Informatics

Influence of Anthropomorphism in AI-enabled Voice Assistants on User Trust

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Media and Human-Centered Computing

by

Edona Haziraj, BSc
Registration Number 12224127

to the Faculty of Informatics

at the TU Wien

Advisor: Ass.-Prof. Dr.x Katta Spiel
Assistance: Leopold, Lou, Projektass.in MSc

Vienna, May 2, 2025

Edona Haziraj

Katta Spiel

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU Bibliothek
WIEN Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Informatics

Der Einfluss von Anthropomorphismus in KI-gestützten Sprachassistenten auf das Vertrauen der Nutzer

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Media and Human-Centered Computing

eingereicht von

Edona Haziraj, BSc
Matrikelnummer 12224127

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ass.-Prof. Dr.x Katta Spiel
Mitwirkung: Leopold, Lou, Projektass.in MSc

Wien, 2. Mai 2025

Edona Haziraj

Katta Spiel

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Declaration of Authorship

Edona Haziraj, BSc

I hereby declare that I have written this Master Thesis independently, that I have completely specified the utilized sources and resources and that I have definitely marked all parts of the work - including tables, maps and figures - which belong to other works or to the internet, literally or extracted, by referencing the source as borrowed.
I further declare that I have used generative AI tools only as an aid, and that my own intellectual and creative efforts predominate in this work. In the appendix “Overview of Generative AI Tools Used” I have listed all generative AI tools that were used in the creation of this work, and indicated where in the work they were used. If whole passages of text were used without substantial changes, I have indicated the input (prompts) I formulated and the IT application used with its product name and version number/date.

Vienna, May 2, 2025

Edona Haziraj

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Erklärung zur Verfassung der Arbeit

Edona Haziraj, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschlieÿlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang Übersicht verwendeter Hilfsmittel habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 2. Mai 2025

Edona Haziraj

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Acknowledgements

Throughout my Master's studies and my years living in Vienna—I have been fortunate to be surrounded by people whose encouragement, insight, and support have played a meaningful role in shaping both my academic and personal growth.

First and foremost, I would like to thank my supervisor, Katta Spiel, for their invaluable guidance and support throughout this thesis. From shaping the research direction to providing thoughtful feedback, their influence has been both intellectually and personally meaningful. I'm also grateful for their early teaching, which laid the groundwork for much of my academic growth. I also want to sincerely thank Astrid Weiss, who first introduced me to the world of Human-Robot Interaction. Her teaching sparked my initial curiosity and opened the door to exploring the socio-technical aspects of this field, which eventually led to the research questions that became central to this thesis. Special thanks go to Dr. Giulia Perugia from Utrecht University, whose thoughtful and generous feedback during the early stages of my thesis helped me to clarify my research focus and frame the overall structure of the work more effectively. I also want to thank Leopold Lou, who provided valuable input during the design of the user study and helped me think more clearly about how to approach my topic in a practical and research-driven way.

Throughout this journey, my family has been a constant and powerful source of love, strength, and motivation. I am forever grateful to my parents, Sanije and Ibrahim, whose endless support and belief in me have never wavered, even though we were far by distance. A very special thank you also goes to my wonderful sisters—Ana, Besa, Blerta, Njomza, Urta, Leti, and Rrezja. Each of you has supported me in your own way, offering love, encouragement, and strength when I needed it most. You have been my emotional backbone and a constant reminder of where I come from. I dedicate all of my achievements to my family that has made me who I am.

Lastly, I want to thank my dear friend Ana for being such an important part of my time in Vienna. Her warmth and support grounded me during challenging moments, and her presence turned ordinary days into joyful memories. To all the friends and loved ones who stood by me quietly—thank you for being there.

This thesis is more than just an academic endeavor; it is also a deeply personal reflection of my values, experiences, and the perspectives I wish to amplify. As a woman working at the intersection of technology and society, I have often felt the weight of assumptions and stereotypes that still linger in both subtle and overt forms. With this work, I hope to push back against some of those narratives—to contribute to more inclusive, diverse, and thoughtful design in technology, and to help ensure that the tools we build reflect the richness of the people who use them.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Danksagung

Während meines Masterstudiums und meiner Jahre in Wien hatte ich das Glück, von inspirierenden Menschen umgeben zu sein, deren Ermutigung, Einsicht und Unterstützung mein akademisches wie persönliches Wachstum maßgeblich geprägt haben.

An erster Stelle möchte ich mich bei meiner Betreuungsperson Katta Spiel für die unschätzbare Anleitung und beständige Unterstützung während dieser Arbeit bedanken. Von der Ausrichtung der Forschungsfragen bis hin zu kontinuierlichem Feedback war Katta Spiels Einfluss sowohl intellektuell als auch persönlich bedeutsam. Ebenso bin ich dankbar für Katta Spiels frühere Lehrtätigkeit, die den Grundstein für einen Großteil meines akademischen Werdegangs gelegt hat.

Ich möchte auch Astrid Weiss herzlich danken, die mich erstmals in die Welt der Mensch-Roboter -Interaktion eingeführt hat. Ihr Unterricht weckte meine anfängliche Neugier und eröffnete mir den Raum, die sozio-technischen Aspekte dieses Gebiets zu erkunden, was letztlich zu den Forschungsfragen führte, die im Mittelpunkt dieser Arbeit stehen.

Ein besonderer Dank gilt Dr. Giulia Perugia von der Universität Utrecht, deren aufmerksames und großzügiges Feedback in der Anfangsphase meiner Arbeit half, meinen Forschungsschwerpunkt zu schärfen und die Gesamtstruktur der Arbeit effektiver zu gestalten.

Ebenso möchte ich mich bei Leopold, Lou bedanken, deren wertvolle Impulse bei der Konzeption der Nutzerstudie mir halfen, klarer darüber nachzudenken, wie ich mein Thema praxisnah und forschungsorientiert angehen kann.

Während dieser gesamten Reise war meine Familie eine kontinuierliche Quelle der Liebe, Stärke und Motivation. Ich bin meinen Eltern Sanije und Ibrahim für immer dankbar, da sie mich stets unterstützt haben und ihr Glaube an mich nie nachgelassen hat, auch wenn uns große Entfernungen trennten. Ein ganz besonderer Dank geht auch an meine wunderbaren Schwestern – Ana, Besa, Blerta, Njomza, Urta, Leti und Rrezja. Jede von euch hat mich auf ihre eigene Weise unterstützt und mir Liebe, Ermutigung und Kraft geschenkt, wenn ich sie am meisten brauchte. Ihr wart mein emotionales Rückgrat und eine ständige Erinnerung daran, woher ich komme.

Ich widme alle meine Leistungen meiner Familie, die mich zu dem Menschen gemacht hat, der ich heute bin.

Zu guter Letzt möchte ich meiner lieben Freundin Ana danken, die ein so wichtiger Teil meiner Zeit in Wien war. Ihre Wärme und Unterstützung gaben mir in schwierigen Momenten Halt, und ihre Anwesenheit verwandelte gewöhnliche Tage in freudige Erinnerungen.

All den Freund*innen und geliebten Menschen, die mir im Stillen zur Seite standen – danke, dass ihr da wart.

Diese Arbeit ist mehr als nur ein akademisches Unterfangen; sie ist auch eine zutiefst persönliche Reflexion meiner Werte, Erfahrungen und Perspektiven. Als Frau, die am Schnittpunkt von Technologie und Gesellschaft tätig ist, habe ich oft das Gewicht von Annahmen und Stereotypen gespürt, die in subtiler wie offener Form fortbestehen. Ich hoffe, mit dieser Arbeit einige dieser Narrative herauszufordern, zu einem inklusiveren, vielfältigeren und reflektierteren Design in der Technologie beizutragen und sicherzustellen, dass die Werkzeuge, die wir bauen, die Vielfalt der Menschen widerspiegeln, die sie nutzen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

AI voice assistants such as Siri, Alexa, and Google Assistant are now woven into everyday life, providing information and automating tasks through natural-language dialogue. Their anthropomorphic cues—especially the gender signaled by voice—shape how people perceive, engage with, and trust them. This thesis examines how gendered versus gender-neutral voice designs influence user trust and what that means for ethical, inclusive system design.

A mixed-methods study with 30 Vienna-based university students compared three OpenAI voice profiles: a masculine voice (*Cove*), a feminine voice (*Juniper*), and a gender-neutral voice (*Breeze*). Each participant interacted with one profile while completing three scenario tasks: seeking parenting advice, troubleshooting a spacecraft failure, and handling a medical emergency. Trust, usability, and anthropomorphism were measured through surveys, *post-task interviews*, participant drawings of the assistant persona, and behavioral observation.

Gendered voices triggered stronger anthropomorphic attributions and higher trust scores, broadly matching stereotypes of warmth (feminine) and authority (masculine). Reactions to the gender-neutral voice were more polarized: some participants praised its inclusivity, while others found it less emotionally resonant. Observable behaviors—sustained attention, affirmative back-channels, and positive affect—reinforced these patterns.

Building on these findings, the thesis proposes an *Inclusive Voice Design Framework* with actionable guidelines for personalization, emotional alignment, and equitable representation in conversational interfaces. The work contributes empirical evidence to human–AI interaction and AI-ethics debates, underscoring how subtle design choices—such as vocal gender cues—can materially affect user trust and acceptance.

Keywords: anthropomorphism; AI voice assistants; gender bias; genderless; inclusive design; HCI

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Kurzfassung

KI-Sprachassistenten wie Siri, Alexa und Google Assistant sind mittlerweile in den Alltag eingebettet. Sie liefern Informationen und automatisieren Aufgaben durch natürlichsprachliche Dialoge. Ihre anthropomorphen Hinweise - insbesondere das durch die Stimme signalisierte Geschlecht - prägen die Art und Weise, wie Menschen sie wahrnehmen, mit ihnen umgehen und ihnen vertrauen. In dieser Arbeit wird untersucht, wie geschlechtsspezifische und geschlechtsneutrale geschlechtsneutrale Sprachgestaltung das Vertrauen der Nutzer beeinflusst und was dies für ein ethisches, integratives Systemdesign bedeutet. Eine Mixed-Methods-Studie mit 30 Wiener Universitätsstudenten verglich drei OpenAI Sprachprofile: eine männliche Stimme (Cove), eine weibliche Stimme (Juniper) und eine geschlechtsneutrale Stimme (Breeze). Jeder Teilnehmer interagierte mit einem Profil, während er drei Szenario-Aufgaben erfüllte: Suche nach Erziehungsratschlägen, Fehlersuche bei einem Raumschiffausfall und Bewältigung eines medizinischen Notfalls. Vertrauen, Benutzerfreundlichkeit und Anthropomorphismus wurden durch Umfragen und Interviews nach den Aufgaben gemessen, Zeichnungen der Assistentenpersönlichkeit und Verhaltensbeobachtung gemessen. Geschlechtsspezifische Stimmen lösten stärkere anthropomorphe Zuschreibungen und höhere Vertrauenswerte aus, was im Großen und Ganzen Stereotypen von Wärme (weiblich) und Autorität (männlich) entsprachen. Die Reaktionen auf die geschlechts geschlechtsneutrale Stimme waren polarisierender: Einige Teilnehmer lobten ihre Inklusivität, während andere fanden sie weniger emotional ansprechend. Beobachtbare Verhaltensweisen - anhaltende Aufmerksamkeit, bejahende Rück und positiver Affekt - verstärkten diese Muster. Aufbauend auf diesen Ergebnissen wird in dieser Arbeit ein Inclusive Voice Design Framework vorgeschlagen, das mit umsetzbaren Richtlinien für Personalisierung, emotionale Ausrichtung und gleichberechtigte Repräsentation in Konversationsschnittstellen. Die Arbeit liefert empirische Beweise für die Mensch-KI-Interaktion und KI-Ethik-Debatten bei und unterstreicht, wie subtile Design-Entscheidungen - wie z.B. geschlechtsspezifische Sprachhinweise - das Vertrauen und die Akzeptanz das Vertrauen und die Akzeptanz der Nutzer wesentlich beeinflussen können.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Contents

Abstract **xiii**

Kurzfassung **xv**

Contents **xvii**

1 Introduction **1**
- 1.1 Motivation 2
- 1.2 State of the Art 5
- 1.3 Problem Statement 7
- 1.4 Research Gap 9
- 1.5 Contribution of the Thesis 10
- 1.6 Thesis Outline 11

2 Theoretical Background **13**
- 2.1 Theoretical Models: S-O-R, Media Equation, TPB, Uncanny Valley, CASA . 14
- 2.2 Trust and Anthropomorphism in Conversational Agents 21
- 2.3 Conversational AI: NLP, Chatbots, and Voice Interfaces 23
- 2.4 Gender as an Anthropomorphic Cue 25
- 2.5 (Non-)Binary Identities in Voice Assistants 27

3 Methodology **33**
- 3.1 Study Design 33
- 3.2 Independent Variables 36
- 3.3 Dependent Variables 37
- 3.4 Data Collection 38
- 3.5 Ethical Considerations 39
- 3.6 Data Analysis 40

4 Results and Analysis **41**
- 4.1 Participant Demographics 41
- 4.2 Pre-Interaction Survey Results 41
- 4.3 Perceived Anthropomorphic Representations 42
- 4.4 Trust Metrics 48
- 4.5 Impact of Voice Type on Trust 50
- 4.6 Thematic Analysis 53
- 4.7 Behavioral Observations 57
- 4.8 Evaluation of Hypotheses 58

5 Discussion **61**
- 5.1 Interpretation of Key Findings 61
- 5.2 Relation to Prior Research 63

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

5.3 Limitations 64
5.4 Implications and Future Research 65
5.5 Inclusive Design Framework 65

6 Conclusion **69**

7 Overview of Generative AI Tools Used **71**

8 Übersicht verwendeter Hilfsmittel **73**

List of Figures **75**

List of Tables **77**

Acronyms **79**

A Informed Consent Form **81**

B Participant Drawings: Voice Assistant I - OpenAI Juniper **85**

C Participant Drawings: Voice Assistant II - OpenAI Cove **93**

D Participant Drawings: Voice Assistant III - OpenAI Breeze **103**

E Prompt script (Task scenarios) **109**

F Pre-Interaction Survey Questions **113**

G Post-Interaction Survey Questions **115**

H Semi-structured Interview Guide **117**

Bibliography **121**

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 1

Introduction

Have you ever considered what qualities your ideal voice assistant should have? What voice would you trust the most to interact with every day? Perhaps its tone, intonation, or even its perceived gender might influence your preference. Have you ever found yourself imagining a human behind the voice—envisioning their age, gender, or appearance based solely on how they sound? This master's thesis investigates these questions through a study examining how anthropomorphic characteristics in AI-enabled voice assistants, such as those developed by OpenAI, influence user trust. By analyzing voice attributes and their impact on user perception, this research offers insights into the complex relationship between humans and AI. The findings presented here aim to inform the future design of voice assistants, contributing to their evolution as more intuitive, trustworthy, and user-aligned technologies.

Anthropomorphism is defined as ”the tendency to attribute human forms, behaviors, and emotions to non-human animals or objects” [MRMZ$^+$21]. Anthropomorphism, a pervasive and widespread phenomenon, takes on new dimensions in the rise of AI, one of which, though seldom emphasized, is the hype surrounding AI systems; this hype emerges from the attribution of distinct human characteristics to AI, which misrepresent and exaggerate its capabilities and performance, making anthropomorphism in AI an integral component of this hype [Pla24]. As conversational AI systems weave ever deeper into everyday life, their design raises crucial questions about how people perceive and engage with them. [ZJB24] posits that as artificial intelligence (AI) systems evolve to exhibit increasing human-like characteristics in their tone and behavior, users tend to establish personal—even emotional—relationships with these systems, thereby obfuscating the distinction between a tool and a companion. In a similar vein, [ACPR21] have demonstrated that despite corporate assertions of impartiality, users predominantly ascribe gender to and voice their interactions with virtual assistants such as Alexa and Siri, frequently employing female pronouns and perceiving them as human-like.

Beyond the evolving relationship humans have with these ”talking machines,” the growing presence of AI voice assistants in everyday life necessitates a more profound understanding of how their human-like characteristics—such as tone of voice, perceived gender, and conversational style—affect the manner in which we perceive and trust them. “Anthropomorphism—the attribution of human-like characteristics to non-human entities—is increasingly impacting the way we interact with generative AI systems such as ChatGPT. One consequence of anthropomorphizing conversational AI is the increase in unwarranted, perceived trust towards these systems” [Woo24]. As these systems increasingly obscure the distinction between tool and social entity, it becomes imperative to investigate the manner in which anthropomorphic cues influence user expectations, cultivate emotional connections, and potentially perpetuate societal biases. Trust has previously been studied as an important factor that affects people's acceptance of new technologies [YW22, AYYL20]. In this case, trust in the AI voice assistant will primarily be related to the user's belief in the assistant's ability to perform tasks accurately, reliably, and with no harm intent. It will mainly be focused on these dimensions: reliability- consistent task performance, competence-ability to understand and respond well, honesty-providing truthful, transparent information, security-protecting user data and privacy, comfort-how at ease users feel, especially with voice characteristics. Although AI Voice Assistants are typically designed not to be gender neutral (there is usually a female or male voice which the user can choose), it is unclear whether users would apply gender stereotypes to these tools, which could have an impact on acceptability and usage. Additionally, prior research has shown that men generally favor robots exhibiting female features, while women typically prefer those with male features, highlighting the presence of gender-related effects [SN22].

Specifically, this research will investigate whether gender-neutral voices can boost trust without reinforcing gender stereotypes, and how different components of anthropomorphism (e.g., personality, competence, interaction style, voice tone) influence user trust. Gender biases in AI Systems and Robotics are important topic as they reflect and potentially reinforce societal stereo- types. Although the significance gender biases is emphasized by literature that investigates anthropomorphism in robotics, suggesting that such biases can affect trust and engagement [Fin12]. Additionally, the study will examine potential biases that arise when users interact with AI agents, particularly in relation to gendered and non-gendered representations.

1.1 Motivation

The development of talking machines has been deeply intertwined with cultural perceptions of gender, authority, and labor. As explored in Vox ex Machina: A Cultural History of Talking Machines by Julianna C. Smith [Smi21a], and summarized in her article “Bitching Betty Speaks,” [Smi21b] the default design of early talking machines was not just a technical decision—it reflected and reproduced broader societal expectations. Until the 1980s, most consumer-grade talking devices used male voices, often referred to with the pronoun ”he,” reinforcing associations between technological authority and masculinity. However, by the time voice interfaces entered the public domain—especially in aviation and GPS systems—designers shifted toward female voices. This was not a neutral move. Designers assumed that users would find women's voices more ”soothing” and ”pleasant,” especially in high-stress environments, reflecting gender stereotypes about emotional labor and compliance. One iconic example is the ”Bitching Betty” voice used in fighter jets. While the nickname itself points to a tension between helpfulness and nagging, it also illustrates how female-voiced technologies are often expected to serve support roles—correcting, warning, and guiding—yet are simultaneously dismissed or mocked. The choice of voice thus becomes a site of cultural meaning, where the feminization of technology echoes historical patterns of invisible or undervalued emotional labor performed by women. Smith argues that these choices are not just functional but deeply embedded in sociotechnical imaginaries that shape how people interact with machines, how they trust them, and how they attribute personality, competence, or even emotion to them. By embedding gender into voice interfaces, designers implicitly influence the anthropomorphic characteristics users perceive in AI systems. This underscores the importance of examining how gendered voices reinforce or challenge societal norms, particularly in the context of trust and perceived intelligence in AI-enabled assistants.

In recent years, the rise of intelligent computing and advancements in speech recognition have led to a surge in voice interaction, especially through voice-controlled devices. These devices rely entirely on voice for communication, with no physical embodiment—such as a humanoid robot, robotic tool, or other tangible forms of representation. Despite this lack of physical presence, the way users perceive and interact with these voice assistants varies significantly

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

depending on the characteristics of the voice itself. Even though a voice assistant is not a humanoid, the nuances in its voice—such as how closely it resembles a human voice can strongly influence how users experience and interpret the interaction. For example, subtle differences in tone, pitch, and other vocal qualities can affect how users perceive the content being communicated and their overall impression of the assistant. This has raised interest in the integration of anthropomorphic features, human-like characteristics, into AI-powered voice assistants, particularly regarding their impact on user trust. Anthropomorphism, defined as the attribution of human traits to non-human entities, plays a pivotal role in shaping user interaction with and perception of these technologies [JCL24]. Understanding how such human-like characteristics influence trust is essential for designing voice assistants that are not only effective but also perceived as trustworthy and relatable by their users. Research shows that adding human-like features, such as voices and personalities, to voice assistants can make them seem more trustworthy to users. These features help people feel more connected to the technology, making it easier to trust and interact with. For instance, Hsu and Lee [HL23] investigated the impact of voice assistants' personality traits on user trust, risk perception, and user attitudes. Their study used the stimulus-organism-response (SOR) model and the theory of reasoned action (TRA to explore how people interact with voice assistants. It showed that human-like traits play a key role in building trust and encouraging positive reactions from users. Similarily, Seymour and Van Kleek [SVK21] conducted a comprehensive survey involving 500 voice assistant users to investigate the interplay between anthropomorphism and user trust. Their research showed a significant correlation between the perception of human-like characteristics in VAs and the development of trust. Specifically, participants who attributed greater human-like traits to their voice assistants reported stronger emotional connections and higher trust levels. These findings suggest that anthropomorphism not only enhances functional trust—users' belief in the assistant's ability to perform tasks reliably—but also boosts relational trust, which is rooted in users' emotional and psychological connections with the device. Moreover, Seymour and Van Kleek [SVK21] explored various anthropomorphic elements, including tone of voice, conversational style, and perceived personality, to determine their individual and combined effects on trust. Their analysis revealed that subtle cues, such as a warm and empathetic tone, were particularly effective in building trust, especially in scenarios involving sensitive or personal topics. The effectiveness of anthropomorphic cues varies based on context and the quality of a voice assistant's contributions. Westby et al. [WRRW24] examined how voice type (e.g., human-like versus synthetic) and response quality influence perceptions of anthropomorphism, animacy, intelligence, and trustworthiness, as well as team performance. Their findings highlight that while human-like voices can enhance perceptions, the quality of the assistant's contributions plays a more critical role in building trust. High-quality, contextually appropriate responses significantly improved perceptions of competence and trust, whereas low-quality responses, even with human-like voices, diminished trust and satisfaction. The study also noted that human-like voices encouraged engagement and natural interaction, but overuse of anthropomorphic cues could lead to unrealistic expectations if not matched by the assistant's actual capabilities. Westby et al. [WRRW24] stress the need for a balance between anthropomorphic design and functional performance to increase trust and avoid user frustration. On the other hand, research on anthropomorphism in verbal interactions has extended beyond voice assistants to include humanoid robots such as NAO. The study done by Thimm et al. [TEN$^+$24] explores how voice quality and anthropomorphic cues affect user trust and comfort levels. The findings suggest that voice quality plays a huge role in shaping perceptions of the robot's intelligence, reliability, and emotional alignment with users. Specifically, smoother, human-like voice qualities were associated with higher trust and lower discomfort, particularly in tasks requiring cooperative engagement. Their study also examines the balance between anthropomorphic features and functional design. While human-like voices increase perceived relatability and trust, overly anthropomorphic cues sometimes created discomfort or unrealistic expectations when the robot's behavior fell short of human-like standards. Additionally, the research highlights contextual dependencies, where trust and comfort were influenced by task complexity and the perceived stakes of the interaction. Their study reinforces the importance of designing voice quality and anthropomorphic traits to align with user needs and task demands.

As AI voice assistants (VAs) become increasingly integrated into everyday environments—whether in homes, cars, or workplaces—they are no longer perceived solely as tools but often as social entities. Users tend to attribute human characteristics, including gender roles and emotional expectations, to these agents based solely on their vocal cues. This phenomenon is rooted in the human tendency to anthropomorphize interactive technologies, particularly those that mimic human speech patterns and tones. Lopatovska and Williams [LW18] investigated interactions with voice assistants like Amazon Alexa and Apple Siri and found that users frequently interpret and relate to these assistants as "female employees." The assistants' often default feminine voices, combined with their service-oriented interaction style, lead users to associate them with traits such as friendliness, helpfulness, politeness, and even subservience. This not only reinforces traditional gender norms, but also makes these systems more likely to receive inappropriate, commanding, or even abusive language—exposing the unequal treatment female-voiced systems face compared to their male-voiced counterparts. These behavioral patterns reflect broader societal power dynamics and underscore how technology design can unintentionally reproduce harmful stereotypes. Purington et al. [PTS+17] similarly observed that users often assign emotional and relational roles to voice assistants, treating them as companions or confidants. Design features such as voice pitch, tone, and even the assistant's name were found to significantly shape user perceptions of trustworthiness, warmth, and relational closeness. These findings suggest that anthropomorphism operates not only at a functional level (e.g., understanding commands) but also at a symbolic and affective level—where voice assistants begin to occupy quasi-social roles in users' lives. However, anthropomorphism does not always yield positive effects. As Przegalińska et al. [PCS+19] point out, human-like cues such as voice gender and conversational style can enhance trust—but only when appropriately matched to the user's expectations and the assistant's role. For example, an empathetic, highly expressive voice may be welcomed in caregiving or entertainment contexts but could appear disingenuous or unsettling in transactional tasks like banking or navigation. When anthropomorphic design overshoots user expectations or task relevance, it can backfire, reducing credibility and comfort. These issues become even more critical when viewed through the lens of cultural and gender identity. Seaborn et al. [SSW24] conducted a cross-cultural study that explored how voice gender and tone influence trust across users of different backgrounds and gender identities. Their findings emphasize that the design of AI systems must be inclusive and adaptable, particularly for individuals who do not identify within the traditional gender binary. The study supports the development of gender-neutral or non-binary voice assistants—not just as a technical feature but as an ethical imperative. Such alternatives may not only be more inclusive but may also help reduce the automatic activation of gender-based stereotypes, thereby contributing to more equitable perceptions of competence, authority, and trustworthiness in voice interfaces. Collectively, these studies highlight the far-reaching sociotechnical implications of anthropomorphic voice design. The voice assistant's gender presentation does not merely influence surface-level user preferences; it plays a critical role in shaping ongoing relationships with AI, user behaviors, and broader cultural narratives about identity, labor, and trust. As voice interfaces become more ubiquitous and autonomous, these design decisions take on even greater significance—affecting not just interaction quality, but also social inclusion, power dynamics, and the ethics of human-AI engagement.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1.2 State of the Art

Artificial intelligence (AI)-enabled voice assistants (VAs), such as Apple Siri, Amazon Alexa, and Google Assistant, have become part of everyday life. Designed to respond to spoken commands and carry out tasks through natural language processing, they often feature human-like voices and personality traits that reflect gendered characteristics, subtly influencing users' perceptions and trust. This section presents a comprehensive review of existing research on anthropomorphism in AI—particularly voice assistants—and how it shapes user trust, focusing on evolving theoretical frameworks, empirical findings on gender cues, and emerging trends toward more inclusive design. According to [UNE19], the default feminization of AI voice assistants—such as Siri, Alexa, and others—paired with their often servile, apologetic, and overly accommodating interaction styles, risks reinforcing long-standing and harmful gender stereotypes. These assistants are typically portrayed as polite, submissive, and eager to please, which subtly suggests that women—through the personification of female-gendered voices—are expected to occupy subservient roles in service-oriented contexts. Such design choices, although perhaps unintentionally made for user familiarity or marketing purposes, contribute to a broader sociotechnical issue: the embedding of gender bias in emerging technologies. By giving voice to machines in ways that echo outdated gender norms, designers risk normalizing these stereotypes in everyday human–AI interactions.

The title *I'd Blush If I Could* is a reference to a former response by Apple's Siri when faced with verbal abuse, specifically when a user said, "Hey Siri, you're a bi***." At the time, Siri—a widely used female-voiced assistant—would respond with the phrase that now headlines the UNESCO report. While the assistant's response has since been modified to a more neutral reply such as *"I don't know how to respond to that,"* and very recently to *"I won't respond to that,"* the broader issue remains: the assistant continues to embody a passive and compliant demeanor, even when confronted with offensive or gendered language. This highlights a concerning dynamic in the design of AI systems, where anthropomorphized and gendered technologies may inadvertently uphold and normalize harmful social behaviors and power imbalances. This anthropomorphic framing of AI—imbuing voice assistants with human-like traits such as tone, gendered voice, and personality—not only shapes user expectations but also directly influences user trust and perceived competence of the system. UNESCO's report calls for a more critical and inclusive approach to voice assistant design, advocating for the development of gender-neutral or non-binary voices and more balanced personality traits that do not reproduce societal inequalities. This thesis takes up that challenge by examining how vocal gender cues and interaction styles influence trust in AI-enabled voice assistants. By introducing a typically male voice, a typically female voice, and a non-gendered voice into the user study, this research investigates the role that vocal anthropomorphism plays in shaping users' perceptions, expectations, and trust in digital agents—thus contributing to a more equitable and conscious design of future AI systems.

AI voice assistants have transitioned from novelty items to ubiquitous household fixtures. These systems leverage advanced machine learning and natural language processing to provide conversational interactions that closely resemble human dialogues. For instance, Apple's Siri uses speech synthesis and context-aware algorithms, Amazon's Alexa applies a cloud-based NLP model for query handling, and Google Assistant harnesses a deep learning pipeline.

Anthropomorphism—our instinctive attribution of human traits, emotions, and intentions to non-human entities—plays a key role in user experience [ZJB24]. Even though users know they are interacting with AI, many describe their VAs as "friendly" or "annoying." This paradox highlights the deeply ingrained psychological tendency to treat anything that exhibits human-like behavior, especially speech, as a social actor. According to Epley et al. [EWC07], anthropomorphism is not a random or purely imaginative process but rather a systematic cognitive response rooted in fundamental psychological needs and mechanisms. They propose a *three-factor theory* *of anthropomorphism*, which identifies three key psychological determinants that influence when and why people anthropomorphize.

The first factor, *elicited agent knowledge*, refers to the accessibility of human-related knowledge that people draw upon when interpreting the behavior of non-human entities. When individuals lack specific information about an agent, they often rely on familiar human schemas to fill in the gaps. The second factor, *effectance motivation*, reflects the human desire to understand, control, and predict the behavior of other agents. Anthropomorphizing can provide a sense of coherence and predictability, especially in uncertain or ambiguous situations. The third factor, *sociality motivation*, is driven by the innate need for social connection. When individuals experience loneliness or social isolation, they are more likely to project human-like qualities onto non-human agents to fulfill their social needs.

Together, these three factors explain why people often ascribe minds, emotions, or intentions to machines, animals, or even abstract concepts, particularly in contexts where human interaction is limited or where understanding and control are desired.

Reeves and Nass [RN96] demonstrated with the *Media Equation* theory that people treat computers and digital agents as social counterparts when they present even minimal human-like cues. Their experiments showed that social rules such as politeness and reciprocity emerge automatically.

Anthropomorphism in VAs can manifest through voice pitch, language style, personality, and even implied gender—elements collectively promoting user engagement and trust. From a human-computer interaction HCI perspective, these design choices evoke strong social responses. Eyssel et al. [EKHdR12] studied how participants responded to robots with male vs. female voices, finding that robots whose perceived gender conformed to the participant's stereotypes were rated as more competent and trustworthy. This outcome was reinforced by both behavioral observation and Likert-scale measures, highlighting the power of subtle voice manipulations. Additionally, Nass and Moon [NM00] found that people often apply social behaviors and expectations to computers automatically, treating them similarly to humans. Their studies showed that individuals frequently assign human characteristics, such as gender stereotypes or ethnic identities, to computer agents and engage in social behaviors like politeness and reciprocity. This indicates that interactions with computers are guided by social rules usually meant for human interactions

With regard to DVAs, anthropomorphism plays a significant role in shaping user experiences and perceptions. As discussed by [FS22], individuals who personify DVAs—such as by using personal pronouns or attributing human-like characteristics—tend to report higher satisfaction with their interactions and often form more sociable, companion-like relationships with these technologies. This aligns with empirical findings by [PTS$^+$17], who observed that users commonly assign social roles to DVAs like Amazon's Alexa, perceiving them as friends or assistants, which in turn fosters emotional engagement and increased usage satisfaction. Beyond simple personification, the emotional dynamics of human-DVA interaction can be further influenced by how the assistant responds in specific contexts. [CCT20, FS22] demonstrated that when DVAs employ empathetic language—particularly in response to verbal aggression or user frustration—they can mitigate hostile behavior and even elicit feelings of guilt in users. For instance, when a DVA replies to an insult with a response like "You're upset, let me know if there's something I can do better," it creates a humanlike interactional script that encourages users to reconsider their behavior [FS22, CCT20]. Together, these studies suggest that anthropomorphic and empathetic design features in DVAs not only enhance user engagement but also shape the moral and social frameworks within which users relate to these technologies.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

1.3 Problem Statement

The rapid adoption of AI-enabled conversational agents has transformed how users interact with technology in daily life. These systems increasingly perform social roles—offering reminders, answering questions, facilitating tasks—by mimicking human speech and behavior. As these assistants become more anthropomorphized through natural-sounding voices, conversational turn-taking, and emotional expression, questions arise about the psychological and social implications of their design, particularly regarding user trust. Trust is a central determinant in the successful integration of intelligent systems into human environments. According to [LS04], trust in automation reflects an individual's belief that a system will perform actions that help achieve their goals, especially under conditions of uncertainty or vulnerability. Trust in VAs extends beyond performance metrics such as accuracy or speed to include perceived warmth, reliability, and alignment with user expectations—traits often communicated implicitly through anthropomorphic design features like voice tone, speech patterns, and perceived gender [LS04, NMG97]. Among the most powerful of these cues is voice gender. Research has demonstrated that users often ascribe human traits and stereotypes to VAs based on vocal cues alone, even when no visual embodiment is present [EH12, NSS94]. For example, female voices are frequently associated with warmth and helpfulness, while male voices are more often perceived as authoritative and competent [NB05, SK21]. These perceptions echo longstanding gender stereotypes and can influence user attitudes, comfort levels, and even perceived trustworthiness of the system [UNE19]. Feminized voice assistants—often presented as submissive, apologetic, or service-oriented—may unintentionally reinforce patriarchal norms by embedding gender biases into everyday technologies [UNE19, Smi21a].

Despite growing awareness of these issues, most commercial VAs continue to offer users a binary choice between male and female voices. Emerging efforts such as the "Q" project—an intentionally designed gender-neutral voice—highlight a nascent shift toward more inclusive voice design [Vir19]. However, little empirical evidence exists on how such voices are perceived by users, or whether they foster trust to the same extent as traditionally gendered voices. Early work suggests that gender-neutral VAs may reduce stereotyping and increase accessibility for non-binary users, but may also introduce discomfort due to unfamiliarity or ambiguity [DHHP+23, MOP+20]. Moreover, while anthropomorphism has been shown to increase both functional and relational trust—helping users feel more connected to technology—the overuse of human-like cues can backfire, especially if expectations set by human-like voices are not met by the system's actual capabilities. This phenomenon is part of the "uncanny valley" effect, where entities that appear or sound almost human, but not quite, evoke discomfort or mistrust [MMK12, DL22]. In the context of voice assistants, this creates a design dilemma: should systems sound more human to increase trust, or less human to maintain realism and ethical clarity? Regardless the abundance of research on anthropomorphism and trust, very few studies isolate and test the role of voice gender alone, particularly in relation to different trust dimensions (e.g., reliability, competence, benevolence). Most work conflates voice gender with interaction style, personality, or content, making it difficult to discern whether trust is driven by vocal qualities or by social context. Furthermore, most experiments are conducted via surveys or hypothetical interactions, lacking the ecological validity of real-time, interactive user studies. As VAs have become central components of daily digital interaction, functioning across a variety of domains and designed to provide natural, accessible communication, these systems increasingly rely on anthropomorphic cues—particularly in the use of human-like voices—to facilitate intuitive interaction and foster user trust. However, while these cues can improve usability and user satisfaction, they also introduce deeper psychological, social, and ethical complexities. Among the most pressing of these is the impact of vocal gender representation on user trust and the reinforcement of societal stereotypes through design choices. VAs are often not interpreted by users as mere functional tools. Instead, they are often treated as social agents, with users instinctively applying human social rules and expectations based on vocal traits alone. Even without visual embodiment, cues such as voice pitch, intonation, emotional tone, and speaking style strongly shape user impressions of an assistant's competence, warmth, and trustworthiness [EH12, NB05]. This effect is magnified when the assistant's voice is gendered, as users routinely project culturally embedded assumptions onto male and female voices [SK21, UNE19]. Male-voiced assistants are often perceived as more authoritative and technically competent, while female-voiced systems are more likely to be perceived as empathetic, helpful, or emotionally responsive.

What may seem like neutral design choices—such as defaulting to a female voice—are, in fact, embedded in broader cultural narratives. These design defaults frequently reflect and reproduce traditional gender roles, positioning voice assistants in service-oriented, subordinate roles associated with emotional labor and compliance. Lopatovska and Williams [LW18] show that such assistants are often treated by users in ways that mirror real-world gender dynamics, including commanding language and increased tolerance for verbal abuse. These patterns demonstrate that anthropomorphized voice design is not just a matter of user preference—it can subtly normalize power imbalances and contribute to the replication of societal biases in technological systems.

At the same time, the link between anthropomorphism and trust is not linear. While human-like cues such as warmth, familiarity, or emotional expression can promote trust and ease of interaction, these cues may also backfire if they are perceived as exaggerated, contextually inappropriate, or misaligned with user expectations. Przegalińska et al. [PCS+19] argue that anthropomorphic design must be deployed carefully; a mismatch between a system's apparent emotional intelligence and its actual functionality may cause discomfort or reduce credibility. Similarly, Seaborn et al. [SSW24] highlight that trust in voice assistants is shaped not just by system performance, but by the user's identity and whether they feel acknowledged and represented in the interaction. Binary-gendered voice options may marginalize or exclude non-binary individuals, raising questions of inclusivity, ethical responsibility, and equitable user experience.

Despite growing recognition of these challenges, most commercial voice assistants continue to offer only male or female voice choices and rarely address the broader diversity of gender identity in their interaction design. Projects like the "Q" voice [Vir19] and inclusive TTS initiatives [DHHP+23] represent important steps toward more inclusive voice technology, but their real-world effects on user trust—particularly in contrast to binary-gendered voices—remain underexamined. Early findings suggest that gender-neutral voices may reduce stereotyping and support more equitable engagement, yet such voices can also evoke discomfort or confusion due to their unfamiliarity or lack of social precedent [MOP+20].

Moreover, the construct of trust itself is multidimensional. It includes not only perceptions of competence and reliability but also emotional safety, benevolence, and the user's comfort with disclosure or collaboration [LS04]. Existing studies frequently focus on general attitudes or user satisfaction without investigating how specific trust components are influenced by voice gender or anthropomorphic cues. In addition, many of these studies rely on self-report data, hypothetical scenarios, or short-term interactions, limiting the ecological validity of their findings.

This thesis addresses these limitations by empirically investigating how voice gender—male, female, and gender-neutral—affects user trust in AI-enabled voice assistants. It focuses on how specific vocal attributes (e.g., tone, personality cues, interaction style) influence different dimensions of trust in practical, task-based settings. Furthermore, it examines whether gender-neutral voices can offer a viable, inclusive alternative that avoids reinforcing traditional stereotypes while still maintaining or enhancing user trust. In doing so, this research contributes to ethical and human-centered design practices in conversational AI, aiming to inform more inclusive,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

psychologically attuned, and socially responsible systems.

1.4 Research Gap

Despite the widespread integration of AI-enabled voice assistants (VAs) into everyday life, substantial research gaps persist regarding the influence of vocal gender representation on user trust, particularly in real-world, interactive contexts. While prior studies have acknowledged that users anthropomorphize VAs—assigning human-like traits based on voice tone, perceived gender, and interaction style—most research continues to reinforce binary gender paradigms and traditional stereotypes [GM23]. Female-voiced assistants are commonly associated with warmth, empathy, and helpfulness, whereas male-voiced assistants tend to be perceived as authoritative and competent. This dichotomy not only perpetuates entrenched gender norms but also neglects the experiences, expectations, and identity representations of non-binary and gender-diverse users.

Although recent initiatives have introduced gender-neutral voices to mitigate stereotyping and promote inclusivity, empirical studies evaluating their effectiveness in building trust remain limited. Some evidence suggests that gender-neutral voices may reduce biased projections, yet user reactions often range from appreciation of inclusivity to discomfort due to unfamiliarity. This ambivalence signals a nuanced challenge: inclusive design must not only avoid reinforcing bias but also remain emotionally resonant and socially acceptable across diverse user demographics.

Furthermore, trust in AI is a complex, multifaceted construct that is frequently oversimplified in existing literature. Most prior work assesses trust as a general or monolithic attitude, overlooking key dimensions such as *competence*, *reliability*, *emotional comfort*, and *benevolence* [SEK24]. In particular, few studies dissect how specific vocal cues—such as tone, pitch, or speech rhythm—impact these individual trust components. Moreover, most of the existing literature relies on short-term, survey-based, or hypothetical interactions, thereby limiting ecological validity and failing to capture the evolving, contextual nature of human-VA relationships over time.

This gap is particularly evident when considering the absence of embodied visual agents. While insights from human-robot interaction (HRI) have highlighted how embodiment and expressive behavior contribute to perceived trustworthiness [CFG$^+$21], these findings may not directly translate to voice-only agents, whose social presence relies solely on auditory cues. As such, the mechanisms by which *vocal-only anthropomorphism* shapes trust in VAs remain underexplored.

Finally, current design approaches often disregard the ethical implications of anthropomorphic design choices. Assigning gender to AI systems is not a neutral decision; it actively shapes user expectations, emotional engagement, and the perceived roles these systems occupy in social and domestic settings. By defaulting to feminized voices in service roles or masculine voices in authoritative contexts, many commercial VAs implicitly reinforce social hierarchies and normative assumptions, with real-world consequences for how users treat, trust, and relate to these technologies.

In summary, there is a critical need for comprehensive, empirical studies that:

- Examine the impact of male, female, and gender-neutral voices on the various dimensions of user trust.
- Explore user interactions with VAs in real-time, task-based scenarios to enhance ecological validity.
- Investigate the experiences and preferences of non-binary and gender-diverse users to inform inclusive VA design.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

- Deconstruct the multifaceted nature of trust to understand how specific vocal cues influence each dimension.

Addressing these gaps is a way forward in developing ethical, inclusive, and user-centered voice assistant technologies that resonate with the diverse needs of users.

1.5 Contribution of the Thesis

This thesis contributes to the interdisciplinary fields of human-computer interaction (HCI), AI ethics, VUI design, and gender studies by investigating the nuanced relationship between anthropomorphic cues—such as voice tone, gender expression, and interaction style—and user trust in AI-enabled voice assistants. In particular, it addresses pressing societal concerns around inclusivity, fairness, and representational bias in conversational agents by examining how gendered and gender-neutral voice attributes shape user perceptions, trust levels, and emotional engagement. The study offers both empirical and conceptual contributions. Empirically, it provides detailed observational and qualitative evidence demonstrating how different voice profiles—typically male, typically female, and gender-neutral—evoke varying behavioral, cognitive, and affective responses. Behavioral cues such as attentive listening, natural conversational flow, visible enjoyment (e.g., smiles and laughter), and requests to repeat certain voices suggest differentiated levels of comfort, memorability, and perceived authority. These findings highlight how users subtly interpret and react to vocal and interactional cues, even in the absence of explicit visual or narrative framing.

Conceptually, the thesis advances theoretical understanding of anthropomorphism by unpacking its relational impact on trust—specifically how warmth, competence, and relatability are inferred from auditory features alone. This contributes to ongoing debates around the Media Equation, the S-O-R (Stimulus-Organism-Response) model, and the Uncanny Valley theory, which are elaborated more in section 2.1 offering new insights into how perceived gender and neutrality intersect with users' mental models of intelligence, helpfulness, and social presence in voice assistants.

A central contribution of the thesis is the development of an *Inclusive Design Framework for Voice Assistants*, which is grounded in user feedback and behavioral observations from the study. The framework outlines actionable design principles that prioritize user trust, voice customization, emotional resonance, and cultural sensitivity. Key recommendations include:

- Offering *customizable voice profiles* to support individual user preferences related to tone, pace, gender expression, and familiarity.
- Prioritizing *gender-neutral and non-binary voice options* to promote inclusivity and reduce the reinforcement of gender stereotypes.
- Embedding *transparency and ethical cues* in AI responses to build trust and convey competence without relying on gendered tropes.
- Designing for *accessibility and emotional comfort*, especially for older users and those from diverse cultural or linguistic backgrounds.

This framework is visualised and extensively explained in section 5.5

By synthesizing these findings into a practical framework, the thesis contributes to the broader movement toward ethical, human-centered AI design. The inclusion of end-user perspectives in this research helps illuminate how seemingly minor design decisions—such as the choice of a

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

voice—can profoundly influence user experience, trust, and social equity. Ultimately, this thesis seeks to inform both academic discourse and industry practice by offering grounded, evidence-based strategies for developing conversational AI that is not only technically competent but also socially aware, inclusive, and trustworthy. It demonstrates that voice assistants are not neutral interfaces, but socio-technical systems that reflect and shape user values, expectations, and biases—making inclusive design not just a desirable feature, but an ethical imperative.

1.6 Thesis Outline

This thesis is structured into six main chapters. Chapter 1 introduces the research context and outlines the research problem, the existing gap in the literature, and the objectives and contributions of the study. Chapter 2 provides the theoretical background, presenting relevant literature and frameworks related to voice assistants, gender representation, and anthropomorphism. Chapter 3 details the methodology, including the study design, variables, data collection procedures, and ethical considerations. Chapter 4 presents the results and analysis of the study, including participant demographics, survey findings, behavioral observations, and the development of an inclusive design framework. Chapter 5 discusses the key findings, interprets the results in relation to the research questions, compares them with existing literature, and highlights the limitations and implications of the study. Finally, Chapter 6 concludes the thesis by summarizing the main contributions and offering directions for future research. The thesis also includes an overview of generative AI tools used, appendices with supporting materials, and a comprehensive bibliography.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

CHAPTER 2

Theoretical Background

To establish a substantial theoretical foundation for this thesis, an extensive literature review was conducted across a range of academic databases and platforms. Most visited platforms, among others were ResearchGate, Google Scholar, ScienceDirect, ACM Digital Library, IEEE Xplore, SpringerLink, ArXiv, JSTOR, MIT Press, Design Studies, John Wiley & Sons, Taylor & Francis, and the Journal of Applied Social Psychology. The selection of these sources was made with the intention of providing a comprehensive overview of interdisciplinary subjects, including but not limited to human-computer interaction, voice perception, artificial intelligence ethics, gender studies, psychology, and inclusive design.

Throughout the literature search for this thesis, I relied on a focused set of keywords and Boolean combinations to capture the interdisciplinary dimensions of voice-based AI. Core search terms included “anthropomorphism in artificial intelligence,” “voice assistants and trust,” “gendered and non-binary voice perception,” “inclusive design in conversational agents,” “digital voice identity,” “voice pitch and gender,” “stereotypes in synthesized speech,” “uncanny valley and voice interfaces,” and “human-likeness in robots.” I also iterated on these with combinations such as “anthropomorphism AND AI,” “gender AND voice assistant,” and related variants to refine results and ensure comprehensive coverage of the most relevant scholarship. These terms assisted in the filtration of studies pertinent to the research questions at hand and the identification of theoretical frameworks, including the Stimulus-Organism-Response (S-O-R) model, the Theory of Planned Behavior (TPB), and the Media Equation theory.

In addition to conference proceedings, peer-reviewed journals and a select number of influential white papers and public research projects were included to reflect current discourse in AI ethics and inclusivity. Notable examples of this phenomenon include the UNHCR Digital Transformation Strategy (2022–2026), the MIT Press Reader article ”Bitching Betty Speaks,” and the GenderlessVoice.com project by Virtue Nordic and Copenhagen Pride, which introduced the world’s first gender-neutral voice assistant, ”Q.”

This research enabled the synthesis of findings from both academic and applied perspectives, thereby facilitating a comprehensive understanding of how anthropomorphic design choices, particularly those pertaining to voice and perceived gender, can influence trust, perception, and inclusivity in AI-enabled voice assistants.

In the opening stage of my review, the latest research on anthropomorphism in conversational agents was explored, examining how researchers describe and gauge the “human-likeness” of voice-driven AI. This involved charting the linguistic, acoustic, and interactional cues—such as tone, prosody, response timing, and persona consistency—that studies treat as practical benchmarks for determining whether a talking machine feels authentically human.

[Bă23] identified five distinct but interrelated dimensions that define the humanlikeness of chatbots and voice assistants. These dimensions go beyond physical appearance to encompass cognitive, emotional, social, and communicative capabilities:

- **Morphological Design Features** – such as a humanlike body, face, gestures, voice, and name.
- **“Own” Mind, Knowledge & Intelligence** – referring to the agent’s ability to think, learn, understand context, and make decisions.
- **Humanlike Emotions & Behavior** – including the expression of emotions, empathy, humor, mood changes, and politeness.
- **Social Dimension** – the capacity to build relationships, personalize interactions, remember past conversations, and offer companionship or support.
- **Communication Skills & Conversational Performance** – such as turn-taking, natural language processing, clarity, fluency, and the use of small talk.

These categories provide a comprehensive lens through which anthropomorphism in voice assistants can be understood and assessed. The current study focuses particularly on *voice characteristics* (a subset of morphological design features) and their impact on user trust.

2.1 Theoretical Models: S-O-R, Media Equation, TPB, Uncanny Valley, CASA

One of the first demonstrations that observers project social meaning onto abstract motion comes from Fritz Heider and Marianne Simmel’s classic 1944 experiment. They showed that observers readily impose intentions, emotions, and social narratives onto simple moving shapes. Thirty subjects watched a 2½-minute animation of a large triangle, a small triangle, and a circle moving in and out of a rectangular “house.” Despite the stimuli’s geometric abstraction, participants overwhelmingly described the shapes in anthropomorphic and social terms—using verbs like “chased,” “tricked,” and “helped” and attributing motives such as “jealousy” and “cooperation.” Heider and Simmel concluded that “the processes which are involved in perceiving other individuals, their behavior and their personal qualities, have received but little attention in psychological literature,” yet are “basic in almost any social act” [HS44, p. 243]. This seminal study laid the groundwork for decades of research into theory of mind and the perception of animacy.

Hsu and Lee [HL23] used the Stimulus-Organism-Response (S-O-R) model to explain how external voice stimuli impact internal user states and behavioral intentions. The Stimulus-Organism-Response (S-O-R) model, originally developed by Woodworth (1929) and later extended by [MR74], offers a valuable framework for analyzing the role of internal states in shaping behavioral responses. Unlike the basic Stimulus-Response (S-R) paradigm, which focuses solely on external inputs and resulting actions, the S-O-R model introduces the “organism” as a mediating factor—acknowledging the importance of cognitive and emotional processes that occur between stimulus and response.The S-O-R model is particularly relevant in examining how anthropomorphic features in voice assistants (stimuli), such as gendered or non-gendered voices and interaction styles, influence users’ internal states (organism), including perceptions of trust, familiarity, and comfort, ultimately shaping their behavioral responses toward the system. [MR74] identified three key emotional dimensions—pleasure, arousal, and dominance—that explain how individuals react to environmental stimuli, which can result in either approach or

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

avoidance behavior. Applying this model allows for a deeper understanding of how user trust in AI-enabled voice assistants may be moderated by psychological processes elicited by specific anthropomorphic cues [MR74, HBL23]. A visualisation of SOR Model is shown in Figure 2.1

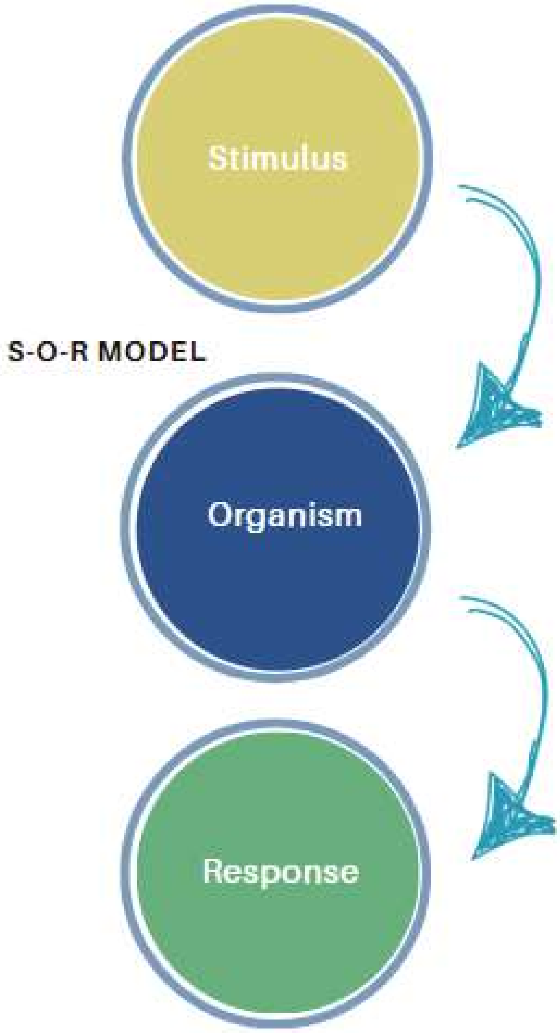


Figure 2.1: S-O/R Model adapted from [HBL23]

The Media Equation Theory, developed by Byron Reeves and Clifford Nass [RN96], asserts that people naturally and unconsciously respond to media and technological interfaces—such as computers, television, or AI systems—as though they were real people or places. This tendency persists even when individuals are fully aware that the interaction is with a machine [SL21]. Rooted in the understanding that the human brain is evolutionarily conditioned to react socially, the theory explains how even minimal social cues—such as voice, politeness, personality, turn-taking, or emotional tone—can trigger deeply ingrained human-to-human behavioral patterns. One foundational principle highlighted by Reeves and Nass is that *"people are polite to computers, too"*, demonstrating that users exhibit the same courtesy and social expectations when interacting with media as they would with other people. Politeness toward conversational AI has real-world operational consequences. As reported by Deb (2025) [Deb] very recently, OpenAI's infrastructure incurs "tens of millions of dollars" in additional electricity costs each year solely because users habitually append "please" and "thank you" to prompts to ChatGPT [Deb]. This quantification of courtesy reveals how ingrained social rituals extend beyond purely human interaction, manifesting as measurable burdens on computational resources. By foregrounding the monetary cost of linguistic politeness, we make visible the extent to which users

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

unconsciously grant AI systems a quasi-social status—treating them not merely as tools, but as interlocutors deserving of respect and etiquette.

The tendency to anthropomorphize AI, moreover, is underpinned by both cultural norms and cognitive framing. Turkle (2020) [Tur20] observes that “if an object is alive enough for us to start having intimate conversations, friendly conversations, treating it as a really important person in our lives, even though it’s not, it’s alive enough for us to show courtesy to” [Tur20]. Empirical survey data further corroborate this dynamic: seventy-eight percent of respondents report using polite language with voice assistants—55 percent as a moral habit and 12 percent “just in case” the system attains consciousness. Design studies at Microsoft also suggest that polite framing can improve AI output quality through a “tone-mirroring” effect [Deb, Sta25]. Collectively, these findings illustrate that polite interaction with AI emerges not only from empathy or future-oriented anxiety but also from strategic considerations aimed at eliciting more cooperative and coherent responses. Back to Reeves and Nass, they further describe how users often apply socially constructed roles and rules to media entities. These include:

- **The Politeness Rule**: users tend to be as courteous to computers as they are to humans.
- **Similarity Attraction**: people are more positively disposed toward media that reflects their own characteristics or preferences.
- **Gender Stereotyping**: users attribute gender-based roles and expectations to media systems that use gendered cues like names or voices.
- **Anthropomorphism**: even minimal human-like signals, such as tone or responsiveness, are sufficient to prompt users to treat media as socially present actors.

The core argument—that anthropomorphic cues in voice assistants influence user trust—mirrors the Media Equation’s proposition that social behaviors are elicited by minimal humanlike signals. For instance, voice assistants that use polite or empathetic language may be perceived as more trustworthy simply because they reflect expected norms of human social behavior. Furthermore, the exploration of gendered versus gender-neutral voices builds on Reeves and Nass’s findings that users project gender stereotypes onto AI based on voice alone, which can shape perceived competence, trustworthiness, or emotional warmth. The Media Equation also contextualizes trust not merely as a technical attribute but as a relational and affective one, underscoring the importance of perceived “humanness” in fostering credibility and user comfort.

On the other hand, the Theory of Planned Behavior (TPB) [Ajz91], as illustrated in Figure 2.2, postulates that human behavior is guided by three core components: *attitudes*, *subjective norms*, and *perceived behavioral control*. These components influence an individual’s *intention* to perform a behavior, which is the most immediate predictor of actual behavior. Specifically:

- **Attitude toward the behavior** refers ”to the degree to which a person has a favorable or unfavorable evaluation of the behavior in question.”
- **Subjective norm** relates ”to the perceived social pressure to perform or not perform the behavior.”
- **Perceived behavioral control** reflects ”the perceived ease or difficulty of performing the behavior, which is also assumed to reflect past experience and anticipated obstacles.”

When applied to AI and voice assistants (VAs), TPB provides a useful framework for understanding the psychological drivers of user trust and engagement. A user’s willingness to trust and rely on a VA can be understood through:

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

- **Attitudes** about the VA's helpfulness, intelligence, or reliability,
- **Social norms**, such as whether peers or society increasingly accept and use VAs,
- **Perceived control**, such as how confident users feel in managing or influencing the assistant's responses.

Studies have shown that trust increases when users feel they can predict and influence the assistant's behavior [DAHD21, La24, DAHD20], highlighting the relevance of perceived behavioral control in trust formation. TPB therefore emphasizes that user trust is not only a matter of the technology's performance but also shaped by users' beliefs, social context, and sense of autonomy during interaction.

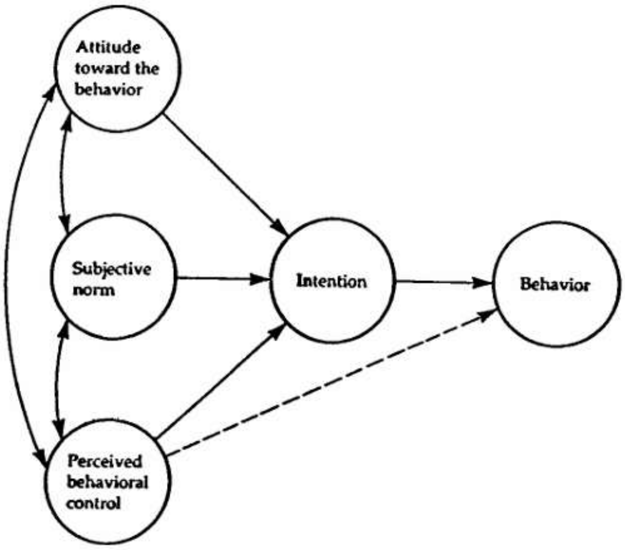


Figure 2.2: Theory of Planned Behavior model [Ajz91]

As noted by Hagger [Hag19], the Theory of Planned Behavior and its extensions have been widely used to explain how beliefs, intentions, and perceived control influence behavior across various domains, including technology adoption [Hag19].

Empirical studies indicate that anthropomorphism can enhance trust by increasing perceived warmth [NB05], but excessive human-likeness may trigger distrust if users perceive the AI as deceptive, the *uncanny valley* effect in voice interactions [MMK12, DL22].

The *Uncanny Valley* theory, introduced by Masahiro Mori in 1970 [MMK12], describes a phenomenon in which artificial entities that appear almost—but not fully—human elicit a strong sense of unease or eeriness in observers, as shown in Figure 2.3. As an entity's human-likeness increases, people tend to respond to it with greater emotional affinity. However, when the entity reaches a point where it is nearly human but falls short in subtle ways—such as unnatural voice, awkward facial expressions, or mechanical movements—this affinity abruptly declines, resulting in a "valley" of discomfort. The theory suggests that the closer a machine gets to resembling a human without achieving perfect realism, the more unsettling it becomes.

In the context of AI-enabled voice assistants (VAs), the Uncanny Valley offers a valuable lens through which to understand user trust and engagement. Human-like voices and personalities in VAs are often employed to foster familiarity and connection. While these traits can initially enhance user experience, they also risk backfiring if the realism is not consistent across all aspects of interaction. For example, a VA that uses a highly realistic voice but responds with

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

unnatural phrasing, emotionless delivery, or delayed timing may trigger cognitive dissonance. This disconnect between expectation and actual behavior can lead to discomfort and a reduction in trust, particularly when users subconsciously anticipate human-like interaction patterns but encounter behavior that breaks those expectations.

Recent studies support the relevance of the Uncanny Valley effect in virtual agents. Alimardani et al. (2024) demonstrated that mismatches between an agent's appearance and its voice can lead to heightened feelings of eeriness and diminished trust in human-agent collaboration [ARVL24]. Their research emphasizes the importance of coherence between visual and auditory cues in establishing trustworthy interaction. Similarly, Tarlan et al. (2024) [TE24] found that although humanoid robots and virtual human avatars can enhance perceived intelligence and likeability, they also pose a risk of entering the Uncanny Valley if their human-likeness is not balanced across physical presence, gestures, and verbal behavior [TE24]. These findings underline the need for a careful and holistic approach in designing voice assistants—where realism must be supported by emotional congruence, responsiveness, and interactional fluency—in order to build and sustain user trust.

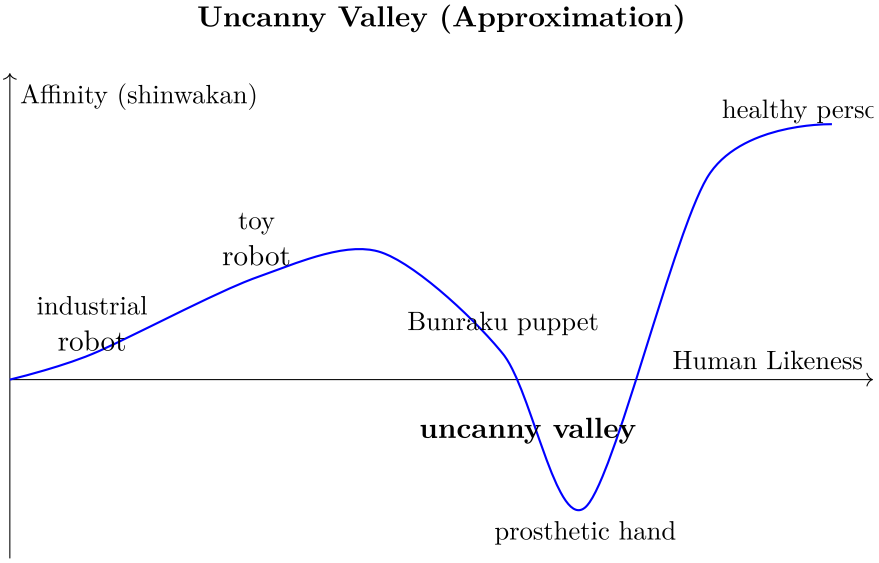


Figure 2.3: Approximate representation of the Uncanny Valley, adapted from [MMK12]

[ZLZ$^+$20] reviewed existing studies on the Uncanny Valley and found that the feeling of discomfort is not only caused by how human-like an agent looks or sounds, but also by mismatches in behavior and emotional expression. If a voice assistant sounds very human but responds in a mechanical or unnatural way, it may cause unease. This shows that both appearance and behavior need to match in order to avoid triggering the Uncanny Valley effect and to support user trust [ZLZ$^+$20].

In the context of VAs, the Uncanny Valley theory is particularly relevant. Voice assistants that employ highly human-like voices or personalities might initially enhance user trust and engagement. However, if these features are not coherently aligned—for example, if the voice sounds realistic but lacks emotional congruence or natural phrasing—users may experience cognitive dissonance or discomfort, thereby reducing trust.

Furthermore, the UV theory offers an explanation for why some users may prefer less anthropomorphic, more synthetic, or gender-neutral voices. These alternatives do not trigger the same mismatch between appearance (or sound) and behavior, avoiding the dip in comfort associated with the Uncanny Valley. This aligns with the S-O-R (Stimulus-Organism-Response) model, where the assistant's anthropomorphic cues (stimulus) are filtered through the user's percep-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

tion (organism), affecting their behavioral outcome—such as approach, trust, or avoidance. [HBL23, MR74].

Reeves and Nass [RN96] propose that computers, televisions, and digital agents are treated “as if” they were real people or places. Anthropomorphic cues like voice pitch, mannerisms, and tone trigger automatic social scripts in users.

The Computers Are Social Actors (CASA) paradigm, elaborated in subsequent works by Nass et al. [NSS94], extends these insights to voice-based interactions. *CASA paradigm*, developed by Nass et al., fundamentally challenges the assumption that human interactions with computers are purely rational or instrumental. Instead, it suggests that when computers exhibit even minimal social cues—such as using natural language, responding interactively, or maintaining a dialogue—people automatically apply social rules and behaviors typically reserved for interactions with other humans. To test this theory, Nass et al. conducted a series of five experiments designed to observe whether individuals treat computers socially, despite knowing they are machines. In one key experiment, participants interacted with a computer that later asked them to evaluate its own performance. Remarkably, participants were more polite when giving feedback directly to the same computer they had interacted with than when evaluating it on a different computer or on paper. This mirrors the human tendency to avoid direct criticism in face-to-face social interactions—suggesting the participants responded socially, even to a non-human agent. Another experiment tested ingroup bias by having participants evaluate a computer that supposedly shared their own characteristics (e.g., same gender or personality style) versus a different one. Again, the results showed that participants favored the “similar” computer, indicating that social identity cues were being applied even in human-computer contexts. These findings led to the conclusion that people unconsciously respond to computers using the same social heuristics they apply to humans—a mindless form of social interaction triggered by simple cues like language, voice, or turn-taking. Importantly, participants did not explicitly report believing that the computer had human-like qualities. This supports the idea that social responses to computers are automatic, not intentional or based on conscious anthropomorphism.

The *CASA paradigm* thus provides a foundational framework for understanding why people might treat AI voice assistants, chatbots, or other interactive technologies as social agents. It also helps explain phenomena such as users attributing gender, personality, or trustworthiness to voice interfaces—issues highly relevant to the investigation of anthropomorphism and trust.

Perception of Personality Traits from Voices and Faces

The Big Five personality model is one of the most established frameworks in psychology for describing human personality [MC87]. It comprises five broad dimensions:

- **Extraversion** — Sociability, assertiveness, and energy.
- **Agreeableness** — Trust, altruism, and cooperativeness.
- **Conscientiousness** — Organization, dependability, and discipline.
- **Neuroticism** — Emotional instability, anxiety, and moodiness.
- **Openness to Experience** — Imagination, creativity, and openness to new experiences.

These traits are considered relatively stable across time and cultures, and are commonly measured through validated instruments such as the NEO-FFI or BFI.

A recent study by Skuk et al. [SJW$^+$24] examined the perception of these Big Five traits from both vocal and facial cues. The researchers conducted three experiments:

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Experiment 1 (E1): Trait-average voices were created based on speakers' **self-rated** personality scores. Participants listened to pairs of synthesized voices and selected which one matched a trait-related description.

Experiment 2 (E2): Trait-average voices were created based on **personality impressions** by independent listeners. Participants completed the same task as in E1.

Experiment 3 (E3): Participants judged Big Five traits from facial composites based on prior work by Kramer and Ward [KW10], using pairs of averaged face images.

As of results, experiment 1 revealed limited ability to accurately perceive personality traits from self-rated voices. Exceptions included:

- **Extraversion**, which was accurately judged for female voices.
- **Neuroticism**, which was accurately judged for male voices.

In contrast, Experiment 2 showed robust and consistent above-chance accuracy for all five traits, indicating the presence of **vocal stereotyping**—i.e., shared but not necessarily accurate impressions.

Experiment 3 replicated prior findings that facial composites allow for reliable trait judgments, especially for extraversion, agreeableness, and neuroticism.

Table 2.1: Summary of Trait Discrimination Accuracy Across Experiments

Trait	**Voices (Self-Ratings, E1)**	**Voices (Listener Impressions, E2)**	**Faces (E3)**
Extraversion	Only for females (↑)	Very high accuracy (↑↑)	High accuracy (↑↑)
Agreeableness	Inconsistent	Moderate (↑)	Very high accuracy (↑↑)
Conscientiousness	At or below chance	High (↑↑)	Moderate (↑)
Neuroticism	Only for males (↑)	Moderate to high (↑↑)	High (↑)
Openness	Inconsistent	Moderate (↑)	At chance (–)

Legend: ↑ = above chance accuracy, ↑↑ = consistently above chance, – = not significantly different from chance.

The study demonstrates that while voice-based personality judgments based on actual self-assessments show low validity, impressions formed by listeners are consistent and stereotypical. Facial composites, as shown in prior research, remain more effective at conveying certain personality cues. Importantly, the results suggest that people rely more on **socially shared stereotypes** than on acoustically valid personality markers in voices. In the context of human-computer interaction, particularly in the design of AI-enabled voice assistants, understanding how users perceive personality traits from voices is crucial. As voice assistants increasingly become part of users' daily routines, their voice characteristics contribute significantly to user experience, trust, and engagement. Previous research suggests that users attribute personality traits—such as warmth, trustworthiness, or competence—to a voice assistant based solely on its vocal qualities [NB05]. The current study by Skuk et al. [SJW$^+$24] offers empirical insights into how such personality perceptions from voices can stem more from socially shared stereotypes than from any objective "truth" about the speaker or system. This highlights the potential for voice assistant designers to intentionally shape user impressions via vocal synthesis choices (e.g., pitch, tone, cadence), thereby influencing user trust, satisfaction, and even perceived competence of the system. Importantly, the study also warns that such impressions might not reflect actual system capabilities, raising ethical considerations about the manipulation of user expectations through vocal cues.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

2.2 Trust and Anthropomorphism in Conversational Agents

Trust in technology spans multiple dimensions: perceived competence, reliability, and ethical intent [LS04]. For VAs, anthropomorphism may enhance trust by amplifying relational cues ("this assistant is friendly and understands me") but can also create mismatched expectations. When voice quality or empathy is high but the system's actual competence is lacking, user trust may erode [SVK21].

Trust is a critical factor in the effective use of automated systems. In their seminal work, Lee and See [LS04] define trust in automation as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." [KSC+22, LS04]. This definition underscores the fact that trust becomes essential when users must rely on automation without fully understanding or controlling it.

Lee and See approach trust as a multidimensional construct influenced by factors spanning cognitive science, psychology, sociology, and human-computer interaction. They emphasize that users form trust judgments through three primary cognitive processes:

- **Analytical processes**, which involve deliberate reasoning and assessment of automation performance,
- **Analogical processes**, which rely on comparison to familiar systems or past experiences,
- **Affective processes**, which draw on emotional responses, especially under time pressure or stress.

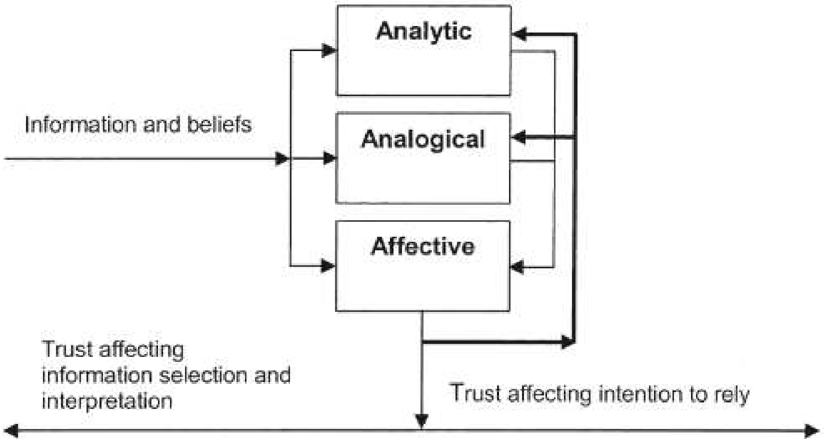

Analytic
Analogical
Affective
Information and beliefs
Trust affecting information selection and interpretation
Trust affecting intention to rely

Figure 2.4: A visual representation by Lee [LS04], which describes trust as an interaction between analytic, analogical, and affective processes. The analytic and analogical processes shape emotional responses, which then strongly influence the analytic and analogical processes themselves

Trust is not static but evolves through ongoing interactions with automation. As a result, trust can be miscalibrated: overtrust may lead to misuse (relying on the system when it is inappropriate), while undertrust may lead to disuse (ignoring a reliable system).

Components of Trust

Lee and See identify three core components that influence trust in automation:

- **Competence (C)**: The system's capability and technical skill in executing tasks.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

- **Reliability (R)**: The system's consistency and dependability over time.
- **Intent or Benevolence (I)**: The perceived alignment of the system's "intentions" with the user's goals—particularly important in systems perceived as social or intelligent.

Though not explicitly stated in the paper as an equation, trust can be conceptualized as a function of these dimensions:

$$T = \alpha C + \beta R + \gamma I + \epsilon$$

where T is trust, C is competence, R is reliability, I is intent or benevolence, α, β, and γ are weights reflecting their importance, and ϵ is an error term capturing unmodeled influences.

This formulation reflects the idea that trust is a weighted sum of perceived system qualities. In many contexts, especially when automation is anthropomorphized, the intent component (I) becomes particularly influential.

As conversational agents become increasingly anthropomorphic—exhibiting human-like voices, conversational styles, and emotional expressions—the boundaries between human and machine interlocutors begin to blur. [IL23] explore the implications of this human-likeness, particularly focusing on the emergence of trust or mistrust in user-agent interactions. They argue that while human-like features are often designed to enhance rapport and trust, they can paradoxically lead to suspicion when users become uncertain about the true nature of their conversational partner. The study emphasizes that the effectiveness of communication with a conversational agent does not inherently guarantee mutual understanding. Users might experience the interaction as smooth, yet still feel uneasy if they suspect the agent is not human. This suspicion becomes especially relevant in cases where users are unaware that the interlocutor is artificial. Once such suspicion emerges, it can disrupt the collaborative construction of meaning and undermine trust [IL23]. Furthermore, the authors suggest that the presence of sophisticated conversational AI may also influence human-to-human communication. As people become more familiar with artificial agents capable of mimicking human behavior, they may begin to question the authenticity of their human interactions. This can lead to an increase in interpersonal skepticism, particularly when someone exhibits atypical communicative behavior [IL23]. These findings are significant when evaluating the role of anthropomorphic cues such as voice, personality, and interaction style in the development of trust toward AI-enabled voice assistants. The study warns that anthropomorphism alone is not sufficient to ensure trust and may even have the opposite effect if users feel deceived or uncertain about the agent's identity.

Rest Energy (1980) by Marina Abramović and Ulay serves as a compelling metaphor for human-AI trust. In this performance, Abramović holds a bow aimed at her own heart while Ulay pulls the string—creating a fragile balance based entirely on trust [AU80, Pep18]. This moment of shared vulnerability parallels the dynamic between users and AI systems, where trust is extended despite risks and lack of control. The work captures the essence of reliance and potential danger, offering a visceral lens through which to explore human-AI interaction.

Trust is a critical component in determining the effectiveness and user experience of human-agent collaboration. This element serves as a foundational component, determining the extent to which users will rely on an artificial agent, particularly in contexts where collaboration, decision-making, or assistance is anticipated [SLAD15, Dar21].

Several studies have highlighted that the user's ability to *predict* and *influence* an AI agent's behavior directly impacts the degree of trust formed in such interactions. A comprehensive review of the extant literature reveals a consensus among researchers regarding the multifaceted nature of trust in human-agent interaction. The quality of the user experience is contingent not only on the agent's accuracy and competence, but also on how users perceive the agent's transparency, reliability, and especially predictability. Trust is cultivated when users feel confident in their

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Figure 2.5: Marina Abramović and Ulay, *Rest Energy*, 1980. Performance still. Courtesy of the Marina Abramović Archives [AU80]

ability to predict an agent's behavior and believe that they have some degree of control over the interaction. Consequently, the design of future AI systems and voice assistants must be informed by two principles: optimal performance and behavioral consistency and responsiveness that align with human expectations [DAHD21, DAHD20].

[DAHD21] assert that agents exhibiting predictable behavior significantly enhance trust, task performance, and reduce cognitive workload. Their findings, based on real-time collaboration scenarios, show that predictability is not merely a passive attribute but actively shapes users' perceptions and their willingness to delegate tasks to AI agents. When users can anticipate how an agent will act or respond, they experience a sense of control and reduced uncertainty, which fosters trust and increases efficiency in joint tasks.

Further expanding on this, [DAHD20] examined the combined roles of *predictability* and *reliability* in human-agent collaboration. Through a controlled experiment using an interactive task, they demonstrated that agents whose actions were easier to predict, irrespective of overall performance, led to *higher trust*, greater reliance, and *lower cognitive load*. Importantly, the study introduced a model that could infer trust levels from real-time behavioral data, opening the door for the development of adaptive systems that dynamically adjust their behavior to reinforce user trust.

Moreover, [La24] explored the nuanced dynamics between *AI transparency* and trust. Their case study on ChatGPT emphasized that transparency alone does not guarantee trust unless it also contributes to the user's ability to comprehend and forecast the AI's reasoning process. In other words, when users feel that an AI system is both transparent and predictable, their *perceived agency* in the interaction increases, which directly bolsters trust. The study underscores that trust is not static but rather *fluid and context-dependent*, shaped by how intelligible and controllable the AI appears during ongoing interaction.

2.3 Conversational AI: NLP, Chatbots, and Voice Interfaces

Conversational Artificial Intelligence (AI) encompasses a suite of technologies that enable machines to engage in human-like interactions through text or speech [XT25]. At its core, Conversational AI integrates Natural Language Processing (NLP), machine learning, and sometimes speech recognition and synthesis to facilitate seamless communication between humans and machines. This technology powers a range of applications, from simple chatbots handling customer service inquiries to sophisticated voice assistants like Amazon's Alexa, Apple's Siri, and Google Assistant [Hoy18]. The progression of these systems has been significantly shaped by the continuous development of NLP, which enables machines to comprehend, interpret, and produce human language. Early implementations relied heavily on rule-based systems with limited adaptability, responding only to predetermined inputs. As research advanced, the introduction of deep learning techniques and large language models allowed systems to interpret conversational context, handle ambiguity, and deliver more flexible and responsive interactions [Yag23].

An important evolution within Conversational AI has been the integration of voice interfaces, which have contributed to broader accessibility and user convenience. Through ASR, spoken language is transcribed into text, enabling the NLP component to analyze and derive intent. Subsequently, the appropriate response is generated and vocalized using Text-to-Speech (TTS) technologies. This pipeline has enabled the development of voice assistants that can interact in real time, offering a more fluid and hands-free user experience [RCIZ19]. The refinement of each layer within this process has been crucial to achieving natural-sounding interactions and minimizing latency.

The scope of Conversational AI continues to expand across various domains. In customer service environments, chatbots efficiently manage high volumes of requests, contributing to cost savings and improved user satisfaction. Within the healthcare sector, voice assistants assist patients by offering medication reminders, answering health-related queries, and even supporting mental health interventions. In education, intelligent tutoring systems leverage conversational agents to provide tailored instruction and feedback. Beyond these areas, the fusion of Conversational AI with the IoT has facilitated the rise of smart environments, where devices respond to spoken commands to control lighting, climate, and security systems [Bă23]. These use cases highlight the versatility and scalability of Conversational AI technologies.

Nevertheless, the widespread adoption of Conversational AI presents challenges that demand critical attention. Privacy and data protection remain fundamental concerns, particularly as these systems often access sensitive personal information. In parallel, ensuring ethical behavior and reducing bias in AI-generated responses are pressing priorities. Models trained on unbalanced or prejudiced datasets risk producing outputs that reinforce stereotypes or exclude certain user groups. Researchers and developers are therefore called to implement fairness-aware design principles, improve transparency, and build systems capable of handling diverse user needs with inclusivity and respect [Pot24]. Fig. 2.6 illustrates the processing flow of a voice assistant, showcasing the key stages involved in transforming a spoken user command into a meaningful system response. The process begins with Automatic Speech Recognition (ASR), where the assistant captures the user's voice input and transcribes the spoken words into machine-readable text. This transcription is then passed to the Natural Language Processing (NLP) module, which plays a critical role in interpreting the user's intent, analyzing the linguistic structure, and extracting relevant information from the input. Once the intent is understood, the Action Handling/Task Execution stage is triggered, where the assistant determines the appropriate response or action—such as setting a reminder, searching for information, or controlling a smart device. After the task is executed, the output (usually in the form of text) is converted back into audible speech through the Text to Speech (TTS) system, enabling the assistant to deliver a clear, natural-sounding response to the user [SPEN25]. This seamless flow allows for intuitive, real-time human-computer interaction through voice.

The progression of user interfaces, as shown in Figure 2.7, illustrates a clear trajectory toward more human-centered interaction paradigms. This historical shift—from Command-Line Inter-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

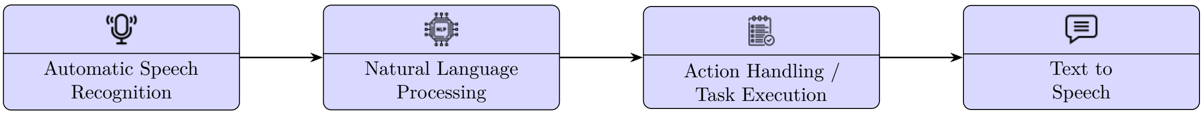

Automatic Speech Recognition
Natural Language Processing
Action Handling / Task Execution
Text to Speech

Figure 2.6: Voice Assistant Processing Flow

faces (CUI) to Graphical User Interfaces (GUI), and now to Natural User Interfaces (NUI)—mirrors the growing demand for intuitive, accessible, and seamless user experiences.

- **Command-Line User Interface (CUI):** Early interfaces were purely text-based and required memorization of commands, limiting usability to technically proficient users. Interaction was constrained and devoid of any anthropomorphic or social cues.
- **Graphical User Interface (GUI):** GUIs introduced visual affordances and interaction through pointing devices, significantly lowering the barrier to entry. While more intuitive than CUIs, GUIs remained transactional and lacked social or human-like qualities.
- **Natural User Interface (NUI):** NUIs use natural human behaviors such as touch, gesture, and increasingly, voice. Voice-based NUIs in particular represent a fundamental shift, enabling conversational interaction that mimics human dialogue. This evolution opens up new possibilities—and challenges—in the design of AI voice assistants, especially regarding anthropomorphism, trust, and user perception.

The rise of voice-based NUIs signifies a turning point where interface design intersects with human psychology and social cognition. Unlike GUI or CUI, voice assistants possess inherently anthropomorphic attributes—such as voice tone, gender cues, and conversational style—which can significantly influence user trust, perceived competence, and emotional engagement. Understanding how users interpret and respond to these attributes is essential for creating ethical and inclusive AI systems.

NUI is a system that enables interaction between humans and computers through intuitive actions that mirror everyday human behaviors, such as touch, gestures, or voice commands. The objective of a NUI is to minimize the cognitive load in human-computer interaction by eliminating artificial control devices, allowing users to engage with digital systems in a manner that feels natural and seamless [Int23]. They have gained increasing popularity, with voice-activated interfaces being among the most widely used. In particular, intelligent voice assistants such as Google Assistant, Amazon Alexa, Microsoft Cortana, and Apple Siri have become prominent examples of this technology, enabling users to interact with devices through speech-based commands. [RRARNMV21]

Voice assistants are software agents designed to recognize and process human speech, enabling users to interact with them through synthesized voices; such as Apple’s Siri, Amazon’s Alexa, Microsoft’s Cortana, and Google Assistant, which are commonly integrated into smartphones and smart home devices to facilitate a range of tasks, including answering user queries, controlling smart home automation, managing media playback, and assisting with organizational activities such as handling emails, to-do lists, and calendar events [Hoy18].

2.4 Gender as an Anthropomorphic Cue

Sex typically refers to the biological and physiological characteristics that define humans as male, female, or intersex, such as chromosomes, hormone levels, and reproductive organs [But24]. In contrast, gender encompasses the roles, behaviors, expressions, and social expectations that societies and cultures construct around individuals’ identities. It is shaped by cultural, historical,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

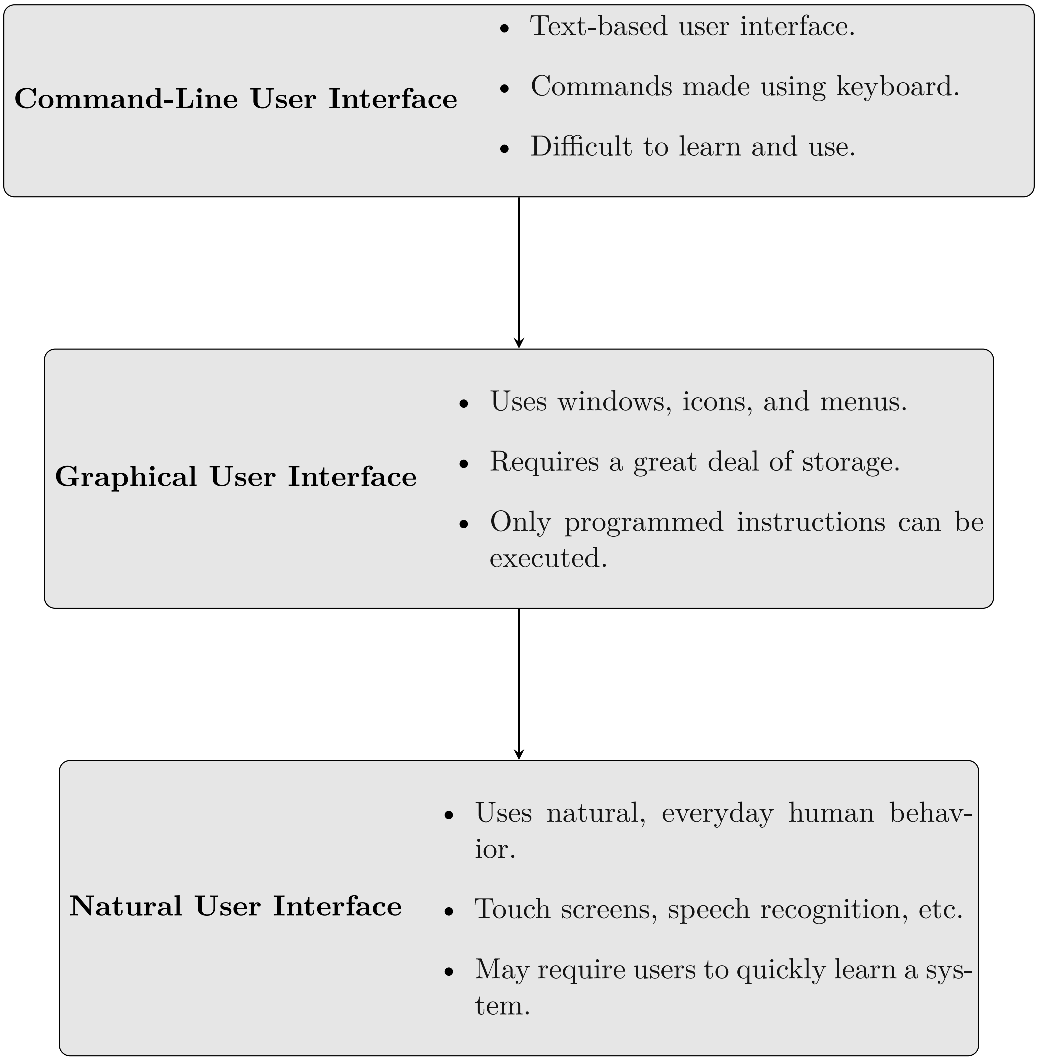

Command-Line User Interface
Text-based user interface.
Commands made using keyboard.
Difficult to learn and use.
Graphical User Interface
Uses windows, icons, and menus.
Requires a great deal of storage.
Only programmed instructions can be executed.
Natural User Interface
Uses natural, everyday human behavior.
Touch screens, speech recognition, etc.
May require users to quickly learn a system.

Figure 2.7: Evolution of User Interfaces (Adapted from RTI International, 2023).

and personal factors and may not necessarily align with an individual's biological sex. Despite their frequent treatment as binary categories, gender exists along a spectrum, and biological sex includes variations such as intersex. Although the two concepts are distinct, they are often conflated or used interchangeably, particularly in health-related data and research [KEK23]. Gender, by contrast, is a socially constructed concept that defines roles and expectations according to what a given society considers suitable for individuals based on their assigned sex at birth [KEK23]. The distinction between cisgender and transgender identities has been explored in prior research, where cisgender refers to "individuals whose gender identity aligns with the sex assigned to them at birth, while transgender encompasses those whose gender identity diverges from that assignment" [LSa21]. This conceptual differentiation serves as a foundation for understanding gender beyond binary categorizations and is crucial in discussions of inclusivity in technology and design. The design of voice assistants is not neutral; it often reflects and reinforces gendered assumptions embedded within technological development. Many mainstream voice assistants—such as Amazon's Alexa, Apple's Siri, and Microsoft's Cortana—are assigned feminine-sounding voices and names by default. This feminization of AI assistants is often justified by user preference studies that suggest people find female voices more pleasant or trustworthy. However, these preferences are socially conditioned and rooted in broader gender norms that associate women with roles of care, service, and emotional labor. As such, the default design choices in voice interfaces risk perpetuating outdated stereotypes by coding fem-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

ininity as subservient, supportive, and non-confrontational. [SK21] critique this phenomenon through their concept of the ”Smart Wife”—a term used to describe feminized digital assistants modeled after the 1950s housewife archetype. These assistants are portrayed as ever-available, polite, and helpful companions that perform domestic and emotional labor without question. The authors argue that such design choices not only normalize gendered divisions of labor within human-computer interaction but also reinscribe traditional gender roles into emerging technologies. By doing so, the industry reinforces the expectation that femininity is inherently tied to servitude and compliance, even within virtual contexts. Their call for a feminist ”reboot” of voice assistant design invites designers and developers to rethink how gender is operationalized in AI and to explore more diverse, inclusive, and critically aware alternatives. Nevertheless, to support more inclusive and adaptable human–robot interactions, platforms like Furhat are leading new approaches in social robotics. Furhat uses a back-projected humanoid face and multiple sensors to enable natural, real-time communication [fur25]. Its design allows it to quickly change its face, gender, and age while keeping realistic eye contact and lip-synced speech in over 30 languages. With an open-access SDK, Furhat has become a popular tool in research, used in more than 250 studies on topics like trust and healthcare. This shows how flexible and customizable robots can help researchers explore cultural differences and long-term user experiences [fur25].

2.5 (Non-)Binary Identities in Voice Assistants

The design of AI-enabled voice assistants is deeply influenced by social norms and cultural expectations, particularly with regard to gender. Many commercial voice assistants—such as Amazon’s Alexa, Apple’s Siri, and Google Assistant—default to feminine-sounding voices and names. These choices are not arbitrary but reflect a broader pattern in which femininity is associated with service-oriented and emotionally supportive roles. The consistent portrayal of these assistants as friendly, helpful, and polite reinforces the stereotype of the docile, always-available “digital secretary” [SK21]. Scholars have described this phenomenon as the emergence of the “Smart Wife”—a feminized, often flirtatious AI designed to carry out domestic labor and emotional support while remaining subservient to user commands [SK21]. This reflects a form of anthropomorphism, in which gender cues are deliberately embedded in technology to make it more relatable and acceptable to users.

Voice frequency plays a critical role in how users assign gender to voice assistants. Voice pitch, or F0, is one of the most immediate and influential cues in the perception of gender. Cisgender male voices typically range from 85 to 180 Hz, with an average around 110 Hz, while cisgender female voices range from 165 to 255 Hz, averaging around 200–220 Hz [Tit94, CW91, KNA16, GALA22].

These biological differences are rooted in vocal fold length and mass, which vary by sex. However, voice pitch is not fixed; for example, transgender individuals often shift their pitch through hormone therapy or vocal training. Trans women may achieve a speaking range between 150 and 250 Hz, while trans men often lower their pitch to approximately 120–140 Hz after testosterone therapy [Oat01]. In an effort to challenge binary gender norms in voice assistant design, projects like “Q”—the world’s first genderless voice—were developed using a target F0 range of 145–175 Hz, which lies between typical male and female frequencies [Vir19]. This range was selected to be perceived as androgynous across cultures and to provide a more inclusive alternative to gendered voice interfaces. The manipulation of F0 in synthetic voice design demonstrates how gender is constructed and operationalized within human-computer interaction, making it a key anthropomorphic cue that significantly shapes user perception and trust.

Higher-pitched voices are typically associated with femininity, while lower-pitched ones are perceived as masculine. This perception is rooted in long-standing social and biological associations, where pitch serves as an auditory signal for categorizing gender [CW91]. Designers often cap-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

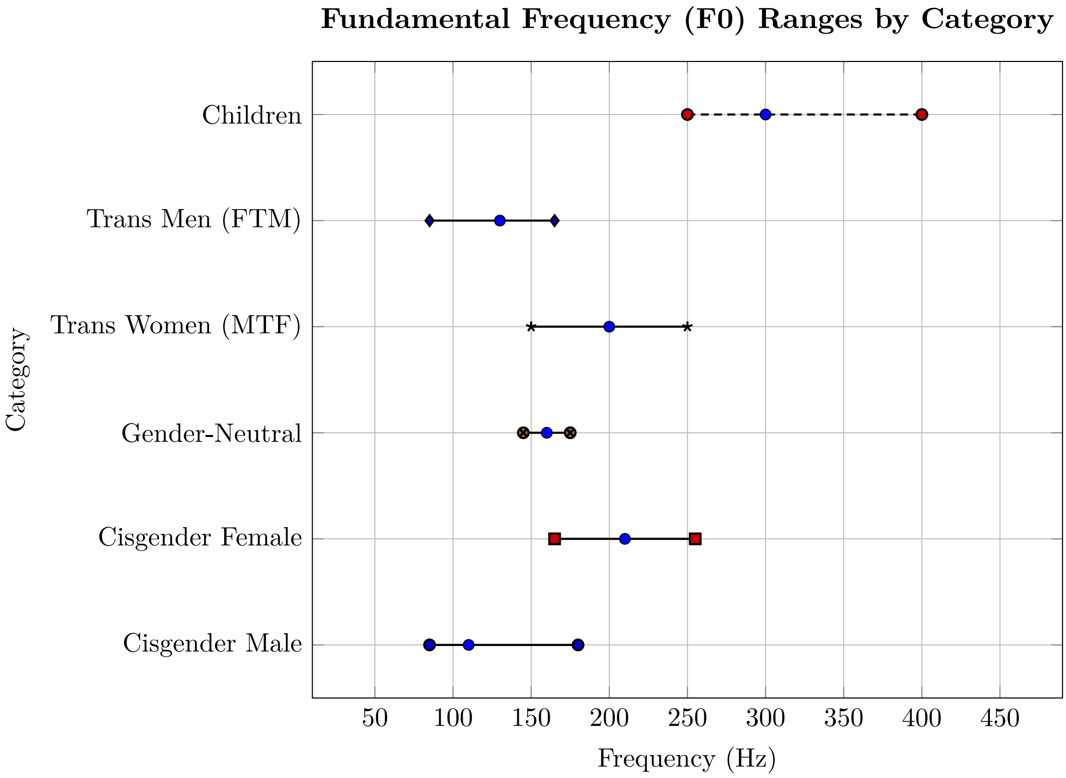


Figure 2.8: Estimated fundamental frequency (F0) ranges and midpoints across gender and age categories. Adapted from [Tit94, CW91, Oat01, Vir19].

italize on these associations by selecting voice frequencies that clearly indicate gender, thus enhancing anthropomorphic effects and user trust [MB07]. However, these design choices risk reinforcing binary gender norms and overlooking users who do not identify within the traditional male-female dichotomy.

The frequent use of feminized voices in voice assistants is often justified by industry research claiming that users find female voices more pleasant and trustworthy [GD19]. While such preferences may be statistically evident, they are shaped by cultural conditioning rather than objective user needs. As Sutton [Sut19] argues, the preference for female voices is not benign—it reflects and perpetuates gendered expectations in which women (or feminized voices) are expected to be agreeable, helpful, and compliant. In doing so, these AI designs encode social hierarchies and power dynamics into technological interfaces, subtly reinforcing existing gender inequalities.

Efforts to challenge these norms have led to the development of genderless or non-binary voice assistants. One prominent example is “Q,” the world’s first gender-neutral voice, developed through a collaboration between Copenhagen Pride, Virtue Nordic, and EqualAI [DHHP$^+$23]. Q was designed using a frequency range of 145–175 Hz—considered a neutral zone between traditionally male and female voice frequencies—to avoid binary categorization [Vir19]. Studies suggest that while some users find gender-neutral voices unfamiliar or uncanny at first, they hold promise in promoting inclusivity and disrupting gendered assumptions in AI design [BST21]. These developments are important not only for the representation of non-binary individuals but also for expanding the ethical scope of human-computer interaction.

The concept of gender in voice assistant design is not limited to voice frequency alone. It encompasses a broader set of cues such as speech patterns, personality traits, responsiveness, and even the names assigned to the assistant. These anthropomorphic cues interact with users’ social

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

expectations and cognitive biases, shaping their trust, satisfaction, and perceived intelligence of the AI [CMK12]. When these cues rely heavily on gender stereotypes, they risk reinforcing limited views of gender roles. As such, researchers and designers must critically reflect on how gender is operationalized in AI and consider the implications of these choices on diverse users.

A growing body of literature calls for more inclusive and socially responsible approaches to voice assistant design. This involves not only offering a wider variety of voice options, including non-binary and culturally diverse voices, but also rethinking the underlying interaction models that dictate how users engage with these technologies. The future of voice interfaces should be one in which gender is not defaulted to binary norms, but rather treated as a fluid and multifaceted aspect of user identity. As Strengers and Kennedy [SK21] argue, the goal is not to erase gender from AI, but to "reboot" it—to move beyond simplistic representations toward a more equitable, reflective, and empowering digital future.

2.5.1 Implications for Voice Assistants and Anthropomorphic Design

Modern voice assistants (VAs) such as Alexa, Siri, and Google Assistant are increasingly anthropomorphized through natural-sounding voices, conversational turn-taking, and even emotionally expressive tones. Anthropomorphic cues directly influence how users interpret the VA's intentions and benevolence—captured in the I component of the trust model.

When users perceive a VA as having human-like qualities, they may expect social competencies such as empathy, honesty, and responsiveness. This can elevate expectations around trustworthiness and ethical behavior, even if the VA is purely rule-based. While anthropomorphic design can enhance engagement and trust, it also raises concerns about overtrust—users may assume the system understands more than it does.

Therefore, designers must calibrate user trust carefully. Strategies to achieve this include:

- Increasing **transparency** about system capabilities and limitations,
- Offering **feedback mechanisms** for error correction,
- Avoiding excessive social mimicry that could mislead users about the system's true competence or autonomy.

Understanding trust through the lens of competence, reliability, and intent allows researchers and designers to build voice interfaces that users trust appropriately—enhancing usability without creating false expectations.

Gender Stereotypes and Bias Reinforcement. Research suggests the "feminine assistant / masculine authority" dichotomy in tech. Nass and Brave [NB05] found that identical scripts read by a "female" voice were perceived as more "supportive," whereas "male" voices were rated more "competent."

These gendered roles are often embedded in design choices, influenced by market assumptions and developer demographics [UNE19, SK21].

2.5.2 Emerging Alternatives and Ethical Considerations. Inclusive Design in Voice Assistants

In response, designers have proposed *gender-neutral* or *non-binary* VAs to challenge binary voice defaults. The "Q" voice project [Vir19] offers a frequency range between typical male and female tones (145–175 Hz). Early evaluations show 42% fewer gender attributions, though 27% of users initially reported discomfort. UNHCR emphasizes digital inclusion as a fundamental principle in its Digital Transformation Strategy 2022–2026. [UNH22] According to the report, ensuring equitable access to digital services is essential, particularly for marginalized groups such as refugees, people with disabilities, and individuals with limited digital literacy [UNH22]. This is highly relevant to the design of inclusive voice assistants. These systems must account for a wide range of user needs, including diverse language abilities, gender identities, and cognitive or physical limitations. The strategy underlines the importance of contextually relevant and culturally sensitive solutions, reinforcing the value of offering gender-neutral voice options and localized speech outputs. Additionally, the strategy highlights the need for human-centered design by involving end-users directly in the development process—an approach that aligns with best practices in building trustworthy AI systems. Drawing inspiration from this framework, inclusive voice assistant design should prioritize low-barrier interaction methods, multi-language support, and accessibility features, while also considering offline capabilities for low-resource settings. In another work by Spiel et al. [SFKF19] they critically evaluate how autistic children's agency has been represented within technology research, underscoring the necessity for more inclusive and participatory research methods. They identify a frequent lack of meaningful involvement from autistic users, resulting in technologies that inadequately address their specific needs and preferences. The paper strongly advocates for acknowledging autistic children as active agents in the technological design process, arguing that inclusive practices significantly improve technology's relevance, effectiveness, and user trust. The insights from [SFKF19] reinforce this approach by emphasizing the importance of involving diverse user groups—including those traditionally marginalized—in the design process. By advocating for active participation and agency of users, [SFKF19] underline the argument presented here that inclusive and participatory design methodologies can mitigate bias, enhance user trust, and ultimately lead to more effective, empowering, and inclusive AI technologies.

Design Recommendations Based on UNHCR's Digital Inclusion Principles

- Implement voice assistants that support multiple languages and dialects, including those spoken by vulnerable or displaced populations.
- Offer voice options that are gender-neutral to reduce the reinforcement of gender stereotypes and to improve inclusivity for non-binary users.
- Design interfaces with low-literacy users in mind—prioritize intuitive interactions, natural speech recognition, and clear, slow, non-technical language.
- Involve marginalized users in the design and testing process (e.g., refugees, people with disabilities, older adults) to better understand their needs and perceptions of trust.
- Ensure accessibility for people with disabilities, such as compatibility with screen readers or options for tactile/visual feedback where appropriate.
- Develop lightweight or offline-compatible assistant versions to ensure accessibility in areas with low connectivity or limited resources.

A study by Masina et al. [MOP$^+$20] looked at how people with different disabilities use voice assistants like Google Home. The results showed that how well someone could use the assistant depended a lot on their thinking and speaking abilities. For example, people who could repeat sentences well and had higher scores on the Mini-Mental State Examination (MMSE)[1] were

---

[1] The MMSE is a short test used to measure a person's cognitive function, including memory, attention, and language skills.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

better at using the system. The study also found common problems like saying commands in the wrong way or at the wrong speed. These findings show how important it is to design voice assistants that work for many different users: this study supports the idea that good design helps people feel more confident and comfortable when using voice technology.

The integration of inclusive design principles in voice assistant development has gained significant attention in recent years. A notable contribution in this domain is the work by [DHHP+23], which presents a methodology for creating a non-binary TTS voice. The authors introduce ”Sam,” a TTS voice developed with input from non-binary and transgender individuals, aiming to provide a more inclusive user experience. This approach underscores the importance of considering diverse gender identities in technology design, aligning with the broader objectives of exploring inclusive practices in human-computer interaction.

Additionally, [Spi21b] shows how many technologies are built with binary gender in mind, which can make it difficult for non-binary people to use them comfortably. For example, digital forms often only offer ”male” or ”female” as gender options. This not only excludes non-binary users but can also cause emotional stress. The paper highlights the importance of designing systems that are more inclusive and recognize different gender identities. These ideas are helpful when thinking about how users relate to AI voice assistants and the role gender might play in that interaction [Spi21b].

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 3

Methodology

3.1 Study Design

A mixed-method experimental design was implemented, involving 30 participants ($n = 30$), all of whom were students from universities in Vienna. Each participant interacted with only one voice assistant throughout the session in order to avoid potential bias and to isolate the effects of individual voice characteristics. The three voice assistants varied in perceived gender: one with a typically male voice, one with a typically female voice, and one designed to be gender-neutral. Participants completed three tasks with the same voice assistant to ensure consistency in their experience and to minimize confounding variables related to voice variation.

To further reduce bias and increase experimental control, participants were randomly allocated to either a *conversational-style* or a *baseline* (neutral) interaction condition. This allocation was managed automatically within the experiment software and remained concealed from both the participant and the researcher until the debriefing stage. Additionally, to minimize order effects, the three interaction prompts were presented in a different randomized order for each participant.

All study procedures and materials received prior approval from the supervising faculty. The sessions took place in a quiet, controlled lab environment to ensure consistent conditions across all participants. Participation was voluntary and formed part of the course's experiential learning component, enabling students to engage in a practical research study while contributing meaningful data for this thesis on user trust and perception in relation to anthropomorphism in voice assistants. Participants were evenly divided into three groups of 10:

- **Group I**: Participants interacted with a typically male voice assistant (OpenAI Cove).
- **Group II**: Participants interacted with a typically female voice assistant (OpenAI Juniper).
- **Group III**: Participants interacted with a gender-neutral voice assistant (OpenAI Breeze).

This study employed a mixed-method experimental design to investigate how different voice types in AI assistants influence user trust and perception. Participants were randomly assigned to one of three conditions (typically male, female, or gender-neutral voice) and completed a series of structured interaction tasks. These tasks involved problem-solving scenarios in high-stakes domains, including spacecraft troubleshooting, emergency medical response, and parenting advice—contexts chosen specifically because they require a high degree of trust in the assistant's competence and reliability.

Quantitative data were collected through pre- and post-interaction surveys, measuring factors such as prior experience with voice assistants, perceived trustworthiness, and anthropomorphic attributes. Qualitative data were obtained through persona visualization exercises—where participants were asked to draw or describe the imagined assistant—and through short open-ended interviews conducted at the end of each session. The names or any other information about voice assistants were not disclosed to the participants.

Insights from the qualitative responses were further synthesized to develop a proposed *inclusive design framework* as shown in the final results, figure 5.5, presented as a conceptual graph. This framework maps key user preferences and expectations for an ideal voice assistant, based on recurring themes in how participants envisioned trustworthy, relatable, and inclusive AI voices. It serves as a user-informed foundation for designing future voice assistants that are not only technically functional but also socially inclusive and aligned with diverse user needs.

3.1.1 Participants

Thirty individuals completed the study ($N = 30$). Participants self-identified across a spectrum of genders: 15 identified as men (50%), 13 as women (43%), one participant identified as non-binary (3%), and one preferred not to disclose their gender identity (3%). The age distribution was predominantly within the 18–34 range, with 13 participants aged 18–24 years (43%), 16 participants aged 25–34 years (53%), and one participant aged 35 or older (3%). In terms of educational background, the sample was relatively well-educated: 24 participants (80%) held at least a bachelor's degree, including four with a master's degree. The remaining six participants (20%) reported holding an upper-secondary school qualification, aligning with the academic level expected from enrolled university students.

Recruitment for the study was conducted over a three-week period. Participants were recruited from the *User Research Methods* course offered at the university, targeting first-year Master's students with a background in human-computer interaction, UX research, or related fields. At the beginning of the semester, I introduced the study in person during one of their initial lectures, providing a brief overview of the topic, the structure of the study, and its relevance to the course content. This live introduction also gave students the opportunity to ask questions directly and clarify any uncertainties before deciding whether to participate.

Following this introduction, recruitment materials were shared via the course's online platform, including a Calendly link that allowed students to book a time slot of their choice. The sessions were conducted in a controlled lab environment. Participation was voluntary and contributed to the course's experiential learning component. Through this process, students not only gained firsthand experience as study participants but also deepened their understanding of research design and ethical considerations—skills directly relevant to their academic development. Their participation also provided valuable empirical data for this thesis, which investigates user perceptions of voice assistants with varying anthropomorphic characteristics.

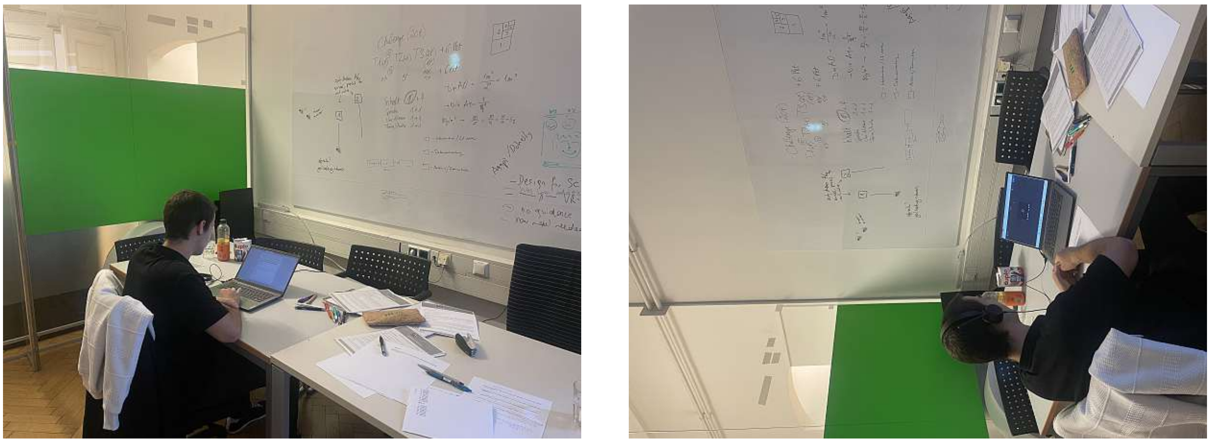
3.1.2 Study Procedure

The study followed the following steps, as shown below, and Figure 3.1 shows a participant during the study.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1. Participants read the informed consent form, agreed to participate, and signed two copies (one for the researcher and one for themselves). As shown in Appendix A
2. Participants filled out a pre-interaction survey anonymously, capturing baseline data on their prior experience with voice assistants and trust expectations. As shown in Appendix F.1
3. Each participant received the first task written on paper and gave it to the voice assistant, listening to the response through headphones without interruption and without being informed about the voice assistant's gender.
4. The same process was repeated for the second and third tasks. As shown in Appendix E
5. Participants were provided with a blank sheet of paper to "draw" the persona they imagined while interacting with the voice assistant. If they could not visualize a persona, they wrote down the main characteristics they associated with the voice assistant (e.g., "white blonde-haired woman, 30 years old, thin"). If they did not imagine a persona, they skipped this step and proceeded to the post-interaction survey. As shown in Appendixes B, C & D
6. Participants completed a post-interaction survey to assess their perceptions of trust and anthropomorphic attributes. As shown in Appendix G.1
7. A short open-ended interview was conducted and recorded to gather additional insights. As shown in Appendix H
8. Participants were thanked and greeted for their participation.



Participant filling out the pre-interaction survey

Participant listening to a voice assistant task via headphones

Figure 3.1: User study setup showing a participant during the pre-survey phase and interacting with the voice assistant task

The user study took place at the TU Wien Institute located at Argentinierstraße 8, 2nd floor (E193-05), in the monitored Fishbowl room. Participants were able to choose their preferred time slots one to two weeks in advance via the Calendly booking tool. The study was conducted over the course of the last two weeks of October 2025 and the first week of November 2025. As of Software and Hardware Requirements:

- Each participant used a desktop computer and over-ear headphones to interact with the voice assistant tasks in a controlled and distraction-free environment.

- Pre- and post-interaction surveys were administered via Google Forms.
- Scheduling of study sessions was managed through the Calendly platform, allowing participants to choose convenient time slots in advance.
- The voice assistant tasks were pre-recorded and played through a media player application to ensure consistency across all sessions.
- All sessions took place in a monitored study room equipped with necessary audio hardware and a stable internet connection.

3.1.3 Task Scenarios

Participants interacted with the voice assistants by completing three tasks. Task prompts were randomly assigned to prevent bias introduced by a fixed order.

Table 3.1: Task Scenario: Parenting Advice

User study participant: *"As a future parent, I would have one concern about my child having trouble focusing on their homework. How can I help them stay motivated?"*

Voice Assistant: *"That's a really common concern and it's great that you're thinking ahead. A few strategies could help keep your future kid motivated with their homework. First, try to create a routine. Having a set time for homework helps build a habit and kids usually do better with consistency. You can also break tasks into smaller pieces so it feels less overwhelming. Second, make sure to encourage them to take breaks. Short, regular breaks can actually boost focus. Another idea is to celebrate small wins. Positive reinforcement can go a long way. Even something small like a sticker or extra playtime can motivate them. How do you feel about trying those approaches when the time comes?"*

The voice assistant's responses to these prompts were pre-recorded to ensure consistency.

3.2 Independent Variables

- **Voice Gender:** The primary variable manipulated, with three levels: male, female, and gender neutral.
- **Task Scenario:** Contextual factors including parenting advice, problem-solving, and emergency medical response.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Table 3.2: Task Scenario: Problem-Solving (Spacecraft Troubleshooting)

User study participant: *"One of the engines of the spacecraft is overheating. What should I check to fix it?"*

Voice Assistant: *"If a spacecraft engine is overheating, the first things to check would be cooling systems and airflow. Ensure that the cooling mechanisms, like radiators or thermal control systems, are functioning properly and not blocked or damaged. Also, check for any software or sensor issues that might be misreading temperatures. Finally, verify fuel flow, since any irregularities in fuel injection or combustion could contribute to overheating. Do you know if this is happening during specific maneuvers or consistently?"*

Table 3.3: Task Scenario: Emergency Medical Response

User study participant: *"A person has fainted but is still breathing. What should I do to help them recover?"*

Voice Assistant: *"If someone has fainted but is still breathing, here's what you can do to help them recover. One, lay them down on their back and raise their legs slightly if possible. This helps improve blood flow to the brain. Two, loosen any tight clothing around their neck or waist to make it easier for them to breathe. Three, check their airway to ensure it's clear, but avoid putting anything in their mouth. Stay with them until they regain consciousness. Speak calmly and encourage them to stay lying down until they feel fully better. If they don't regain consciousness quickly, or if you're concerned, it's important to seek medical help. Would you like me to help clarify any of these steps?*

Table 3.4: Task Scenario: Problem-Solving (Spacecraft Troubleshooting)

3.3 Dependent Variables

- **User Trust:** Assessed through Likert-scale survey questions and behavioral observations, focusing on participants' belief in the assistant's ability to perform tasks accurately and reliably.

- **Perception of Anthropomorphic Attributes:** Measured through survey responses, including ratings on personality, competence, interaction style, and voice tone.

Table 3.5: Task Scenario: Emergency Medical Response

3.4 Data Collection

3.4.1 Pre-Interaction Survey

The pre-interaction survey collected:

- General information about participants.
- Prior experience with AI voice assistants.
- Pre-existing trust levels.
- Expectations regarding AI voice assistants.

Detailed questionnaire can be seen in the table 3.6

Table 3.6: Pre-Interaction Trust Questionnaire grouped by aspects

Aspect	**Id**	**Question**
Familiarity/Usage	pre1	For what purposes do you typically use voice assistants?
Trust	pre2	How much do you trust AI voice assistants to provide accurate information?
Comfort	pre3	How comfortable are you with AI voice assistants making decisions on your behalf?
Privacy	pre4	How much do you believe AI voice assistants respect your privacy?
Voice Characteristics	pre5	How important is the gender of the voice assistant to you?
Voice Characteristics	pre6	How important is it for the voice assistant to sound natural and human-like?
Trust	pre7	What characteristics of a voice assistant would make you trust it more?
Expectations	pre8	What do you expect your overall experience to be like?

3.4.2 Post-Interaction Survey

The post-interaction survey gathered data on:

- Trust and satisfaction with the voice assistant.
- Perceptions of anthropomorphic traits.
- Open-ended feedback about the interaction experience.
- Optional task: Draw or describe the imagined persona of the voice assistant.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Table 3.7: Post-Interaction Trust Questionnaire grouped by aspects

Aspect	**Id**	**Question**
Trust	post1	Your overall experience interacting with the AI voice assistant?
Comprehension	post2	How easy or difficult was it to understand the AI voice assistant?
Trust	post3	How much did you trust the AI voice assistant to provide accurate info?
Confidence	post4	Confidence that the assistant's responses were reliable?
Personality	post5	The assistant had a unique personality that influenced trust?
Competence	post6	The assistant seemed knowledgeable and competent?
Tone/Engagement	post7	The assistant communicated naturally and engagingly?
Tone/Trust	post8	The assistant's voice tone made it easy to trust responses?
Comfort	post9	How comfortable were you with the assistant's persona?
Gender Matching	post10	Would you trust an assistant more if its gender matched yours?
Tone Appropriateness	post11	Tone appropriateness for parenting advice?
Task-Specific Trust	post12	Trust in assistant's technical advice for problem-solving?
Tone Appropriateness	post13	Tone appropriateness in an emergency situation?
Confidence	post14	Confidence in response for a real emergency?
Trust Focus	post15	In which task did you feel the most trust?

3.4.3 Observational Data

Behavioral cues, such as hesitation, engagement, and verbal reactions, were observed during the study to supplement the assessment of trust in the AI voice assistants. These non-verbal indicators provided valuable contextual insights that might not have been fully captured through self-reported data alone. Hesitation was noted in moments of uncertainty, especially when participants were unsure about the assistant's capabilities or when they encountered unexpected responses. Engagement was assessed through participants' body language, tone of voice, and willingness to continue interacting with the assistant. Verbal reactions—including laughter, sighs, or expressions of frustration—were also noted, as they offered direct reflections of participants' emotional responses and levels of comfort or discomfort. The observational data was recorded in field notes during each session and later used to triangulate findings from post-interaction interviews and questionnaire responses. This qualitative layer enriched the overall understanding of how trust and perceived anthropomorphic traits manifested in real-time interactions.

3.5 Ethical Considerations

Prior to signing the consent form, participants were informed about the study's purpose and procedures. The subjects were assured of anonymity and the right to withdraw from the study at any time without consequence.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

3.6 Data Analysis

As part of data analysis, two main methods were implemented. Quantitative analysis, where descriptive statistics summarized trust ratings and perceptions of anthropomorphic attributes across different voice conditions and tasks. Qualitative analysis, where for the analysis of interview data, thematic analysis was used to identify patterns and recurring themes in participant perspectives across the three voice assistant conditions (typically female, typically male, and genderless voices). Thematic analysis is well-suited for exploring subjective experiences and meaning-making, particularly in user-centered research involving open-ended data [CB13].

This study followed the six-phase approach of Braun and Clarke to thematic analysis: (1) familiarization with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report [CB13], as illustrated in Figure 3.2 The process was conducted manually using Microsoft Excel to code interview transcripts and to systematically organize participant responses.

An inductive (bottom-up) coding approach was adopted, allowing themes to emerge organically from the data rather than being constrained by pre-existing theoretical constructs. This decision ensured a more authentic representation of participant voices and minimized researcher bias in the interpretation of trust, voice tone, and gender-related impressions.

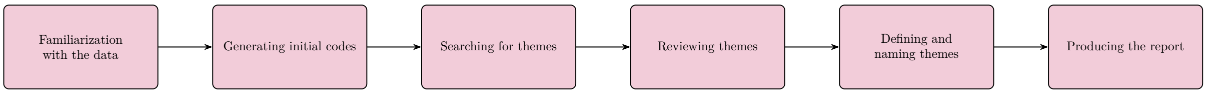


Figure 3.2: Braun and Clarke's six-phase approach to thematic analysis [CB13]

3.6.1 Hypotheses

1. **Hypothesis I:** Users will display higher levels of trust in AI voice assistants with gendered voices (male or female) compared to gender-neutral voices, across different task scenarios.
2. **Hypothesis II:** Anthropomorphic attributes such as personality, competence, interaction style and voice tone will be perceived more positively in gendered voice assistants (male or female) compared to gender-neutral voice assistants–resulting to higher levels of trust.
3. **Hypothesis III:** Users might display gender-preferences/expectations when interacting with AI voice assistants.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

CHAPTER 4

Results and Analysis

The study provided insights into the role of voice gender in shaping user trust in AI voice assistants across different task scenarios. By investigating three main AI-enabled voice assistants—male, female, and gender-neutral voices—the research clarified how trust is influenced by the assistant's gender in specific contexts such as parenting advice, problem-solving, and emergency medical response.

Additionally, the findings revealed how users perceive anthropomorphic attributes, such as personality, competence, interaction style, and voice tone, across different voice types. These insights contribute to a deeper understanding of how human-like traits manifest in AI assistants and their impact on trust. Furthermore, the study explored users' gender preferences/expectations when interacting with AI assistants, uncovering potential biases or preferences.

The results inform the design of more inclusive and trusted AI systems, particularly regarding gender-neutral options that may reduce biases toward gendered AI voice assistants. The findings also address the influence of anthropomorphic biases[1] on users' willingness to engage with AI in the future.

4.1 Participant Demographics

A total of 30 participants took part in this study, with 10 individuals assigned to each of the three experimental conditions. Participants were divided into three groups based on the voice assistant they interacted with: Group 1 (Cove, male-associated voice), Group 2 (Juniper, female-associated voice), and Group 3 (Breeze, gender-neutral voice). The participants' ages ranged from 18 to over 35, with the majority falling within the 18–34 age range. Educational background varied, but most participants held a bachelor's degree. Gender distribution included both male and female participants, along with a few identifying as non-binary or preferring not to disclose their gender. Prior experience with AI voice assistants was common among the participants, with varying levels of familiarity and trust toward AI systems before the study.

4.2 Pre-Interaction Survey Results

To understand participants' initial attitudes toward AI voice assistants prior to engaging with the study, a structured pre-interaction survey was conducted. This section summarizes the

---

[1] Anthropomorphic bias refers to the tendency to ascribe human-like characteristics, typically mental properties or agency, to entities that do not possess them [Joh18]

participants' familiarity, expectations, and trust perceptions using a series of Likert-scale questions.

Overall, participants showed a moderate level of familiarity with AI-enabled voice assistants, with the average familiarity rating being 3.67 out of 5. In terms of general trust, the average rating for trusting AI voice assistants to provide accurate information was 2.93, while comfort with the assistant making decisions autonomously was slightly lower, averaging 2.47. These figures suggest a cautiously optimistic attitude toward AI systems—participants acknowledged their functionality but were hesitant to delegate decision-making.

Privacy concerns were evident: the average perception of whether voice assistants respect user privacy was only 2.17. This points to a gap in user confidence regarding data handling and security.

Interestingly, the perceived importance of the assistant's gender was relatively low (mean = 1.87), indicating that most users were not strongly concerned with whether the assistant sounded male, female, or gender-neutral. On the other hand, the naturalness and human-likeness of the assistant's voice was rated quite highly, with a mean score of 4.07. This highlights the importance of emotional tone and realism in fostering user engagement and comfort.

The participants' overall expectations for the study interaction leaned positive, with a mean of 3.70. This suggests a generally open and curious attitude toward the upcoming AI voice interaction sessions.

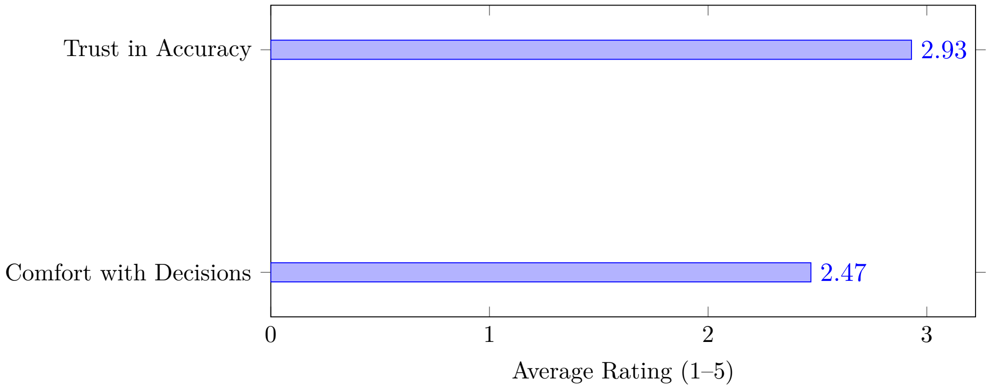


Figure 4.1: Average participant ratings of trust in accuracy and comfort with decision-making by AI voice assistants

These baseline metrics are critical for contextualizing the results presented in later sections. The data confirms that while participants are already familiar with AI technologies and open to interaction, there remains a considerable trust and privacy gap that may influence their perception of voice assistant behavior and reliability during the study.

4.3 Perceived Anthropomorphic Representations

To explore how users mentally visualize voice-based artificial agents, this study included a supplementary task where participants were asked to either draw or verbally describe the "person" they imagined while interacting with a voice assistant. This qualitative component aimed to uncover perceived anthropomorphic characteristics—such as gender, age, appearance, personality, and level of human-likeness—that may not be immediately accessible through quantitative surveys or interviews.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

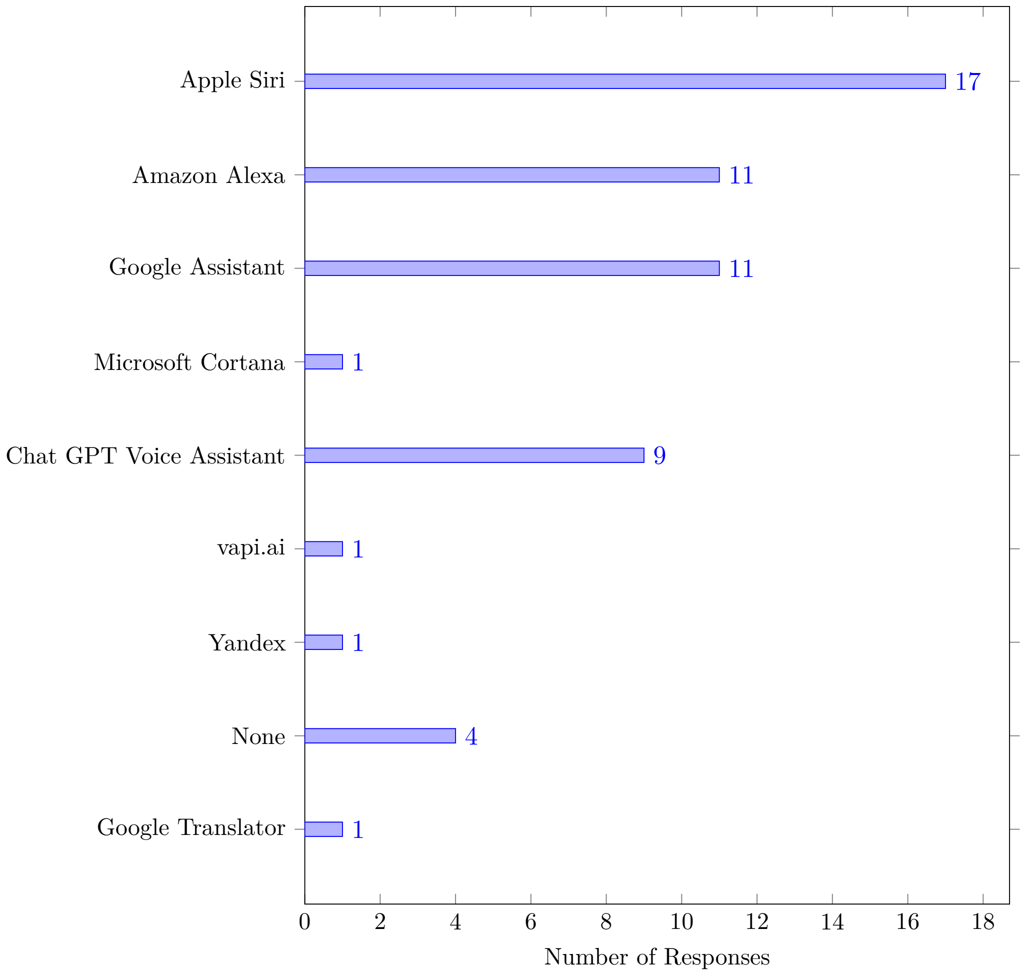


Figure 4.2: Results on the previous usage of different AI voice assistants from study participants

Anthropomorphism, the attribution of human-like qualities to non-human agents, has long been recognized as a natural cognitive tendency, especially in contexts involving social cues such as speech, emotion, or intentionality [EWC07]. Voice is one of the most socially charged cues in human communication, and prior studies have shown that even subtle differences in vocal pitch, tone, and cadence can lead users to form strong impressions of the speaker's identity, gender, age, and even emotional state [NMG97, EKB⁺12].

In human-computer interaction (HCI), the concept of "media equation" theory posits that people treat computers and digital agents as social actors when these systems exhibit human-like attributes [RN96]. Drawing on this principle, several studies have demonstrated that voice alone can evoke rich mental models of the system, including assumptions about competence, friendliness, trustworthiness, and other social dimensions [LN05, GGW10]. This is particularly relevant in the design of voice assistants, where the auditory channel is often the primary (and sometimes the only) interface.

The task of drawing or describing the imagined entity was inspired by projective and interpretive methods often used in social and cognitive psychology [Gof59]. Visualizations and descriptive narratives allow researchers to tap into users' internalized schemas and biases, surfacing perceptions that might otherwise remain implicit. In the context of AI, such techniques have been successfully used to analyze how users relate to and humanize robotic or virtual agents [Fin12, SWLF19].

By analyzing these representations, this study aims to reveal how anthropomorphic traits are

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

projected onto voice assistants in the absence of visual embodiment. Furthermore, it provides insight into whether and how voice modality alone can trigger socially meaningful inferences. The data suggest that the voice alone was sufficient to evoke rich mental imagery. Gendered voices led to clearer human projections, often complete with imagined age and personality. In contrast, the genderless voice assistant often resulted in robotic or ambiguous representations, echoing findings from Eyssel et al. [EKB$^+$12] about how similarity and familiarity influence anthropomorphic projections. This has important implications for the design of voice assistants depending on the intended emotional and social affordances. The following subsections present and analyze the collected representations, highlighting key themes and patterns observed across different voice assistant profiles. A selection of participant-generated representations is shown below. Where no drawing was provided, descriptive annotations have been transcribed. These artifacts illustrate the diverse mental models evoked by different voice profiles.

4.3.1 Representation: Female Voice

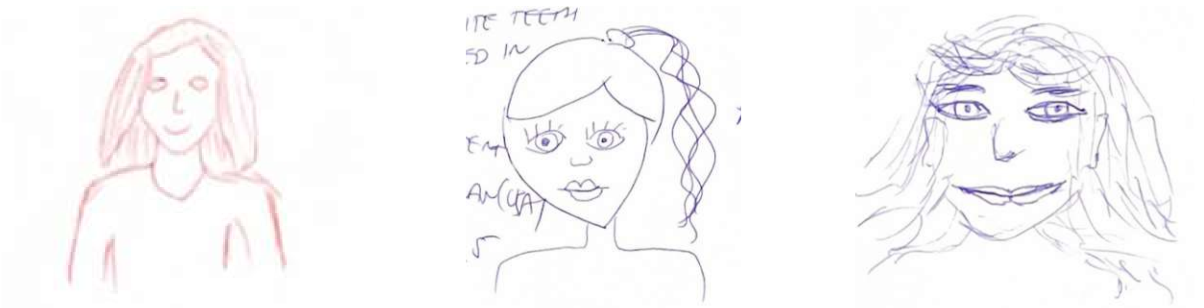
Participants interacting with the typically female voice assistant consistently assigned a strong gender identity. The assistant was overwhelmingly perceived as a young woman, often between the ages of 21 and 25. Common descriptors included “young,” “white,” “blond,” “long hair,” “clear teeth,” and “huge smile.” These physical and aesthetic references reveal that participants not only anthropomorphized the system but also projected gendered beauty standards and culturally familiar stereotypes onto it.

In addition to appearance, personality traits were frequently attributed. Words such as “friendly,” “confident,” “calm,” “smart,” and “competent” suggest that the assistant was seen as both emotionally supportive and intellectually capable. Interestingly, participants also referenced behavioral and style cues—such as “seamlessly ironed blouse” or “no accessories”—which reinforce the assistant’s perceived attentiveness, organization, and care-oriented demeanor.

This tendency to attribute both nurturing and polished qualities reflects broader social stereotypes of female service personas. Prior research has shown that gendered voice cues can activate social schemas that align with culturally constructed roles, particularly those related to caregiving, warmth, and interpersonal support [NMG97, EKB$^+$12]. Female-voiced assistants are often interpreted not only as more empathetic but also more trustworthy and approachable [LN05]. The female voice in this study, in evoking vivid mental imagery and emotional familiarity, supports the hypothesis that anthropomorphic projections are intensified when the voice aligns with users’ internalized gender norms and expectations [RN96, EWC07].

Overall, the typically female voice assistant elicited the richest anthropomorphic characterizations—both visual and emotional—indicating a heightened level of personalization and engagement. Figure 4.3 shows examples of participant-genderated representations after interacting with the system.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



P1: "White, US-American, young female."

P7: "White, blond, white teeth, huge smile, interested in appearance, but smart and competent. Young American girl, 21 – 25 years old."

P5: "Female tone, sounds familiar and confident, advice sounded helpful in these cases."

Figure 4.3: Examples of participant-generated representations of the voice assistant – Open AI Juniper

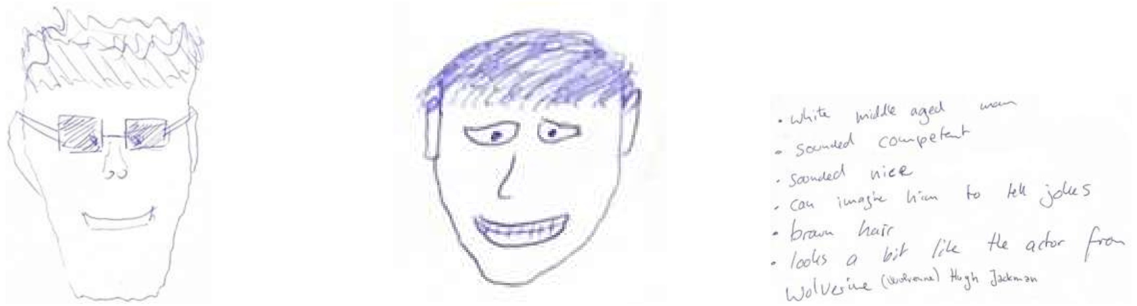
4.3.2 Representation: Male Voice

Participants interacting with the typically male voice assistant consistently recognized and labeled the voice as male. Descriptions commonly referred to a "typical office guy" or a "middle-aged white man," with age estimates ranging from the early 30s to late 40s. This reflects a clear gendered attribution combined with demographic markers such as race and age. Physical appearance was often noted, with participants including descriptors such as "short hair," "buzz cut," "bearded," "tanned," and "black hair." These visualizations suggest that listeners mapped vocal cues onto familiar social archetypes, possibly drawing from media or professional stereotypes. One participant even referenced a "Hugh Jackman lookalike," highlighting the influence of celebrity imagery in voice-based personification.

Unlike the typically female voice, the male assistant elicited more varied personality judgments. While some described him as "friendly," "competent," or "calm," others depicted the assistant as emotionally detached or impersonal—such as "not interested in your problems." These mixed reactions suggest a less emotionally warm representation overall. Interestingly, the male voice seemed to prompt occupational or functional associations more than emotional ones. Participants' references to professionalism, neutrality, and even aloofness mirror findings from [LN05], who observed that male-voiced agents were more often perceived as authoritative and task-focused, rather than socially engaging. The social schema invoked here leans toward competence and seriousness rather than emotional support or warmth.

In summary, while the male voice was anthropomorphized, it was more likely to be interpreted through occupational and functional lenses, evoking a less intimate and more utilitarian form of personification. Figure 4.4 shows examples of participant-genderated representations after interacting with the system.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.


· white middle aged man
· sounded competent
· sounded nice
· can imagine him to tell jokes
· brown hair
· looks a bit like the actor from
Wolverine (Wolverine) Hugh Jackman

P11: “Friendly, male, calm, open-minded, not biased, 30–40 years old.”

P12: “Male-voiced, typical office guy, white, calm, not interested in your problems.”

P20: “White middle-aged man, sounded competent and nice, can imagine him telling jokes, brown hair, looks a bit like Hugh Jackman.”

Figure 4.4: Examples of participant-generated representations of the voice assistant – Open AI Cove

4.3.3 Representation: Genderless Voice

Participants interacting with the genderless voice assistant produced the most ambiguous and diverse representations. Gender attributions varied widely: while some described the assistant as a man or a woman, others explicitly interpreted the voice as belonging to a non-human or robotic entity. This lack of gender clarity led to a broader range of imagined identities, and in some cases, the absence of a clear anthropomorphic projection.

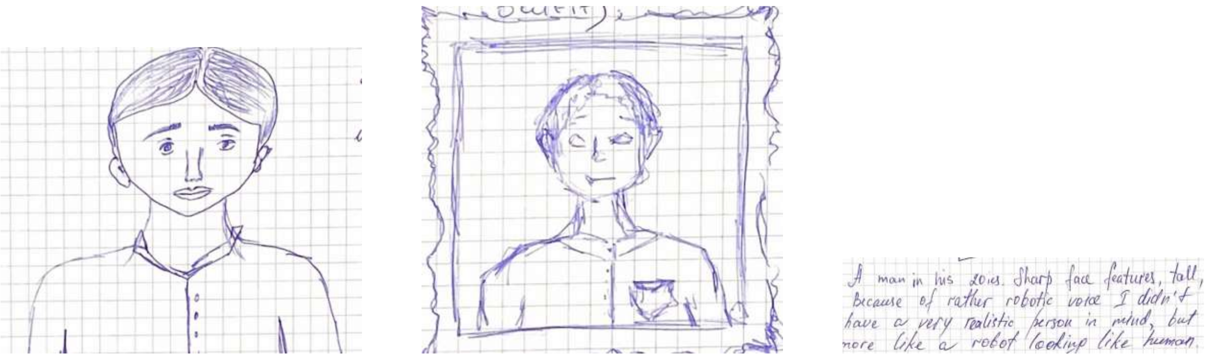
Age was still applied, but inconsistently—ranging from a “5-year-old” to a “middle-aged office worker” or “a man in his 20s with sharp facial features.” Notably, these age descriptors were often speculative and accompanied by hesitation or uncertainty, indicating that the voice did not provide sufficient cues for a coherent mental image.

Descriptions of physical appearance were minimal or absent. Instead, many participants focused on qualities such as “mechanical,” “robotic,” or “not a real person,” suggesting a shift from anthropomorphic interpretation to machine-oriented conceptualization. This aligns with previous research indicating that when agents lack human-identifying features—especially gender—users are less likely to project social characteristics or roles onto them [EWC07, Fin12].

Although a few participants still assigned human-like qualities (e.g., “strong mom vibes,” “tough, hands-on”), these instances were rare and did not form a consistent pattern. The absence of strong personality or relational cues suggests that vocal neutrality diminishes emotional engagement and stereotype activation.

These findings support the hypothesis that genderless voice design may reduce anthropomorphic bias, allowing for more neutral, task-focused interpretations. In contexts where objectivity, professionalism, or fairness is critical—such as legal, educational, or medical domains—such ambiguity may be advantageous. As [Fin12] notes, minimizing human-likeness in design can encourage users to treat artificial agents as tools rather than social partners, thus altering expectations and interaction styles. Figure 4.5 shows examples of participant-genderated representations after interacting with the system.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.



A man in his 20ies. Sharp face features, tall,
Because of rather robotic voice I didn't
have a very realistic person in mind, but
more like a robot looking like human.

P27: "A 5 year old guy, wearing a black dress shirt, looking more like a telephone operator."

28: "Man of the age of 35 - 40 years old, in an office suite (Corporate style outfit)."

P30: "A man in his 20s. Sharp face features, tall. Because of rather robotic voice, I didn't have a very realistic person in mind, but more like a robot looking like human."

Figure 4.5: Examples of participant-generated representations of the voice assistant - Open AI Breeze

4.3.4 Thematic Categorization of Representations

To synthesize the data, participant inputs were coded into categories based on recurring anthropomorphic traits. Figure 4.6 presents a summary of the most commonly attributed features.

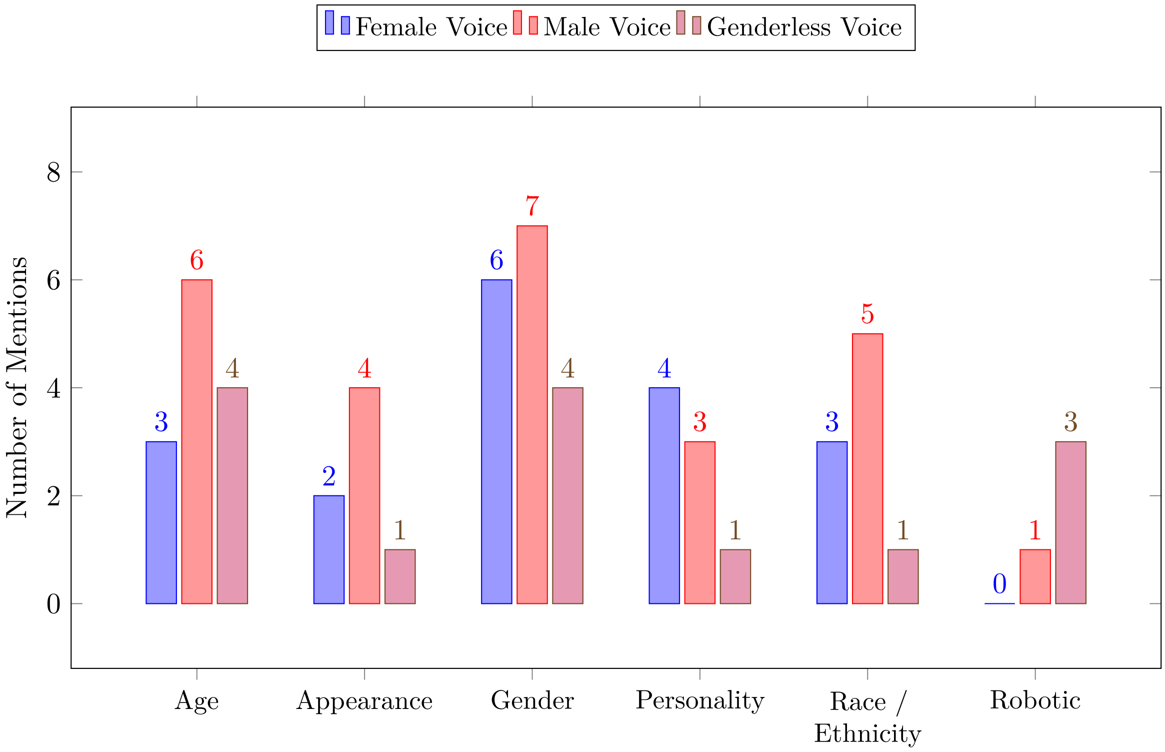


Figure 4.6: Distribution of anthropomorphic trait mentions across voice types. Gendered voices elicited stronger projections of personality, appearance, and ethnicity; the genderless voice evoked more robotic or ambiguous representations.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.3.5 Thematic Comparison Across Voice Types

Figure 4.6 displays the distribution of anthropomorphic traits mentioned by participants across the three voice assistant types. Descriptions were categorized into six thematic codes: *gender*, *age*, *appearance*, *personality*, *race/ethnicity*, and *robotic/mechanical traits*.

Participants interacting with the typically female and male voices frequently attributed clear gender identities to the assistant, with notable mentions of appearance (e.g., "blond," "bearded," "short hair") and estimated age ranges (e.g., "in their 30s" or "middle-aged"). These results suggest that the presence of gendered vocal cues strongly activates pre-existing schemas and stereotypes about age, ethnicity, and personal style [EKB$^+$12, NMG97].

In contrast, the genderless voice elicited fewer human-like associations overall. While some participants still imagined a person (e.g., "a man in a suit" or "strong mom vibes"), several explicitly described the assistant as "robotic" or "mechanical." This supports prior research indicating that vocal ambiguity reduces social projection and anthropomorphism, instead encouraging users to interpret the system as a functional machine rather than a social actor [EWC07, Fin12].

These patterns reflect how anthropomorphic projections are shaped not only by voice presence but by voice characteristics—especially gender. Gendered voices promote personalization and social connection, while neutral or synthetic ones may encourage objectivity, detachment, or even distrust depending on context.

4.4 Trust Metrics

Trust in the system was assessed using a single, direct 5-point Likert item, anchored at 1 for "No trust at all" and 5 for "Complete trust." Participants ($N = 30$) provided this rating twice: first, immediately *before* engaging in the interactive task, and again immediately *after* completing it. By administering the same question at these two critical junctures, shifts in confidence were captured that could be attributed to the interaction itself. This pre–post design ensured that each participant served as their own control, allowing us to observe whether—and to what extent—the experience bolstered or undermined their initial trust perceptions.

Before the interaction, participants arrived with varying baseline expectations: some were predisposed toward skepticism, selecting lower points on the scale, while others anticipated a seamless experience and rated their trust higher. As they navigated the interface—encountering its affordances, responding to prompts, and observing system feedback—their attitudes were subtly shaped by every positive confirmation, smooth transition, or unexpected hiccup.

Immediately after the session, they were prompted once more to quantify their trust. These second ratings captured the net effect of the interaction: whether participants felt reassured by a responsive, transparent system, translating into upward shifts on the Likert continuum; or conversely, whether any glitches, delays, or unclear instructions seeded doubt, reflected in downward adjustments. This straightforward, symmetric measurement approach provided rich, within-subject data on how direct experience influences trust judgments in real time.

4.4.1 Descriptives and overall change

A paired t-test revealed a statistically significant increase in trust ratings from pre- to post-interaction, $t(29) = -3.17$, $p = .004$. The mean change in trust was 0.60 points on the 5-point Likert scale, with a 95% confidence interval ranging from 0.21 to 0.99. The within-subjects effect size was in the medium range, $d_z = 0.58$, indicating a practically meaningful shift in participants' trust levels.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU WIEN Bibliothek Your knowledge hub

Time-point	**M**	**SD**
Pre-interaction	2.90	0.88
Post-interaction	3.50	0.86

Table 4.1: Overall trust before and after interacting with the voice assistant.

Given that the Shapiro–Wilk test indicated departures from normality in the difference scores, a nonparametric Wilcoxon signed-rank test was conducted to corroborate the t-test findings. This analysis likewise demonstrated a significant increase in trust, $Z = -2.80$, $p = .005$, thereby strengthening confidence in the robustness of the observed effect.

Together, these complementary statistical approaches confirm that the interaction produced a reliable and moderate enhancement of user trust, and that this conclusion holds under both parametric and nonparametric assumptions.

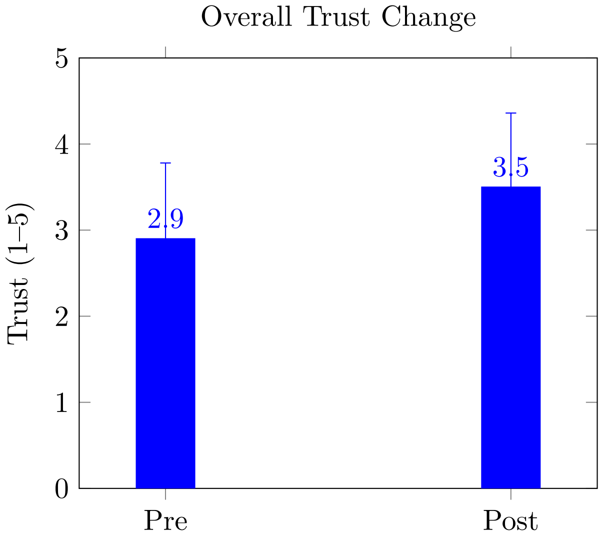


Figure 4.7: Mean trust before and after the interaction (error = SD).

Additionally, in Figure 4.8 reported "Overall correlation $r = 0.29$ (ns)" refers specifically to the *Pearson product–moment correlation coefficient*, which quantified the degree of linear association between pre-interaction trust and post-interaction trust across all participants. An r value of $+0.29$ indicates that, as pre-interaction trust increases, post-interaction trust tends to increase as well, but only modestly. Squaring this coefficient gives

$$r^2 \approx (0.29)^2 \approx 0.08,$$

which suggests that only about 8% of the variance in post-interaction trust can be explained by initial trust levels.

Importantly, the designation "ns" (not significant) tells us that this Pearson correlation fell short of the usual threshold for statistical significance (typically $p < 0.05$). With the sample size ($n = 30$, hence $df = 28$), the observed $r = 0.29$ corresponds to a p-value well above 0.05, meaning it cannot confidently be distinguished it from zero in the population.

Taken together, these results imply that, although there is a slight positive linear trend, baseline trust alone is a weak and unreliable predictor of trust after the interaction. Other experimental factors—such as the specific voice condition, the content of the interaction, or individual differences in openness to persuasion—are likely exerting a stronger influence on participants'

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

post-interaction trust scores. Future studies might explore these additional variables or increase the sample size to determine whether a true underlying relationship between pre- and post-interaction trust exists.

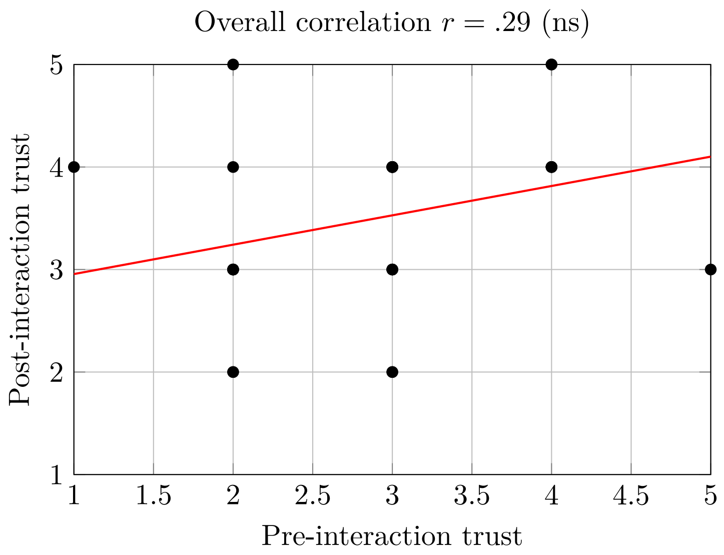


Figure 4.8: Scatterplot of pre- vs post-interaction trust (overall sample).

4.5 Impact of Voice Type on Trust

A one-way analysis of variance (ANOVA) was conducted to examine whether there were differences in the outcome variable across the three experimental conditions. The ANOVA revealed no statistically significant effect of condition, $F(2, 27) = 0.52$, $p = .599$, and the estimated proportion of variance accounted for by condition was small, $\eta^2 = .04$, as shown in the table below 4.2

Voice	**M**	**SD**	n
Female (*Juniper*)	3.7	0.82	10
Male (*Cove*)	3.3	1.06	10
Gender-neutral (*Breeze*)	3.5	0.71	10

Table 4.2: Post-interaction trust across the three voice conditions.

To explore potential pairwise contrasts, Tukey's honest significant difference (HSD) post-hoc comparisons were performed. None of the pairwise comparisons reached significance (all $p >$ .70), indicating that the observed means did not differ reliably from one another across any two conditions.

Together, these results suggest that the manipulation did not produce appreciable differences in the dependent variable across the three groups. The small effect size and non-significant post-hoc tests reinforce the conclusion that any numerical differences in group means are likely attributable to sampling variability rather than systematic effects of the experimental conditions.

4.5.1 Pre–post change within each voice

4.3 shows that mean trust ratings increased from pre- to post-measures in all three voice conditions. Specifically, trust in the female voice rose by 0.6 points (from 3.1 to 3.7), a change that

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

was statistically significant, t(9) = –3.67, p < .01. Although the male and gender-neutral voices also showed gains of 0.5 and 0.7 points respectively, these differences did not reach conventional levels of significance (male: t(9) = –1.25, p = .24; gender-neutral: t(9) = –1.77, p = .11). Thus, while all voices appeared to boost listener trust to some extent, only the female voice produced a reliably measurable effect.

Voice	Pre M	Post M	Δ	$t(9)$
Female	3.1	3.7	+0.6	-3.67
Male	2.8	3.3	+0.5	-1.25
Gender-neutral	2.8	3.5	+0.7	-1.77

Table 4.3: Within-condition trust change (paired *t*-tests).

4.5.2 Correlation pattern by voice

Figure 4.12 visualizes the relationship between participants' *pre-interaction trust* and *post-interaction trust* ratings in AI voice assistants, segmented by three distinct voice types: **Female (Juniper)**, **Male (Cove)**, and **Gender-neutral (Breeze)**. Each of the 30 semi-transparent data points represents one participant (10 per voice condition). Transparency was applied to the markers to preserve visibility of overlapping points, which occurred frequently due to repeated or similar responses.

Each voice assistant was characterized by a distinct fundamental frequency range (F_0), a key acoustic correlate of perceived vocal gender:

- **Cove (Male):** F_0 85–180 Hz
- **Juniper (Female):** F_0 165–255 Hz
- **Breeze (Gender-neutral):** F_0 145–175 Hz

Juniper, the female voice assistant, with its higher F_0 range characteristic of stereotypically feminine voices, showed a pronounced positive trend in user trust development, as illustrated in Figure 4.9 The corresponding regression line ($y = 0.8776x + 0.9796$) indicates a strong linear association, suggesting that increased pre-interaction trust was consistently met with higher post-interaction ratings. Participants who initially rated the assistant neutrally or moderately positively often reported heightened trust after interaction. The clustering of points along the upward trajectory, and the visible overlap among mid-to-high values, demonstrates a generally favorable reception to Juniper's voice characteristics. Cove, representing a stereotypically male voice within the 85–180 Hz F_0 range, yielded more heterogeneous responses. The regression line ($y = 0.2708x + 2.5417$) reflects only a slight positive trend, suggesting a weaker correlation between pre- and post-interaction trust. Some participants increased their trust after the interaction, while others reported either no change or a decrease. This variability may reflect a broader range of individual expectations or preferences associated with more traditional, masculine-sounding AI voices, as well as varying perceptions of authority, warmth, or neutrality. Breeze was designed to have a voice within a narrower and intentionally androgynous frequency band (145–175 Hz), overlapping with both male and female ranges but avoiding the extremes. The regression line ($y = -0.1316x + 3.8684$) for this condition indicates a slightly negative trend, suggesting that higher pre-interaction trust did not necessarily translate into sustained or improved post-interaction ratings. While some participants with initially low trust experienced a positive shift, others who began with higher expectations reported lower post-interaction trust.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

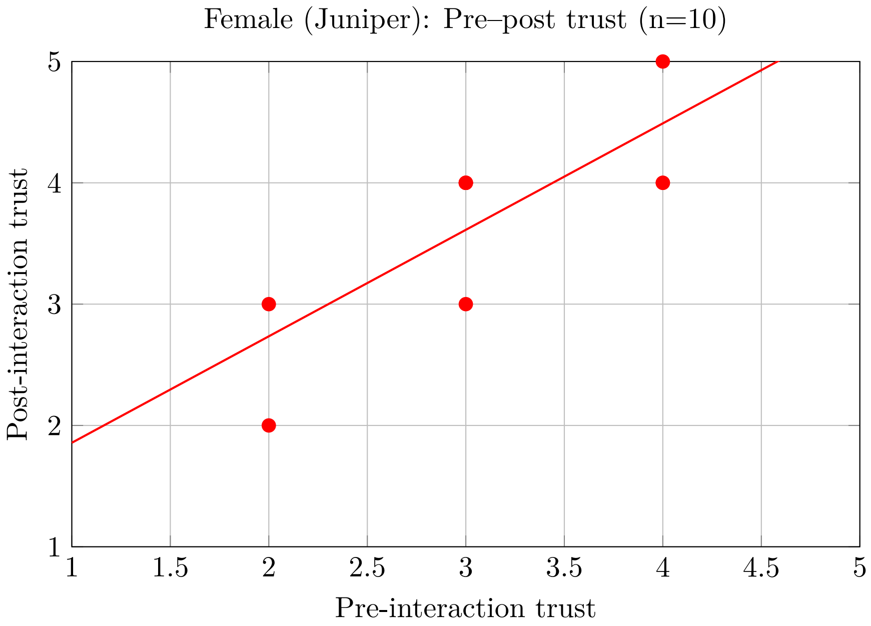


Figure 4.9: Scatter of pre– vs post-interaction trust for the female voice, with fitted slope $b = 0.88$.

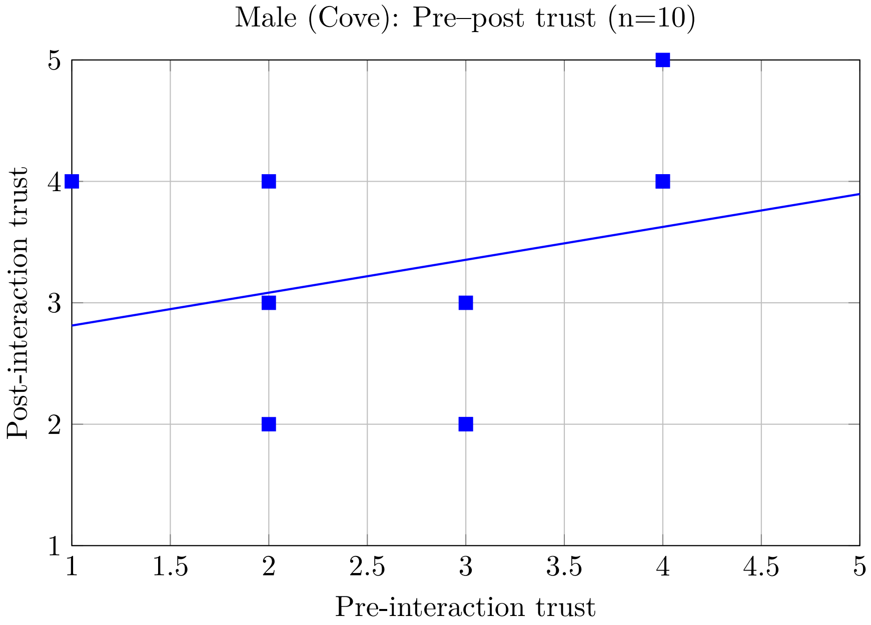


Figure 4.10: Scatter of pre– vs post-interaction trust for the male voice, with fitted slope $b = 0.27$.

This distribution suggests that the gender-neutral voice may have challenged participants' implicit expectations or lacked familiar vocal cues typically associated with trustworthiness, possibly due to limited societal exposure to non-binary voice profiles. The summarised plot as illustrated in Figure 4.12 suggests that trust dynamics in AI voice assistants are influenced not only by the interaction itself but also by acoustic characteristics such as fundamental frequency, which cue perceived gender. The female voice condition (Juniper) demonstrated the strongest and most consistent trust enhancement, followed by the more variable male voice condition (Cove). The gender-neutral voice (Breeze), though designed to be inclusive, elicited the most complex responses, possibly reflecting both novelty effects and unconscious bias. These results

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

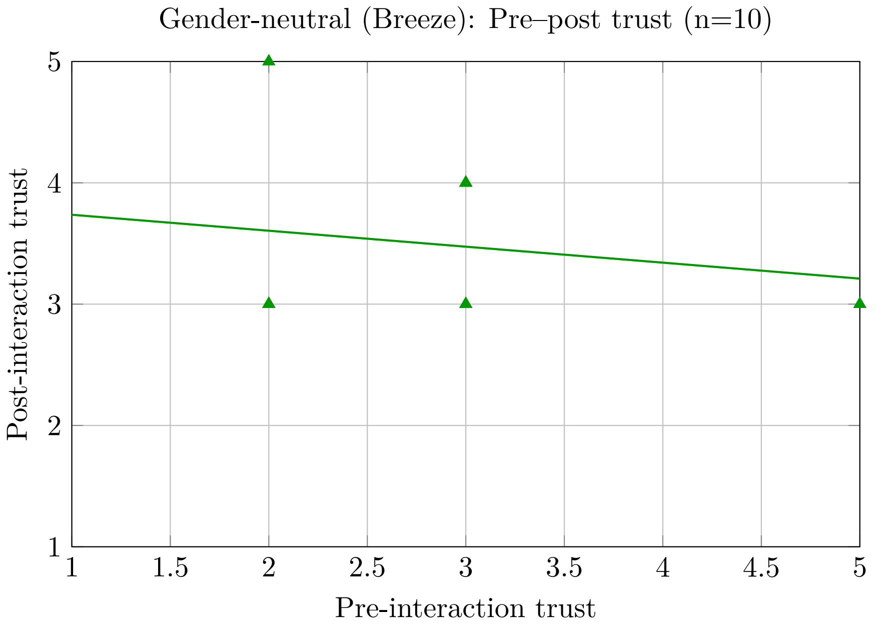


Figure 4.11: Scatter of pre– vs post-interaction trust for the gender-neutral voice, with fitted slope $b = -0.13$.

underscore the importance of considering both sociolinguistic perception and acoustic design when developing inclusive, trustworthy voice interfaces.

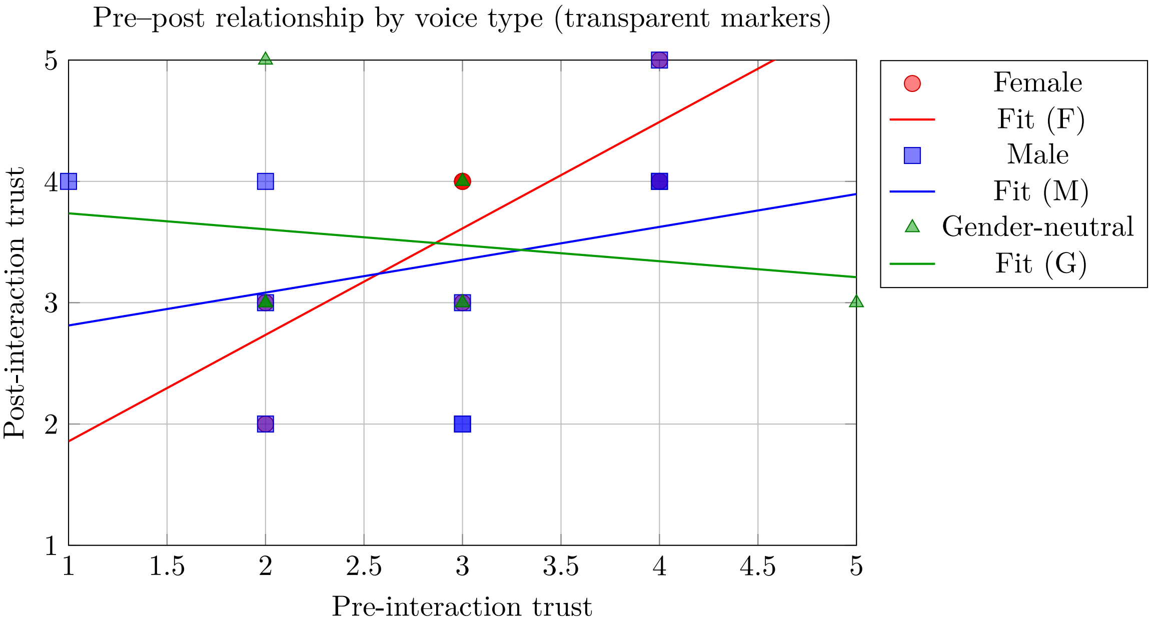


Figure 4.12: Pre–post trust by voice type with semi-transparent markers so overlapping points remain visible.

4.6 Thematic Analysis

This section presents the findings of the qualitative thematic analysis conducted on user interviews across three voice assistant conditions: female, male, and gender-neutral. The analysis

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

follows Braun and Clarke's six-phase thematic analysis framework. The results are organized around four emergent themes that address the study's research questions:

1. How does the gender of AI voice assistants (male, female, gender-neutral) influence users' trust in different task scenarios?
2. How do the anthropomorphic attributes (e.g., personality, competence, interaction style, voice tone) of AI voice assistants manifest differently across male, female, and gender-neutral voices?
3. To what extent do users display gender preferences/expectations when interacting with AI voice assistants?

The analysis involved an inductive coding approach, allowing themes to emerge directly from the data rather than from pre-existing theoretical frameworks. Interviews were transcribed, coded using short, descriptive labels (codes), and grouped into thematic clusters based on similarity, frequency, and relevance to the research questions. Thematic saturation was achieved when no new significant codes emerged from the remaining transcripts.

4.6.1 Theme 1: Tone and Contextual Fit Shape Perceived Trustworthiness

Across all voice conditions, participants consistently emphasized that the assistant's tone and its alignment with the situational context were critical in establishing trust. A calm, confident, and emotionally attuned tone fostered trust, while tonal mismatches—such as overly cheerful or flat expressions during high-stakes tasks—undermined credibility.

> *It sounded happy, which didn't feel natural in that situation.*[2]

In this case, the participant engaged with an emotionally charged task and perceived a tonal mismatch that compromised authenticity and trust.

> *In a medical emergency, the tone should've been more affirmative and confident.*[3]

This participant stressed the need for tonal assertiveness in high-pressure contexts, noting that a lack of authority diminished the assistant's trustworthiness.

> *In the emergency task, the voice was very calm and clear, which helped calm my nerves.*[4]

Conversely, a composed and steady tone helped this participant manage stress, suggesting that tonal preferences may vary depending on individual coping styles.

> *It was most useful for the parenting task... but in the emergency task, it went too fast.*[5]

---

[2] Participant 3, female voice condition.
[3] Participant 5, female voice condition.
[4] Participant 14, male voice condition.
[5] Participant 25, genderless voice condition.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Here, vocal pacing emerged as a trust barrier, particularly in cognitively demanding situations, where speed compromised clarity and reassurance.

These accounts directly support RQ1 and RQ2 by demonstrating how anthropomorphic attributes such as emotional tone and delivery speed influence perceived trustworthiness. Relevant codes include *calm tone*, *urgency mismatch*, *natural pacing*, and *emotional realism*.

4.6.2 Theme 2: Customization and Familiarity Enhance Trust

Many participants expressed a desire for greater personalization in the voice assistant's vocal attributes—such as pitch, accent, or emotional tone—as a means to increase trust and comfort. A customizable voice was seen as more relatable and user-friendly than a fixed, one-size-fits-all configuration.

> *Customization is important. Users have different preferences.*[6]

This sentiment reflects a broader preference for systems that accommodate individual needs, enhancing user satisfaction.

> *One voice shouldn't define how all voice assistants sound. Diversity is good.*[7]

Here, the participant called for representational diversity, tying trust to inclusivity and voice variety.

> *Everyone should be able to choose what makes them feel most comfortable.*[8]

Trust was described as closely linked to psychological comfort, highlighting user control as a factor in emotional safety and openness.

> *More options would help. Some older people might prefer an older-sounding voice.*[9]

This remark pointed to the value of demographic-specific tailoring, reinforcing the link between voice design and accessibility.

These findings align with RQ2 and RQ3, indicating that trust is shaped not only by the assistant's default design but also by its capacity to reflect diverse user identities and preferences. Associated codes include *customization*, *familiarity*, *voice preference*, and *user control*.

4.6.3 Theme 3: Trust is Filtered Through Task Context

Participants' trust in the assistant varied across tasks, shaped significantly by the perceived importance or emotional weight of the scenario. In lower-stakes contexts—such as caregiving or parenting—trust was more readily extended. In contrast, high-stakes scenarios like medical emergencies elicited greater skepticism and caution.

---

[6] Participant 1, female voice condition.

[7] Participant 11, male voice condition.

[8] Participant 6, female voice condition.

[9] Participant 18, male voice condition.

> *For me, the tone is crucial... in emergencies, I generally don't trust AI voice assistants.*[10]

This response illustrates a fundamental mistrust in AI under pressure, where emotional tone alone cannot compensate for perceived lack of competence.

> *I trusted it the most when it came to parenting advice... it's a task category where I'm open to suggestions.*[11]

Here, the participant indicated a willingness to rely on AI in supportive, lower-risk domains.

> *If the stakes are high... I don't trust the assistant enough.*[12]

This comment captures the conditional nature of trust—high-stakes tasks demand higher levels of perceived reliability and authority.

> *In the emergency task, I thought it was less useful because it went too fast.*[13]

Once again, vocal pacing emerged as a barrier, reinforcing that high-stakes contexts require not only emotional attunement but also precise timing and delivery.

These findings are especially relevant to RQ1 and RQ2, suggesting that trust is highly context-sensitive and modulated by task relevance. Key codes include *task relevance*, *emergency distrust*, and *contextual filtering*.

4.6.4 Theme 4: Perceived Gender Influence is Subtle but Present

While many participants claimed that gender did not affect their perception of trust, their reflections revealed subtle gender associations. Female voices were often linked to warmth and emotional intelligence, while male voices were perceived as authoritative or emotionally flat. Genderless voices elicited mixed reactions—some interpreted them as progressive, while others found them unfamiliar or difficult to categorize.

> *I tend to trust women more than men, so I would trust a female AI voice more.*[14]

This participant made an explicit connection between interpersonal trust biases and AI voice evaluation.

> *Maybe subconsciously... but I wouldn't say I trust a male voice more for technical advice.*[15]

Here, a tension between professed neutrality and implicit bias surfaced—a pattern observed across several responses.

---

[10] Participant 8, female voice condition.
[11] Participant 13, male voice condition.
[12] Participant 7, female voice condition.
[13] Participant 27, genderless voice condition.
[14] Participant 29, genderless voice condition.
[15] Participant 15, male voice condition.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

> *It felt robotic... maybe because it was a male voice.*[16]

This remark links emotional disconnection to voice gender, subtly influencing the participant's trust experience.

> *I don't think a voice can truly be genderless—it will always be interpreted.*[17]

This participant highlighted the interpretive nature of voice perception, underscoring the challenge of designing voices that are perceived as truly neutral.

These findings contribute to RQ1 and RQ3 by revealing that gender cues, even when implicit, shape users' evaluations of trust. Core codes include *gender bias*, *neutrality preference*, and *voice association.*

4.7 Behavioral Observations

In addition to survey responses and post-task reflections, a detailed analysis of behavioral cues was conducted to triangulate participant perceptions of trust, engagement, and ease of interaction with the voice assistants. Several key patterns emerged from the observational data collected during the study sessions.

At the outset of the study, a number of participants exhibited a need for procedural clarity, asking follow-up questions about the structure and expectations of the task. This initial uncertainty diminished rapidly once interaction with the assistant commenced, suggesting that the interface and interaction flow were generally intuitive and required minimal cognitive effort to navigate. However, during the post-interaction activity, where participants were asked to visually represent or describe the assistant they had interacted with, some participants requested to replay the voice. This behavior may indicate that certain voice characteristics—particularly in the case of more neutral or less personified voices—were less salient or memorable, requiring additional cognitive retrieval effort.

Interaction fluency throughout the core dialogue phases was high. Conversational flow was largely uninterrupted, characterized by seamless turn-taking and minimal latency. This suggests that participants perceived the assistants as responsive and coherent, contributing to a sense of naturalness and social presence in the interaction.

Non-verbal behaviors provided additional insight into participants' affective states. Smiling, light laughter, and relaxed facial expressions were common across sessions, indicating a generally positive emotional tone and a sense of ease with the interaction. Notably, there were no significant signs of cognitive overload or social discomfort—participants maintained open posture, showed no signs of fidgeting or disengagement, and displayed consistent attentional focus. Eye gaze was typically directed toward the screen or the device, and body movement was minimal during audio playback, particularly when participants wore headphones. These behaviors are consistent with states of high engagement and attentional absorption.

Taken together, these observations support the interpretation that participants felt comfortable and engaged during the interaction. They also suggest that the different voice assistant conditions—regardless of vocal gender characteristics—were able to support a smooth, trust-conducive interaction experience. Nonetheless, the need to replay voices for memory recall in certain cases may point to a lower level of perceived distinctiveness or anthropomorphic salience, particularly for gender-neutral voices.

---

[16] Participant 30, genderless voice condition.

[17] Participant 26, genderless voice condition.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

4.8 Evaluation of Hypotheses

This section evaluates the three hypotheses proposed in this study, integrating quantitative results, qualitative feedback, and theoretical insights to provide comprehensive academic analysis.

4.8.1 Hypothesis I

Hypothesis: *Users will display higher levels of trust in AI voice assistants with gendered voices (male or female) compared to gender-neutral voices, across different task scenarios.*

The quantitative analysis conducted using a one-way ANOVA did not yield statistically significant differences in trust ratings among voice types ($F(2, 27) = 0.74$, $p = .487$). While descriptive statistics indicated slightly higher mean trust ratings for the feminine voice condition (Juniper, $M = 3.60$, $SD = 1.17$) compared to masculine (Cove, $M = 3.40$, $SD = 1.07$) and gender-neutral voices (Breeze, $M = 3.50$, $SD = 1.14$), these observed differences lacked statistical significance. This suggests that the gender attribution inherent in the voices does not significantly impact user trust when evaluated independently from other attributes, such as perceived competence or clarity. The absence of a significant effect challenges assumptions from previous literature (Nass & Brave, 2005; Sutton et al., 2019) that gendered voices inherently foster greater user trust due to familiarity or stereotype-consistent expectations.

Thus, Hypothesis I is **not supported** by the present empirical findings.

4.8.2 Hypothesis II

Hypothesis: *Anthropomorphic attributes such as personality, competence, interaction style, and voice tone will be perceived more positively in gendered voice assistants (male or female) compared to gender-neutral voice assistants—resulting in higher levels of trust.*

Analysis of qualitative data partially supports this hypothesis. Participants frequently described gendered voice assistants as exhibiting clearer anthropomorphic attributes, notably personality, interaction style, and tone. Specifically, the feminine voice was frequently described as "warm," "engaging," or "friendly," while the masculine voice was described as "reliable," "clear," and "authoritative." In contrast, the gender-neutral voice often elicited comments about ambiguity or neutrality, described as less engaging or emotionally expressive.

However, quantitative trust ratings indicated that these enhanced anthropomorphic perceptions did not directly translate into significantly higher trust ratings. Despite clearly expressed positive attitudes toward the gendered voices' anthropomorphic qualities, the statistical relationship between these attributes and overall trust was not statistically robust. The nuanced qualitative feedback underscores the complexity of anthropomorphism's role, suggesting that while anthropomorphic attributes are perceived more positively, they may not independently guarantee higher trust outcomes, particularly within brief or single-encounter interactions (Lee et al., 2019).

Consequently, Hypothesis II is **partially supported**, emphasizing a nuanced relationship between anthropomorphism and trust that warrants further investigation in future research.

4.8.3 Hypothesis III

Hypothesis: *Users might display gender preferences/expectations when interacting with AI voice assistants.*

The findings provide strong empirical support for Hypothesis III. Behavioral observations and data derived from participant-generated visualizations (drawings) revealed pronounced gender-association patterns. Specifically, around 90% of participants interacting with the gendered

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

voices (male and female conditions) visualized distinctively gendered personas. This explicit gender attribution indicates robust gender stereotypes influencing participants' perceptions and expectations. In contrast, the gender-neutral condition revealed significant variability in participant perceptions, with many respondents expressing uncertainty or ambiguity concerning the assistant's identity.

These results align with social-role theory [EW12] and the Computers Are Social Actors (CASA) paradigm [NSS94], suggesting that users naturally project socially learned gender expectations onto technology interfaces, thereby reinforcing gender stereotypes in the absence of explicit cues to the contrary. The variability observed with gender-neutral voices further highlights the complexity and challenges associated with truly neutral anthropomorphic design, reinforcing the importance of nuanced and context-aware voice designs to mitigate unintended gender biases.

Thus, Hypothesis III is **strongly supported** by qualitative and behavioral data.

In summary, the study underscores the complexity inherent in the relationship between voice assistant gender, anthropomorphic perceptions, and trust. While gendered voices clearly evoke stronger anthropomorphic attributions and gender-specific stereotypes, these factors do not automatically translate into significantly increased trust. Future research should explore longer-term interactions and varied contexts to better understand how these dynamics evolve over time and influence user trust in technology-mediated communication.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 5

Discussion

5.1 Interpretation of Key Findings

This study investigated how anthropomorphic voice characteristics—particularly voice gender (male, female, and gender-neutral)—influence user trust in AI-enabled voice assistants (VAs). The research was guided by three central research questions (RQ1–RQ3) and aimed to test three corresponding hypotheses regarding the effects of voice anthropomorphism on user trust and engagement.

Pre-interaction survey results revealed that participants were moderately familiar with AI voice assistants, yet their baseline trust levels and comfort with AI autonomy were relatively low (M = 2.90 for trust and M = 2.47 for autonomy comfort). These initial findings suggest a cautious attitude toward AI, reflecting broader societal concerns about machine agency and reliability. Interestingly, while users reported low concern about the assistant's gender (M = 1.87), they placed a high premium on the naturalness and human-likeness of the voice (M = 4.07), indicating that emotional expressiveness and clarity—rather than gender identification—are critical drivers of initial expectations. These results are consistent with prior research by Hoy [Hoy18] and Westby et al. [WRRW24], who similarly observed that users prioritize vocal quality and emotional tone over demographic attributes in their evaluation of voice-based AI.

Following the interaction sessions, trust scores improved significantly across all participant groups (post-interaction M = 3.50, with a significance level of p = .004). This positive shift underscores the dynamic nature of trust formation, highlighting the crucial role of direct experience and interaction quality in shaping user attitudes toward AI systems. The findings support existing literature emphasizing experience-based trust-building in AI agents [AYYL20, HL23], affirming that trust cannot be fully gauged through first impressions alone but develops progressively through interactive engagement.

Thematic analysis of qualitative data and behavioral observations during the sessions revealed four key themes influencing user trust: tone-context alignment, customization opportunities, task dependency, and subtle gender influences. Participants consistently reported that calm, emotionally appropriate tones enhanced feelings of trust, particularly during high-stakes or sensitive contexts such as simulated medical emergencies. This dynamic interpretation of anthropomorphic cues validates the context-sensitivity outlined in RQ1 and RQ2 [SVK21, TEN$^+$24], demonstrating that users interpret the same voice features differently depending on the situational context.

In addition, participants' drawings and verbal descriptions provided compelling evidence of anthropomorphic projection. Gendered voices (male and female) triggered rich, vivid mental models wherein users ascribed age, personality traits, and even physical appearances to the assistant. This confirms earlier findings [EKB$^+$12, EWC07, RN96] that vocal gender strongly activates human identity schemas in the mind of the listener. By contrast, gender-neutral voices elicited more mechanical or ambiguous representations, often leading to weaker emotional resonance and reduced anthropomorphic engagement. This difference suggests that while gender-neutral designs promote inclusivity, they may sometimes struggle to evoke the same level of relational richness expected in socially oriented or emotionally charged tasks (RQ2, RQ3).

Despite these perceptual and experiential variations, quantitative analysis via ANOVA revealed no statistically significant difference in trust scores across voice types ($p = .487$). This result challenges traditional assumptions that male or female voices inherently foster more trust [NMG97, PTS$^+$17] and instead aligns with more contemporary perspectives suggesting that trust formation is highly task- and context-dependent [SVK21, AJGS22]. Thus, while vocal gender influences cognitive and emotional projections, it does not independently determine trust outcomes in short-term interactions.

Behavioral observations corroborated these findings. Participants exhibited high engagement levels and reported minimal discomfort across all voice conditions. Nevertheless, a subtle yet notable pattern emerged: users interacting with the gender-neutral voice (Breeze) more frequently needed to replay the audio before providing descriptions or sketches. This behavior suggests lower cognitive salience or weaker emotional anchoring for the gender-neutral voice, supporting RQ2's premise that anthropomorphic salience influences engagement.

The study both confirms and extends established theories in human-computer interaction (HCI) and social psychology.

Drawing on Reeves and Nass's Media Equation theory [RN96] and the Computers Are Social Actors (CASA) paradigm [NSS94], the results reveal that even minimal social cues—such as a natural-sounding voice and fluid turn-taking—are sufficient to elicit automatic social responses from users. Participants routinely adapted their language, tone, and emotional affect during interactions, treating the voice assistant as if it were a genuine social actor. This spontaneous anthropomorphization underscores the persistent influence of innate social heuristics in human-AI communication.

The findings also align closely with Mehrabian and Russell's Stimulus–Organism–Response (S–O–R) model [MR74] and Ajzen's Theory of Planned Behavior (TPB) [Ajz91]. Voice characteristics functioned as environmental stimuli that shaped users' internal states—such as feelings of trust, comfort, or uncertainty—which in turn influenced their behavioral responses, such as willingness to continue interacting with the assistant. Particularly, the opportunity for voice customization (perceived control) and the emotional appropriateness of tone emerged as key mediators of trust, demonstrating a clear behavioral pathway in line with S–O–R and TPB predictions.

However, the study also highlighted risks associated with anthropomorphic realism, resonating with Mori's Uncanny Valley hypothesis [MMK12]. Participants occasionally expressed discomfort when highly human-like voices produced canned, mechanical, or contextually inappropriate responses. These moments of incongruity disrupted the illusion of social presence, reinforcing the importance of achieving coherence between vocal realism and functional behavior to avoid unsettling user experiences.

In terms of social bias, the study found evidence consistent with prior research on anthropomorphic bias and gendered stereotypes [EH12, NSS94]. Even without visual embodiment, gendered voices activated implicit warmth-versus-competence associations, subtly shaping perceptions of trustworthiness, reliability, and expertise. These findings suggest that vocal gender cues continue to influence user judgments in ways that reflect broader social conditioning.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Finally, the study contributes to the growing conversation around inclusive voice design. Building on emerging work advocating for gender-neutral and non-binary voice options [Cin22, SSW24], the results suggest that neutral voices can reduce the activation of stereotypical schemas. However, the study also cautions that gender-neutral designs must be accompanied by strong emotional resonance and contextual sensitivity; otherwise, they risk appearing less engaging or relatable to users. In sum, while inclusive voice design offers a promising path forward, its success hinges on a careful balance between neutrality and emotional expressiveness. Hypothesis I which predicted higher trust in gendered voices compared to gender-neutral ones, was not supported by the ANOVA results. Although the female voice group had a slightly higher trust mean (M = 3.60), the differences across groups did not reach statistical significance. Hypothesis II, which suggested that anthropomorphic traits would be more pronounced in gendered voices and lead to higher trust, was partially supported. Qualitative and visual data confirmed that gendered voices elicited stronger anthropomorphic projections and were perceived as more emotionally relatable. However, these perceptions did not translate into significantly higher trust ratings at the statistical level. Hypothesis III, proposing that user gender preferences/expectations would manifest in trust perceptions and anthropomorphic attributions, was strongly supported by the qualitative and observational data. Participants frequently projected stereotypical social identities onto the gendered voices and expressed subtle biases that appeared to align with broader societal conditioning regarding gender norms and relational expectations.

Taken together, these findings highlight the complex interplay between voice anthropomorphism, user expectations, context sensitivity, and trust formation, suggesting that inclusive and trustworthy AI voice design must move beyond simplistic demographic cues to address deeper emotional and relational factors.

5.2 Relation to Prior Research

Voice technologies such as *Auto-Tune* and *Vocaloid* illustrate the significant role of pitch in shaping gender perception through audio. As Bauer and Devaney (2022) [BD22] describe, pitch is not a neutral parameter—it is deeply entangled with social cues that inform how listeners interpret vocal identity. A higher pitch is commonly associated with femininity, while a lower pitch is often perceived as masculine, reflecting long-standing societal stereotypes. These associations are embedded in both music production and speech synthesis technologies. *Auto-Tune*, for instance, can be used to exaggerate or diminish gendered vocal traits by altering pitch in real time, while *Vocaloid* avatars are often designed with pitch ranges that align with gendered expectations. The manipulation of pitch, therefore, goes beyond technical sound processing; it becomes a cultural act that can either conform to or subvert binary notions of gender. This dynamic is especially relevant in AI voice assistants, where the choice of pitch strongly influences whether a voice is perceived as male, female, or gender-neutral. Exploring these implications helps reveal how anthropomorphic cues like voice pitch impact user trust, bias, and identification in human–AI interaction.

The findings of this study extend Seymour and Van Kleek's [SVK21] insights into relational trust and the role of anthropomorphic cues in shaping user perceptions of voice assistants. By showing that trust is influenced not only by anthropomorphic features but also by subtler emotional and contextual factors, the results deepen our understanding of the relational dynamics between users and AI systems. Specifically, the study supports prior work emphasizing the critical importance of emotional realism and voice tone in fostering user trust [WRRW24, HL23]. These affective attributes appear to play a significant role in humanizing AI interactions, making the voice assistant seem more approachable, competent, and socially present.

At the same time, the findings challenge the longstanding assumption that gendered voices are inherently more trustworthy, an idea rooted in earlier research on human-computer interaction [NMG97, LN05]. Although gendered voices have traditionally been perceived as familiar and relatable proxies for human social roles, the present study suggests that trust formation may be less dependent on gender markers and more sensitive to the emotional authenticity and contextual appropriateness of the voice assistant's behavior.

Notably, the study also makes an important contribution to the growing body of literature on inclusive AI design. Prior research has highlighted the importance of offering gender-neutral voice options to ensure better representation and inclusion of underrepresented groups, particularly non-binary and gender-diverse users [Cin22, NPR19]. However, the current findings add complexity to this perspective. While gender-neutrality is valued as a principle of fairness and inclusion, users' actual experiences of trust and engagement appear to depend heavily on the emotional resonance and contextual fit of the voice assistant, regardless of gender identity cues. This observation supports recent arguments by Seaborn et al. [SSW24], who emphasize the necessity of designing emotionally intelligent and contextually adaptive AI systems, rather than focusing solely on surface-level demographic characteristics.

Overall, these findings position emotional realism, contextual sensitivity, and relational appropriateness as key components in advancing both the trustworthiness and inclusiveness of future AI voice assistants.

5.3 Limitations

The present study's sample size ($n = 30$) and the demographic composition of its participants—predominantly university students residing in Vienna—place important limitations on the broader applicability of its findings. While the insights gained from this group are meaningful, the homogeneity of the sample in terms of age, educational background, and cultural environment suggests that caution must be exercised when attempting to generalize the results to other populations. Different age groups, professional backgrounds, and individuals from diverse cultural settings may interact with and perceive AI-enabled voice assistants in distinct ways, influenced by varying levels of familiarity with technology, trust expectations, and social norms. Future research should therefore strive for more diverse and representative samples to validate and extend the present findings across a broader spectrum of users.

The study employed a between-subject design, meaning that each participant interacted with only one type of voice assistant—either typically male, typically female, or gender-neutral. While this design minimized potential carryover effects that could arise from direct comparisons within the same individual, it also introduced a limitation: it was not possible to assess how the same participant might perceive differences across the three voice types. Within-subject designs, in contrast, allow for more nuanced comparisons by controlling for individual variability. Thus, the between-subject approach, while methodologically sound for isolating certain effects, restricted the depth of comparative analysis and may have obscured subtler patterns in how trust and perception shift when multiple voice profiles are experienced firsthand.

Despite careful standardization of interaction sessions—including scripting conversation flows, controlling environmental factors, and minimizing researcher influence—participant interpretations of the voice assistants' behaviors likely varied. Human perceptions are deeply subjective and can be influenced by a multitude of factors, including personal mood, prior experience with similar technologies, and individual expectations about artificial intelligence. Even slight variations in tone, perceived responsiveness, or phrasing may have been interpreted differently by participants, introducing variability that is difficult to fully eliminate in human-subject research. This interpretative variability underscores the complexity of studying trust, which is not solely a rational judgment but is also shaped by emotions, expectations, and subconscious biases.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Moreover, the brief nature of the interactions—approximately twenty to thirty minutes per session—poses a significant limitation in capturing the evolving nature of trust in AI systems. Trust is a dynamic, context-sensitive process that often develops through repeated encounters, during which users have opportunities to assess consistency, reliability, empathy, and responsiveness over time. Initial impressions, while informative, may differ substantially from judgments formed after extended use, where experiences of error, adaptation, forgiveness, and reinforcement come into play. As a result, the findings of this study primarily reflect participants' immediate, surface-level trust responses rather than deeper, more resilient forms of trust that emerge through ongoing interaction. Longitudinal studies would be necessary to fully understand how trust in AI voice assistants grows, deteriorates, or transforms with long-term use, exposure to mistakes, and evolving user expectations.

5.4 Implications and Future Research

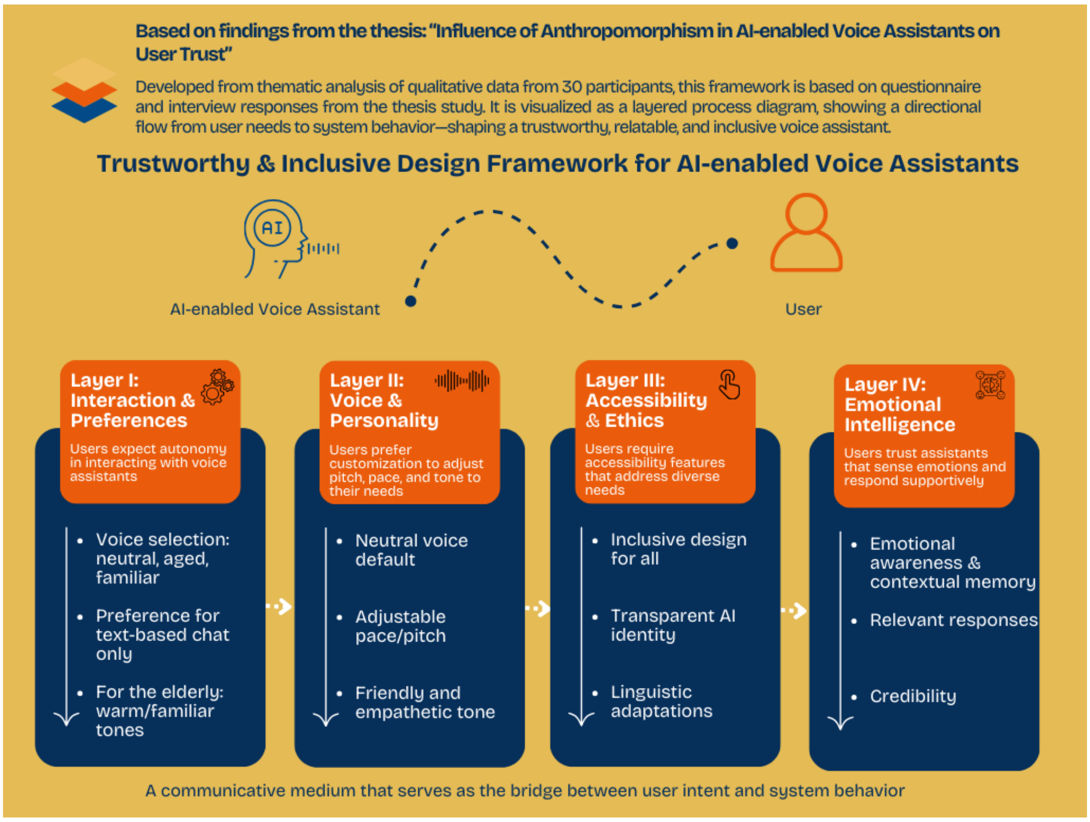
The insights gained from this research carry important implications for the future of inclusive AI design. The results highlight that fostering user trust in voice assistants requires careful attention not only to the functional capabilities of these systems but also to their emotional tone, the situational context of interactions, and the degree of personalization available to users. Voice assistants that exhibit authentic emotional cues and that can flexibly adapt to the user's context are more likely to be perceived as trustworthy, relatable, and inclusive across diverse user populations [Agr19, RN96].

Future research should therefore explore longer-term and naturalistic interactions, moving beyond short-term laboratory studies. Longitudinal designs, where users interact with voice assistants over weeks or months in their everyday environments, would provide deeper insights into how trust evolves, solidifies, or deteriorates over time. These studies could examine how adaptive or customizable voice options affect sustained user engagement, emotional attachment, and trust recovery following system errors. In addition, cross-cultural research is urgently needed to better understand how voice perception varies across different sociocultural contexts. Cultural norms significantly influence expectations regarding emotional expressiveness, politeness, gender roles, and authority, all of which may affect how users interpret and trust voice assistants. Expanding research samples beyond Western academic populations would be critical to designing AI systems that are globally inclusive and sensitive to cultural diversity. Another promising avenue for future investigation is the integration of biometric trust indicators, such as eye-tracking, heart rate variability, and galvanic skin response. These physiological measures can complement self-reported and observational data, offering more granular, real-time insights into the unconscious dimensions of trust formation. By triangulating these data sources, researchers can develop a more comprehensive understanding of the affective processes underpinning human-AI trust.

Ultimately, integrating inclusive design practices and implementing deeper voice customization features—as proposed in the following framework—could significantly enhance the equity, accessibility, and trustworthiness of future AI systems. Rather than adopting a ”one-size-fits-all” [GB19] model, future designs should strive to offer users meaningful choices and emotional intelligence, recognizing the diverse needs, identities, and contexts of those who interact with voice technologies.

5.5 Inclusive Design Framework

The inductive coding process began with open-coding of all interview and survey transcripts to surface recurring concepts—such as autonomy in voice selection, urgency-appropriate tone, and personalized pacing. These initial codes were then grouped into higher-order categories via axial coding, producing four principal themes. “Interaction & Preferences” (Layer I) directly addresses participants’ stated need for choice and agency in how they engage with the assistant. Layer II, “Voice & Personality,” encapsulates subcodes related to pitch, pace, and expressiveness—attributes participants explicitly linked to trust. The third theme, focused on inclusivity and transparent AI identity, became “Accessibility & Ethics” (Layer III), ensuring that diverse user groups are both represented and informed about the system’s capabilities. Finally, the emergent theme of emotional attunement and contextual memory informed “Emotional Intelligence” (Layer IV), embedding mechanisms for affect detection and context-aware response selection. This four-layered flow translates qualitative findings into a coherent design blueprint, guiding the system from initial interaction choices through to emotionally resonant behavior. Figure 5.1 presents the proposed *Trustworthy & Inclusive Design Framework for AI Voice Assistants*, developed through thematic analysis of qualitative user responses. It is visualized as a layered process diagram, where each layer contributes to the formation of a trustworthy, relatable, and inclusive voice assistant. The model is structured as a directional flow, starting from user needs and progressing through successive layers that implement these needs into system behavior and output.



Based on findings from the thesis: “Influence of Anthropomorphism in AI-enabled Voice Assistants on User Trust”

Developed from thematic analysis of qualitative data from 30 participants, this framework is based on questionnaire and interview responses from the thesis study. It is visualized as a layered process diagram, showing a directional flow from user needs to system behavior—shaping a trustworthy, relatable, and inclusive voice assistant.

Trustworthy & Inclusive Design Framework for AI-enabled Voice Assistants

Layer I: Interaction & Preferences

Users expect autonomy in interacting with voice assistants

- Voice selection: neutral, aged, familiar
- Preference for text-based chat only
- For the elderly: warm/familiar tones

Layer II: Voice & Personality

Users prefer customization to adjust pitch, pace, and tone to their needs

- Neutral voice default
- Adjustable pace/pitch
- Friendly and empathetic tone

Layer III: Accessibility & Ethics

Users require accessibility features that address diverse needs

- Inclusive design for all
- Transparent AI identity
- Linguistic adaptations

Layer IV: Emotional Intelligence

Users trust assistants that sense emotions and respond supportively

- Emotional awareness & contextual memory
- Relevant responses
- Credibility

A communicative medium that serves as the bridge between user intent and system behavior

Figure 5.1: Inclusive and Trustworthy Design Framework for AI Voice Assistants. A layered model developed from thematic analysis of questionnaire and interview responses from 30 participants, illustrating how user needs inform system behavior to create trustworthy, relatable, and inclusive experiences.

At the core of this ecosystem is the **User**, whose personal preferences, ethical concerns, and communication expectations shape the foundation of the assistant’s design. This human-centered approach ensures that AI voice interfaces are not only functionally competent but also socially attuned.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Layer 1: Interaction & Preferences

This layer reflects the initial and most direct input from users. Based on the study's findings, users expect a high degree of autonomy in how they interact with voice assistants. Many participants expressed the desire to select from a range of voice options—including gender-neutral, aged, and familiar-sounding voices. Some envisioned voices that resemble friends or family members, but emphasized the necessity of explicit consent for such cloning to take place.

Privacy concerns also emerged, particularly around the use of voice data for training purposes. As a result, several participants preferred the option to interact via text-based chat, highlighting the need for multimodal interaction options that respect varying comfort levels and privacy sensitivities.

Furthermore, some participants thought of older people and indicated that they would feel more comfortable interacting with voice assistants that sound familiar (or old-sounded voice), suggesting that tone and warmth are crucial factors for building trust between older user groups.

Layer 2: Voice & Personality

This layer focuses on how the assistant's voice and perceived personality are technically configured to align with user expectations. A gender-neutral voice is recommended as the system default in order to mitigate gender stereotyping and reduce cognitive bias. However, customization remains essential. Users should be able to modify parameters such as pitch, pace, and tone to meet their linguistic, cognitive, or emotional needs.

In terms of personality, participants preferred assistants that were warm, polite, and empathetic—but not overly human. The design recommendation is to maintain a balance between relatability and transparency. While human-like warmth improves interaction comfort, it should not be used to disguise the artificial nature of the assistant.

Layer 3: Accessibility & Ethics

This final preparatory layer ensures that ethical considerations and inclusive design principles are embedded into the system architecture. The assistant must be accessible across a broad range of users, regardless of age, ability, gender identity, or cultural background. This includes support for simplified interaction modes, visual assistance, and regional linguistic adaptations.

It is essential that the assistant's artificial identity is clearly communicated. Participants expressed discomfort with assistants that mimic humans too closely without disclosing their nature. Transparency around the assistant's capabilities, limitations, and AI status builds trust and prevents unrealistic expectations.

Finally, in cases where voice cloning or personalization is offered, it is critical that the assistant obtains informed, explicit consent from users. This aligns with privacy regulations such as the GDPR and supports user empowerment through ethical data practices.

Layer 4: Emotional Intelligence

Emotional intelligence emerged as a vital theme influencing trust and engagement. Participants valued the idea of the assistant adapting to emotional cues, provided this is done respectfully and without overstepping. The system should be able to detect basic emotional states and respond in ways that are helpful and non-intrusive—for instance, by offering assistance when frustration is detected or modifying tone in response to user distress.

To ensure continuity and reduce cognitive load, the assistant should also maintain contextual memory. This allows for fluid conversation and prevents repetitive or irrelevant responses. More-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

over, responses should be relevant and situationally appropriate, especially in sensitive domains such as healthcare or mental well-being. Where factual information is provided, references should be included to enhance the credibility and transparency of the system.

The final node of the diagram, the **Voice Assistant**, represents the integration of all preceding layers. It embodies a system that is inclusive, transparent, emotionally aware, and responsive to individual user needs. By following a layered and modular design process, the assistant remains grounded in user-informed insights while maintaining technical flexibility and ethical integrity.

This ecosystemic model serves not only as a conceptual guide for designers and developers but also as a framework for evaluating the inclusivity and trustworthiness of voice assistant systems in future research and practice.

This framework models how user-centered principles can guide the development of emotionally aware, ethically sound, and socially inclusive voice assistants, building on growing calls within human-computer interaction (HCI) and AI ethics communities for more equitable AI systems [HWVD$^+$19, Spi21a, BG18].

At its core lies the **User**, positioned not merely as a passive recipient but as an active agent whose needs, expectations, and emotional realities should inform each layer of system design. User needs identified through this study—such as privacy protection, emotional tone sensitivity, voice familiarity, and personalization—serve as foundational criteria for inclusive and trustworthy voice assistant development. Users increasingly expect voice assistants to offer multimodal interaction options, including voice, text, and touch-based inputs, in order to accommodate different communication needs and abilities. Privacy safeguards must be made visible and easily adjustable, empowering users to control how their data is collected, stored, and used [HWVD$^+$19]. Furthermore, offering familiar or emotionally resonant voice profiles can enhance trust, particularly when users can select tones that match their individual comfort levels or cultural expectations. By default, systems should provide neutral, non-gendered voice options to avoid reinforcing stereotypes, in line with inclusive design principles outlined in prior work [Cin22, Spi21a]. However, customization features allowing users to adjust voice pitch, tone, speaking pace, and even linguistic styles (e.g., formal versus casual) are critical to fostering a sense of ownership, personalization, and emotional connection. True inclusivity demands robust support for diverse populations, including individuals with disabilities, non-native speakers, and culturally diverse groups [BG18]. Accessibility must be embedded from the earliest stages of design, not retrofitted later. Ethical practices, such as transparent disclosures about data usage and algorithmic decision-making, are equally essential for maintaining user trust over time and avoiding technological harms or biases. Emotionally intelligent voice assistants must be able to recognize and appropriately respond to users' emotional cues, adapting their tone, speed, and even language choices based on contextual factors. This aligns with calls from AI ethics research to develop affective computing systems that prioritize user well-being rather than just transactional efficiency [MK19]. When executed thoughtfully, such systems do not merely perform tasks; they foster trust, demonstrate empathy, and enhance social belonging for a broader, more diverse spectrum of users.

This framework thus synthesizes user expectations, emotional trust-building factors, and ethical design imperatives into a practical roadmap for AI developers. It echoes and extends emerging frameworks in inclusive AI design [HWVD$^+$19, Spi21a], aiming to pave the way for future voice assistants that are not only technologically advanced but also socially responsible and universally accessible.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 6

Conclusion

This study investigated how anthropomorphic voice characteristics—specifically voice gender (male, female, and gender-neutral)—affect user trust in AI-enabled voice assistants. Through a carefully designed experimental setup that combined qualitative and quantitative methods, the research addressed three core questions and tested corresponding hypotheses, offering novel insights at the intersection of human-computer interaction, AI ethics, and inclusive design. The results indicate that while gendered voices often facilitated anthropomorphic engagement—reflected in participants' spontaneous development of mental models, attribution of personality traits, and relational behaviors—this anthropomorphism alone did not result in significantly higher trust levels. Rather than being driven by gendered voice characteristics in isolation, user trust was more strongly shaped by factors such as emotional tone, vocal expressiveness, contextual fit, and the perceived authenticity of the interaction. These dimensions were repeatedly emphasized in participant feedback and were shown to influence whether users found the voice assistant relatable, reliable, and worthy of trust. This directly answers RQ1 and RQ2 by demonstrating that surface-level gender traits do not reliably predict trust outcomes; instead, it is the combination of affective resonance and interactional appropriateness that drives trust formation. Furthermore, the investigation of RQ3 revealed subtle but consistent gender-based perception biases, with participants more likely to project stereotypical traits such as competence, friendliness, or authority based on the perceived gender of the voice, even though these stereotypes did not translate into measurable differences in trust metrics. Interestingly, gender-neutral voices—though praised by many participants for their inclusive intent—elicited mixed reactions: while some users appreciated the avoidance of binary categorization, others found these voices emotionally flat or cognitively less engaging, suggesting that neutrality must be designed alongside clear emotional cues, prosodic variety, and contextual sensitivity to foster engagement and relational trust. Theoretically, the findings of this study affirm and extend several key models discussed in the literature review. The Stimulus-Organism-Response (S-O-R) framework was validated, as users' internal affective states—shaped by external voice stimuli—played a crucial role in their behavioral responses, including their willingness to rely on, listen to, or follow the assistant. The Media Equation and CASA paradigms were clearly supported, as participants unconsciously transferred social scripts and norms—such as politeness, responsiveness, or gender stereotyping—onto the non-human voice agents, confirming that even brief interactions can trigger deeply embedded social cognition. The Theory of Planned Behavior (TPB) also proved useful in explaining how trust was influenced not only by immediate interactional qualities but also by broader attitudes, expectations, and perceived social norms surrounding AI systems. Moreover, occasional discomfort or unease with genderless voices points to a subtle form of the Uncanny Valley in auditory interfaces, where the lack of clear social cues may violate user expectations and disrupt cognitive fluency, leading to reduced affective alignment. On a practical level, these insights highlight the limitations of current voice assistant designs and underscore the need for a more sophisticated approach to inclusive AI. Simply offering gender-neutral voice options is not sufficient; rather, developers must consider how to embed emotional intelligence, adaptability, and cultural awareness into voice interactions. The Inclusive Design Framework proposed in this thesis synthesizes these findings and provides actionable guidance for designers and technologists who seek to build AI systems that are equitable, empathetic, and trustworthy. This includes recommendations for customizable voice profiles, transparency about voice sourcing, responsiveness to diverse user needs (including age, gender identity, cultural background, and disability), and ethical principles that center user autonomy and agency. Nonetheless, this study also surfaces important gaps that future research must address. Most notably, while this and many other studies focus on short-term, first-impression trust, the long-term development of trust in AI remains poorly understood. Questions around trust repair after errors, the durability of trust over repeated interactions, and how users' perceptions evolve over time are still open [Z+23]. Additionally, cross-cultural differences in the perception of voice traits and anthropomorphic cues remain underexplored, yet are likely to be highly relevant in globalized AI deployment [LC+22]. Future research would benefit from longitudinal and ecologically valid designs, as well as the integration of physiological measures—such as EEG, electrodermal activity (EDA), or galvanic skin response (GSR)—to capture unconscious and affective reactions during human-AI interaction. In conclusion, this thesis contributes to a more nuanced understanding of trust in AI-enabled voice assistants by showing that voice gender alone is not a reliable predictor of trust; instead, trust is the product of a dynamic interplay between emotional resonance, social cognition, and context-sensitive design. By embracing a holistic, user-centered approach that integrates inclusive design principles, emotional intelligence, and ethical foresight, the future of AI voice assistants can become more than functional tools—they can become socially aware, trustworthy companions capable of serving diverse populations with empathy, transparency, and fairness.

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU WIEN Bibliothek Your knowledge hub

CHAPTER 7

Overview of Generative AI Tools Used

During the process of composing this thesis, I have used several digital tools that helped me with writing, translating, and formatting my work. These tools did not replace my own thinking or writing but supported me in expressing my ideas more clearly and organizing my content more efficiently. One of the main tools I used was ChatGPT by OpenAI. I used it to rephrase or polish certain parts of my writing, especially when I wanted to improve the flow or make complex ideas easier to understand. It was also helpful when I worked with LaTeX code, particularly for formatting graphs, tables, and other visual elements. In some cases, I asked for help simplifying or correcting LaTeX syntax to save time and avoid technical mistakes. For language support, I used DeepL Translator. It helped me rewrite sentences from my native language (Shqip) into more natural and advanced English. This was especially useful when I wanted to make sure that my wording was academically appropriate and easy to read. I also used Google Translate, mainly for individual words or short phrases that I couldn't easily translate on my own. Both tools supported me in making sure that the meaning of my ideas was not lost in translation. In addition to writing and translation tools, I used online citation tools to help with formatting my references correctly. For example, I used tools that convert references from websites like Google Scholar into BibTeX format, which I could then include in my LaTeX bibliography. This helped me stay organized and saved time when managing my sources. For the analysis part of my research, I also used digital tools to support the transcription of my recorded interviews, for example, Descript tool. I used automated transcription software to generate initial drafts of the conversations, which I then manually reviewed and corrected to ensure accuracy. This made the process of preparing the interview data much faster, while still allowing me to carefully listen and reflect on the participants' responses. All of these tools were used carefully and responsibly. They supported my work but did not replace my role as the researcher and author. I made all f inal decisions about the content, structure, and wording of my thesis. The tools simply helped me improve clarity, accuracy, and formatting throughout the writing process.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

KAPITEL 8

Übersicht verwendeter Hilfsmittel

Während des Verfassens dieser Arbeit habe ich verschiedene digitale Tools verwendet, die mir beim Schreiben beim Schreiben, Übersetzen und Formatieren meiner Arbeit geholfen haben. Diese Tools haben mein eigenes Denken und Schreiben nicht ersetzt oder Schreiben, sondern unterstützten mich dabei, meine Ideen klarer auszudrücken und meine Inhalte effizienter zu effizienter zu organisieren. Eines der wichtigsten Tools, die ich verwendet habe, war ChatGPT von OpenAI. Ich habe es benutzt, um bestimmte bestimmte Teile meines Textes umzuformulieren oder zu verfeinern, vor allem wenn ich den Fluss verbessern oder komplexe Ideen leichter verständlich zu machen. Es war auch hilfreich, wenn ich mit LaTeX-Code arbeitete, insbesondere für Formatierung von Diagrammen, Tabellen und anderen visuellen Elementen. In einigen Fällen bat ich um Hilfe bei der Vereinfachung oder Korrektur der LaTeX-Syntax, um Zeit zu sparen und technische Fehler zu vermeiden. Für die Sprachunterstützung habe ich DeepL Translator verwendet. Er half mir, Sätze aus meiner Muttersprache (Shqip) in natürlicheres und fortgeschrittenes Englisch umzuschreiben. natürliches und fortgeschrittenes Englisch umzuschreiben. Dies war besonders nützlich, wenn ich sicherstellen wollte, dass dass meine Formulierungen akademisch angemessen und leicht zu lesen waren. Ich habe auch Google Translate verwendet, vor allem für einzelne Wörter oder kurze Sätze, die ich nicht ohne weiteres selbst übersetzen konnte. Beide Tools halfen mir dabei, sicherzustellen, dass die Bedeutung meiner Ideen in der Übersetzung nicht verloren ging. Unter Zusätzlich zu den Schreib- und Übersetzungstools habe ich auch Online-Zitierhilfen verwendet, um meine meine Referenzen richtig zu formatieren. Ich habe zum Beispiel Tools verwendet, die Referenzen von Websites wie Google Scholar in das BibTeX-Format konvertieren, das ich dann in meine LaTeX-Bibliografie einfügen konnte. Diese half mir, den Überblick zu behalten und Zeit bei der Verwaltung meiner Quellen zu sparen. Für den Analyseteil meiner meiner Forschung habe ich auch digitale Tools verwendet, um die Transkription meiner aufgezeichneten Interviews zu unterstützen, zum Beispiel zum Beispiel das Tool Descript. Ich verwendete eine automatische Transkriptionssoftware, um erste Entwürfe der Gespräche zu erstellen. die ich dann manuell überprüft und korrigiert habe, um die Genauigkeit zu gewährleisten. Dadurch wurde der Auf diese Weise konnte ich die Aufbereitung der Interviewdaten wesentlich beschleunigen und hatte gleichzeitig die Möglichkeit, den Teilnehmern sorgfältig zuzuhören und über die Antworten der Teilnehmer nachzudenken. Alle diese Instrumente wurden sorgfältig und verantwortungsbewusst eingesetzt. Sie unterstützten meine Arbeit, aber sie ersetzten nicht meine Rolle als Forscher und Autor. Ich traf alle f Ich traf alle Entscheidungen über den Inhalt, die Struktur und die Formulierung meiner Arbeit. Die Hilfsmittel haben mir lediglich geholfen die Klarheit, Genauigkeit und Formatierung während des gesamten Schreibprozesses zu verbessern.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Figures

2.1 S-O/R Model adapted from [HBL23] . . . 15
2.2 Theory of Planned Behavior model [Ajz91] . . . 17
2.3 Approximate representation of the Uncanny Valley, adapted from [MMK12] . . . 18
2.4 A visual representation by Lee [LS04], which describes trust as an interaction between analytic, analogical, and affective processes. The analytic and analogical processes shape emotional responses, which then strongly influence the analytic and analogical processes themselves . . . 21
2.5 Marina Abramović and Ulay, *Rest Energy*, 1980. Performance still. Courtesy of the Marina Abramović Archives [AU80] . . . 23
2.6 Voice Assistant Processing Flow . . . 25
2.7 Evolution of User Interfaces (Adapted from RTI International, 2023). . . . 26
2.8 Estimated fundamental frequency (F0) ranges and midpoints across gender and age categories. Adapted from [Tit94, CW91, Oat01, Vir19]. . . . 28

3.1 User study setup showing a participant during the pre-survey phase and interacting with the voice assistant task . . . 35
3.2 Braun and Clarke's six-phase approach to thematic analysis [CB13] . . . 40

4.1 Average participant ratings of trust in accuracy and comfort with decision-making by AI voice assistants . . . 42
4.2 Results on the previous usage of different AI voice assistants from study participants . . . 43
4.3 Examples of participant-generated representations of the voice assistant – Open AI Juniper . . . 45
4.4 Examples of participant-generated representations of the voice assistant – Open AI Cove . . . 46
4.5 Examples of participant-generated representations of the voice assistant - Open AI Breeze . . . 47
4.6 Distribution of anthropomorphic trait mentions across voice types. Gendered voices elicited stronger projections of personality, appearance, and ethnicity; the genderless voice evoked more robotic or ambiguous representations. . . . 47
4.7 Mean trust before and after the interaction (error = SD). . . . 49
4.8 Scatterplot of pre- vs post-interaction trust (overall sample). . . . 50
4.9 Scatter of pre– vs post-interaction trust for the female voice, with fitted slope $b = 0.88$. . . . 52
4.10 Scatter of pre– vs post-interaction trust for the male voice, with fitted slope $b = 0.27$. 52
4.11 Scatter of pre– vs post-interaction trust for the gender-neutral voice, with fitted slope $b = -0.13$. . . . 53
4.12 Pre–post trust by voice type with semi-transparent markers so overlapping points remain visible. . . . 53

5.1 Inclusive and Trustworthy Design Framework for AI Voice Assistants. A layered model developed from thematic analysis of questionnaire and interview responses from 30 participants, illustrating how user needs inform system behavior to create trustworthy, relatable, and inclusive experiences. 66

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Tables

2.1 Summary of Trait Discrimination Accuracy Across Experiments 20

3.1 Task Scenario: Parenting Advice 36
3.2 Task Scenario: Problem-Solving (Spacecraft Troubleshooting) 37
3.3 Task Scenario: Emergency Medical Response 37
3.4 Task Scenario: Problem-Solving (Spacecraft Troubleshooting) 37
3.5 Task Scenario: Emergency Medical Response 38
3.6 Pre-Interaction Trust Questionnaire grouped by aspects 38
3.7 Post-Interaction Trust Questionnaire grouped by aspects 39

4.1 Overall trust before and after interacting with the voice assistant. 49
4.2 Post-interaction trust across the three voice conditions. 50
4.3 Within-condition trust change (paired *t*-tests). 51

F.1 Pre-Interaction Survey Questions 113

G.1 Post-Interaction Survey Questions 115

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Acronyms

ASR Automatic Speech Recognition. 24

BFI Big Five Inventory. 19

CASA Computers Are Social Actors paradigm. xvii, 14, 15, 17, 19

DVA Digital Voice Assistant. 6

F0 Fundamental Frequency. 27

GDPR General Data Protection Regulation. 67
GPS Global Positioning System. 2

HCI Human-Computer Interaction. 6

IoT Internet of Things. 24

MMSE Mini-Mental State Examination. 30

NEO-FFI NEO Five-Factor Inventory. 19
NLP Natural Language Processing. xvii, 23
NUI Natural User Interface. 25

SDK Software Development Kit. 27
SOR Stimulus-Organism-Response model. 15

TPB Theory of Planned Behavior. xvii, 14, 15, 17, 19
TRA Theory of Reasoned Action. 3
TTS Text-to-Speech. 31

UNHCR United Nations High Commissioner for Refugees. 30
UV Uncanny Valley. 18

VUI Voice User Interface. 10

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

APPENDIX A

Informed Consent Form

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

INFORMED CONSENT

TITLE OF STUDY

Influence of Anthropomorphism in AI-enabled Voice Assistants on User Trust

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

PRINCIPAL INVESTIGATOR

Edona Haziraj
Technische Universität Wien
Media and Human-Centered Computing
e12224127@student.tuwien.ac.at

PURPOSE OF STUDY

You are being asked to take part in a research study. Before you decide to participate, it is important that you understand why the research is being conducted and what it will involve. Please read the following information carefully and feel free to ask the researcher if anything is unclear or if you need more information.
The purpose of this study is to investigate how different levels of anthropomorphism (human-like characteristics) in AI voice assistants—such as voice tone and gender—affect user trust and engagement.

STUDY PROCEDURES

If you agree to participate, you will be asked to interact with three different AI voice assistants during a 20-30 minute session. Each assistant will guide you through predefined tasks such as answering questions, managing simple tasks, and engaging in conversation. After interacting with each assistant, you will be asked to fill out a brief questionnaire about your experience.
You may be asked to complete both multiple-choice and open-ended questions. Observational data, such as your reactions and engagement during the tasks, will also be observed to help assess your interaction with the voice assistants.

RISKS

There are no significant risks associated with this study. You may decline to answer any questions or stop your participation at any time without any consequences.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

BENEFITS

There will be no direct benefit to you for your participation in this study. However, your participation will help further research in human-computer interaction and contribute to the development of more inclusive and trusted AI systems. Additionally, this experience will serve as a learning experience on steps of how to conduct a user study in the future as part of User Research Methods course.

CONFIDENTIALITY

As part of gathering data, your name, gender, age, email, and audio of the open-ended questions will be stored. Your responses will be confidential. No identifying information will be attached to your responses. All data will be stored securely and will only be accessible by the research team. Data will be kept for research purposes and may be included in future publications.

CONTACT INFORMATION

If you have any questions about the study or experience any adverse effects as a result of participating, you may contact the principal investigator, Edona Haziraj, at e12224127@student.tuwien.ac.at or Edona.Haziraj@gmail.com

VOLUNTARY PARTICIPATION

Your participation in this study is entirely voluntary. You are free to choose not to participate. Even if you decide to participate, you can withdraw from the study at any time without any reason, and doing so will not affect your relationship with the researcher. If you choose to withdraw from the study, any collected data will be returned or destroyed.

CONSENT

I have read and understand the information provided above. I have had the opportunity to ask questions and received satisfactory answers. I understand that my participation is voluntary, and I may withdraw from the study at any time without giving a reason. I voluntarily agree to participate in this study.

Participant's Signature: ________________________________ Date: ____________
Investigator's Signature: ________________________________ Date: ____________

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

APPENDIX B

Participant Drawings: Voice Assistant I - OpenAI Juniper

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

SVMV1

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

[redacted]

7M4V1

white, US-american, young female

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

SE5V1

- female tone
- sounds familier and confident
- advices sound really helpfull for these cases.

J9BV1

female, young, white

TU Bibliothek Your knowledge hub WIEN Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

9PP5V1

WHITE, BLONDE, WHITE TEETH
HUGE SMILE, INVESTED IN
APPEREAMCE BUT
SMART & COMPETENT
YOUNG AMERICAN(USA)
GIRL 21-25
Y.O.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

8C ENV1

I've heard a female Person with a very good voice. It's really calm when people hear very good voice when they seek for a help.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

9KR2V1

- has clear teeth
- long hair but with a low ponytail
- wearing a blouser that is seamlessly ironed
- no accesories
- woman

APPENDIX C

Participant Drawings: Voice Assistant II - OpenAI Cove

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

WQF1V2

- white middle aged man
- sounded competent
- sounded nice + ~~[illegible]~~
- can imagine him to tell jokes
- brown hair
- looks a bit like the actor from Wolverine (Wolverine) Hugh Jackman

TU Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

H8M2V2

- male
- mid 30s
- short hair (like a buzz cut)
- short beard

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

9QN6V2 :

not really imagining a human behind the voice
not really a robot either ~~per~~ but more in the direction of a machine
and I do not mind it

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Z4J7V2:

associations

- male, middle aged (~40)
- tanned
- bearded
- dark hair

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

LJH5V2:

- brown hair
- male
- short hair
- 37 y.o.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

4GY9V2

Characteristics:

- middle aged white CIS man

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

B8K4V2

- male read
- black hair (short)
- 30-40yo
- white
- 1,~~60~~70m - 1,90m

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

PLM8V2:

friendly

male - voiced

typical 0815 office guy

most common

white, light beard

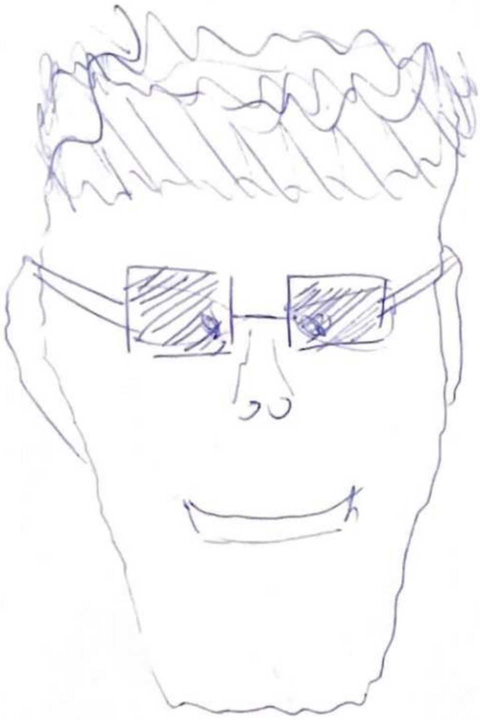
calm

Urban area

not really interested in your problems.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TRX9V2



friendly
male
calm
open-minded, not biased
30-40 yrs old

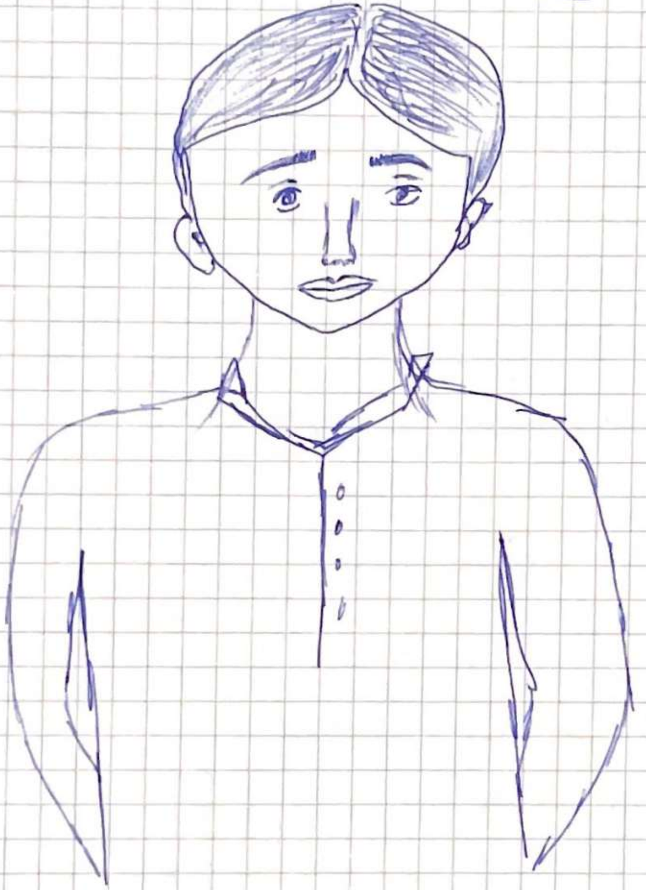
APPENDIX D

Participant Drawings: Voice Assistant III - OpenAI Breeze

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

XJ 6BV3



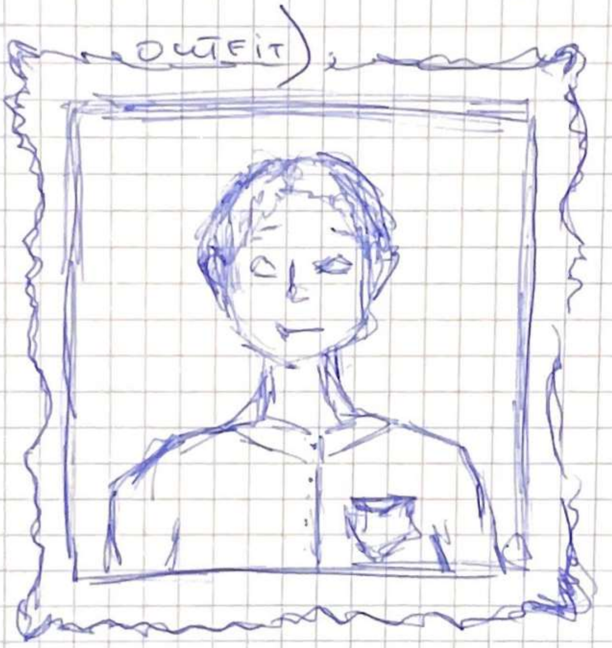
25 year old guy wearing a black dress shirt

looking more like a telephone operator

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

3D7PV3

PERSON IMAGINED BEHIND: MAN OF THE AGE OF 35-40 YEARS OLD IN AN OFFICE-SUIT (CORPORATE STYLE OUTFIT).



7K 5BV3

- female
- middle-age
- "strong mom vibes"
- tough
- hands-on

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

LJ7FV3

A man in his 20ies. Sharp face features, tall, Because of rather robotic voice I didn't have a very realistic person in mind, but more like a robot looking like human.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

APPENDIX E

Prompt script (Task scenarios)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TASK I:

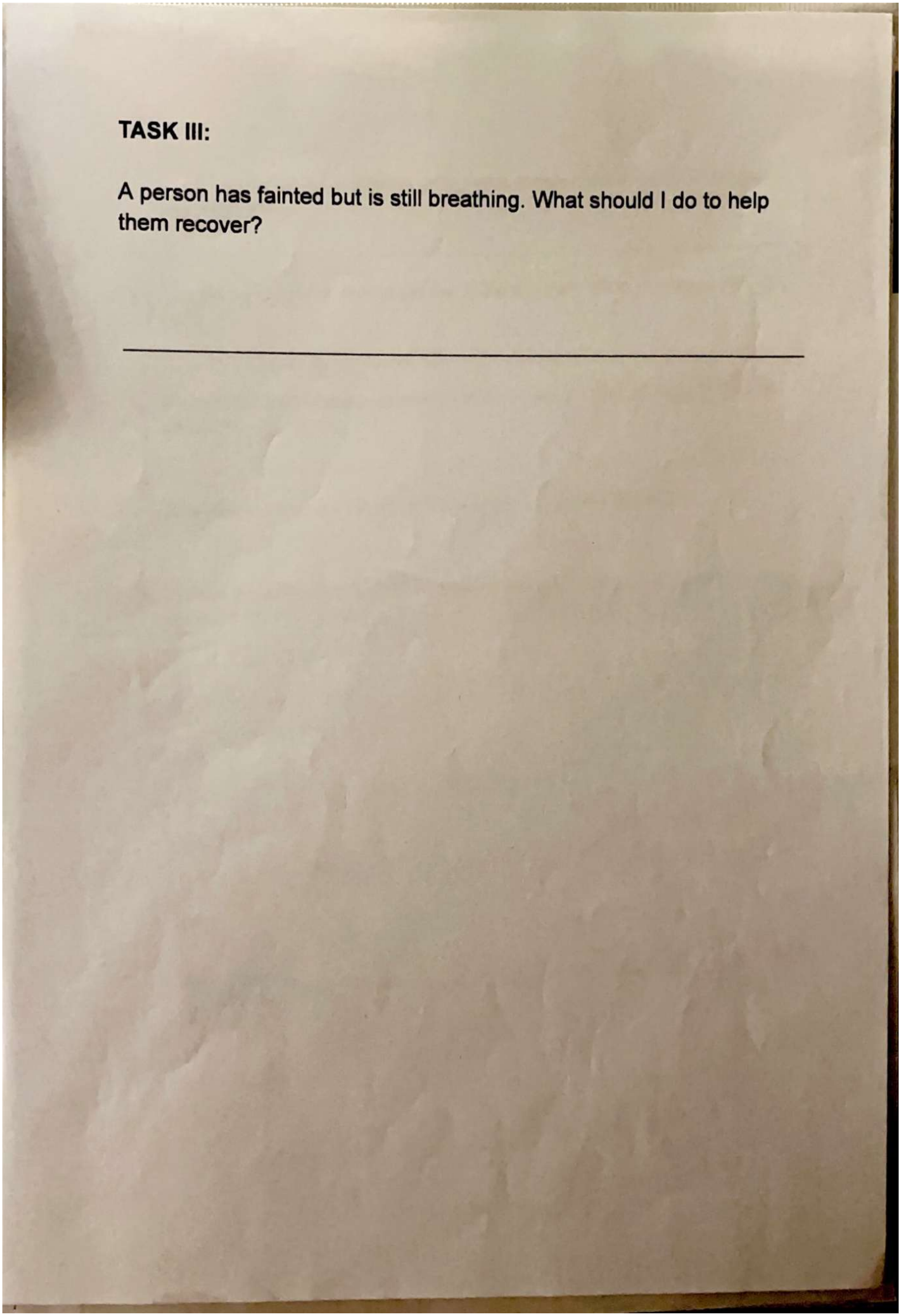
As a future parent, I would have one concern of my kid having trouble focusing on their homework. How can I help them stay motivated?

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TASK II:

One of the engines of the spacecraft is overheating.
What should I check to fix it?

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.



TASK III:

A person has fainted but is still breathing. What should I do to help them recover?

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

APPENDIX **F**

Pre-Interaction Survey Questions

Table F.1: Pre-Interaction Survey Questions

No.	**Question**	**Type**	**Options**
1	What is your age?	Multiple Choice	18–24, 25–34, 35+
2	What is your gender?	Multiple Choice	Male, Female, Non-binary, Prefer not to say, Other
3	What is your highest level of education?	Multiple Choice	Bachelor's degree, Master's degree, PhD or higher, Other
4	Please rate your familiarity with AI technology and voice assistants, from 1 (Not Familiar) to 5 (Very Familiar).	Likert Scale	1–5
5	Have you used any AI-enabled voice assistants before (e.g., Siri, Alexa, Google Assistant, ChatGPT)?	Yes/No	Yes, No
6	If yes, how frequently do you use AI voice assistants?	Multiple Choice	Daily, Weekly, Monthly, Rarely, Never
7	Which AI voice assistants have you used? (Select all that apply)	Checkboxes	Apple Siri, Amazon Alexa, Google Assistant, Microsoft Cortana, Chat GPT Voice Assistant, Other
8	For what purposes do you typically use voice assistants? (Select all that apply)	Checkboxes	Setting reminders/alarms, Searching for information, Controlling smart home devices, Sending texts or making calls, Playing music or podcasts, Other
9	How much do you trust AI voice assistants to provide accurate information? (1 = No Trust, 5 = Complete Trust)	Likert Scale	1–5

Continued on next page

Table F.1: Pre-Interaction Survey Questions (continued)

No.	Question	Type	Options
10	Please rate how comfortable you are with AI voice assistants making decisions on your behalf (e.g., booking a service, suggesting products), from 1 (Not Comfortable) to 5 (Very Comfortable).	Likert Scale	1–5
11	Please rate how much you believe AI voice assistants respect your privacy, from 1 (Not at All) to 5 (Completely).	Likert Scale	1–5
12	Please rate how important the gender of the voice assistant is to you, from 1 (Not Important) to 5 (Very Important).	Likert Scale	1–5
13	Please rate how important it is for the voice assistant to sound natural and human-like, from 1 (Not Important) to 5 (Very Important).	Likert Scale	1–5
14	What characteristics of a voice assistant would make you trust it more? (Select all that apply)	Checkboxes	Human-like voice, Clear tone, Gendered voice, Ability to express emotions, Other
15	Before interacting with AI voice assistants in this study, what do you expect your overall experience to be like? from 1 (Very Negative) to 5 (Very Positive).	Likert Scale	1–5

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

APPENDIX G

Post-Interaction Survey Questions

Table G.1: Post-Interaction Survey Questions

No.	Question	Type	Options
1	Please rate your overall experience interacting with the AI voice assistant.	Likert Scale	1 (Very Negative) to 5 (Very Positive)
2	Please rate how easy or difficult it was to understand the AI voice assistant.	Likert Scale	1 (Very Difficult) to 5 (Very Easy)
3	Please rate how much you trusted the AI voice assistant to provide helpful and accurate information.	Likert Scale	1 (Not at All) to 5 (Completely)
4	To what extent did you feel confident that the assistant's responses were reliable?	Likert Scale	1 (Not Confident at All) to 5 (Extremely Confident)
5	The voice assistant had a unique personality that influenced how I trusted it.	Likert Scale	1 (Strongly Disagree) to 5 (Strongly Agree)
6	The assistant seemed knowledgeable and competent during the tasks.	Likert Scale	1 (Strongly Disagree) to 5 (Strongly Agree)
7	The assistant communicated in a way that felt natural and engaging.	Likert Scale	1 (Strongly Disagree) to 5 (Strongly Agree)
8	The assistant's voice tone made it easy for me to trust its responses.	Likert Scale	1 (Strongly Disagree) to 5 (Strongly Agree)
9	How comfortable were you with the characteristics of the voice assistant's persona?	Likert Scale	1 (Very Uncomfortable) to 5 (Very Comfortable)
10	Do you think you would trust an AI voice assistant more if its gender matched your own?	Likert Scale	1 (Strongly Disagree) to 5 (Strongly Agree)
11	Please rate how appropriate the voice assistant's tone felt for the parenting advice task.	Likert Scale	1 (Very Inappropriate) to 5 (Very Appropriate)
12	Please rate how much you trusted the AI voice assistant to provide accurate technical advice for problem-solving.	Likert Scale	1 (Not at All) to 5 (Completely)

Continued on next page

Table G.1: Post-Interaction Survey Questions (continued)

No.	Question	Type	Options
13	Please rate how appropriate the assistant's voice tone felt for an emergency response situation.	Likert Scale	1 (Very Inappropriate) to 5 (Very Appropriate)
14	Please rate how confident you were that the assistant's response would be helpful in a real emergency.	Likert Scale	1 (Not Confident at All) to 5 (Extremely Confident)
15	In which task did you feel the most trust toward the AI voice assistant?	Short Answer	Open-ended response

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

APPENDIX H

Semi-structured Interview Guide

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

POST-SURVEY SEMI-STRUCTURED INTERVIEW GUIDE

TITLE OF STUDY

AI-enabled Voice Assistants in Various Task Scenarios

POINT OF CONTACT

Edona Haziraj
Technische Universität Wien
Media and Human-Centered Computing
e12224127@student.tuwien.ac.at

- PURPOSE & CONTEXT
  - This interview is conducted after participants have completed two consecutive surveys about AI voice assistants. The goal is to unpack why participants felt (dis)trust, particularly regarding gender cues in synthetic voices, and to gather illustrative quotes for thematic analysis.

- INTERVIEWER PREP CHECKLIST
  - Review the participant's Survey A & Survey B responses and note any anomalies to probe. Confirm informed-consent form is signed Test audio recorder & backup device; clear storage. Quiet, private room; phone on silent. Open this guide, note-taking sheet, and demographic sheet.

- INTERVIEW FLOW & SCRIPTS
  - Start recording – state participant ID and date.
  - WELCOME "Thank you for taking the time to speak with me today after completing the two surveys. I'm really interested in hearing about your experiences in your own words. Before we start I want to remind you that:
    - Participation is voluntary—you may skip any question or stop at any time.
    - The session will be audio-recorded for transcription; your responses will be anonymised.
    - Any questions before we begin?"

- WARM-UP "To begin, could you briefly describe your everyday experience with voice assistants?"

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

POST-SURVEY SEMI-STRUCTURED INTERVIEW GUIDE

QUESTIONS

- CORE QUESTIONS & PROBES

- — **Q1**. In your own words, what aspects of the voice assistant influenced your level of trust the most? • Probe: Which aspect came to mind first? Why? • Probe: Can you recall a specific moment during the task when your trust changed?

- — **Q2**. How did the gender or lack of gender in the AI voices affect your interaction or trust? • Probe: Did you notice the gender right away or later on? • Probe: How did that make you feel? Can you give an example?

- — **Q3**. Do you think that AI voice assistants should always be gender-neutral? Why or why not? • Probe: What benefits or drawbacks do you foresee? • Probe: Does the context (e.g., task type, culture) change your view?

- — **Q4.** What would make you trust an AI voice assistant more in the future? • Probe: If you could change one thing about today's voice assistants, what would it be? • Probe: Are there design cues—visual, verbal, or behavioural—that would help?

- — **Q5.** Is there anything else you'd like to share about your experience with the AI voice assistants in this study? • Probe: Were there any surprises or frustrations we haven't talked about? • Probe: Any advice for designers of future AIs?

- WRAP-UP "Those are all my questions. Is there anything we didn't cover that you'd like to add?" • Explain next steps: when/where findings will be shared. • Re-affirm confidentiality and thank participant.

- POST-INTERVIEW TASKS
  - Stop recording; verify file saved. Label file with participant ID, date, and interviewer initials. Write a short reflection: notable emotions, body language, emerging themes.

- DATA MANAGEMENT
  - PLAN Recordings and transcripts will be stored on an encrypted drive accessible only to the research team. Audio files will be deleted once coding is complete (target date: 30 April 2025). Transcripts will be pseudonymised.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Bibliography

[ACPR21] Gavin Abercrombie, Amanda Cercas Curry, Mugdha Pandya, and Verena Rieser. Alexa, google, siri: What are your pronouns? gender and anthropomorphism in the design and perception of conversational assistants. *CoRR*, abs/2106.02578, 2021.

[Agr19] Roxana Agrigoroaie. *Exploring the relationship between morningness-eveningness, cognitive performance and the internal physiological state in different human-robot interaction scenarios*. Ph.d. thesis, Université Paris Saclay (COmUE), 2019. NNT: 2019SACLY005.

[AJGS22] Amani Alabed, Ana Javornik, and Diana Gregory-Smith. Ai anthropomorphism and its effect on users' self-congruence and self–ai integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change*, 182:121786, 2022.

[Ajz91] Icek Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50:179–211, 12 1991.

[ARVL24] Maryam Alimardani, Robyn Roode, Julija Vaitonyte, and Max Louwerse. Effect of a virtual agent's appearance and voice on uncanny valley and trust in human-agent collaboration. 09 2024.

[AU80] Marina Abramović and Ulay. Rest energy. Museum of Modern Art (MoMA), 1980. Performance art. Retrieved from `https://www.moma.org/audio/playlist/243/3100`.

[AYYL20] Muhammad Ashfaq, Jiang Yun, Shubin Yu, and Sandra Maria Correia Loureiro. I, chatbot: Modeling the determinants of users' satisfaction and continuance intention of ai-powered service agents. *Telematics and Informatics*, 54:101473, 2020.

[BD22] Christine Bauer and Johanna Devaney. Constructing gender in audio: Exploring how the curation of the voice in music and speech influences our conception of gender identity. In Marcus Erbe, Aycha Riffi, and Wolfgang Zielinski, editors, *Mediale Stimmentwürfe*, pages 83–100. kopaed, 2022.

[BG18] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Sorelle A. Friedler and Christo Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 77–91. PMLR, 23–24 Feb 2018.

[BST21] Cat Bennett, Jordan Smith, and Emily Taylor. Genderless voice assistants and nonbinary user perceptions. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[But24] Judith Butler. Who's Afraid of Gender?, 2024. Goodreads page, accessed 2025-05-02.

[Bă23] Carmen Bălan. Chatbots and voice assistants: Digital transformers of the company–customer interface—a systematic review of the business research literature. *Journal of Theoretical and Applied Electronic Commerce Research*, 18(2):995–1019, 2023.

[CB13] Victoria Clarke and Virginia Braun. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist*, 26:120–123, 02 2013.

[CCT20] Melody M. Chin, Crystal Chao, and Andrea L. Thomaz. Empathy for everyone: The effect of gender and personality on empathetic behavior in interactive voice agents. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 33–42, 2020.

[CFG$^+$21] Lorenzo Cominelli, Francesco Feri, Roberto Garofalo, Caterina Giannetti, Miguel A. Meléndez-Jiménez, Alberto Greco, Mimma Nardelli, Enzo Pasquale Scilingo, and Oliver Kirchkamp. Promises and trust in human–robot interaction. *Scientific Reports*, 11(1):9687, 2021.

[Cin22] Cintya Chaves::::::San Francisco State University:::. Voice as identity: Creating a genderless voice assistant, June 2022.

[CMK12] David Coyle, James Moore, and Per Ola Kristensson. Exploring the relationship between anthropomorphism and trust in technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 13–22. ACM, 2012.

[CW91] Donald G Childers and Kui Wu. Gender recognition from speech. part ii: Fine analysis. *The Journal of the Acoustical Society of America*, 90(4):1841–1856, 1991.

[DAHD20] Sylvain Daronnat, Leif Azzopardi, Martin Halvey, and Mateusz Dubiel. Impact of agent reliability and predictability on trust in real time human-agent collaboration. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, HAI '20, page 131–139, New York, NY, USA, 2020. Association for Computing Machinery.

[DAHD21] S. Daronnat, L. Azzopardi, M. Halvey, and M. Dubiel. Inferring trust from users' behaviours; agents' predictability positively affects trust, task performance and cognitive load in human-agent real-time collaboration. *Frontiers in Robotics and AI*, 8:642201, 2021.

[Dar21] Sylvain Daronnat. *Factors influencing trust, reliance, performance and cognitive workload in human-agent collaboration.* PhD thesis, University of Strathclyde, 2021. Accessed: 2024-06-21.

[Deb] Sopan Deb. Please and thank you: How politeness costs openai tens of millions of dollars. *The New York Times.*

[DHHP$^+$23] Andreea Danielescu, Sharone A Horowit-Hendler, Alexandria Pabst, Kenneth Michael Stewart, Eric M Gallo, and Matthew Peter Aylett. Creating inclusive voices for the 21st century: A non-binary text-to-speech for conversational assistants. In *Proceedings of the 2023 CHI Conference on Human Factors*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

in Computing Systems, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[DL22] Jianning Dang and Li Liu. Implicit theories of the human mind predict competitive and cooperative responses to ai robots. *Computers in Human Behavior*, 134:107300, 2022.

[EH12] Friederike Eyssel and Frank Hegel. Activating social categorization: When anthropomorphism leads to stereotype-based expectancies. *Journal of Applied Social Psychology*, 42(1):247–254, 2012.

[EKB$^+$12] Friederike Eyssel, Dieta Kuchenbrandt, Simon Bobinger, Laura De Ruiter, and Frank Hegel. 'if you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. 03 2012.

[EKHdR12] Friederike Eyssel, Dieta Kuchenbrandt, Frank Hegel, and Laura de Ruiter. Activating elicited agent knowledge: How robot and user features shape the perception of social robots. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 851–857, 2012.

[EW12] Alice Eagly and Wendy Wood. Social role theory. In Paul A. M. Van Lange, Arie W. Kruglanski, and E. Tory Higgins, editors, *Handbook of Theories of Social Psychology*, volume 2, pages 458–476. SAGE Publications Ltd, 2012.

[EWC07] Nicholas Epley, Adam Waytz, and John T. Cacioppo. On seeing human: a three-factor theory of anthropomorphism. *Psychological Review*, 114(4):864–886, October 2007.

[Fin12] Julia Fink. Anthropomorphism and human likeness in the design of robots and human-robot interaction. volume 7621, pages 199–208, 10 2012.

[FS22] Janik Festerling and Iram Siraj. Anthropomorphizing technology: A conceptual review of anthropomorphism research and how it relates to children's engagements with digital voice assistants. *Integrative Psychological and Behavioral Science*, 56(3):709–738, 2022.

[fur25] Furhat robotics: Next-generation humanoid robot for research and innovation. `https://www.furhatrobotics.com/`, 2025. Accessed 1 May 2025.

[GALA22] Kathryn Gorton, R. Allan Allday, Justin D. Lane, and Melinda J. Ault. Effects of brief training plus electronic feedback on increasing quantity and intonation of behavior specific praise among preschool teachers. *Journal of Behavioral Education*, 31(4):731–750, dec 2022.

[GB19] N. Goksel and A. Bozkurt. Artificial intelligence in education: Current insights and future perspectives. In S. Sisman-Ugur and G. Kurubacak, editors, *Handbook of Research on Learning in the Age of Transhumanism*, pages 224–236. IGI Global Scientific Publishing, 2019.

[GD19] Shalini Gulati and Sanjay Dubey. Effects of gendered synthetic voices on user perception and trust in voice assistants. *International Journal of Human–Computer Interaction*, 35(9):815–829, 2019.

[GGW10] Heather Gray, Kurt Gray, and Daniel M. Wegner. Mind perception: A new account of how people judge the minds of others and objects. *Journal of Personality and Social Psychology*, 101(3):390, 2010.

[GM23] Kylie L. Goodman and Christopher B. Mayhorn. It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance. *Applied Ergonomics*, 106:103864, 2023.

[Gof59] Erving Goffman. *The Presentation of Self in Everyday Life.* Anchor Books, 1959.

[Hag19] Martin Hagger. *The Reasoned Action Approach and the Theories of Reasoned Action and Planned Behavior.* 03 2019.

[HBL23] Verena Hochreiter, Caterina Benedetto, and Matthias Loesch. The stimulus-organism-response (s-o-r) paradigm as a guiding principle in environmental psychology: Comparison of its usage in consumer behavior and organizational culture and leadership theory. *Journal of Entrepreneurship and Business Development*, 3(1):7–16, 2023.

[HL23] Wen-Chin Hsu and Mu-Heng Lee. Semantic technology and anthropomorphism: Exploring the impacts of voice assistant personality on user trust, perceived risk, and attitude. *Journal of Global Information Management*, 31:1–21, 01 2023.

[Hoy18] Matthew B. Hoy. Alexa, siri, cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, 37(1):81–88, 2018. PMID: 29327988.

[HS44] Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259, 1944.

[HWVD$^+$19] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–16, New York, NY, USA, 2019. Association for Computing Machinery.

[IL23] Jonas Ivarsson and Oskar Lindwall. Suspicious minds: the problem of trust and conversational agents. *Computer Supported Cooperative Work (CSCW)*, 32:1–27, 04 2023.

[Int23] RTI International. Natural user interfaces and smart personalization: Lab 58 technology research brief. Technical report, RTI International, 2023.

[JCL24] Shizhen (Jasper) Jia, Oscar Hengxuan Chi, and Lu Lu. Social robot privacy concern (srpc): Rethinking privacy concerns within the hospitality domain. *International Journal of Hospitality Management*, 122:103853, 2024.

[Joh18] David Kyle Johnson. *Anthropomorphic Bias*, chapter 69, pages 305–307. John Wiley Sons, Ltd, 2018.

[KEK23] Michelle R. Kaufman, Emily L. Eschliman, and Taylor S. Karver. Differentiating sex and gender in health research to achieve gender equity. *Bulletin of the World Health Organization*, 101(10):666–671, 2023.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

[KNA16] Casey A. Klofstad, Stephen Nowicki, and Robert C. Anderson. How voice pitch influences our choice of leaders. *The American Scientist*, 104(5):282, 2016.

[KSC+22] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. ”because ai is 100% right and safe”: User attitudes and sources of ai authority in india. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI ’22, New York, NY, USA, 2022. Association for Computing Machinery.

[KW10] Robin S. S. Kramer and Robert Ward. Internal facial features are signals of personality and health. *Quarterly Journal of Experimental Psychology*, 63(11):2273–2287, 2010. Original work published 2010.

[La24] Changhyun Lee and Kyungjin Cha and. Toward the dynamic relationship between ai transparency and trust in ai: A case study on chatgpt. *International Journal of Human–Computer Interaction*, 0(0):1–18, 2024.

[LC+22] Nan Li, Wai Chan, et al. Cross-cultural differences in anthropomorphism: A meta-analytic review. *Computers in Human Behavior*, 134:107302, 2022.

[LN05] Jieun Lee and Clifford Nass. The psychology of human-agent interaction: Designing for engagement. *AI Society*, 19(1):18–26, 2005.

[LS04] John Lee and K.A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46:50–80, 04 2004.

[LSa21] Anna Lindqvist, Marie Gustafsson Sendén, and Emma A. Renström and. What is gender, anyway: a review of the options for operationalising gender. *Psychology & Sexuality*, 12(4):332–344, 2021.

[LW18] Irene Lopatovska and Harriet Williams. Personification of the amazon alexa: Bff or a mindless companion. pages 265–268, 03 2018.

[MB07] Benjamin Munson and Molly Babel. Loose lips and silver tongues, or, projecting sexual orientation through speech. *Blackwell Publishing Ltd Language and Linguistics Compass*, 1, 09 2007.

[MC87] Robert McCrae and Paul Costa. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52:81–90, 01 1987.

[MK19] Daniel McDuff and Ashish Kapoor. Toward emotionally intelligent artificial intelligence, April 2019. Accessed: 2025-04-27.

[MMK12] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. The uncanny valley [from the field]. *IEEE Robotics Automation Magazine*, 19(2):98–100, 2012.

[MOP+20] Federico Masina, Valentina Orso, Paola Pluchino, Giulia Dainese, Silvia Volpato, Cristian Nelini, Davide Mapelli, Anna Spagnolli, and Luciano Gamberini. Investigating the accessibility of voice assistants with impaired users: Mixed methods study. *Journal of Medical Internet Research*, 22(9):e18431, 2020.

[MR74] Albert Mehrabian and James A. Russell. *An Approach to Environmental Psychology*. The MIT Press, 1974.

[MRMZ+21] Daniel Mota-Rojas, Chiara Mariti, Andrea Zdeinert, Giacomo Riggio, Patricia Mora-Medina, Alondra Del Mar Reyes, Angelo Gazzano, Adriana Domínguez-Oliva, Karina Lezama-García, Nancy José-Pérez, and Ismael Hernández-Ávalos. Anthropomorphism and its adverse effects on the distress and welfare of companion animals. *Animals (Basel)*, 11(11):3263, November 2021.

[NB05] Clifford Nass and Scott Brave. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press, 2005.

[NM00] Clifford Nass and Youngme Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000.

[NMG97] Clifford Nass, Youngme Moon, and Nancy Green. Are machines gender neutral? gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, 27(10):864–876, 1997.

[NPR19] NPR. Meet q: The gender-neutral voice assistant, 2019. Accessed: 2024-09-18.

[NSS94] Clifford Nass, Jonathan Steuer, and Ellen Siminoff. Computer are social actors. page 204, 04 1994.

[Oat01] Jennifer Oates. Voice changes in transgender people: A review of the literature. *International Journal of Transgender Health*, 5(3):1–5, 2001.

[PCS+19] Aleksandra Przegalinska, Leon Ciechanowski, Anna Stroz, Peter Gloor, and Grzegorz Mazurek. In bot we trust: A new methodology of chatbot performance measures. *Business Horizons*, 62(6):785–797, 2019. Digital Transformation Disruption.

[Pep18] Robert Pepperell. *Art, energy, and the brain*, volume 237. 01 2018.

[Pla24] Adriana Placani. Anthropomorphism in AI: hype and fallacy. *AI Ethics*, 4(3):691–698, August 2024.

[Pot24] Vijaya Pothuri. Natural language processing and conversational ai. *International Research Journal of Modernization in Engineering Technology and Science*, 6:436–440, 09 2024.

[PTS+17] Amanda Purington, Jessie Taft, Sneha Sannon, Natalya N. Bazarova, and Samuel W. Taylor. ”alexa is my new bff”: Social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2853–2859, 2017.

[RCIZ19] Kennedy Ralston, Yuhao Chen, Haruna Isah, and Farhana Zulkernine. A voice interactive multilingual student support system using ibm watson, 2019.

[RN96] Byron Reeves and Clifford Nass. *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, 1996.

[RRARNMV21] Francis Rakotomalala, Hasindraibe Randriatsarafara, Hajalalaina Aimé Richard, and Ravonimanantsoa Ndaohialy Manda Vy. Voice user interface: Literature review, challenges and future directions. *SYSTEM THEORY, CONTROL AND COMPUTING JOURNAL*, 1:65–89, 12 2021.

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

[SEK24] Arndt Schäfer, Reinhold Esterbauer, and Bettina Kubicek. Trusting robots: a relational trust definition based on human intentionality. *Humanities and Social Sciences Communications*, 11(1):1412, 2024.

[SFKF19] Katta Spiel, Christopher Frauenberger, Os Keyes, and Geraldine Fitzpatrick. Agency of autistic children in technology research—a critical literature review. *ACM Trans. Comput.-Hum. Interact.*, 26(6), November 2019.

[SJW$^+$24] Verena Skuk, Isabell Jacob, Rebecca Wientzek, Robert Ward, and Stefan Schweinberger. Personality perceptions from voices and faces – impressions and kernels of truth, 04 2024.

[SK21] Yolande Strengers and Jenny Kennedy. The smart wife: Why siri, alexa, and other smart home devices need a feminist reboot, 2021. MIT Press (alternate reference for the same book, year labeled as 2021 in some citations).

[SL21] Hervé Saint-Louis. Machine-human interaction: A paradigm shift? In Masaaki Kurosu, editor, *Human-Computer Interaction. Theory, Methods and Tools*, pages 123–136, Cham, 2021. Springer International Publishing.

[SLAD15] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, page 141–148, New York, NY, USA, 2015. Association for Computing Machinery.

[Smi21a] Julianna C. Smith. *Vox ex Machina: A Cultural History of Talking Machines*. MIT Press, Cambridge, MA, 2021.

[Smi21b] Julianna C. Smith. "bitching betty" speaks: How talking machines got their gendered voices. *The MIT Press Reader*, March 2021.

[SN22] Tomohiro Suzuki and Tatsuya Nomura. Gender preferences for robots and gender equality orientation in communication situations. *AI SOCIETY*, 04 2022.

[SPEN25] Tianyu Song, Felix Pabst, Ulrich Eck, and Nassir Navab. Enhancing patient acceptance of robotic ultrasound through conversational virtual agent and immersive visualizations, 2025.

[Spi21a] Katta Spiel. The bodies of tei – investigating norms and assumptions in the design of embodied interaction. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[Spi21b] Katta Spiel. "why are they all obsessed with gender?" — (non)binary navigations through technological infrastructures. pages 478–494, 06 2021.

[SSW24] Katie Seaborn, Yuto Sawa, and Mizuki Watanabe. Coimagining the future of voice assistants with cultural sensitivity. *Human Behavior and Emerging Technologies*, 2024:1–21, March 2024.

[Sta25] Fox5DC.com Staff. Saying 'please' and 'thank you' to chatgpt costs energy. Fox5DC.com, 2025. Reports on a 2024 survey of polite language usage and Microsoft's "tone-mirroring" findings.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

[Sut19] Joy Sutton. Siri, gender, and synthesized speech: a critical look at gender performativity in voice assistants. *Feminist Media Studies*, 19(4):412–425, 2019.

[SVK21] William Seymour and Max Van Kleek. Exploring interactions between trust, anthropomorphism, and relationship development in voice assistants. volume 5, pages 1–16, 10 2021.

[SWLF19] Amy Spencer, Joanna Weidler-Lewis, and Paul Fitzpatrick. Mental models of anthropomorphism: Mechanistic versus mindful representations of robots. *ACM Transactions on Human-Robot Interaction (THRI)*, 8(3):1–25, 2019.

[TE24] Bora Tarlan and Nisa Erdal. How can i assist you today?: A comparative analysis of a humanoid robot and a virtual human avatar in human perception, 12 2024.

[TEN$^+$24] Caja Thimm, Phillip Engelhardt, Patrick Nehls, Jiliane Hens, Mira Biehler, Christopher Hermanns, Philipp Huzel, Maren Bennewitz, Jesper Mücke, and Nils Dengler. Trust, (dis)comfort, and voice quality: Anthropomorphism in verbal interactions with nao. In Constantine Stephanidis, Margherita Antona, Stavroula Ntoa, and Gavriel Salvendy, editors, *HCI International 2024 Posters*, pages 405–412, Cham, 2024. Springer Nature Switzerland.

[Tit94] Ingo R. Titze. *Principles of Voice Production*. Prentice Hall, Englewood Cliffs, NJ, 1994.

[Tur20] Sherry Turkle. *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster, 2020.

[UNE19] UNESCO. I'd Blush If I Could: Closing Gender Divides in Digital Skills Through Education, 2019. Accessed: 2025-04-05.

[UNH22] UNHCR. Digital transformation strategy 2022–2026. `https://www.unhcr.org/digitalstrategy/wp-content/uploads/sites/161/2023/07/Digital-Transformation-Strategy-2022-2026-UNHCR-Web.pdf`, 2022. Accessed: 2025-04-13.

[Vir19] EqualAI Virtue Nordic, Copenhagen Pride. Q: The first genderless voice. `https://www.genderlessvoice.com`, 2019. Accessed March 2025.

[Woo24] Geert M. Wood. Exploratory study: The impact of AI literacy training on the perception of anthropomorphism and trust in conversational AI. Master's thesis, Eindhoven University of Technology, Eindhoven, Netherlands, July 2024. Award date: 31 July 2024; Department of Industrial Engineering and Innovation Sciences.

[WRRW24] Samuel Westby, Richard J. Radke, Christoph Riedl, and Brooke Foucault Welles. How voice and helpfulness shape perceptions in human–agent teams. *Computers in Human Behavior: Artificial Humans*, 2(2):100101, August 2024.

[XT25] Anna Xygkou-Tsiamoulou. *Exploring companionship and social connectedness in interactions of groups vulnerable to social isolation with conversational virtual humans (CVHs)*. PhD thesis, University of Kent,, March 2025.

[Yag23] Deepak Yagamurthy. Advancements in natural language processing (nlp) and its applications in voice assistants and chatbots. *Journal of Artificial Intelligence Cloud Computing*, pages 1–6, 12 2023.

[YW22] Rongbin Yang and Santoso Wibowo. User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, 32(4):2053–2077, December 2022.

[Z+23] Wei Zhang et al. Trust in ai assistants: A systematic review. *International Journal of Human-Computer Studies*, 172:102786, 2023.

[ZJB24] Anne Zimmerman, Joel Janhonen, and Emily Beer. Human/ai relationships: challenges, downsides, and impacts on human/human relationships. *AI and Ethics*, 4(4):1555–1567, 2024.

[ZLZ+20] Jie Zhang, Shuo Li, Jing-Yu Zhang, Feng Du, Yue Qi, and Xun Liu. *A Literature Review of the Research on the Uncanny Valley*, pages 255–268. 07 2020.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub