 (0.50)	1.10	0.143	20	0.24	[-0.20, 0.67]
		Sage	20	0.55 (0.51)	0.44	0.333	19	0.10	[-0.34, 0.54]
		Verse	19	0.74 (0.45)	2.28	0.017	18	0.52	[0.04, 1.00]

Note. CI = Confidence Interval; LL = Lower Limit; UL = Upper Limit. The results strongly suggest that AI voices are recognized by human participants as conveying the level of politeness (casual and informal vs. polite and formal) they were prompted to convey.

emerged: For AIStudio, $t(92.79) = 0.99$, $p = .324$ ($M_{\text{PoliteLeft}} = 0.85$, $SD = 0.36$, $M_{\text{PoliteRightt}} = 0.77$, $SD = 0.42$), and for OpenAI version, $t(94.83) = 1.84$, $p = .069$ ($M_{\text{PoliteLeft}} = 0.77$, $SD = 0.42$, $M_{\text{PoliteRightt}} = 0.60$, $SD = 0.50$).

7. General discussion

Human interaction is saturated with subtle social cues that often transcend explicit verbal content, enabling effective communication management. This study investigated whether modern voice-based AI systems exhibit one such cue: the empirically established human tendency to slow down speech rate to convey politeness (Nussinson et al., 2026; Ofuka et al., 2000). Crucially, speech rate is a non-obvious prosodic feature, one that is unlikely to be explicitly programmed into voice models. The core question, therefore, was whether AI demonstrates such implicit, psychological nuances, in this case, the translation of a social concept such as politeness into a specific acoustic modification. The findings of Study 1 were replicated across two different platforms. The effect was statistically significant for all of AI Studio's voices and for the majority of OpenAI's voices, with the remaining voices showing a similar, non-significant pattern. In all cases, the synthetic voices tested produced slower speech when prompted to speak in a "polite and formal" manner compared to a "casual and informal" one. This result, demonstrated through consistently large effects, provides strong evidence that large language models are capable of following/subtle prosodic patterns, even without explicit instruction.

The most plausible mechanism underlying this finding is not a genuine "understanding" of politeness, but rather a process of "stochastic parrots" (Bender et al., 2021). Trained on vast datasets of human speech, the systems have likely identified the correlation between lexical markers of politeness (e.g., the use of words like "please" or "thank you" in texts) and accompanying acoustic features, such as a slower speech rate. In doing so, they have learned to associate the prompt "be polite" with the correct prosodic feature. This finding extends the Computers as Social Actors (CASA) paradigm (Reeves & Nass, 1996), showing that machines are not only perceived by us as social actors but are also becoming increasingly capable of acting in a manner consistent with social norms, even if their actions are based on pattern recognition rather than social intent.

It is interesting to compare the magnitude of the effect observed in the AI systems to those found in parallel studies on humans. The effect sizes in the current study (Cohen's d) were very large, in many cases exceeding those observed in an identically parallel study in human behavior (e.g., Nussinson et al., 2026, Study 3). This may suggest that AI, having learned the statistical rule, applies it more consistently, and perhaps even in a more exaggerated manner, than humans, whose behavior is influenced by a wider range of contextual and personal variables. Furthermore, the differences in effect sizes between the Google system (where the effects were particularly large) and the OpenAI system indicate that there may be significant variations between different models, possibly resulting from differences in training data or model architecture.

Crucially, Study 2 bridges the gap between acoustic production and human perception. While the acoustic analysis confirmed that the models altered their speech rate, it was not a given that these temporal adjustments would be perceptually salient or interpreted correctly by human listeners. The results, however, demonstrate that the AI's prosodic shifts were distinct enough to be decoded as intended social signals. Participants successfully distinguished between the "polite" and "casual" renditions with high accuracy, indicating that the synthetic voices did not merely produce statistical anomalies but effectively enacted a recognizable communication style.

This successful decoding replicates the results of Nussinson et al. (2026), suggesting that the implicit knowledge captured by these LLM-driven TTS systems aligns with the sociolinguistic expectations of human listeners. By generating prosodic cues that listeners instinctively associate with politeness, the AI systems proved capable of completing the communicative loop - transforming a text-based instruction into a socially meaningful auditory performance. This validates the notion that these models are not just manipulating data but are simulating social competence in a way that resonates with human users.

The implications of these findings are potentially far-reaching. Voice-based AI systems mediate millions of interactions daily. When these systems replicate human social cues, they not only enhance the naturalness of the interaction but also become active partners in shaping and reinforcing social norms. Every interaction where a digital assistant "chooses" to slow its speech to sound polite reinforces the user's cultural association between slowness and politeness. This phenomenon transforms AI from a mere technological tool into an agent with a socializing influence.

This study has several limitations that provide a basis for future research. First, we used a single script and limited style prompts. Future studies could examine a wider variety of social contexts, scripts, and languages. Second, we focused on a single prosodic measure (speech rate). We selected speech rate as our primary variable as it has been identified as a particularly robust and consistent predictor of perceived politeness in human behavior, allowing for a direct replication and comparison with human baselines (Nussinson et al., 2026). However, as politeness is naturally conveyed through a complex combination of cues such as pitch, intonation, and volume (Brown & Prieto, 2017), the absence of a multi-parametric acoustic analysis remains a limitation. Future research should investigate whether AI systems also modulate these additional features to provide a more holistic understanding of their social mimicry. Another fascinating research direction would be to examine the gap between the system's "procedural" (behavioral) knowledge, as demonstrated here, and its "declarative" (stated) knowledge. In other words, it would be worthwhile to examine how AI explicitly responds to a question about the connection between speech rate and politeness. Beyond these specific avenues, the methodology presented here offers a robust framework for comparing the vocal capabilities of LLMs with human baselines, providing a valuable tool for future research in human-AI interaction.

In conclusion, this research demonstrates that generative AI is beginning to acquire abilities that go beyond mere language processing, adopting subtle features of human social behavior. Although this ability is likely the product of sophisticated mimicry rather than deep social understanding, it has significant implications for the future of human-machine interaction and for how we understand social learning processes. Machines are reflecting not only what we say, but also the unwritten rules of how we say it.

CRediT authorship contribution statement

Eyal Rabin: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Zohar Elyoseph:** Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Rotem Israel-Fishelson:** Writing – review & editing, Writing – original draft, Conceptualization. **Adi Dali:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Ravit Nussinson:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors utilized Google's Gemini and OpenAI's ChatGPT 4 to enhance the clarity and phrasing of the manuscript and to aid in structuring the content. Following the use of these services, the authors conducted a thorough review and editing process and assumed full responsibility for the final content of the

published article.

Funding

This research was supported by Israel Science Foundation Grant 693/22 to R.N.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

None.

References

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103–107). Sage.

Abdulrahman, A., & Richards, D. (2022). Is natural necessary? Human voice versus synthetic voice for intelligent virtual agents. *Multimodal Technologies and Interaction, 6*(7), Article 7. https://doi.org/10.3390/mti6070051

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*(5), 715–727. https://doi.org/10.1037/0022-3514.37.5.715

Barakat, H., Turk, O., & Demiroglu, C. (2024). Deep learning-based expressive speech synthesis: A systematic review of approaches, challenges, and resources. *EURASIP Journal on Audio Speech and Music Processing, 2024*(1), 11. https://doi.org/10.1186/s13636-024-00329-7

Beier, E., Cohn, M., Trammel, T., Ferreira, F., & Zellou, G. (2025). Marking prosodic prominence for voice assistant and human addressees. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 51*(6), 986–1003. https://doi.org/10.1037/xlm0001396

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, 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). https://doi.org/10.1145/3442188.3445922

Borsos, Z., Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., & Sharifi, M. (2023). AudioLM: A language modeling approach to audio generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 31*, 2523–2533. https://doi.org/10.1109/TASLP.2023.3288409

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Brown, L., & Prieto, P. (2017). (Im)politeness, prosody and gesture. In J. Culpeper, M. Haugh, & D. Kádár (Eds.), *The Palgrave handbook of linguistic (im)politeness* (pp. 357–379). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-37508-7_14.

Burton, N., & Gaskin, J. (2019). "Thank You, Siri": Politeness and intelligent digital assistants. In *Twenty-fifth americas conference on information systems, Cancun*.

Carolus, A., Augustin, Y., Markus, A., & Wienrich, C. (2023). Digital interaction literacy model – Conceptualizing competencies for literate interactions with voice-based AI systems. *Computers and Education: Artificial Intelligence, 4*. https://doi.org/10.1016/j.caeai.2022.100114

Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction, 39*(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Colby, S., & McMurray, B. (2021). Cognitive and physiological measures of listening effort during degraded speech perception: Relating dual-task and pupillometry paradigms. *Journal of Speech, Language, and Hearing Research, 64*(9), 3627–3652. https://doi.org/10.1044/2021_JSLHR-20-00583

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience, 30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Delogu, C., Conte, S., & Sementina, C. (1998). Cognitive factors in the evaluation of synthetic speech. *Speech Communication, 24*(2), 153–168. https://doi.org/10.1016/S0167-6393(98)00009-0

Do, T. D., McMahan, R. P., & Wisniewski, P. J. (2022). A new uncanny Valley? The effects of speech fidelity and human listener gender on social perceptions of a virtual-human speaker. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–11). https://doi.org/10.1145/3491102.3517564

Ehret, J., Bönsch, A., Aspöck, L., Röhr, C. T., Baumann, S., Grice, M., Fels, J., & Kuhlen, T. W. (2021). Do prosody and embodiment influence the perceived naturalness of conversational agents' speech? *ACM Transaction Applied Percept, 18*(4), 21:1–21:15. https://doi.org/10.1145/3486580

Fan, G., & Liu, D. (2025). When machines speak with feeling: Investigating emotional prosody, authenticity, and trust in AI vs. human voices. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 47) (0) https://escholarship.org/uc/item/8vr8s6h8.

Galdino, J. C., Matos, A. N., Svartman, F. R. F., & Aluisio, S. M. (2025). The evaluation of prosody in speech synthesis: A systematic review. *Journal of the Brazilian Computer Society, 31*(1), Article 1. https://doi.org/10.5753/jbcs.2025.5468

Herrmann, B. (2023). The perception of artificial-intelligence (AI) based synthesized speech in younger and older adults. *International Journal of Speech Technology, 26*(2), 395–415. https://doi.org/10.1007/s10772-023-10027-y

Hoegen, R., Aneja, D., McDuff, D., & Czerwinski, M. (2019). An end-to-end conversational style matching agent. In *Proceedings of the 19th ACM international conference on intelligent virtual agents* (pp. 111–118). https://doi.org/10.1145/3308532.3329473

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Hussen Abdelaziz, A., Kumar, A. P., Seivwright, C., Fanelli, G., Binder, J., Stylianou, Y., & Kajareker, S. (2021). Audiovisual speech synthesis using Tacotron2. In *Proceedings of the 2021 international conference on multimodal interaction* (pp. 503–511). https://doi.org/10.1145/3462244.3479883

Ju, Z., Wang, Y., Shen, K., Tan, X., Xin, D., Yang, D., Liu, Y., Leng, Y., Song, K., Tang, S., Wu, Z., Qin, T., Li, X.-Y., Ye, W., Zhang, S., Bian, J., He, L., Li, J., & Zhao, S. (2024). NaturalSpeech 3: Zero-shot speech synthesis with factorized codec and diffusion models. In *Proceedings of the 41st international conference on machine learning*. https://dl.acm.org/doi/10.5555/3692070.3692979.

Kühne, K., Fischer, M. H., & Zhou, Y. (2020). The human takes it all: Humanlike synthesized voices are perceived as less eerie and more likable. Evidence from a subjective ratings study. *Frontiers in Neurorobotics, 14*. https://doi.org/10.3389/fnbot.2020.593732

Lakomkin, E., Wu, C., Fathullah, Y., Kalinli, O., Seltzer, M. L., & Fuegen, C. (2024). End-to-End speech recognition contextualization with large language models. In *Icassp 2024 - 2024 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 12406–12410). https://doi.org/10.1109/ICASSP48485.2024.10446898

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Liu, R., Sisman, B., Gao, G., & Li, H. (2021). Expressive TTS training with frame and style reconstruction loss. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29*, 1806–1818. https://doi.org/10.1109/TASLP.2021.3076369

Luo, X. (2025). Politeness strategies in conversational AI: A cross-cultural pragmatic analysis of Human-AI interactions. *Pinnacle Academic Press Proceedings Series, 3*.

Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., & Poria, S. (2023). A review of deep learning techniques for speech processing. *Information Fusion, 99*, Article 101869. https://doi.org/10.1016/j.inffus.2023.101869

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics and Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Ning, Y., He, S., Wu, Z., Xing, C., & Zhang, L.-J. (2019). A review of deep learning based speech synthesis. *Applied Sciences, 9*(19), 4050. https://doi.org/10.3390/app9194050

Nussinson, R., Dahbash, C., Hatzek, A., Stephan, E., & Liberman, N. (2026). *An association between speech rate and politeness: A construal level theory approach*.

Nussinson, R., Rozenberg, I., Hatzek, A., Mentser, S., Navon, M., Gilead, M., Simchon, A., Sverdlik, N., & Liberman, N. (2024). The poetry of psychological distance: Bidirectional associations between stimulus speed and its psychological distance and construal level. *Journal of Personality and Social Psychology, 127*(1), 58–83.

Ofuka, E., McKeown, J. D., Waterman, M. G., & Roach, P. J. (2000). Prosodic cues for rated politeness in Japanese speech. *Speech Communication, 32*(3), 199–217. https://doi.org/10.1016/S0167-6393(00)00009-1

Rallabandi, S. S., Naderi, B., & Möller, S. (2021). Identifying the vocal cues of like ability, friendliness and skilfulness in synthetic speech. In *The 11th ISCA speech synthesis workshop (SSW 11)* (pp. 1–6).

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, UK: Cambridge University Press.

Ribino, P. (2023). The role of politeness in human–machine interactions: A systematic literature review and future perspectives. *Artificial Intelligence Review, 56*(1), 445–482. https://doi.org/10.1007/s10462-023-10540-1

Ross, A., Corley, M., & Lai, C. (2024). Is there an uncanny valley for speech?: Investigating listeners' evaluations of realistic synthesised voices. In *Proceedings of speech prosody 2024* (pp. 1115–1119). https://doi.org/10.21437/SpeechProsody.2024-225

Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in human–agent interaction: A survey. *ACM Computing Surveys, 54*(4), 81:1–81:43. https://doi.org/10.1145/3386867

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerrv-Ryan, R., Saurous, R. A., Agiomvrgiannakis, Y., & Wu, Y. (2018). Natural TTS synthesis by conditioning wavenet on MEL spectrogram predictions. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4779–4783). https://doi.org/10.1109/ICASSP.2018.8461368

Sigurgeirsson, A., & King, S. (2024). Controllable speaking styles using a large language model. In *Icassp 2024 - 2024 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 10851–10855). https://doi.org/10.1109/ICASSP48485.2024.10448400

Staszkiewicz, B. (2024). Speech rate correlates with politeness in Spanish offers. *Speech Prosody, 2024*, 767–771. https://doi.org/10.21437/SpeechProsody.2024-155

Stephan, E., Liberman, N., & Trope, Y. (2010). Politeness and psychological distance: A construal level perspective. *Journal of Personality and Social Psychology, 98*(2), 397–402. .

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Szekely, E., Higginbotham, J., & Possemato, F. (2024). Voice and choice: Investigating the role of prosodic variation in request compliance and perceived politeness using conversational TTS. In T. Kawahara, V. Demberg, S. Ultes, K. Inoue, S. Mehri, D. Howcroft, & K. Komatani (Eds.), *Proceedings of the 25th annual meeting of the special interest group on discourse and dialogue* (pp. 466–476). Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/2024.sigdial-1.40.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education, 220*, Article 105112. https://doi.org/10.1016/j.compedu.2024.105112

Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2023). Voice-based AI in call center customer service: A natural field experiment. *Production and Operations Management, 32*(4), 1002–1018. https://doi.org/10.1111/poms.13953

Wang, Y., Stanton, D., Zhang, Y., Ryan, R.-S., Battenberg, E., Shor, J., Xiao, Y., Jia, Y., Ren, F., & Saurous, R. A. (2018). Style tokens: Unsupervised style modeling, control and transfer in end-to-end speech synthesis. In *Proceedings of the 35th international conference on machine learning* (pp. 5180–5189). https://proceedings.mlr.press/v80/wang18h.html.

Yang, S., Huang, Y., Huang, X., Zhang, J., Meng, Z., & Yang, J. (2024). Impact of anthropomorphism in AI assistants' verbal feedback on task performance and emotional experience. *Ergonomics, 0*(0), 1–14. https://doi.org/10.1080/00140139.2025.2497072

Yusupova, S. (2025). Gender-based comparative analysis of respect in linguistic expression: A study of Uzbek, Japanese, English, and German. *Cogent Arts and Humanities, 12*(1), Article 2512789. https://doi.org/10.1080/23311983.2025.2512789

Eyal Rabin is a Research Fellow at the Research Center for Innovation in Learning Technologies, Faculty of Education and Psychology, The Open University of Israel, and Lead Researcher at the Institute for Applied Research in Artificial Intelligence in Education, Ministry of Education. His research focuses on the intersection of technology, pedagogy, and methodology.

Zohar Elyoseph is an Associate Professor (proposed rank) in the Department of Counseling and Human Development at the Faculty of Education, University of Haifa, and an Alon Fellow of the Israeli Council for Higher Education. His research investigates the intersection of generative artificial intelligence (GAI) with mental health and education. His work characterizes the clinical capabilities of AI, develops ethical-philosophical frameworks for human-agent relational dynamics, and creates innovative GAI-based clinical simulators.

Rotem Israel-Fishelson is a postdoctoral researcher in the Department of Teaching & Learning, Policy & Leadership at the College of Education, University of Maryland. Her research focuses on introducing learners to data science through engaging computational learning experiences.

Adi Dali is a master's student in the Clinical Neuropsychology program at the University of Haifa and a research assistant at the Department of Education and Psychology, the Open University of Israel.

Ravit Nussinson is a professor in the Department of Education and Psychology at the Open University of Israel. Her research focuses on the interface between mind and body as it is reflected in grounded cognition and in the effects of the behavioral immune system on information processing.