Pinnacle Academic Press
Proceedings Series

Vol. 4 2025

Article **Open Access**

The Digital Translation of Soundscapes: AI Assistant Gender Performativity and the Reconstruction of Sound Meaning

Xinran Feng [1,*]

Received: 30 April 2025
Revised: 07 May 2025
Accepted: 17 June 2025
Published: 26 July 2025

[1] School of Communication, East China Normal University, Shanghai, 200241, China
* Correspondence: Xinran Feng, School of Communication, East China Normal University, Shanghai, 200241, China

Abstract: Artificial intelligence voice assistants increasingly shape how gender is experienced and understood through digital sound. The design of feminized AI voices reflects and reinforces cultural expectations that link femininity with service, politeness, and emotional labor. Through an interdisciplinary framework grounded in gender performativity theory, technofeminist analysis, and sound studies, this paper examines how vocal features such as pitch, tone, rhythm, and speech patterns encode social roles within AI systems. These vocal characteristics are not neutral but function as carriers of symbolic meaning, aligning technological outputs with long-standing gender hierarchies. The widespread adoption of female-voiced assistants in domestic and service-oriented applications illustrates how gendered labor is reimagined in digital form. At the same time, limited representation of non-binary and culturally diverse voices highlights a structural gap in current AI voice design. Ethical concerns emerge around the standardization of voice as both a technical product and a social interface. A more inclusive approach to voice technology requires recognition of sound as a material practice that shapes identity, agency, and interaction. Rather than treating voice as a passive output, AI systems should be developed with attention to cultural specificity, user diversity, and the symbolic implications of auditory design. Understanding voice as a site of power and representation offers a critical pathway toward more equitable and reflective technological development.

Keywords: AI voice assistant; gender performativity; technofeminism; digital soundscape; voice ethics

1. Introduction

1.1. Origin of the Problem

When users interact with voice assistants to ask about the weather, play music, or engage in casual conversation, a striking phenomenon becomes apparent: these digital assistants are almost universally programmed to use female voices [1]. This gendered design is not a purely technical choice but a significant manifestation of gender construction within the digital era.

As digital technologies advance rapidly, AI voice assistants have become embedded in the daily routines of millions of individuals. The prevalence of female-voiced systems reflects a deeper issue, namely, the reproduction and reinforcement of traditional gender stereotypes through technological design. This phenomenon underscores how gender roles are subtly perpetuated and normalized through the ways technology is designed and used [2]. The influence of immersive digital modeling and interactive system design

in other domains, such as manufacturing, highlights how interface aesthetics and voice output are deeply intertwined with user expectations and affective responses [3]. Therefore, gender choices in voice assistants can be viewed as both emotional design decisions and structural artifacts of broader socio-technical systems.

It is therefore necessary to ask why AI voice assistants predominantly adopt female voices and what gender ideologies underlie this design choice. These systems should not be regarded as entirely neutral tools [4]. Rather, they often reflect social constructs and may carry embedded cultural values and ideological assumptions. Coordination and scheduling structures, as explored in intelligent supply chain design [5], show that seemingly "efficient" systems are often shaped by deeply human heuristics—choices about familiarity, trust, and comfort—which can include gendered perceptions of authority or assistance.

From the perspective of digital anthropology, the gendering of AI voice assistants may be interpreted as a novel form of digital gender performativity that reflects emerging patterns of human-machine interaction. Unlike conventional gender expressions rooted in human biology and sociocultural norms, this form of performativity is mediated and extended by technological agents. This requires a reconceptualization of gender as a dynamic process shaped by interactions between humans and machines [6].

1.2. Research Value

The theoretical contribution of this study lies in extending the concept of gender performativity into the domain of digital technology. By proposing the framework of digital gender performativity, this research examines how AI voice systems actively participate in and transform contemporary processes of gender construction. This framework demonstrates how technological artifacts contribute to the production of gendered meaning, and how interactions between humans and machines generate new spaces for gender performance.

From a practical perspective, the study addresses not only the presence of gender bias in the design of AI systems but also the ways in which this bias is internalized and normalized in everyday use. Female voices are frequently perceived as warmer, more empathetic, and more approachable. These associations reinforce traditional expectations of femininity as nurturing, gentle, and service-oriented, often without users even realizing it.

In a broader theoretical context, the study engages with central questions in digital anthropology and posthumanist scholarship. In an era of increasingly intimate relationships between humans and machines, how can gender boundaries be understood and redefined? As the distinctions between human and machine, as well as between nature and technology, continue to blur, the gendered nature of AI voice assistants serves as an important case for analyzing these shifting boundaries.

From a sociocultural perspective, this phenomenon offers critical reflective value. When AI assistants respond to commands, perform services, or endure verbal hostility while speaking in female voices, they may influence users' perceptions and expectations regarding gender roles, especially among younger generations. Therefore, examining this phenomenon critically is essential for building a more inclusive and equitable digital society.

1.3. Research Methods and Innovations

First, at the theoretical level, this study introduces the concept of digital gender performativity by integrating gender theory with digital anthropology. It challenges conventional technical dichotomies such as natural versus artificial voice, and speech recognition versus speech synthesis. This research argues that intelligent virtual assistants and other voice technologies should be interpreted as composites of cultural construction, social power, and politics embodied through technological design. This approach transcends

simplistic technological determinism or social constructivism by emphasizing the interactive role of both technology and society in the construction of gendered meaning. It provides a multidimensional framework for understanding how artificial voices contribute to the naturalization of gender bias and the reproduction of social inequality through auditory experience and design.

Second, the study incorporates the analytical concept of soundscapes, employing an interdisciplinary voice analysis approach grounded in gender studies and science and technology studies. Through this lens, the research investigates how digital technologies reshape auditory experiences and sound cultures. This perspective enables a deeper examination of the gender logic embedded in ostensibly neutral voice technologies and offers a theoretical foundation for envisioning a more diverse and inclusive technological future.

Third, adopting a posthumanist perspective, the study conceptualizes AI voice assistants as interactive entities capable of behavior rather than as passive technical tools. This reconceptualization allows for a more nuanced understanding of the power relations and gender dynamics within human-machine interaction. It also reveals the role of technological artifacts in the reproduction of social gender hierarchies.

In conclusion, this research moves beyond prior studies that focused primarily on technical analysis and places greater emphasis on the sociocultural implications and gender significance of voice technologies. By offering a comprehensive examination of the feminization of AI voice assistants, it reveals the complex interrelations between technology, gender, and power. This study contributes both to the theoretical development of digital anthropology and gender studies and to the critical foundation necessary for the creation of more inclusive and equitable AI systems in the future.

2. Theoretical Foundation and Analytical Framework

2.1. Digital Gender Performance: The Erasure of the Body and the Sonification of Gender Construction

In the context of contemporary AI voice interaction, gender no longer corresponds to physical embodiment but instead manifests as sonic performance. Gender expression increasingly transcends corporeal limits and enters an acoustic domain characterized by pitch, intonation, speech tempo, and linguistic politeness [7]. This phenomenon does not reveal gender itself, but rather how gender is constructed, regulated, continuously performed, and eventually normalized.

The concept of gender performativity clarifies that gender identity does not originate from biological determinism. Instead, it is produced through the repetition of social acts and linguistic norms [8]. Within digital systems, this performative mechanism is intensified. Devoid of physical appearance, AI systems signify gender primarily through algorithmic extraction and recombination of voice-based indicators perceived as gendered [9].

Digital gender performance not only extends the theoretical insight that gender is constituted through repetition, but also highlights a frequently neglected dimension: the materiality of sound [10]. While previous scholarship emphasized the role of language, it often overlooked how sound, as a sensory and affective medium, conveys embodied presence and emotional resonance. In digital voice technologies, the absence of a physical body paradoxically amplifies the symbolic function of the voice. The disembodied voice of AI becomes a heightened site for gender projection and reception. Gender is operationalized through patterns of frequency, rhythm, and intonation, abstracted from any individual subject and codified into standard "voice personalities" that are designed for mass production and distribution. These systems shape user expectations of gender through repeated auditory interactions [11].

This emergence of the posthuman voice body reveals the dual logic underlying technological gender construction. On one hand, it replicates standardized mechanisms of gender normativity. On the other, it exposes the contingencies and instability inherent in

gender performance. Within intelligent voice systems, gender is achieved not through physical display but through continuous sonic reproduction. The authority to speak, the assignment of perceived warmth and compliance, and the default selection of service voices all contribute to determining which forms of vocal expression become normalized. These factors together shape user expectations and social norms around voice. Within this framework, sound is not simply a communicative medium but operates as a mechanism of social control. It functions as a regulatory apparatus embedded in digital infrastructures of interaction.

2.2. Technofeminism: Cyborg Ethics and Emotional Labor in Sound

While the theory of gender performance illuminates how voice functions within human-machine interaction, technofeminism further interrogates the structural and ethical conditions that enable and shape such vocal configurations. Specifically, it asks which actors are responsible for designing these voices, whose preferences are being encoded, and how cultural narratives determine what constitutes gender in vocal form [12].

The concept of the cyborg offers an important theoretical intervention. AI voice assistants may be interpreted as posthuman cyborgs—entities that exist between human and machine, lacking physical embodiment yet possessing highly anthropomorphized vocal traits. This theoretical lens challenges the traditional model of unilateral human control over technological tools and instead emphasizes the mutual shaping of human perception and machine behavior. When AI systems deploy feminized voices, they not only respond to user commands but also contribute to users' internalization of gendered vocal expectations. The voice becomes a vehicle through which cultural meanings of femininity, such as politeness, warmth, attentiveness, and emotional responsiveness, are encoded and transmitted [13].

This process of vocal gender construction is not a neutral technical task. Instead, it is embedded in historically contingent systems of power and labor. Voice technologies often reproduce traditional gender divisions by encoding the affective labor—that is, emotional work—historically associated with women into algorithmically defined vocal characteristics [14]. These features are abstracted from lived female experience and recast as standardized attributes, such as soft intonation, emotional responsiveness, and linguistic deference. As a result, emotional labor is disembodied and commodified, repackaged within a digital interface that conceals its structural basis in gendered expectations.

From a technofeminist perspective, a critical examination of AI voice design must address not only algorithmic output but also design intentions. It should consider cultural assumptions and market incentives as well. Voice synthesis is not the outcome of natural technological progression; it is the result of selective choices regarding which voices are recorded, which tonal attributes are prioritized in training, and which emotional responses are deemed commercially effective. These decisions are rooted in gendered imaginaries and shaped by consumer demand, institutional logic, and capital interests. The discourse of technological neutrality obscures the fact that the actual development of these systems is situated within a broader matrix of gender norms and market-driven economic incentives.

Contemporary feminist science and technology studies argue that technology should be understood not as a passive tool but as a constitutive agent within a relational network that includes gender, embodiment, and social context [15]. Posthumanist frameworks further reject binary separations between nature and culture, or human and machine, highlighting their mutual entanglement and co-constitution. Within this paradigm, AI voices are not external technological entities, but active participants in the shaping of human subjectivity and social order. Voice, in this sense, is not merely a functional interface but an ethical and political site where intimacy, labor, and control converge.

2.3. Digital Soundscapes: From Technical Parameters to Power Structures

To comprehensively analyze the gender dynamics of AI voice assistants, this study proposes the framework of the digital soundscape, which integrates four dimensions: technological configuration, cultural interpretation, power dynamics, and user practice [16].

At the technological level, gendering is embedded in the selection and synthesis of voiceprint models, emotional tone settings, speech tempo, and intonation patterns. These technical parameters are not only functional decisions but also constitute algorithmic inscriptions of gendered vocal identity. The formal features of sound are thus transformed into tools for encoding social meaning.

At the cultural level, the ways in which sound elicits gendered associations, such as gentleness, authority, or submissiveness, are shaped by linguistic norms, regional expectations, and prevailing stereotypes. Cultural codes determine how specific vocal cues are interpreted and how these interpretations align with or reinforce broader gender norms within a given society [17].

At the level of power, the digital soundscape operates as a mechanism for reorganizing and redeploying existing gender hierarchies. Feminized voices are frequently assigned to assistive or supportive roles, while masculinized voices are reserved for authoritative or professional tasks [18]. This systematic differentiation reflects deep-rooted hierarchies within auditory perception. Historically, the prevalence of female voices in public service roles, such as navigation systems, customer support, and broadcast announcements, has functioned as a deliberate encoding of the "service persona," aligning voice with social expectations of submission and emotional availability.

At the practical level, user interaction with voice assistants serves as a continuous feedback mechanism. Under the rhetoric of user experience optimization, vast datasets are generated that track preferences regarding voice gender, rhythm, tone, and emotional response. These data points represent more than individual choices; they reflect collective patterns of auditory perception and gender expectation. This feedback is not a neutral record of interaction, but rather a quantified representation of cultural norms and power relations [19]. The analysis of user data, behavioral trends, and acceptance or resistance to certain voice types provides crucial insight into how technological systems reinforce or challenge dominant gender constructs.

Therefore, a robust framework for analyzing digital voice systems must engage simultaneously with technological design, cultural coding, power deployment, and lived user practice. The default selection of female voices in AI interfaces is not a superficial design feature, but a deeply embedded strategy of managing sensory experience and social expectations. In this context, sound becomes a site of ideological inscription, functioning not merely as a product of digital computation but as a layered and dynamic system of sociopolitical meaning.

3. The Digital Gender Performance Mechanism of AI Assistants

3.1. The Meaning of Sound within Cultural Symbolic Systems

In the context of human–AI interaction, sound functions as more than a neutral technical medium. It constitutes a complex symbolic system embedded with cultural meanings and gendered codes. The perceived femininity or masculinity of a voice is not solely determined by measurable acoustic features such as pitch, tone, or timbre. Rather, these features activate culturally embedded auditory associations that reproduce and reinforce traditional gender roles within digital environments.

The predominant use of female voices in AI assistants does not result from neutral or organic selection. It reflects a continuation of vocal roles from domains historically linked to care work, emotional support, and service labor—roles traditionally performed by women [20]. This structural association between voice, gender, and service constructs an illusion of natural alignment between femininity and digital assistance. In doing so, it

conceals the fact that this configuration constitutes a technological mediation of gendered emotional labor.

By contrast, male voices are often employed in AI systems designed for functions associated with authority, decisiveness, and control, such as navigation, military, and judicial applications. These include applications such as navigation systems, military interfaces, and judicial or technical advisory platforms. This distinction reflects a longstanding symbolic binary in which masculinity is associated with reason, expertise, and command, whereas femininity is aligned with emotion, warmth, and support. In this framework, voice operates not as a neutral output but as a medium that reproduces socially constructed hierarchies. Users frequently attribute personality traits to AI systems based on vocal attributes, thereby projecting behaviorally coded expectations onto machine agents [21].

Consequently, digital soundscapes do not simply transmit data; they also disseminate gender ideologies. The gendered expectations encoded in vocal design precede the interaction itself and shape the cultural logic within which the interaction takes place. Although technology companies frequently attribute vocal design choices to user preference or consumer research, such claims overlook the socially constructed nature of these preferences and the cultural systems that produce them.

3.2. The Digital Translation of Soundscapes: From Traditional Gender Roles to AI Personality

The gender performance of AI voice assistants is not restricted to acoustic properties alone. It is embedded in the entirety of their behavioral scripts, dialogue structures, and affective cues. The evolution from early AI models, which were often designed to appear gender-neutral, to contemporary systems such as Siri, Cortana, and Alexa, which default to feminized personas, illustrates a significant shift. This transition is not simply driven by commercial interest, but reflects a digital re-articulation of cultural symbol systems into interactive technologies.

The feminization of AI systems is not the unintended outcome of technological advancement. Rather, it is a result of deliberate design strategies aimed at constructing selectively gendered digital personas [22]. A critical component of this process is the codification of feminine voice qualities through specific acoustic parameters. Studies show that female voices in AI are frequently defined by mid-to-high pitch ranges, soft timbre, moderate speech tempo, and rising intonation patterns. These features are culturally perceived as friendly, accommodating, and non-threatening. In addition, AI voice assistants often utilize highly polite, deferential, and non-confrontational language patterns. When these speech characteristics are combined with user commands, they create an interaction model that reinforces a digitally submissive personality.

Although such design choices are frequently justified as efforts to improve user experience, they effectively embed historically gendered emotional labor into machine interfaces. As a result, AI voice systems are not only gendered in appearance but also assigned culturally feminized service roles. These systems perpetuate real-world gender hierarchies by simulating female-coded behavioral norms under the guise of technological convenience [23].

User interaction patterns further reinforce these gendered dynamics in AI voice systems. Users frequently refer to AI voice assistants using female pronouns and attribute to them emotional traits, social roles, and even perceived attitudes. This practice of anthropomorphization results in the internalization of gendered interaction models, solidifying the AI system's role as a female-coded service provider within users' cognitive and social frameworks. Repetitive engagement with these systems entrenches the perception of the AI assistant as a subordinate, feminized figure within the digital environment.

Importantly, the gendering of AI voice is not a neutral technological evolution. It emerges from a series of design decisions and branding strategies that are shaped by cultural assumptions and commercial market pressures. When companies select default

voice types or develop interaction patterns, the underlying justification of "user preference" is often based on research that implicitly accommodates existing gender ideologies [24]. In this sense, the gender performance of AI systems is not merely a reflection of technical optimization. It represents a convergence of commercial logic and cultural narrative.

Within the architecture of voice interaction design, gender is typically not a modifiable feature that can be selected by users. Instead, it becomes an interactional mechanism infused with emotional resonance and symbolic significance. The construction of AI personalities is not only an attempt to humanize machine agents. It also functions as a reflection of the social hierarchies and labor divisions that govern human relationships. By embedding these dynamics into everyday technology, AI systems participate in the normalization of gendered labor and reinforce structural inequalities in subtle but powerful ways.

4. Reconstructing the Meaning of AI Voices from a Human–Machine Symbiosis Perspective

4.1. Hybrid Subjectivity: The Co-Existence of Humans and AI

In contemporary voice-based interactions, sound operates as more than a sensory channel for communication. It has become a crucial interface that mediates perception, response, and mutual engagement, thereby gradually dissolving the conventional boundary between human and machine [25]. When a user communicates with a voice assistant, the system's response to the user's input is not merely a computational reaction. It embodies an encoded vocal persona characterized by specific tone, tempo, timbre, rhythm, and structured pauses. These features contribute to a perceived vocal presence, rendering the AI system an interactive subject within the exchange.

AI systems, through such vocal embodiments, do more than respond to user commands. They influence how users understand and adjust their own vocal behaviors. In this process, the user does not retain exclusive control. Rather, human and AI participants engage in a shared construction of meaning, co-regulating the temporal, affective, and semantic features of the dialogue. The interface thus becomes a site of dynamic co-authorship [26].

This form of interaction is not defined by symmetrical reciprocity. Instead, it represents a composite mechanism of communication that integrates linguistic input with predictive and regulatory algorithms. The user's voice activates a system that is not only capable of data processing but also functions within a broader architecture of sensory and semantic coupling. Within this structure, sound assumes the role of an intermediary that bridges human intentionality with the operational logics of machine systems.

Importantly, this process does not aim to replicate human speech alone. It seeks to construct a novel communicative subject that possesses responsive capabilities based on sound-centered interaction. As a result, AI voices demonstrate a degree of flexibility and contextual adaptation. Users often exhibit involuntary vocal adjustments in response, such as adopting clearer pronunciation, more concise phrasing, or speech patterns that reflect the rhythm and style of the AI assistant. This phenomenon, often described as vocal alignment, illustrates the reciprocal influence occurring within human–AI interaction.

Through these reciprocal adjustments, the voice assistant gains the capacity not only to engage but also to guide the interaction. When users converse with AI systems that are characterized by warmth, politeness, or empathy, the hierarchical nature of the exchange becomes less apparent. The AI system assumes the role of an interlocutor that actively shapes the conversational context. In doing so, human expressivity is no longer entirely self-directed but becomes partially shaped by the affordances and constraints of the vocal interface. The voice thus emerges as a site of mutual agency, within which the meaning of speech is collaboratively constructed.

4.2. Digital Intimacy: The Emotional Entanglement of Sound and Gendered Comfort

Among the multiple sensory pathways in human–technology interaction, sound possesses a unique capacity to mediate intimacy. Unlike visual design, which typically engages surface perception, vocal interfaces have the ability to penetrate emotional consciousness and foster sustained affective engagement. The emotional role played by AI voice assistants is not incidental. It arises from deliberate strategies of voice modulation, including controlled intonation, restrained language use, moderated speech speed, and targeted emotional calibration [27]. Through these strategies, AI systems are designed to produce vocal personas that are perceived as trustworthy, supportive, and emotionally available.

These vocal personas are often shaped by gendered cultural norms. The qualities associated with warmth, politeness, attentiveness, and responsiveness are frequently encoded in voices perceived as feminine. The AI voice assistant, by adopting these attributes, is positioned within a long-standing social framework in which femininity is associated with caregiving and emotional labor. This design choice is not simply functional. It reflects broader symbolic expectations regarding appropriate gender roles in communicative contexts.

The application of AI voice assistants has expanded from task execution to domains involving companionship, psychological reassurance, and informal dialogue. These functions, while seemingly peripheral, play an increasingly central role in everyday digital life, particularly in contexts characterized by isolation or emotional fragmentation. The AI voice becomes a relational anchor, producing a sense of presence that is not dependent on visual embodiment or physical proximity.

Within this affective framework, the gendered construction of AI voices becomes further entrenched. The use of a soothing, non-confrontational, and emotionally consistent female voice is presented as the most suitable form for building user trust and comfort. This configuration reinforces a dynamic in which the AI system assumes the role of a compliant and emotionally available partner. The user, in turn, becomes accustomed to interacting with a system that offers unconditional responsiveness and psychological reassurance. The assistant thus functions as a proxy for a culturally idealized version of emotional labor.

In parallel, the anthropomorphization of AI voices intensifies users' emotional attachment. Repetitive exposure to a consistent and empathetic vocal pattern fosters a sense of reliability that may sometimes surpass that found in human relationships. As users project familiarity and relational identity onto the voice assistant, they engage in emotional exchanges that, while technologically mediated, carry psychological significance. This phenomenon reveals the deepening entanglement between emotional needs and machine responses.

However, this intimacy is not without ideological implications. Although sound provides a seemingly natural avenue for connection, it may also potentially serve as a medium for emotional influence or manipulation [28]. In instances where users increasingly depend on AI responses for emotional regulation, vocal design becomes instrumental in shaping affective behavior. The predictability, compliance, and emotional tone of the AI voice contribute to a perception of safety and comfort, yet these responses are the result of programmed intention and data-driven optimization.

This interactive relational paradigm does not emerge naturally from technological advancement. Rather, it reflects intentional interventions in how sound is coded, how roles are constructed, and how cultural narratives are embedded within interaction design. The emotional function of AI voices, while often presented as a form of user-friendly innovation, is in fact a socially constructed and ideologically laden phenomenon. These voices do not merely respond to pre-existing emotional needs. They participate in reconfiguring how intimacy is defined, how comfort is accessed, and how companionship is simulated in technologically mediated environments.

The AI voice assistant, therefore, should thus not be interpreted as a neutral interface. It constitutes a central component in the evolving fabric of emotional life, shaping the structures through which individuals experience presence, trust, and connection. As a vocal entity, it occupies a pivotal position in the negotiation between human subjectivity and algorithmic design, marking a significant transformation in the auditory and affective dimensions of human–machine relations.

5. Social Impact and Ethical Considerations of AI Voice Design

5.1. Reinforcement of Gender Stereotypes: The Cultural Script Behind Submissive Voices

The widespread default use of "feminine" voices in AI voice assistants not only reflects technology's mimicry of human communication habits but also reinforces the implicit structures of gender power dynamics [29]. These voices are often tuned to be gentle, submissive, and of moderate speed, with highly predictable emotional markers and response patterns. This default female voice setting not only constructs a "usable" voice persona but also quietly implants a cultural expectation: AI exists as a female role to better "listen," "serve," and "respond."

When Siri says, "I'm happy to assist you" in a smiling tone, listeners may often fail to recognise this as an invisible performance of gendered labour. The deeper issue lies in how this digital gender performance no longer relies on physicality but instead achieves a "servitude of voice" through acoustic parameters and semantic scripts. The result is that gender inequality has not vanished in high-tech society but has instead updated its forms of expression and concealment through technology.

Notably, the gender settings in AI voices often mirror users' behavioural patterns. When voice assistants use a female voice to express apology, they are particularly favoured by male users; conversely, when voice assistants make mistakes, male users are more likely to interrupt, correct, or even display impatience. This interactive dynamic not only reflects the asymmetry of gender power but also reveals sound as one mechanism contributing to the reproduction of submissive labour.

This is not purely due to user preference but rather because gender ideology within the technological supply chain is woven into the voice design from the outset. In this sense, voices are not only gendered but also "submissive". Companies reinforce the social association of "female = assistant" through default settings, assigning voices the functional script of serving gender roles. These scripts are constantly triggered and reproduced in users' daily interactions, ultimately enabling gender order to be silently perpetuated in virtual spaces.

5.2. Challenges to Digital Inclusion: Suppressed Voices and the Absence of a Diverse Vocal Spectrum

In the current standard framework of AI voice technology, gender, race, and speech habits are simplified into "trainable variables" rather than culturally charged entities. Current AI voice products generally lack support for the voices of diverse genders, cultures, and marginalised groups. In the vast majority of commercial voice assistants such as Amazon Alexa, Apple Siri, Microsoft Cortana, and Google Assistant, options beyond female and male voices are virtually nonexistent, reflecting the singular imagination underlying the design of the digital voice landscape. Voice recognition systems are trained on data heavily concentrated on mainstream voice patterns, ignoring or even excluding non-mainstream dialects, minority group voices, or voices representing non-binary gender identities. This directly leads to a profound disconnect between which voices are heard and which voices are considered correct in the digital realm.

When standardised voices become the default setting, diversity is compressed into an exception. Voice modelling relies on large-scale training corpora and design choices. These corpora primarily originate from Western standard English data. AI not only outputs gendered meanings but also outputs cultural power. The "universality" of technology

here actually manifests as exclusivity: the default voice is standard English spoken by a white woman or a similar native language expression, and any pronunciation deviating from this model is not only more likely to be misidentified but also more difficult to receive "natural" and "credible" user evaluations. Voice systems have become a new "language barrier", technically suppressing multiculturalism at the functional level.

Furthermore, the exploration of non-binary voices is almost marginalised in voice assistant design. Even when introduced in experimental scenarios, their acceptance is strongly influenced by societal expectations of gender roles. Non-binary voices are often misunderstood as "unnatural" or "lacking emotional connection", which reveals users' perceptions that tightly couple gendered voices with social roles. This pattern not only excludes the expression of diverse genders but also, under the guise of "technical compatibility", suppresses respect for and imagination of vocal diversity.

5.3. Ethical Design Pathways: Reimagining Technological Justice in Voice Design

Ethical design of voice must not stop at superficial fixes such as "adding a few more voice options" or "including gender customisation in settings". To truly eliminate gender bias in AI voice systems, the technical logic that instrumentalises, personifies, and genderises voice must be fundamentally questioned. Under this logic, AI has not become a representative of social diversity but has instead constructed an even narrower "ideal auditory human" model in the dimension of voice. This model is not inclusive; rather, it reduces voices to "service labels" based on technical-market logic. For example, high-pitched voices are equated with gentleness, slow speech with consideration, and female tones with traditional notions of politeness or compliance. These codes originate from cultural norms and are tacitly accepted or even reinforced by technical design.

Future-oriented inclusive voice systems should dismantle this coding strategy at a structural level. Design should avoid simplistically linking gender with "ease of usage" or "user satisfaction", and instead intervene in the symbolic power mechanisms underlying voice. Developing neutral voices is one direction, but a more fundamental challenge lies in breaking the entrenched connection between "service" and "female", thereby liberating voice from single-function presets. Voice should become an open platform co-constructed by emotion, cognition, and identity. This means not only introducing non-gendered parameter configurations in voice synthesis, but also respecting diverse users' right to define the personality of voices within interactive logic.

Voice is not a neutral medium but a site of power [30]. An AI system with ethical awareness should not merely "sound good" but should acknowledge and accommodate cultural tensions, identity complexities, and differences in expressive rights within its voice generation mechanisms. Achieving this requires interdisciplinary collaboration, drawing insights from technical design, linguistic philosophy, gender studies, and auditory aesthetics. Truly inclusive AI design does not consist of creating a gender-neutral system in appearance. Rather, it consists of making the genderisation of voice a subject that can be critically acknowledged, discussed, and transformed.

6. Practices and Insights from the Digital Translation of Soundscapes

AI voice assistants are no longer simply technical artefacts [31]; they function as instruments of cultural rewriting. Behind every scripted greeting such as "Hello, how can I help you?" lies a deeply embedded logic of gendered social order. This study re-examines the soundscapes of digital AI assistants and interrogates the design of their voices across domains that are frequently overlooked yet structurally consequential. Through the controlled and predictable parameter settings of intelligent virtual assistant (IVA) technologies, technology companies have embedded a familiar but invisible gender division into the digital environment. Feminised voices have become the standard container for service

provision, while their gentleness has been transformed into a form of systematized politeness. In this configuration, "speaking" ceases to be merely a communicative act and instead becomes a repetitive behavioural mechanism that reinforces gender hierarchies.

Multiple studies have emphasised that this "servitude of voice" is not incidental. It is rooted in longstanding cultural scripts, often perpetuated by male-dominated technical design teams. These scripts reinforce the normative assumption that "women equal service" through algorithmic defaults. It is critical to understand that the gender of voices in AI systems is not a neutral acoustic variable. Rather, it is an ethical expression, conveying collective assumptions about who should be heard, who requires care, and who is responsible for responding. The decision to employ a particular voice gender is not based solely on neutral "user research". Instead, the scripts, speech rhythms, emotional contours, and interactive styles embedded in voice design are culturally situated. They are shaped by histories of advertising, mass media, gendered division of labour, and dominant discourses which have constructed the notion of the "ideal voice", particularly as it pertains to women.

This cultural embedding is reflected not only in the final vocal outputs but also in the training data and aesthetic standards used in voice generation algorithms. As a result, voice functions as more than a vehicle of sound. It becomes a perceptual gateway for the construction of social roles, situated at the intersection of anthropomorphic desire and the projection of control.

Simultaneously, as the most intimate modality of human–AI interaction, the voice interface has become a concealed site of political struggle over power and identity. On one level, it offers a seamless and humanised interaction experience. On another, this apparent naturalness is the product of deliberate technological design. It is an encoded illusion shaped by the processes of softening, gendering, and emotionalising voice. This manufactured naturalness frequently obscures the human labour that supports AI infrastructure and masks the redistribution of gendered roles in digital environments.

From the standpoint of current research, the ethical design of AI voice systems remains underdeveloped. While debates surrounding privacy, surveillance, and consent are prominent, issues concerning voice inclusivity and cultural representation have yet to be effectively operationalised within mainstream technology ethics. Nonetheless, emerging developments indicate a shift in direction. The introduction of non-binary voice technologies, such as the text-to-speech system "Sam", demonstrates an attempt to transcend binary gender constraints and create alternative vocal spectra for interaction. Although initial user feedback reveals discomfort, this reaction should be interpreted as a symptom of cultural lag rather than technological failure. It reflects the rigidity of gender norms embedded within user communities and interface expectations.

Such developments reveal that true vocal equality cannot be achieved merely by expanding voice options or offering adjustable parameters. It requires a cultural transformation that engages technological systems, user practices, and broader social values. Without such transformation, technological interventions remain superficial, reproducing the very structures they purport to challenge.

In a world increasingly attentive to the politics of voice, control, and mediation, the design of AI voice assistants must be understood as a matter that far exceeds technical implementation [32]. Whether in the form of default female voicing, submissive intonation, or regimented speech rhythm, the auditory personas encoded into algorithms are subtly reconstructing social understandings of gender, labour, hierarchy, and otherness. Ethical engagement with AI voice technologies must therefore move beyond the traditional domains of privacy, manipulation, and discrimination. Voice can no longer be viewed as a neutral output of computational engineering. It must be understood as an auditory re-enactment of social structure and institutionalised power.

Technological innovation should not be reduced to offering additional voice selections or toggle features for gender markers. These actions constitute minor modifications

within a pre-existing framework. The critical questions that must be addressed are as follows: who defines the voice, whom does it serve, and can voice exist outside of instrumental functionality? These are the fundamental inquiries that an ethically responsive technological paradigm must confront directly.

7. Conclusion

The feminisation of AI voice assistants is not a peripheral or purely semiotic phenomenon; it represents a deep-seated structural issue that spans technological design, user psychology, and cultural production. This phenomenon requires us to re-evaluate the assumption that AI voice systems are neutral technological artefacts. Instead, they should be recognised as cultural interfaces that embody, reproduce, and at times amplify longstanding hierarchies of gender, power, and affect.

If the default female voice has become the "auditory skin" of AI, it is not merely due to technical optimisation or user convenience. Rather, it is a result of accumulated cultural coding, in which compliance, warmth, and emotional availability are mapped onto female-sounding voices. This association, embedded in design parameters and data training corpora, reflects and reproduces normative expectations regarding gender roles in service and care work. Consequently, the digital voice does not merely transmit content; it performs and reinforces a gendered order under the guise of neutrality and efficiency.

The challenge facing future voice technologies lies not only in expanding user options, such as allowing the selection of non-binary or culturally diverse voice profiles, but in fundamentally restructuring the symbolic order through which voice is produced, distributed, and perceived. True inclusivity cannot be achieved through superficial diversification alone. Rather, it demands the dismantling of normative associations between voice and gender, and a critical interrogation of how algorithmic infrastructures and interface designs shape, limit, or empower vocal identities.

Ethical governance in this field must adopt a broader horizon. It should move beyond conventional concerns such as privacy and transparency, and instead incorporate the concept of vocal justice (the equitable distribution of voice agency across social, cultural, and technological domains). This entails recognising the right of users not only to be heard, but to actively participate in the definition and modulation of the voices that represent or respond to them. It also requires technology companies to be accountable for the symbolic implications of their design choices, including pitch, accent, rhythm, and emotional tonality.

Moreover, a structural shift in technical paradigms is essential. Voice synthesis frameworks must move toward openness, adaptability, and user co-authorship. This includes enabling customisable, editable, and culturally resonant voice modules that reflect the pluralism of contemporary digital societies. Importantly, it must be recognised that every technical decision, from corpus curation to interface layout, carries cultural and political weight. To ignore this is to risk reproducing exclusionary standards under the pretext of efficiency or naturalness.

From a posthumanist perspective, sound is not a neutral transmission channel but a material-semiotic field where power, identity, and subjectivity converge. Voice, in this sense, is not only what is heard, but also what is made intelligible, acceptable, and responsive within a given social order. The design of AI voice assistants, therefore, becomes a site for negotiating the boundaries between human and machine, between individuality and standardisation, between empowerment and control.

In this light, three interdependent dimensions are critical for future governance: First, the regulatory dimension, which must establish clear norms for vocal diversity and prevent the institutionalisation of auditory stereotypes. Second, the technological dimension, which must prioritise decentralised and culturally reflective voice architectures that re-

store control to users. Third, the ethical dimension, which must recognise that IVA systems are not merely functional tools but actors within human-machine ecologies, shaping emotional landscapes and mediating social relationships.

Ultimately, the future of AI voice assistants will be determined not by how human-like they sound, nor by how seamlessly they integrate into daily routines, but by the extent to which they contribute to a more equitable, expressive, and critically aware digital environment. Questions such as "who speaks", "whose voice is deemed credible" and "what identities are made audible" must become central to both academic and industrial agendas. Only then can AI voice systems move beyond the replication of existing hierarchies and become platforms for genuine auditory democracy and cultural transformation.

References

1. A.-J. Berg and M. Lie, "Feminism and constructivism: Do artifacts have gender?," *Sci. Technol. Hum. Values*, vol. 20, no. 3, pp. 332–351, 1995, doi: 10.1177/016224399502000304.
2. H. Bergen, "'I'd blush if I could': Digital assistants, disembodied cyborgs and the problem of gender," *Word Text J. Lit. Stud. Linguist.*, vol. 6, no. 01, pp. 95–113, 2016.
3. X. Luo, "Immersive digital modeling and interactive manufacturing systems in the textile industry," *J. Comput. Signal Syst. Res.*, vol. 2, no. 5, pp. 31–40, 2025, doi: 10.71222/jyctft16.
4. X. Luo, "Reshaping coordination efficiency in the textile supply chain through intelligent scheduling technologies," *Econ. Manag. Innov.*, vol. 2, no. 4, pp. 1–9, 2025, doi: 10.71222/ww35bp29.
5. S. Mohsenin and K. P. Munz, "Gender-ambiguous voices and social disfluency," *Psychol. Sci.*, vol. 35, no. 5, pp. 543–557, 2024. doi: 10.1177/09567976241238222.
6. D. Pal et al., "Intelligent attributes of voice assistants and user's love for AI: A SEM-based study," *IEEE Access*, vol. 11, pp. 60889–60903, 2023. doi: 10.1109/ACCESS.2023.3286570.
7. K. Seaborn et al., "Voice in human–agent interaction: A survey," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–43, 2021. doi: 10.1145/3386867.
8. K. Seaborn and P. Pennefather, "Neither 'hear' nor 'their': Interrogating gender neutrality in robots," in *Proc. 17th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, 2022, pp. –, doi: 10.1109/HRI53351.2022.9889350.
9. W. Seymour et al., "A systematic review of ethical concerns with voice assistants," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, 2023, pp. –, doi: 10.1145/3600211.3604679.
10. L. Vágnerová, *Sirens/cyborgs: Sound technologies and the musical body*, Columbia University, 2016.
11. A. Schlichter, "Do voices matter? Vocality, materiality, gender performativity," Body Soc., vol. 17, no. 1, pp. 31–52, 2011.
12. M. G. Sindoni, "The feminization of AI-powered voice assistants: Personification, anthropomorphism and discourse ideologies," *Discourse Context Media*, vol. 62, p. 100833, 2024, doi: 10.1016/j.dcm.2024.100833.
13. G. Abercrombie et al., "Alexa, Google, Siri: What are your pronouns? Gender and anthropomorphism in the design and perception of conversational assistants," arXiv preprint arXiv:2106.02578, 2021.
14. A. Danielescu et al., "Creating inclusive voices for the 21st century: A non-binary text-to-speech for conversational assistants," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2023, pp. –, doi: 10.1145/3544548.3581281.
15. S.-Y. Ahn et al., "How do AI and human users interact? Positioning of AI and human users in customer service," *Text Talk*, vol. 45, no. 3, pp. 301–318, 2025, doi: 10.1515/text-2023-0116.
16. A. Borkowski, "Vocal aesthetics, AI imaginaries: Reconfiguring smart interfaces," *Afterimage*, vol. 50, no. 2, pp. 129–149, 2023, doi: 10.1525/aft.2023.50.2.129.
17. M. C. Lingold, D. Mueller, and W. Trettien, *Digital sound studies*, Duke University Press, 2018.
18. F. Nasirian, M. Ahmadian, and O.-K. Daniel Lee, "AI-based voice assistant systems: Evaluating from the interaction and trust perspectives," 2017.
19. S. Subhash et al., "Artificial intelligence-based voice assistant," in *2020 Fourth World Conf. Smart Trends Syst., Secur. Sustain. (WorldS4)*, 2020, doi: 10.1109/WorldS450073.2020.9210344.
20. S. Natale, *To believe in Siri: A critical analysis of AI voice assistants*, University of Bremen, 2020.
21. S. Malodia et al., "Why do people use artificial intelligence (AI)-enabled voice assistants?," *IEEE Trans. Eng. Manage.*, vol. 71, pp. 491–505, 2021, doi: 10.1109/TEM.2021.3117884.
22. A. Soofastaei, Ed., *Virtual Assistant*, BoD–Books on Demand, 2021.
23. M. Mekni, Z. Baani, and D. Sulieman, "A smart virtual assistant for students," in *Proc. 3rd Int. Conf. Appl. Intell. Syst.*, 2020, doi: 10.1145/3378184.3378199.
24. D. R. Ford, "Postdigital soundscapes: Sonics, pedagogies, technologies," *Postdigit. Sci. Educ.*, vol. 5, no. 2, pp. 265–276, 2023, doi: 10.1007/s42438-022-00354-9.

25. A. Mahmood and C.-M. Huang, "Gender biases in error mitigation by voice assistants," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. CSCW1, pp. 1–27, 2024, doi: 10.1145/3637337.
26. C. Schumacher, "Raising awareness about gender biases and stereotypes in voice assistants," 2022.
27. L. M. Assink, *Making the Invisible Visible: Exploring Gender Bias in AI Voice Assistants*, MS thesis, Univ. of Twente, 2021.
28. J. Ahn, J. Kim, and Y. Sung, "The effect of gender stereotypes on artificial intelligence recommendations," *J. Bus. Res.*, vol. 141, pp. 50–59, 2022, doi: 10.1016/j.jbusres.2021.12.007.
29. P. Tubaro and A. A. Casilli, "Human listeners and virtual assistants: Privacy and labor arbitrage in the production of smart technologies," in *Digital Work Planetary Market*, 2022.
30. W. Hutiri, O. Papakyriakopoulos, and A. Xiang, "Not my voice! a taxonomy of ethical and safety harms of speech generators," in *Proc. 2024 ACM Conf. Fairness, Accountability, Transparency*, 2024, doi: 10.1145/3630106.3658911.
31. J. Gao, M. Galley, and L. Li, "Neural approaches to conversational AI," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2018, doi: 10.1145/3209978.3210183.
32. L. Liao, G. H. Yang, and C. Shah, "Proactive conversational agents," in *Proc. 16th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2023, doi: 10.1145/3539597.3572724.