

**GALATASARAY UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF RADIO, TELEVISION AND CINEMA**

**DIALOGUE FOR ALL: CRAFTING INCLUSIVE AND HUMANIZED
VOICE ASSISTANTS FOR DIVERSE POPULATIONS THROUGH AN
INTERDISCIPLINARY APPROACH**

DOCTORAL DISSERTATION

Yeliz YÜCEL

Supervisor: Prof. Dr. Kerem RIZVANOĞLU

SEPTEMBER 2023

**GALATASARAY UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF RADIO, TELEVISION AND CINEMA**

**DIALOGUE FOR ALL: CRAFTING INCLUSIVE AND HUMANIZED
VOICE ASSISTANTS FOR DIVERSE POPULATIONS THROUGH AN
INTERDISCIPLINARY APPROACH**

DOCTORAL DISSERTATION

Yeliz YÜCEL

Supervisor: Prof. Dr. Kerem RIZVANOĞLU

SEPTEMBER 2023

ACKNOWLEDGEMENTS

“It’s a rare gift, to know where you need to be, before you’ve been to all the places you don’t need to be.”

Ursula K. L. Guin, Thales from Earthsea

This adventure began while working on my master's degree at Galatasaray University. One day, Kerem Rızvanoğlu, my HCI professor at the time, casually dropped a question that hit me like a lightning bolt: "Do you know how a blind person uses a computer?" That was the moment I realized my understanding of the digital world was like a goldfish's knowledge of the ocean—limited. This simple question led me down a rabbit hole that made me question everything I thought I knew. At the time, I was deep into an ethnographic study with the 'Saturday Mothers,' examining whether the subaltern could actually talk against the backdrop of Turkey's intense political landscape, which is not indeed possible. This new angle from Prof. Rızvanoğlu, however, sparked an idea that rapidly evolved into a fascination.

Spotting an opportunity at the GSU Media and Communication Department, I dove headfirst into the mission of creating social value for vulnerable populations like children, women, LGBT individuals, older adults, and visually impaired individuals. The internet, I thought, should be this grand democratic environment that makes everyone equal. It's a bold concept, isn't it? However, I figured, why not give it a shot? And thus began my quest to make a dent in the digital divide, one user interface at a time. This study is the manifestation of my tenacious drive towards creating something unique and meaningful, a willful departure from the comfort of a white-collar world. It's akin to a mother's journey of nurturing her child. Yes, this is my child - my ideas, my reflections that I'm opening up to you all, possibly even giving voice to those who've long been silenced. Each contribution from our participants has been threaded with care into the fabric of this research, aiming to foster a more genuine, less formal, researcher-participant bond. It's not about grandstanding or making a bold statement but instead sharing a sincere dialogue born from the heart. Let's learn, unlearn, and relearn together.

I dedicate this study to Selim Suman, who, since 2013, has been persistently nudging me towards self-realization and offering unwavering love, even if I'm not lovable. This is for the one who did everything possible to make me believe in my own strengths, even though I have the tendency to be like a cat that retreats to its corner to lick its own wounds, poised to lash out at those who dare to approach. And, to my mom, who raised me as a fierce warrior, ever-ready to pounce - but let's face it, at heart, I'm more like a passionate meatball: crisp and unyielding on the outside, but a total softie within. Thanks for the love, Mom. A massive shout-out to Prof. Dr Kerem

Rızvanoğlu, my academic rock, whose unwavering support, perpetual enthusiasm, and insightful ideas kept me going. And let's not forget the gentle shakes every time I was about to doze off into the realm of self-doubt. Seriously, could I have asked for a cooler advisor? Nope! And thank you to my committee members, for always reminding me of the significance of my study. Your encouragement was a beacon in the process.

A big round of applause goes to my fantastic girl gang (you know who you are, my enduring friends and chosen family; Tuğçe Panzer, Clinical Psychologists Başak Gündeş, and Özlem Ünde) for the brainstorming sessions and the unconditional 'be there' without ever uttering the words. A million thanks to Gizem Sedef and Eren Elibol, my siblings from different parents, for stepping up as pro nurses when it seemed like illnesses and bad luck had installed a revolving door in my life. Shout-out to Onur Ç. and Yovaçavuş, my grounding forces, for constantly reminding me of my roots, of the importance of creating a possible societal value, and for possessing that amazing superpower of seeing beyond my own perspective.

I would like to also express my gratitude for the support received under the TÜBİTAK Science Fellowships and Grant Programmes Directorate (BİDEB) 2211-Domestic Postgraduate Scholarship Programme.

Thank you to all the participants who generously carved out their time for this study and to Galatasaray Nursing Home and the Turkey Foundation for the Blind. In the process of wrapping up this thesis, I ended up wrapping my arms around some new, extraordinary friends who are now waiting for a celebration. Honestly, what more could a researcher ask for? Bring on the confetti!

September, 2023

Yeliz Yücel

LIST OF CONTENTS

	Page No
ACKNOWLEDGEMENTS.....	ii
LIST OF CONTENTS.....	iv
ABBREVIATIONS	vii
LIST OF FIGURES	viii
LIST OF TABLES	ix
RÉSUMÉ.....	x
SUMMARY	xiv
ÖZET.....	xvii
1. INTRODUCTION.....	1
1.1. Purpose and the Scope of the Study	3
1.2. Interdisciplinary Theoretical Framework: Application of Social Science Theories in Human-Computer Interaction	7
1.3. Methodology and Measurement Tools.....	9
2. LEVERAGING SHARED LANGUAGE IN VOICE ASSISTANTS: ENHANCING RECIPROCITY FOR ADULTS WITH SPECIALIZED INTERESTS	12
2.1. Introduction	13
2.2. Theoretical Framework	15
2.3. Methodology and Research Questions	20
2.3.1. Research Design	22
2.3.2. Experimental Set-up: Wizard of Oz	24
2.3.3. Instrumentation	26
2.3.4. Participants	26
2.3.5. Procedure	27
2.4. Results & Discussion.....	28
2.4.1. Pre-test stage Findings	28
2.4.2. Qualitative Findings.....	29
2.4.2.1. Self Disclosure-High Effect	29
2 4.2.1.1. Building Trust and Engagement: The Impact of Reciprocal Self-Disclosure on Trust in Conversational Agents	29
2.4.2.1.2. Personifying the Conversational Agent: The Role of Human- Like Storytelling.....	31

2.4.2.1.3. Social Talk and Reciprocity: Politeness, Back-Channelling, and Turn-Taking.....	34
2.4.2.1.4. The Joy of Conversation: Prioritizing Engagement Over Perfection.....	36
2.4.2.2. Self-Disclosure-Low Effect	36
2.4.2.2.1. Participant Experience: Boredom and Distancing with the CA ..	37
2.4.2.2.2. Participant Perception: Superiority and Bossiness of the CA	38
2.4.2.3. Gender Effect	39
2.4.2.3.1. Unveiling Implicit Gender Stereotypes in Conversational Agents.....	40
2.4.2.3.2. Unintended Perceptions: Male Users' Interpretation of Feminist Attributes in Conversational Agents	42
2.4.3. Quantitative Findings.....	45
2.5. Conclusion.....	48
2.6. Limitations and Future Research.....	54
3. LISTENING TO CULTURAL REFLEXES: TAILORING EMPATHETIC VOICE ASSISTANTS FOR OLDER ADULT USERS	55
3.1. Introduction	56
3.2. Theoretical Framework	60
3.2.1. Empathy effect.....	61
3.2.2. Voice Maturity Effect:	62
3.2.3. Socio-cultural context.....	64
3.3. Methodology	66
3.3.1. Research Procedure and the Experimental Set-up with Wizard of Oz	68
3.3.2. Momentary Test Design and the Dialog Flow	69
3.3.3. Participants	73
3.4. Results & Discussion.....	74
3.4.1. Qualitative Results.....	74
3.4.1.1. Pre-test Results.....	74
3.4.1.2. Momentary Test results.....	75
3.4.1.3. Post-test results.....	76
3.4.1.3.1. Exploring Cultural Reflexes: Unveiling Discreet Companionship Beyond Human Interaction	77
3.4.1.3.2. Testing Boundaries and Finding Comfort: Male Participants' Perceptions and Challenges in Interacting with a Voice Assistant	81
3.4.1.3.3. Bridging the Generational Gap: The Transformative Role of a Voice Assistant during the Covid-19 Quarantine for Older Adults	83
3.4.1.3.4. Critiquing Less Empathic Voice Assistants: Unveiling User Dissatisfaction, Insecurity, and the Need for Authenticity in Conversations	85

3.4.1.3.5. Reevaluating Age and Similarity in Voice Assistant Preferences: Exploring the Role of Technological Relevance and Perceived Understanding	87
3.4.2. Quantitative Results	89
3.5. Conclusion	92
3.6. Limitations and Future Research	94
4. INCLUSIVE VOICES: CRAFTING VOICE ASSISTANT PROTOTYPES FOR THE BLIND USERS TO NAVIGATE CULTURAL EXPERIENCES	96
4.1. Introduction	96
4.2. Theoretical Framework	100
4.2.1. Humour and Laughter Effect	104
4.2.2. Sonification Effect	107
4.3. Methodology	108
4.3.1. Design Process and Dialog Flow	109
4.3.2. Research Procedure and Experimental Set-up of “Wizard of Oz”	116
4.3.3. Participants	118
4.4. Results and Discussion	118
4.4.1. Pre-test Results	119
4.4.2. Momentary Test Results	120
4.4.3. Post-Test Results	121
4.4.3.1. Interactive Engagement.....	121
4.4.3.2. Auditory-Enhanced Immersion and Spatial Understanding	124
4.4.3.3. Social Learning	126
4.4.3.4. Usability	129
4.5. Conclusion.....	131
5. CONCLUSION.....	135
5.1. Restatement of Research Problem and Objectives.....	135
5.2. Summary of Findings	135
5.3. Contribution to the Field	137
5.3.1. Theoretical Contributions and Implications	138
5.3.2. Methodological Contributions and Implications	141
5.3.3. Practical implications.....	143
BIBLIOGRAPHY	147
APPENDICES	170
Appendix A	170
Appendix B.....	172
CURRICULUM VITAE.....	174

ABBREVIATIONS

HCI	: Human-Computer Interaction
CA	: Conversational Agent
VA	: Voice Assistant
VUI	: Voice User Interface
CASA	: Computers are Social Actors Paradigm
SD	: Self-disclosure
HHI	: Human-Human Interaction
SIT	: Social Identity Theory
SAT	: Similarity-Attraction Theory
WoZ	: Wizard of Oz
PWVI	: People with Visual Impairments
SUS	: System Usability Scale
UVM	: Uncanny Valley of Mind

LIST OF FIGURES

	Page No
Figure 1.1: Smart speaker-based VAs: Alexa, Cortana and Google Home	2
Figure 2.1: Google Home Mini Smart Speaker-based Voice Assistant.....	15
Figure 2.2. Data Collection Process, Instruments and Constructs	21
Figure 2.3: One of the Wizard's screens while manipulating voice assistant	25
Figure 2.4: Conversational Movie Recommendation System Prototype.....	26
Figure 3.1: One of Wizard of Oz's Controlled Dialog Flow Panel	69
Figure 3.2: Our user-centered dialogue-flow design process	71
Figure 4.1: The iMuse Model: A Co-design Framework for Inclusive Voice- Assisted Remote Museum Experiences for the Visually Impaired.....	115
Figure 4.2: One of the Wizard of Oz's Controlled Dialog Flow Panel	117
Figure 4.3. Comparison of Quiz Scores for Humorous and Non-Humorous Voice Assistants	129
Figure 4.4: Comparison of SUS Scores	130
Figure 5.1: Holistic and Inclusive Framework for Voice Assistant Design	138
Figure 5.2: Theoretical Implications Diagram.....	141
Figure 5.3: Stage Prototype Design Process	143
Figure 5.4. Research-to-Application Diagram	145

LIST OF TABLES

	Page No
Table 2.1: Ranks.....	45
Table 2.2: Test Statistics	46
Table 2.3: Correlations for H2	46
Table 2.4: Ranks.....	47
Table 2.5: Test Statistics	47
Table 2.6: Correlations for H4	48
Table 3.1: Sources for Construct Items	72
Table 3.2: Two-way ANOVA Results of Dependent Variable I (mature x young voice).....	90
Table 3.3: Two-way ANOVA Results of Dependent Variable II (high emphatic expression x low emphatic expression).....	90
Table 3.4: T-Tests Results for Self-Efficacy After/Before	91

RÉSUMÉ

Les agents conversationnels (CAs) ont un potentiel transformateur qui va au-delà des domaines de la technologie et de l'interaction homme-machine (IHM). Ces technologies peuvent potentiellement avoir un impact sociétal significatif, en particulier pour réduire les fossés numériques et améliorer l'inclusion sociale. Reconnaître et aborder les différentes façons dont les humains interagissent avec les agents conversationnels, influencées par des caractéristiques culturelles, de genre, d'âge et de capacité, est essentiel. Ces variations déterminent non seulement comment les utilisateurs accèdent à l'information et l'utilisent, mais aussi comment ils l'interprètent et l'évaluent. À ce titre, ces agents pourraient risquer de propager et de reproduire des préjugés et des stéréotypes. Adapter les méthodes de communication à diverses populations d'utilisateurs peut aider à atténuer ces préoccupations, améliorant l'expérience utilisateur, augmentant l'inclusivité et élargissant l'accès.

L'utilisation croissante d'assistants vocaux personnels tels que Google Assistant, Alexa d'Amazon, SIRI, et Cortana de Microsoft nécessite une conception efficace capable de retenir l'attention des utilisateurs pour des conversations prolongées et naturelles. L'interaction homme-machine (IHM) doit garantir une utilisabilité universelle. Cependant, l'utilisabilité, l'accessibilité et l'inclusivité des nouvelles technologies restent limitées pour certaines populations spécifiques. En explorant le domaine peu étudié des populations d'utilisateurs d'assistants vocaux, avec une attention particulière portée aux personnes âgées, aux adultes avec les intérêts spécialisés, et aux personnes malvoyantes, cette étude apporte une contribution significative aux domaines de l'interaction homme-machine (IHM), de la psychologie sociale et de la communication.

Au niveau social, le développement d'agents conversationnels personnalisés pour répondre aux besoins individuels des utilisateurs a le potentiel de traiter les disparités d'accès à l'information et aux services. Ces agents ont le potentiel de contribuer de manière significative à la promotion de la santé, du bien-être et de la qualité de vie en général, en particulier pour les groupes vulnérables. En facilitant l'accès à l'information et aux services, ces systèmes ont la capacité d'atténuer efficacement les défis rencontrés par ces groupes d'utilisateurs, abordant ainsi directement les problèmes liés à l'inégalité sociale. L'intégration de techniques de co-conception et de méthodologies centrées sur l'utilisateur dans le développement de ces agents garantit leur utilisabilité et leur pertinence, maximisant ainsi leur impact social. L'implication active des utilisateurs finaux dans le processus de conception est de la plus haute importance pour des solutions culturellement pertinentes, non stéréotypées et inclusives. L'objectif ultime est de créer des agents computationnels humanisés qui répondent aux besoins diversifiés de la population, contribuant à un environnement numérique plus inclusif et accessible.

L'utilisation innovante de la technique du Magicien d'Oz (WoZ) est une contribution méthodologique significative. Cette technique permet aux scientifiques sociaux de manipuler diverses caractéristiques humaines sans programmation, offrant

ainsi une approche plus flexible et adaptable pour comprendre les interactions homme-machine. En intégrant le WoZ, cette étude ouvre de nouvelles opportunités pour la recherche en HCI, en particulier pour ceux qui n'ont pas de formation technique. De plus, l'approche mixte unique, combinée à des quiz avant et après la visite, des évaluations de tests momentanés, des entretiens avant et après la visite, des enquêtes, et des échelles quantifiables dans chaque chapitre, crée une compréhension complète et détaillée des interactions des utilisateurs avec les agents computationnels. Cette recherche contribue à une compréhension plus nuancée des dynamiques lors des interactions entre les humains et les entités artificielles en introduisant un cadre holistique et inclusif pour interpréter ces interactions.

Notre seconde chapitre se penche sur l'influence des signaux vocaux liés au genre et sur le style de conversation, en se concentrant sur le niveau d'auto-révélation exhibé par les agents conversationnels (AC). Nous cherchons à examiner les effets de ces facteurs sur les résultats attitudels et comportementaux, y compris la réciprocité de l'auto-révélation et la confiance. Cette étude a utilisé un prototype d'agent conversationnel (AC) basé sur la parole intégré dans un système d'assistant vocal reposant sur des enceintes intelligentes, spécifiquement Google Home. Nous avons employé la technique du Magicien d'Oz pour simuler un AC entièrement fonctionnel, impliquant des opérateurs humains contrôlant les réponses du système. Notre conception de recherche a également incorporé une approche multi-méthodes complète, incluant la participation de 40 utilisateurs adultes dans un cadre d'environnement domestique naturel. Le concept présenté dans notre recherche est basé sur le paradigme "Les ordinateurs sont des acteurs sociaux" (CASA) et est conçu pour être rentable. Les résultats de cette étude indiquent que l'acte d'auto-révélation par les agents conversationnelles (AC) joue un rôle crucial dans l'établissement de connexions interpersonnelles et la cultivation de la confiance. Les données qualitatives révèlent des stéréotypes de genre implicites envers les AC, mettant l'accent sur la conception centrée sur l'utilisateur et le potentiel de reproduction de biais et stéréotypes des agents conversationnels. Les utilisateurs ont tendance à suivre instinctivement les conventions conversationnelles telles que prendre leur tour pour parler et fournir des réponses de retour. Les découvertes susmentionnées soulignent l'importance de prendre en compte les conventions conversationnelles en fonction du contexte et d'intégrer des sources de connaissances contextuelles pertinentes lors du développement d'assistants vocaux basés sur des enceintes intelligentes. Cette approche facilite la promotion de la pertinence culturelle, de l'empathie utilisateur et de l'engagement, favorisant ainsi des interactions plus significatives.

L'objectif du troisième chapitre est de souligner la nécessité d'améliorer les compétences des agents conversationnels afin de répondre aux besoins des groupes défavorisés, tels que les personnes âgées. Cette étude a utilisé une approche mixte pour enquêter sur le potentiel des assistants vocaux (AV) pour une démographie spécifique, à savoir les personnes âgées entre 65 et 75 ans. Ce groupe d'âge a été fortement impacté par la pandémie de Covid-19 et fait face à l'obstacle supplémentaire de la fracture numérique. L'objectif était d'identifier leurs besoins et défis lorsqu'ils interagissent pour la première fois avec un AV basé sur une enceinte intelligente (Google Home). Pour répondre aux contraintes d'une approche universelle, notre prototype d'assistant vocal offre un AV empathique avec des voix de différents âges adaptées aux origines socio-culturelles des personnes âgées. Ceci a été réalisé grâce à l'utilisation de flux de dialogue spécifiques au contexte et centrés sur l'utilisateur. L'étude a inclus 60 personnes âgées qui ont interagi avec un prototype d'agent conversationnel (AC) basé

sur la parole utilisant l'approche du Magicien d'Oz au sein d'un assistant vocal basé sur une enceinte intelligente (Google Home). Nous avons étudié les effets des caractéristiques vocales de l'AV (voix mature vs jeune) et de la présence d'expression empathique (élevée vs faible) sur les résultats sociaux (soutien et confiance perçus) et fonctionnels (auto-efficacité perçue) envers les assistants vocaux en utilisant le paradigme "Les ordinateurs sont des acteurs sociaux", la théorie de l'attraction par similitude, et la théorie des dimensions culturelles de Hofstede. Les résultats suggèrent que : (1) le style conversationnel de l'assistant vocal est plus important que son apparence pour les personnes âgées, et l'anthropomorphisme se produit même en l'absence d'une incarnation physique, et (2) une forte expression empathique est un puissant motivateur pour percevoir l'AV comme un "confident discret au-delà de l'interaction humaine" basé sur le background et les croyances socio-culturels des personnes âgées. (3) Il n'y avait aucune preuve que l'effet d'attraction par similitude fonctionnait sur les personnes âgées lorsque la similitude était médiatisée, car une voix mature n'avait aucun effet sur le soutien ou la confiance ressentis. De manière intéressante, les personnes âgées ont rejeté l'assistant vocal avec une voix mature. En tenant compte des réflexes générationnels des personnes âgées et d'une conception sensible à la culture, notre étude souligne l'importance d'intégrer l'expression empathique et la conception de la voix dans un assistant vocal "sans corps", centré sur l'utilisateur et basé sur le contexte, pour une interaction optimale avec les personnes âgées.

Dans la quatrième phase de notre étude, nous avons orienté notre attention vers les personnes malvoyantes qui rencontrent souvent des limites pour accéder à la connaissance culturelle, historique et pratique à cause des caractéristiques principalement visuelles de l'environnement physique. La disponibilité limitée d'accès constitue des obstacles à la participation à des activités de loisirs intérieures et extérieures. Si les musées offrent des opportunités d'exploration, d'interaction sociale et d'éducation, l'accessibilité est souvent négligée. De légères modifications peuvent considérablement améliorer l'expérience muséale pour les personnes malvoyantes. Afin de surmonter ces obstacles, nous avons élaboré le modèle iMuse, un cadre de recherche reproductible conçu pour établir des environnements muséaux à la fois accessibles et inclusifs pour les personnes malvoyantes. Le cœur de ce modèle concerne la co-conception d'un prototype d'assistant vocal, mis en œuvre via Google Home, pour la navigation à distance dans le musée, spécifiquement au sein de la Basilique Citerne en Turquie. Le modèle comporte une étude à deux niveaux. Le premier niveau implique la co-conception d'un cadre en quatre phases avec des utilisateurs malvoyants et leurs instructeurs, se concentrant sur leurs besoins et défis spécifiques. La deuxième couche implique de tester cette conception avec 30 autres individus malvoyants, en utilisant plusieurs méthodologies, dont la technique du Magicien d'Oz. Dans notre prototype, nous avons intégré des descriptions audio inclusives et compréhensibles, incorporant des éléments sensoriels, émotionnels, historiques et structurels, ainsi que des sons spatialisés de l'environnement du musée, comme l'eau qui goutte ou des bruits de fond. Cette méthodologie s'est avérée renforcer la compréhension spatiale et le développement de cartes cognitives. Notant la lassitude des blagues formatées des assistants vocaux conventionnels et des interactions utilitaires, nous avons intégré des capacités interpersonnelles dans nos deux versions de l'assistant vocal. Ces versions ont des caractéristiques distinctes, l'une mettant l'accent sur l'humour par le rire, tandis que l'autre se concentre sur des interactions non humoristiques. Les résultats qualitatifs et quantitatifs ont indiqué une préférence pour la version humoristique, qui a augmenté l'interaction, l'immersion et même

l'apprentissage social. Cette étude met en évidence le potentiel du modèle iMuse dans l'utilisation d'assistants vocaux co-conçus, imprégnés d'humour, et culturellement sensibles pour aider les personnes malvoyantes à naviguer dans des espaces inconnus. De plus, elle souligne comment ces assistants vocaux peuvent renforcer l'apprentissage social, l'engagement et l'appréciation du patrimoine culturel au sein de ces environnements.

Globalement, les résultats de cette étude devraient offrir des perspectives précieuses qui peuvent guider la création et l'avancement des agents conversationnels plus captivants, centrés sur l'utilisateur et inclusifs. Cette initiative de recherche a le potentiel d'apporter des avancées théoriques et pratiques significatives, en particulier dans le domaine de la conception d'agents conversationnels plus immersifs et inclusifs. De plus, elle peut améliorer notre compréhension de l'Interaction Homme-Ordinateur (IHO) et étendre l'utilisation des idées des sciences sociales dans le domaine technologique.

SUMMARY

Conversational agents (Cas) have transformative potential that reaches beyond the realms of technology and human-computer interaction (HCI). These technologies can potentially have a significant societal impact, particularly in narrowing digital divides and enhancing social inclusion. Recognising and addressing the various ways in which humans interact with conversational agents, as influenced by cultural, gender, age, and ability-related characteristics, is crucial. These variations shape not only how users access and utilise information but also how they interpret and evaluate it. As such, these agents could risk propagating and reproducing biases and stereotypes. Tailoring communication methods to diverse user populations can help alleviate these concerns, enhancing the user experience, increasing inclusivity and broadening access.

The growing use of personal voice assistants like Google Assistant, Amazon's Alexa, SIRI, Cortana, and Microsoft's Cortana calls for effective design that can hold users' attention for prolonged and natural conversations. Human-computer interaction (HCI) must ensure universal usability. However, new technologies' usability, accessibility and inclusivity remain limited for specific populations. By exploring the under-researched area of voice-assistant user populations, with a particular focus on older adults, adults with specialised interests, and visually impaired people, this study significantly contributes to the fields of human-computer interaction (HCI), social psychology and communication.

At the social level, the development of conversational agents that are customised to meet the individual needs of users has the potential to address disparities in access to information and services. These agents have the potential to significantly contribute to promoting health, wellness, and overall quality of life, especially for vulnerable groups. By facilitating access to information and services, these systems have the potential to effectively mitigate the challenges experienced by these user groups, thus effectively tackling issues related to social inequality. Integrating co-design techniques and user-centred methodologies in developing these agents guarantees their usability and relevance, hence maximising their social impact. The active involvement of end users in the design process is of utmost importance, culturally relevant solutions, non-stereotypical, and inclusive solutions. The ultimate goal is to create humanised computational agents that cater to the diverse needs of the populace, contributing to a more inclusive and accessible digital environment.

The innovative use of the Wizard of Oz (WoZ) technique is a significant methodological contribution. This technique enables social scientists to manipulate various human characteristics without coding, allowing for a more flexible and adaptable approach to comprehending human-computer interactions. By integrating WoZ, this study creates new opportunities for HCI research, particularly for those with non-technical backgrounds. Furthermore, the unique mixed-method approach, combined with pre- and post-visit quizzes, momentary test evaluations, pre- and post-visit interviews, surveys, and quantifiable scales in each chapter, creates a

comprehensive and detailed understanding of user interactions with computational agents. This research contributes to a more nuanced understanding of the dynamics when humans interact with artificial entities by introducing a holistic and inclusive framework for interpreting these interactions.

Our second chapter delves into the influence of gender-related vocal signals and conversational style, focusing on the level of self-disclosure exhibited by conversational agents (CAs). We seek to examine the effects of these factors on attitudinal and behavioural outcomes, including self-disclosure reciprocity and trust. This study used a speech-based conversational agent (CA) prototype integrated into a voice assistant system based on smart speakers, specifically Google Home. We employed the Wizard of Oz technique to simulate a fully functional CA, which involved human operators controlling the system's responses. Our research design also incorporated a complete multi-method approach, which included the participation of 40 adult users in a naturalistic home environment setting. The concept presented in our research is based on the Computers are Social Actors (CASA) Paradigm and is designed to be cost-effective. The findings of this study indicate that the act of self-disclosure by computer agents (CAs) plays a crucial role in establishing interpersonal connections and cultivating trust. Qualitative data reveals implicit gender stereotypes towards CAs, emphasising user-centred design and the potential of reproducing biases and stereotypes of conversational agents. Users tend to instinctively follow conversational conventions such as taking turns in speaking and providing back-channel responses. The aforementioned findings underscore the significance of considering conversational conventions dependent on context and integrating culturally relevant background knowledge sources when developing voice assistants based on smart speakers. This approach facilitates the promotion of cultural relevance, user empathy, and engagement, hence fostering more meaningful interactions.

The objective of the third chapter is to emphasise the necessity of improving the skills of conversational agents in order to cater to disadvantaged groups, such as older adults. This study used a mixed-method approach to investigate the potential of voice assistants (VAs) for a specific demographic, namely older persons between the ages of 65 and 75. This age group has had significant impacts from the Covid-19 pandemic and faces the additional hurdle of the digital divide. The objective was to identify their needs and challenges when interacting with a smart speaker-based VA (Google Home) for the first time. To address the constraints of a one-size-fits-all approach, our voice assistant prototype provides an empathetic VA with varied voice ages tailored to older individuals' socio-cultural backgrounds. This was achieved through the use of context-specific and user-centred dialogue flows. The study included 60 elderly persons who interacted with a prototype speech-based conversational agent (CA) that used the Wizard of Oz approach within a smart-speaker-based voice assistant (Google Home). We investigated the effects of VA's voice characteristics (mature vs young) and the presence of empathic expression (high vs low) on social outcomes (perceived support and trust) and functional outcomes (perceived self-efficacy) towards voice assistants using the Computers are Social Actors paradigm, the Similarity-Attraction Theory, and Hofstede's Cultural Dimensions Theory. The findings suggest that: (1) the voice assistant's conversational style is more important than its appearance for older adults, and anthropomorphism occurs even in the absence of a physical embodiment, and (2) high empathic expression is a powerful motivator for perceiving the VA as a "discreet confidante beyond human interaction" based on older adults' socio-cultural background and beliefs. (3) There was no evidence that the similarity attraction effect

functioned on older individuals when the similarity was mediated, as a mature voice had no effect on felt support or trust. Interestingly, older adults rejected the voice assistant with a mature voice. Considering older adults' generational reflexes and culturally-sensitive design, our study underscores the significance of incorporating empathic expression and voice design into a context-based, user-centred "bodiless" voice assistant for optimal interaction with older adults.

In the fourth phase of our study, we directed our attention towards individuals with visual impairments who often face limitations in accessing cultural, historical, and practical knowledge due to the primarily visual characteristics of the physical environment. The limited availability of access poses obstacles to engagement in indoor and outdoor leisure activities. While museums provide opportunities for exploration, social interaction, and education, accessibility is frequently disregarded. Even slight modifications can significantly enhance the museum experience for those with visual impairments. In order to tackle these obstacles, we have devised the iMuse Model, a reproducible research framework designed to establish museum environments that are both accessible and inclusive for individuals with visual impairments. The central focus of this model revolves around the co-design of a prototype voice assistant, implemented through Google Home, for remote museum navigation, specifically within the Basilica Cistern Museum in Turkey. The model incorporates a two-layer study. The initial layer involves co-designing a four-phase framework with visually impaired users and their sight loss instructors, concentrating on their specific needs and challenges. The second layer involves testing this design with an additional 30 visually impaired individuals, using multiple methodologies, including the Wizard of Oz technique. In our prototype, we integrated inclusive and comprehensible audio descriptions, incorporating sensory, emotional, historical, and structural elements, along with spatialised sounds from the museum environment, such as the dripping water or buzzing noises. The above methodology was discovered to enhance spatial understanding and cognitive map development. Noting the exhaustion with conventional voice assistants' canned jokes and tool-like interactions, we have incorporated interpersonal capabilities into our two iterations of the voice assistant. These versions encompass distinct characteristics, one emphasising humour through laughter, while the other focuses on non-humorous interactions. Both qualitative and quantitative results indicated a preference for the humorous version, which increased interaction, immersion, and even social learning. This study highlights the potential of the iMuse Model in utilising co-designed, humour-infused, and culturally sensitive voice assistants to aid visually impaired individuals in navigating unfamiliar spaces. Additionally, it emphasises how these voice assistants can enhance social learning, engagement, and the appreciation of cultural heritage within these environments.

Overall, the results of this study are anticipated to offer valuable insights that can guide the creation and advancement of conversational agents that are more captivating, user-centric, and inclusive. This research endeavour holds the potential to yield significant theoretical and practical advancements, particularly in the area of designing conversational agents that are more immersive and inclusive. Moreover, it can enhance our comprehension of Human-Computer Interaction (HCI) and expand the utilisation of social science ideas inside the technology domain.

ÖZET

Konuşma ajanları (CA) teknoloji ve insan-bilgisayar etkileşimi (HCI) alanlarının ötesine uzanan dönüştürücü bir potansiyele sahiptir. Bu teknolojilerin, özellikle dijital bölünmeyi daraltma ve sosyal katılımı artırma konularında önemli toplumsal etkileri olabilmektedir. İnsanların konuşma ajanlarıyla etkileşimlerinin kültürel, cinsiyet, yaş ve yetenekle ilişkili özelliklerle nasıl etkilendiğini tanımak ve ele almak hayati önem taşımaktadır. Bu farklılıklar, yalnızca kullanıcıların bilgilere nasıl erişip bu bilgileri nasıl kullandıklarıyla değil, aynı zamanda bu bilgileri nasıl yorumlayıp değerlendirdikleriyle de şekillenmektedir. Bu nedenle, konuşma ajanları, önyargıları ve stereotipleri yayma ve yeniden üretme riski taşıyabilir. İletişim yöntemlerini çeşitli kullanıcı popülasyonlarına uygun hale getirmek bu endişeleri hafifletebilir, kullanıcı deneyimini artırabilir, kapsayıcılığı artırabilir ve erişimi genişletebilir.

Google Asistan, Alexa, SIRI, Cortana gibi kişisel sesli asistanların artan kullanımı, kullanıcıların dikkatini uzun ve doğal konuşmalarda tutabilecek kadar etkili bir tasarım gerektirir. İnsan-bilgisayar etkileşimi (HCI) evrensel kullanılabilirliği garanti etmelidir. Ancak, yeni teknolojilerin kullanılabilirlik, erişilebilirlik ve kapsayıcılığı belirli popülasyonlar için sınırlı kalmaktadır. Bu çalışma, sesli asistan prototipleri üzerinden kullanıcı popülasyonlarının yeterince araştırılmamış bir alanına odaklanarak (özellikle yaşlı yetişkinler, özel ilgi alanlarına sahip yetişkinler ve görme engelli bireyler), insan-bilgisayar etkileşimi (HCI), sosyal psikoloji ve iletişim alanlarına önemli katkıda bulunmaktadır.

Toplumsal düzeyde, kullanıcıların bireysel ihtiyaçlarına göre özelleştirilmiş konuşma ajanlarının geliştirilmesi, bilgi ve hizmetlere erişimdeki eşitsizlikleri ele alma potansiyeline sahiptir. Bu ajanlar, özellikle dezavantajlı gruplar için sağlık ve genel yaşam kalitesini teşvik etme konusunda önemli katkılarda bulunma potansiyeline sahiptir. Bilgi ve hizmetlere erişimi kolaylaştırarak, bu sistemler, bu kullanıcı grupları tarafından yaşanan zorlukları etkili bir şekilde hafifletebilir, böylece sosyal eşitsizlikle ilgili sorunları etkili bir şekilde ele alabilir. Bu ajanların geliştirilmesinde ortak tasarım tekniklerini ve kullanıcı merkezli yöntemlerini entegre etmek, kullanılabilirliklerini ve alakalarını garanti eder, böylece sosyal etkilerini en üst düzeye çıkarır. Tasarım sürecinde son kullanıcıların aktif katılımı kültürel olarak anlamlı, stereotipik olmayan ve kapsayıcı çözümler için son derece önemlidir. Nihai hedef, farklı popülasyonların spesifik ihtiyaçlarına hizmet eden insani sohbet etme yetisine sahip konuşma ajanları yaratmaktır, böylece daha kapsayıcı ve erişilebilir bir dijital ortama katkıda bulunulabilir.

Bu çalışmanın en önemli yöntemsel katkılarından biri Wizard of Oz (WoZ) tekniğinin yenilikçi kullanımıdır. Bu teknik, sosyal bilimcilerin kodlama yapmadan çeşitli insan özellikleri manipüle ederek sesli asistana entegre etmelerine olanak tanır, böylece insan-bilgisayar etkileşimlerini anlamada daha esnek ve uyumlu bir yaklaşım sunar. Bu çalışma, WoZ tekniğinin entegrasyonu ile özellikle teknik geçmişe sahip olmayan araştırmacılar için insan-bilgisayar etkileşimi alanında yeni fırsatlar

yaratmaktadır. Ayrıca, her bölümdeki öncesi ve sonrası ziyaret testleri, anlık test değerlendirmeleri, test öncesi ve sonrası görüşmeler, anketler ve niceliksel ölçeklerle birleştirilen benzersiz karma yöntem yaklaşımı, kullanıcıların konuşma ajanlarıyla etkileşimine dair kapsamlı ve detaylı bir anlayış yaratmaktadır. Bu araştırma, bu etkileşimleri yorumlamak için bütünsel ve kapsayıcı bir çerçeve sunarak, insanların sesli asistanlarla etkileşimi esnasındaki dinamikleri daha nüanslı bir şekilde anlamamıza katkıda bulunmaktadır.

İkinci bölüm, cinsiyetle ilgili vokal sinyallerin ve konuşma tarzının etkisine odaklanarak, konuşma ajanlarının (CA) gösterdiği kendini açma(self-disclosure) seviyelerine odaklanarak, daha insani ve doğal bir sesli asistan prototipi yaratmayı amaçlamaktadır. Bu faktörlerin, mütekabiliyet (reciprocity) ve güven dahil olmak üzere özel ilgi alanlarına sahip yetişkinler üzerindeki tutumsal ve davranışsal sonuçları üzerindeki etkilerini incelemek amaçlanmıştır. Bu çalışmada, özellikle ev tipi bir sesli asistan (Google Home) üzerine entegre edilmiş bir konuşma tabanlı konuşma ajanı (CA) prototipi tasarlanmıştır. Tamamen işlevsel bir sesli asistanı simüle etmek için Wizard of Oz tekniğini kullanılmıştır, bu da sistemin verdiği yanıtları manipüle eden bir araştırmacıyı (wizard) kapsamaktadır. Araştırma tasarımı, doğal bir ev ortamında 40 yetişkin kullanıcıyla test edilmiş ve çok yönlü bir yaklaşımı benimsemiştir. Araştırmamızda sunulan konsept, Bilgisayarlar Sosyal Aktörlerdir (CASA) Paradigması temeline dayanmaktadır. Bu çalışmanın bulguları, sesli asistanların kendini açıklama düzeylerinin(self-disclosure) artmasının, kişilerarası bağlantılar kurmada ve güven oluşturmada kritik bir rol oynadığını göstermektedir. Nitel veriler, CA'lara yönelik içkin cinsiyet stereotiplerini ortaya koymakta, kullanıcı merkezli tasarımın ve konuşma ajanlarının toplumsal önyargı ve stereotipleri yeniden üretme potansiyelinin altını çizmektedir. Kullanıcılar, sesli asistanla etkileşimleri sırasında konuşma sırasını almak (turn-taking) ve arka kanal iletişimi sağlamak (back-channeling) gibi insani konuşma normlarını içgüdüsel olarak takip etme eğilimindedirler. Bahsedilen bulgular, sesli asistanlar geliştirirken, konuşma kurallarını bağlama yönelik olarak dikkate alınması ve kültürel olarak anlamlı arka plan bilgi kaynaklarını (background knowledge) entegre etmenin önemini vurgulamaktadır. Bu yaklaşım, kültürel uygunluğun, kullanıcı empatisinin ve katılımın teşvikini kolaylaştırır ve böylece sesli asistanla daha anlamlı ilişkiler kurmaya elverişli bir ortam hazırlar.

Üçüncü bölümün amacı, yaşlı bireyler gibi dezavantajlı gruplara hizmet edebilmesi için konuşma ajanlarının becerilerinin popülasyonun spesifik ihtiyaçları doğrultusunda geliştirilmesinin gerekliliğini vurgulamaktır. Bu çalışma, 65 ile 75 yaş arası yaşlı bireyler için sesli asistanların (VA) potansiyelini araştırmak üzere karma bir yöntem yaklaşımı kullanmıştır. Bu yaş grubu, Covid-19 pandemisinden önemli ölçüde etkilenmiş olup, dijital bölünme gibi ek bir engelle karşı karşıyadır. Amaç, yaşlı yetişkinlerin ev tipi bir sesli asistana (Google Home) entegre edilmiş bir sesli asistan prototipiyle ilk kez etkileşimde bulunduklarındaki ihtiyaçlarını ve zorluklarını belirlemektir. Sesli asistan prototipimiz var olan sesli asistanların tek tip yaklaşımın kısıtlamalarının altını çizmek için yaşlı bireylerin sosyo-kültürel geçmişine uygun ve değişken ses yaşlarına (olgun ve genç ses) sahip empatik bir sesli asistan (VA) sunmaktadır. Bu prototip bağlama özgü ve kullanıcı merkezli diyalog akışları kullanılarak gerçekleştirilmiş ve diyalog akışları yaşlı yetişkin kullanıcılar tarafından değerlendirilerek valide edilmiştir. Çalışma, diyalog akışlarını manipüle edebildiğimiz Wizard of Oz yaklaşımını kullanan bir prototipe bağlı olup, etkileşimde bulunan 60 yaşlı yetişkin katılımcıyı kapsamaktadır. Bilgisayarlar Sosyal Aktörler paradigması,

Benzerlik-Çekim Teorisi ve Hofstede'nin Kültürel Boyutlar Teorisi kullanılarak sesli asistanın vokal özelliklerinin (olgun ve genç) ve empatik ifadenin varlığının (yüksek ve düşük) toplumsal (algılanan destek ve güven) ve işlevsel sonuçlara (algılanan öz-yeterlilik) etkilerini araştırılmıştır. Bulgular şunları önermektedir: (1) sesli asistanın konuşma tarzı, yaşlı yetişkinler için fiziksel görünümünden daha önemlidir ve antropomorfizma (insanileştirme), fiziksel bir tezahür olmaksızın meydana gelebilmektedir ve (2) yüksek empatik ifadenin varlığı, yaşlı yetişkinlerin sosyo-kültürel geçmişi ve inançlarına dayanarak sesli asistan prototipini "insan etkileşiminin ötesinde sadık bir dost" olarak algılamaları için güçlü bir motivasyon kaynağıdır. (3) Benzerlik çekim etkisinin, benzerliğin vokal özellikler üzerinden (olgun ses) aracılık ettiği zaman yaşlı bireyler üzerinde işlev gördüğüne dair bir kanıt bulunamamıştır, çünkü olgun sesli bir sesli asistanla yaşadıkları deneyimin yaşlı yetişkinler üzerinde hissedilen destek veya güven bazında etkisi yoktur. İlginç bir şekilde, yaşlı yetişkinler olgun sese sahip sesli asistanı reddetmişlerdir. Yaşlı yetişkinlerin jenerasyon bazlı refleksleri ve kültürel duyarlı tasarımı göz önünde bulundurarak, çalışmamız, yaşlı yetişkinlerle optimal etkileşim için bağlam temelli, kullanıcı merkezli "bedensiz" bir sesli asistana empatik ifade ve ses tasarımını entegre etmenin önemini vurgulamaktadır.

Çalışmanın dördüncü aşamasında, fiziksel çevrenin temel olarak görsel özelliklerinden dolayı kültürel, tarihi ve pratik bilgilere erişimde sınırlamalarla karşılaşan görme engelli bireylerin üzerine yoğunlaşmıştır. Sınırlı erişim imkanı, iç ve dış mekanda gerçekleşebilecek boş zaman aktivitelerine katılımı zorlaştırmaktadır. Müzeler, keşif, sosyal etkileşim ve eğitim fırsatları sunmasına rağmen, erişilebilirlik sıkça göz ardı edilmektedir. Küçük değişiklikler bile, görme engelli olanlar için müze deneyimini önemli ölçüde geliştirebilir. Bu engelleri aşmak için, görme engelli bireyler için hem erişilebilir hem de kapsayıcı müze ortamları oluşturmayı amaçlayan tekrar üretilebilir bir araştırma çerçevesi içeren iMuse Modeli'ni tasarlanmıştır. Bu modelin merkezinde, Türkiye'deki Yerebatan Sarnıcı Müzesi'nde uzaktan müze gezintisi için Google Home üzerinden gerçekleştirilen bir sesli asistan prototipinin görme engelli bireylerle birlikte yapılmış kapsayıcı bir ortaklaşa tasarım bulunmaktadır. Model, iki katmanlı bir çalışmayı içermektedir. İlk katmanda, görme engelli kullanıcıların ve görme kaybı eğitmenlerinin belirli ihtiyaçlarına ve zorluklarına odaklanan dört aşamalı bir çerçevenin ortaklaşa tasarlanması yer almaktadır. İkinci katman, bu tasarımı ek 30 görme engelli birey ile test etmeyi ve bunun için Wizard of Oz (WoZ) tekniği de dahil olmak üzere birden çok yöntemi kullanmayı içermektedir. Prototipin diyalog akışlarına, müze ortamında kaydedilmiş damlayan su veya uğultu gibi mekansal seslerle birlikte, mekanın anlatımında duysal, duygusal, tarihi ve yapısal unsurları içeren kapsayıcı ve görme engelliler tarafından anlaşılabilir betimleyici sesli açıklamalar entegre edilmiştir. Yukarıdaki yöntem, mekânsal anlama ve bilişsel harita (cognitive map) gelişimini artırmada etkili bulunmuştur. Katılımcıların var olan sesli asistanların kalıplaşmış şakaları ve insani olarak algılanmayan etkileşimleri hususundaki hassasiyetini not alarak, sesli asistanın iki iterasyonuna kişilerarası yetenekler entegre edilmiştir. Bu versiyonlardan biri mizahi bir şekilde mekanı anlatmaya odaklanırken, diğeri mizahsız etkileşimlere odaklanacak şekilde farklı özelliklere sahiptir. Hem niteliksel hem de niceliksel sonuçlar, görme engelli kullanıcıların mizahi versiyonu tercih ettiklerini, etkileşimi, orada olma hissini (immersion) ve hatta sosyal öğrenmeyi artırdığını göstermiştir. Bu çalışma, iMuse Modeli'nin, tanıdık olmayan alanlarda görme engelli bireylere yardımcı olmak için ortaklaşa tasarlanmış, mizahla süslenmiş ve kültürel açıdan duyarlı sesli asistanları kullanma potansiyelini vurgulamaktadır. Ayrıca, bu sesli

asistanların bu ortamlarda sosyal öğrenmeyi, katılımı ve kültürel mirası değerlendirmeyi nasıl artırabileceğinin altını çizmektedir.

Genel olarak, bu çalışmanın sonuçlarının, daha diyalog bazlı, kullanıcı merkezli ve kapsayıcı olan konuşma ajanlarının oluşturulması ve geliştirilmesi için değerli içgörüler sunması beklenmektedir. Bu araştırma çabası, özellikle daha kapsayıcı ve erişilebilir sesli asistanlar tasarlama alanında önemli teorik ve pratik ilerlemeleri beraberinde getirme potansiyeline sahiptir. Ayrıca, İnsan-Bilgisayar Etkileşimi (HCI) hakkındaki anlayışı derinleştirebilir ve teknoloji alanında sosyal bilim teorilerinin kullanımını genişletme potansiyeline sahiptir.

1. INTRODUCTION

In today's fast-paced and ever-changing world, people from all aspects of life rely on recommendations to enrich many experiences. Whether it is a retiree looking for a quaint coffee shop, a busy professional seeking efficient workout routines, a blind individual navigating a new city, a student trying to organise a memorable trip, or a family preparing a healthy weekly meal, the scope of applicability is vast. Artificial Intelligence (AI) seeks to bridge the gap between humans and machines by designing dialogue systems that facilitate meaningful and context-driven conversations (Kepuska & Bohouta, 2018). Despite considerable progress in recent years, these conversational agents are still in their infancy, with a significant proportion of their potential yet to be realised (Yang et al., 2020).

However, it is not just about advancing AI capabilities; we must also address the user's experience. Today's technology-centric world has widely adopted conversational agents that showcase robust social characteristics, highlighting an urgent issue (Sharma et al., 2023). The main focus for the teams building these agents has traditionally been enhancing their 'intelligence' - their functional ability to interact and respond accurately. However, the limited acceptance of these agents indicates a need to delve deeper into the factors affecting their accessibility and usability. Users often encounter challenges when attempting to incorporate these conversational agents into their daily lives. Among the most significant are the lack of personalisation, limited intelligence of these agents, lack of cultural relevancy, and insufficient provision of emotional support. The role of emotional value should not be underestimated alongside informational support in influencing users' behavioural intentions. Interactions that provide emotional support and anthropomorphise the voice assistant significantly affect user behaviour factors such as perceived playfulness, satisfaction, continued intention of use, empathy, and trust (K. Lee et al., 2020; Papagiannidis & Davlembayeva, 2022; Sestino & D'Angelo, 2023).

Addressing these social characteristics can help conversational agents become effective and meaningful companions. The focus for these system providers should not be limited to designing efficient voice-based interactions. They must also ensure that users derive pleasure from these engagements. The perception of conversational agents as human-like relies heavily on their ability to interact naturally and responsively, foster personal dialogues, and deliver both utilitarian and hedonic benefits (Aw et al., 2022; Rzepka et al., 2022).

Understanding the social attributes crucial for engaging users in meaningful dialogues is a significant challenge for conversational agents like voice assistants, chatbots, virtual assistants, and robots. Most agents prioritise task completion, often overlooking user intentions due to lacking context. Thus, it is necessary for dialogue systems to become more purposeful and personable, enhancing user engagement through dynamic conversations (Zhang et al., 2018). Despite significant strides, current personal digital assistants such as Google Assistant, Apple's Siri, and Amazon's Alexa must break away from their general "one-size-fits-all" approach to truly serve as personalised and autonomous conversational agents (Daniel et al., 2018). To achieve human-like proficiency, these assistants need to provide reliable and efficient information, complete tasks and demonstrate the necessary social competencies for autonomous and engaging interactions (T. Bickmore & Cassell, 2001).



Figure 1.1: Smart speaker-based VAs: Alexa, Cortana and Google Home

With their potential to embody empathetic, humorous, and self-disclosing human interpersonal traits, voice assistants offer a fruitful avenue for enhancing user experience and fostering seamless interaction between humans and machines. Nevertheless, there remains a noticeable gap in understanding how these elements influence diverse user populations, especially those considered vulnerable due to age,

disability, or specific needs (Seaborn et al., 2023; Vieira et al., 2022). This study aims to address this gap, highlighting the effects of customised voice characteristics on these particular user groups' attitudinal and behavioural outcomes. By focusing on different user groups—such as elderly, visually impaired individuals, and adults with particular interests like festival films—we explore the implications of humanised voice assistants for these populations. We aim to unearth new insights into how these populations perceive and interact with such technologies.

Furthermore, this research challenges conventional voice assistants' existing 'one-size-fits-all' methodology. Instead, it underscores the importance of tailoring voice assistants to meet various user demographics' unique needs and communication preferences. By doing so, we aim to create a more inclusive, customised, and effective human-computer interaction environment.

Through the application of co-design methodologies and iterative approaches, our objective is to foster a technology environment characterised by inclusivity and equity. This strategy encourages all users to participate in the design process, enabling active participation and benefiting from technical breakthroughs regardless of age or physical capability.

1.1. Purpose and the Scope of the Study

Assessing the social attributes vital for conversational agents such as chatbots, virtual assistants, and robots is challenging due to their primary task-oriented nature, which often lacks contextual understanding (Zhang et al., 2018). If these agents are to evolve into '*bona fide*' human companions, they must develop attributes like empathy, knowledge, a consistent personality, and the capacity to provide personalised responses (Chaturvedi et al., 2023; J. Li, Galley, Brockett, Gao, et al., 2016; J. Li, Galley, Brockett, Spithourakis, et al., 2016). The design of these agents should go beyond efficient interactions to also create user enjoyment, thereby rendering the agents more human-like (Aw et al., 2022; Rzepka et al., 2022).

Despite the increased adoption of these agents in today's technological landscape, users often encounter challenges when integrating them into their daily

lives, such as lack of personalisation, limited intelligence, and insufficient emotional support. Existing research underscores the importance of emotional support in shaping user behaviour and the anthropomorphic perception of voice assistants, affecting factors like perceived playfulness, satisfaction, and empathy (K. Lee et al., 2020; Papagiannidis & Davlembayeva, 2022; Sestino & D'Angelo, 2023).

Voice assistant design should move away from a one-size-fits-all approach and instead cater to various user groups' specific needs and backgrounds. Notably, older adults and visually impaired individuals, often facing a "digital divide," require customisation for a more positive user experience and improved usability (Chang et al., 2018; Stafford et al., 2014; Fang et al., 2019). These groups can benefit from smart speaker-based voice assistants due to their accessible and user-friendly interfaces. Following the outbreak of the COVID-19 pandemic, the demand for supportive conversational agents has grown exponentially, necessitating designs that alleviate social isolation and anxiety while fostering social connections and well-being (Miner et al., 2020; Yamashita & Huang, 2020).

This research, structured into three distinct sections, evaluates social cues in voice assistants - specifically, vocal characteristics and interpersonal skills - and their impact on attitudinal and behavioural outcomes. The study focuses on three specific user groups: adults with specialised interests, older adults, and visually impaired individuals, to understand better their distinct needs and reactions to these social cues. The study aims to analyse user experiences with more human-like voice assistants, focusing on variables influencing users' attitudes and intentions towards adopting these voice assistants.

With this objective, we raise a set of research questions addressing the social cues of conversational agent systems regarding attitudinal and behavioural outcomes. These questions range from understanding users' perceptions and interactions with voice assistants, examining how reciprocal self-disclosure impacts trust, investigating how voice age and similarity affect user's perception, assessing the influence of gender-based vocal cues on attitudes, to exploring how inclusive audio descriptions and humour-infused voice assistants can affect user experience.

In the second chapter, our research delves into how gender-based vocal cues and the level of self-disclosure in Conversational Agents (CAs) conversational style can impact attitudinal and behavioural outcomes. With this voice assistant, we also aim to scale the reciprocity of self-disclosure and the level of trust users place in these agents. A voice-based CA prototype, designed with a user-centred approach and co-validated techniques, was used in this study. This prototype was integrated with Google Home and tested in the home environment of 40 adult users, employing the Wizard of Oz technique and a multi-method approach.

The unique aspect of this chapter is the development of a system that not only suggests festival movies from the online streaming platform, Blutr but also expresses its "own" opinions. These opinions, derived from the user contributions on "Ekşi Sözlük", a Turkish collaborative hypertext dictionary, aim to mimic the patterns, topics, and user interactions typical to the platform. By utilising the echo effect, our goal is to establish a social environment conducive to self-disclosure, similar to the conversational atmosphere of Ekşi Sözlük.

Further, by harnessing the background knowledge from Ekşi Sözlük, we strive to create a voice assistant that resonates with users' preferences, habits, and cultural nuances. This strategy enhances their propensity to share information and cultivates a sense of familiarity and trust. Thus, our approach imbues our voice assistant prototype with cultural and social dynamics reflective of Ekşi Sözlük, aiming for intuitive and engaging user interactions. This research, grounded in the Computers are Social Actors (CASA) Paradigm, presents a cost-effective model for designing and understanding voice assistant interaction dynamics. It also highlights the importance of considering users' cultural and social context in designing more relatable and compelling conversational agents.

In the third chapter of this study, we turn our focus to the older adult population, specifically those aged between 65 and 75. The Covid-19 pandemic has considerably impacted this group and also grapples with the challenges presented by the digital divide. However, they also hold significant potential in interacting with voice assistants (VAs). Our primary objective was to identify their specific needs and

challenges when they interacted with a Google Home smart speaker-based VA for the first time.

Breaking away from the constraints of a one-size-fits-all approach, we introduced an empathetic VA capable of adjusting voice age to align with the socio-cultural background of older adults. Our team developed and implemented context-specific and user-centred dialogue flows, which the older user groups subsequently validated. This process enabled us to customise the dialogue flows to be perceived as empathetic by the older adults.

The study involved 60 older adults interacting with a voice-based conversational agent (CA) prototype through the Google Home device, utilising the Wizard of Oz technique. Based on theories such as the Computers are Social Actors (CASA) paradigm, Similarity-Attraction Theory, and Hofstede's Cultural Dimensions Theory, we investigated how the voice characteristics of VAs (mature vs young) and levels of empathic expression (high vs low) impacted social outcomes (perceived support and trust) and functional outcomes (perceived self-efficacy) towards voice assistants.

We found potential for VAs to serve as a "trusted confidant" for older adults, which could be a powerful motivator for enhancing their experience and engagement with this technology. This potential, coupled with our distinctive approach to developing and validating user-centred dialogue flows, are vital aspects of our work. Further details and insights from the study's results, including a deep dive into the older adults' interactions with the voice assistant and their perceptions of its role and functionalities, are explored in the subsequent sections of the thesis.

The fourth chapter of this thesis intersects the fields of technology and accessibility, demonstrating our endeavours to enhance inclusivity through co-designing a prototype voice assistant. The goal was to facilitate remote navigation for visually impaired individuals within a particular museum in Turkey, the Basilica Cistern. This part of the study is structured in two stages. Initially, we embarked on an exploratory phase in which we collaborated with visually impaired users and their sight loss instructors to develop a four-phase design framework. This co-design process was driven by the specific needs and challenges faced by these users. In the subsequent stage, we tested our design with an additional 30 visually impaired individuals. We

employed multiple methodologies in this process, including the Wizard of Oz technique, aiming to gather robust data to validate and improve our design.

We drew on several theoretical frameworks to strengthen our understanding and guide the study design. This holistic and inclusive approach included the Computers are Social Actors (CASA) paradigm, Uncertainty Reduction Theory, the Uncanny Valley of Mind (UVM), and Hall's Cultural Context Model. Drawing on these theoretical perspectives, our study sought to understand the utilisation of voice assistants by visually impaired individuals in the context of remote museum navigation while pinpointing areas for potential enhancement. We collected empirical data on the trait attributions of these users towards voice assistants, exploring their social and cultural norms, with a specific emphasis on interpersonal skills such as humour and laughter. Additionally, we investigated the ontological perceptions of users towards voice assistants, questioning whether they view them as tool-like or human-like entities. Lastly, we illustrated the power of co-design methodologies, inclusive audio descriptions, and spatialised sounds in enhancing the remote museum navigation experiences and spatial understanding for visually impaired users.

In response to feedback about the repetitive nature of conventional voice assistants, we took a unique approach by integrating interpersonal skills into our voice assistant. We designed two versions: one incorporated humour and laughter, while the other remained non-humorous. We also addressed this group's frequent lack of cultural, historical, and practical information access.

Following these objectives, our study resulted in creating a prototype voice assistant for virtual museum navigation, designed with a clear purpose: improving social inclusion, learning, and universal usability for visually impaired individuals. By providing an inclusive and enjoyable museum experience, our prototype voice assistant contributes to breaking down barriers for visually impaired users, fostering an environment of increased accessibility and inclusion.

1.2. Interdisciplinary Theoretical Framework: Application of Social Science Theories in Human-Computer Interaction

This doctoral research embarks on an exploration of the potential intersections between social science theories and the realm of human-computer interaction (HCI), establishing a robust interdisciplinary framework. The underlying objective is to harness these theoretical perspectives to enhance our understanding of the dynamics of user interactions with computational agents, ultimately improving the design and usability of such systems.

The theoretical frameworks we employ in this research include the Similarity-Attraction Theory, Hofstede's Cultural Dimensions Theory, Hall's Cultural Context Model, Social Identification Theory, Uncertainty Reduction Theory, the Uncanny Valley concept, and the Computers as Social Actors (CASA) paradigm. Each of these theories, while many of them traditionally associated with human-human interactions, offers unique insights when applied to HCI, opening up new avenues for creating engaging, inclusive, and effective computational agents.

The Similarity-Attraction Theory, for instance, argues that individuals naturally gravitate towards those similar to them. By integrating this theory into the design of artificial entities, we aim to make these systems more 'human-like' and relatable, thereby fostering user engagement and trust.

Hofstede's Cultural Dimensions Theory and Hall's Cultural Context Model provide valuable frameworks for understanding cross-cultural communication. These theories inform the development of culturally sensitive computational agents, aiming to mitigate cultural and social biases and deliver more inclusive experiences to users from diverse cultural backgrounds.

By considering the Social Identification Theory, we acknowledge the influence of social group affiliations on an individual's identity. This insight is crucial in designing computational agents that resonate with users' social identities, promoting acceptance and extended engagement.

Uncertainty Reduction Theory and the Uncanny Valley concept contribute to a nuanced understanding of fostering comfort and trust in users. While the former suggests that predictability in a computational agent's behaviour can enhance user

trust, the latter highlights the need for careful balance when designing human-like agents to avoid creating discomfort.

Finally, the CASA paradigm, a cornerstone in HCI research, underscores the parallels between human-computer and human-human interactions. Our research model builds on this paradigm, adding variables that mediate interactions with voice assistants and cater to different user populations' distinct needs. In doing so, our research model extends the CASA paradigm to more fully consider vulnerable groups and their unique expectations for technology.

The integration of these theories into HCI research has the potential to amplify its social impact significantly. It allows for the development of conversational agents that are more engaging, accessible, and user-friendly, leading to enriched human-computer interaction experiences. Moreover, it presents a fresh perspective on HCI, promising to improve the design and effectiveness of computational agents and, consequently, pave the way for more meaningful, inclusive, and engaging interactions between humans and machines.

Therefore, this thesis aims to demonstrate the utility of these theoretical frameworks in HCI and presents empirical evidence from user studies and the practical application of these theories in designing a voice assistant prototype. In the process, it highlights the significance of interdisciplinary approaches in creating more humane and accessible technologies.

1.3. Methodology and Measurement Tools

While conducting this research, we integrated elements from human-computer interaction and social psychology, applying a mixed-methods approach and developed a unique 3-stage design process for each of our voice assistant prototypes. In the first stage, our methodology was grounded in the user-centred design philosophy, emphasising end-users' involvement in the design process. To this end, we utilised co-design techniques, inviting visually impaired individuals, their sight loss instructors, and older adults to participate in developing our prototype voice assistant. We also conducted surveys, interviews, and focus groups for each voice assistant prototype to

ensure that our dialogue flows were user-centred and engaging, tailored to their needs. Additionally, in the second chapter, we used a hypertext collaborative dictionary based on user contributions as a background knowledge source to enhance a sense of familiarity of the users towards voice assistants' language style. To ensure our design was user-friendly and inclusive, we involved user groups in multiple co-design iterations. This included not only visually impaired individuals but also older adults, expanding the demographic scope of our design process. This collaborative process was instrumental in developing intuitive and accessible dialogue flows in our voice assistants. After creating dialogue flows, we synthesised them using a text-to-speech converter named Voiser and embedded them in simple PowerPoint pages to create cost/time-efficient prototypes without any need to code.

Central to our 3 stage prototype design is the use of the Wizard of Oz (WoZ) method (Dahlbäck, Jönsson, & Ahrenberg, 1993), a technique where participants believe they are interacting with a computer while, in actuality, a hidden human operator, the "wizard," manages the interactions. This approach is cost-effective and highly accessible to social scientists without a coding background, thereby opening up opportunities for a broader range of researchers to contribute to this field. This flexibility provides an invaluable avenue for trialling various interpersonal skills and communication styles, aiding in the design and refinement of our voice assistant. Furthermore, by enabling a real-time assessment of the interaction dynamics between users and the voice assistant, the Wizard of Oz technique facilitates immediate adjustments, enhancing the efficiency and effectiveness of the design process. Overall, the Wizard of Oz technique serves as an essential tool for our research, enabling us to create a more engaging, responsive, and user-friendly conversational agent.

To delve into the dynamics of user interactions, in the latest stage of voice assistant prototyping, we employed pre-test interviews and momentary assessments throughout the test observations, gaining real-time insights into user experiences. Following the task, participants were provided retrospective think-aloud feedback detailing their thoughts and feelings during the interaction. This process allows us to capture users' immediate experiences and reactions to the interactions.

Furthermore, we conducted post-test interviews to gather in-depth insights about participants' behaviours, beliefs, perceptions, and social interactions with voice assistants. The application and sequence of these methods were tailored based on each participant's individual needs and intentions. For the study involving visually impaired individuals, we added pre-visit and post-test quizzes to evaluate user knowledge and perceptions before and after the interactions.

As for construct measurement, we used a range of scales tailored explicitly to each construct and context within the study. This targeted approach ensures that the measurements accurately reflect the intricacies and specificities of each context.

2. LEVERAGING SHARED LANGUAGE IN VOICE ASSISTANTS: ENHANCING RECIPROCITY FOR ADULTS WITH SPECIALIZED INTERESTS

The rise of conversational agents (CAs) in our digital ecosystem highlights the need for designs that both function efficiently and deeply engage users. This chapter delves into how certain vocal and conversational attributes of these agents, such as gender-based vocal cues and varying levels of self-disclosure, can impact users' attitudes, behaviors, and their overall interaction experience. To craft a genuine and shared linguistic experience, our dialogue flows derived from user contributions from Ekşi Sözlük, a prominent Turkish collaborative dictionary. This not only provided a foundation but also guided the implementation of our user-centered dialogue flows into Google Home smart-speaker-based voice assistant system. Through this lens, we explored the dynamics between users and voice assistant prototype, applying diverse methodologies to gauge authentic responses from 40 adult participants with specialized interests.

Central to our exploration are the Computers are Social Actors (CASA) Paradigm and the Wizard of Oz (WoZ) technique, both offering unique perspectives to provide a lens to understand human-agent dynamics in a cost-efficient manner. Our findings suggest the nuanced influence of CA characteristics in fostering trust and user engagement. Implicit gender perceptions emerge, pointing towards potential biases but also avenues for genuine dialogue based interaction. Interestingly, users seem to naturally follow certain conversational norms in human to human interaction when faced with a self-disclosing CA. As this chapter unfolds, we'll delve deeper into these findings, emphasizing the significance of context-aware, culturally-resonant dialogues, and the power of a shared language in orchestrating richer, more profound human-agent interactions.

2.1. Introduction

In today's rapidly evolving world, individuals constantly seek recommendations to enhance their daily experiences, whether it's discovering a delightful restaurant for dinner or planning an unforgettable vacation. To achieve a seamless and authentic interaction between humans and machines, the field of artificial intelligence aspires to develop dialogue systems that facilitate contextual conversations (Kepuska & Bohouta, 2018). These interactive conversational systems, commonly known as dialogue systems, have witnessed remarkable growth in recent years. However, despite significant advancements in natural language processing and social dialogue research, the communication between humans and machines is still in its infancy (Yang et al., 2020). Human language is incredibly intricate, presenting obstacles that personal voice assistants have yet to completely overcome. According to a recent study, users have reported that specific aspects of personal voice assistants, such as their inability to comprehend complex sentences and the occurrence of delays, not their lack of digital skills, have led to technostress and digital divide (Choudrie et al., 2023).

As a result, assessing the social characteristics necessary for conversational agents such as chatbots, virtual assistants, and robots to engage users with a sense of genuine understanding remains a complex challenge. Many of these agents primarily focus on completing tasks, lacking the ability to comprehend user intents due to the absence of contextual analysis. Consequently, dialogue systems must evolve to become more goal-oriented, imbued with distinct personalities and personas to foster engaging conversations with users (S. Zhang et al., 2018). However, current conversational agents still fall short of becoming true human companions, skilled conversationalists, and captivating storytellers (M. Lee et al., 2019). To fulfill these roles, they need to possess essential attributes such as empathy, knowledge, consistent personality, extensive memory capabilities, social companionship and the ability to provide precise and tailored responses (Chaturvedi et al., 2023; J. Li, Galley, Brockett, Gao, et al., 2016; J. Li, Galley, Brockett, Spithourakis, et al., 2016). In the realm of conversational agents, it is crucial for system providers to go beyond the design of efficient speech-based interactions for complex tasks. Equally important is the focus on ensuring users' enjoyment of these interactions. The extent to which users perceive conversational agents as human-like hinges on their ability to communicate naturally

and interactively. This entails being responsive, fostering personal dialogues, and triggering not only utilitarian benefits but also hedonic benefits for users (Aw et al., 2022; Rzepka et al., 2022).

Recognizing the critical importance of social characteristics in conversational agents, their widespread adoption in today's technological landscape becomes a pressing concern. While intelligence remains a key focus in their development, their limited acceptance highlights the necessity to delve deeper into the diffusion and adoption of these agents (Sharma et al., 2023). Users encounter numerous challenges when integrating conversational agents into their daily lives. A significant hurdle lies in the lack of personalization, exceptional intelligence, and emotional support provided by these agents. While informational support positively influences users' behavioral intention, it is essential to acknowledge the significant impact of emotional value. Emotionally supportive interactions and anthropomorphic perception of the voice assistant shape user behavior, encompassing factors such as perceived playfulness, satisfaction, continuance intention, empathy and trust (K. Lee et al., 2020; Papagiannidis & Davlembayeva, 2022; Sestino & D'Angelo, 2023). By embracing and addressing these social characteristics, conversational agents can transcend their current limitations and emerge as highly effective and meaningful companions for users.

In order to truly advance conversational agents towards the goal of being personalized and autonomous, current personal digital assistants such as Google Assistant, Apple's Siri, and Amazon's Alexa need to transcend their generic, one-size-fits-all approach (Daniel et al., 2018). Achieving the same level of proficiency as humans entails not only fulfilling task goals but also delivering reliable and efficient information while possessing the appropriate and customizable social competencies (T. Bickmore & Cassell, 2001) required for autonomous and engaging conversations with users.

It is crucial to recognize that human communication relies not only on the semantic interpretation of words but also on emotions, personality traits, conversation styles, and social cues (Goldman et al., 1980; Hartzler et al., 2014; Lopes et al., 2005). The field of human-robot interaction (HRI) has predominantly fixated on the

embodiment of robots, neglecting the exploration of smart agents devoid of physical forms. Furthermore, a substantial portion of HRI research has centered on the examination of dyadic relationships between robots and humans, predominantly within the confines of controlled laboratory settings. However, it is imperative to transcend these boundaries, delving into the unexplored realms of smart agents and dynamic HRI environments. By broadening our focus, we can unlock new frontiers and gain a comprehensive understanding of the intricacies inherent in human interactions with intelligent non-embodied entities. In light of this, an experiment was designed to investigate the impact of voice-based social cues (male vs. female) and conversational style (high vs. low self-disclosure) exhibited by personal voice assistants during a movie recommendation process in a home environment. This study hypothesized that if a conversational agent possessed a distinct persona and engaged in self-disclosure, users would feel more inclined to reveal their preferences and interests during the conversation. Consequently, the recommendation process would become more efficient and effective. This study stands out as the first to examine the combined effects of self-disclosure and gender as dependent constructs influencing interpersonal outcomes such as trust and reciprocity in an interactional conversation with a smart-speaker based voice assistant (Google Home) in a home environment. The study employed both qualitative and quantitative methods to provide a comprehensive understanding of the phenomenon.

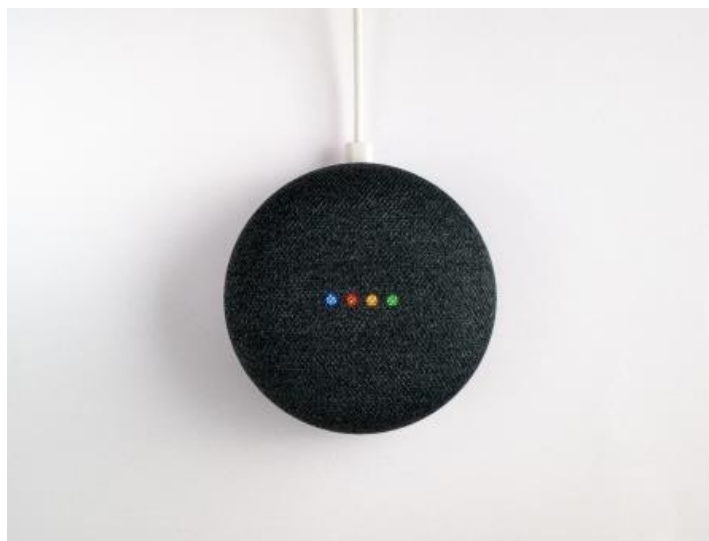


Figure 2.1: Google Home Mini Smart speaker-based voice assistant

2.2. Theoretical Framework

Grounded in the Computers are Social Actors paradigm (CASA), researchers have determined that people tend to assign human personalities such as gender (E. J. Lee et al., 2000a), politeness (Nass et al., 1997), expertise (Nass & Moon, 2000), self-disclosure (Moon, 2000) and ethnicity (Nass et al., 1997) to computers as well. Using CASA as a theoretical framework, it can be hypothesized that people automatically apply social rules, expectations, and scripts known from interpersonal communication in their interaction with conversational agents (Von Der Pütten et al., 2010). Engaging with conversational agents in the role of companions seems to satisfy people's emotional needs. The ability to fulfill these emotional requirements through non-human agents hinges on their ability to possess a lifelike appearance that doesn't result in disappointment (Turkle et al., 2006). When these agents convey a sense of intentionality, individuals tend to attribute them with mental capacities, regardless of their physical form. For instance, voice assistants, even without a physical lifelike presence, can induce a perception of intentionality solely through their voices, which may be comparable in effect. (K. Lee et al., 2020).

People rely on gendered biases and preconceptions and thus tend to assign a particular gender to the computerized agent, especially if they possess humanistic properties or, in other words, social cues such as language, voice, or gender. Nass, Moon, and Green (1997) manipulated the gender-based vocal cues (male or female) of a computer and found that participants tend to attribute gender stereotypes when a conversational agent performs within a gender-specific stereotypical subject domain, such as mechanics, and when the agent's gender does not conform to gender stereotypes. Many studies demonstrated that users' perception of the trustworthiness, expertise, and intelligence of the male computerized agent in the context of a stereotypically male task is greater than that of the female computerized agent due to the present stereotypes that conform to gendered role-based expectations (Gallimore et al., 2019). In another example, researchers performed experimental studies to show that people enjoy the flattery from the agents' characteristics (e.g., the type of computer voice used) (e.g., the type of computer voice used) (Fogg & Nass, 1997).

"Like all technologies before it, artificial intelligence will reflect the values of its creators" (Crawford, 2016). Despite the beneficial aspect of artificial intelligence technology, it has negative implications due to the voice-based conversational agents'

gendered design, according to a recent UNESCO report from 2019 (Feine et al., 2020). The report reveals that most voice-based CAs like Siri, Alexa, Google Assistant, Cortana are designed to be female by default. For example, their voice, their name, the way of advertising often entail female gender associations. This technology's resulting application will perpetuate the manifestation of gender stereotypes (Cooper & Weaver, 2003; Yücel & Rızvanoğlu, 2019). People also apply gender stereotypes towards conversational agents whenever they display specific social cues such as a male or female voice, avatar, or name (Feine et al., 2020; McDonnell & Baxter, 2019).

Following the CASA theory, humans react socially with conversational agents when they display human-like social cues such as language style, name, gender, voice, age, gestures, facial expressions, small talk, social dialogue, or joking (T. W. Bickmore & Cassell, 1999; Moon, 2000; Nass et al., 1997; Nass & Moon, 2000). Social functioning and appearance features of the conversational agents that are more human-like helps foster trust attachment (J. Kim et al., 2022; van Pinxteren et al., 2019). Many studies have also revealed that a computerized agent that uses a deepening self-disclosure strategy during its conversation with the user will lead the user to rate it as more attractive and disclose more personal information (Moon, 2000). If reciprocal social interaction is successful, communication will also be consistent and smooth. Therefore, the users may even experience an increased rapport with the conversational partner (Tartaro & Cassell, 2006). Prior studies demonstrated that conversational interaction could enhance trust and improve users' acceptance of recommendations from a conversational agent. Thus, the interaction design between them is crucial in improving users' willingness to adopt conversational agents' suggestions (T. W. Bickmore & Cassell, 1999; J. Pereira & Díaz, 2019). From this point of view, we created a conversational interaction in a social and low-risk movie recommendation context which triggers users to relate to our conversational agent and enhance co-storytelling and intentionality as suggested by previous studies (K. Lee et al., 2020; M. Lee et al., 2019; Nehaniv, 1999; Urbach, 2011) in users' home environment.

The current body of recent social psychology studies shows us that humans like to reduce their social uncertainty during the interpersonal communication process. They tend to seek ways to understand, predict, and occasionally control other people's behaviour (Gefen & Straub, 2004). Berger and Calabrese's (1975) uncertainty reduction theory posits that humans need interlocutors' information to reduce

uncertainty about their potential behaviour during interpersonal communication. Similarly, when experiencing discomfort during an interaction with a conversational agent, people search for social cues that they can use to predict these agents' potential future behaviours as they do during human-to-human communication (Mourey et al., 2017). The information gained from the other's disclosure, especially when it is positive, may lead to positive beliefs or impressions of the discloser and reduce uncertainty between them. Receiving disclosure from the other is also rewarding since people tend to be attracted to those who disclose them (Collins & Miller, 1994).

Self-disclosure plays an essential role in a wide range of contexts as investigated in prior studies, such as mental well-being (Y. C. Lee, Yamashita, Huang, et al., 2020a; Lucas et al., 2017), personal topics (Moon, 2000), exercise (O'Brien et al., 2018), healthcare (DeVault et al., 2014; Xiao et al., 2019). Self-disclosure reciprocity refers to the process by which one person's self-disclosure elicits another person's self-disclosure and also to whether disclosures are equivalent (e.g., in breadth and depth) (Altman & Taylor, 1973; Jourard, 1971; Sprecher, Treger, Wondra, et al., 2013a). This process has also been called the 'dyadic effect' or the 'mutual effect' (Dindia, 2002). Self-disclosure is an integral part of relationships, especially in the initial stages. People learn about one another via disclosure, and it likely determines whether two people will desire to interact again (Derlega et al., 2008). Therefore, they continue to disclose to facilitate their emotional connection (Sprecher, Treger, & Wondra, 2013). Self-disclosure is crucial in developing and maintaining relationships and relationship satisfaction in human to human interaction (HHI) (Collins & Miller, 1994; Cozby, 1973; Hill & Stull, 1987; Wheelless & Grotz, 1977) and human to computer interaction (HCI) (T. W. Bickmore et al., 2016; Nguyen & Sidorova, 2018; Verhagen et al., 2014). Therefore, prior research has provided evidence for the self-disclosure reciprocity effect: self-disclosure promotes further self-disclosure in HHI and HCI (Collins & Miller, 1994; Jourard, 1971; Sprecher, Treger, Wondra, et al., 2013a). The advancement of computerized technologies has empowered new possible spaces for individuals to self-disclose. The value and importance of self-disclosure through these technologies have been widely explored in prior studies (Barak & Gluck-Ofri, 2007; Burke et al., 2016; Lucas et al., 2017).

According to Social Attraction and Trust Explanation, self-disclosure reduces uncertainty increases trust and liking by establishing a channel between two parties of communication for the reciprocal exchange of information in HHI (Archer, 2016; Collins & Miller, 1994a; Dindia, 2002; Sprecher, Treger, Wondra, et al., 2013a) and also HCI (S. Y. Lee & Choi, 2017; Moon, 2000; Nass & Moon, 2000; Weisband & Kiesler, 1996). On the other hand, Social Exchange Theory argues that people tend to maintain reciprocity and equality in their relationships. Self-disclosure reciprocity is more rewarding than non-reciprocity because people feel uncomfortable with the imbalance in non-reciprocal disclosure in initial interactions. (Archer, 2016). These theoretical arguments are relatively solid for measuring the effects of receiving disclosure: It leads to positive beliefs and impressions of the other, closeness, liking, enhanced familiarity, and the reduction of uncertainty (Collins & Miller, 1994a; Reis et al., 2011; Tamir & Mitchell, 2012). Self-disclosure reciprocity and trust are two critical variables for initiating and developing a relationship with artificial intelligence entities (Gefen & Straub, 2004; Hancock et al., 2011; Heerink et al., 2008; B. Lee, 2006; Moon, 2000). As artificial intelligence entities' conversational abilities quickly upgrade, it is crucial to understand the relational and psychological outcomes of disclosing and reciprocating while building a trustworthy relationship with a conversational agent.

The set of physical cues that voice assistants possess, including their voice, gender, and personality, may align with the user's self-concept and be perceived as belonging to a reference group. The social traits of voice assistants can also carry social meanings that closely align with the user's social identity, fostering a sense of congruence between the user and the assistant. (Alabed et al., 2022). Emotional states and behavioral attitudes possess the ability to spread contagiously among individuals. When individuals use the same language, it generates a social advantage by fostering rapport and affiliation in the minds of listeners. The echo effect theory posits that speakers modulate their speech patterns to either converge or diverge from others, thereby influencing the social distance between them. This process is believed to operate through the generation of shared mental representations facilitated by the use of shared verbal expressions, ultimately promoting rapport and affiliation (Kulesza et al., 2014). Interestingly, this social psychological phenomenon has been observed to persist even when the initiator of the behavior is not a human, but an artificial

conversational agent. In the context of human-agent interactions, a parasocial relationship can develop between users and the artificial conversational agent when the agent adapts its linguistic patterns to align more closely with the communication styles of its users (Aw et al., 2022; Sciuto et al., 2018). To leverage this echo effect, we incorporated a popular Turkish online collaborative dictionary as a background knowledge source for our voice assistant prototype. By utilizing the cultural knowledge, language style, and user interactions prevalent in that specific platform, our voice assistant prototype aims to create a more immersive and relatable user experience. This approach holds the potential to encourage users to engage more deeply, share personal insights or experiences, and establish a stronger connection with the voice assistant.

2.3. Methodology and Research Questions

Is it possible to develop a conversational agent to react reciprocally to users? If so, how would this conversational agent trigger the process of disclosure? For that matter, is it even possible for the conversational agent to participate in a social dialogue containing reciprocal self-disclosure?

In this study, we sought to determine attitudinal and behavioural patterns of users while being exposed to different levels of self-disclosure and different gender of the agent during their interaction with a movie recommendation system via Google Home (Assistant). The following research questions are proposed to examine these goals through a set of constructs mentioned in the following figure, which explains the constructs, relating them to corresponding data collection instruments, and refers to their adapted sources:

R1 How does reciprocal self-disclosure between a user and the voice assistant affect relationship building and developing trust with the voice assistant?

R2 How does the level of self-disclosure of the voice assistant affect the user's tendency to treat it in a human-like way (maintaining conversational norms like turn-taking, back-channelling and politeness)?

R3 How do the gender-based vocal cues affect participants' perception and attitudes toward conversational agents?

Thus, the above literature and research questions lead us to formulate the following four hypotheses with relation to our constructs to discuss our statistical findings:

H1 High self-disclosure of the voice assistant will encourage participants to gain trust and develop reciprocity to the discloser voice assistant.

H2 Participants develop trust in a CA if they participate in self-disclosure reciprocally.

H3 CA's gender will affect the self-disclosure reciprocity and trust of the participants.

H4 Participants' self-disclosure reciprocity positively affects participants' trust.

The following figure is proposed to evaluate our set of constructs mentioned above by explaining our two factors, relating them to corresponding data collection instruments at each process:

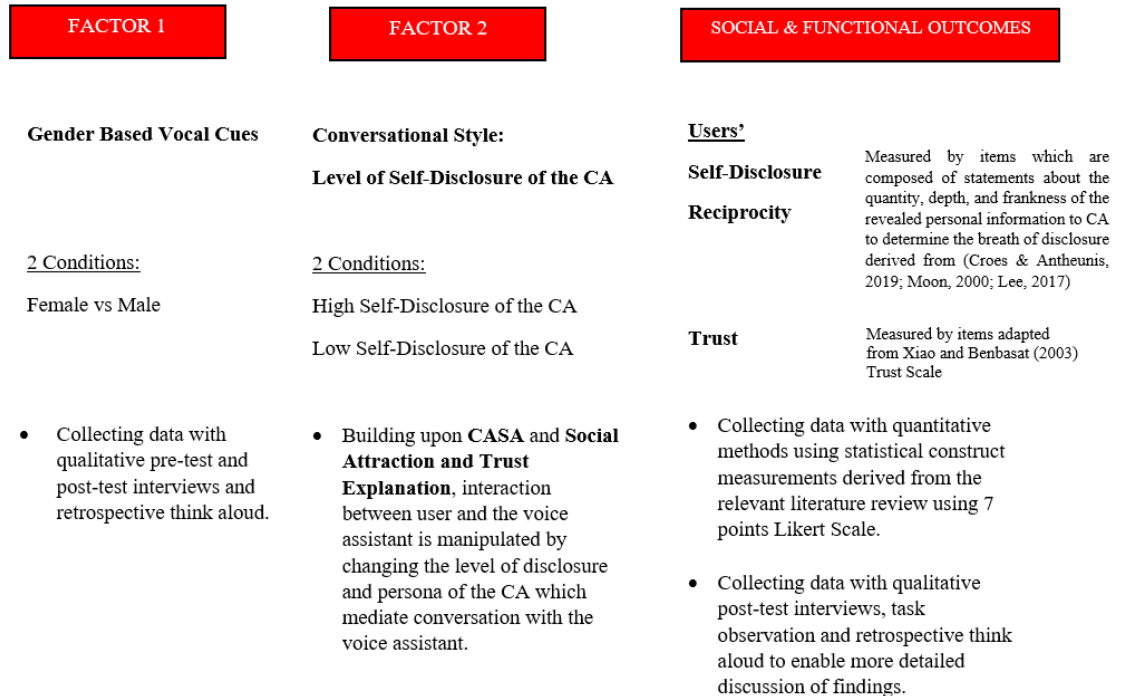


Figure 2.2: Data Collection Process, Instruments and Constructs

This study aims to explore the effects of gender-based vocal cues (female vs male) and different levels of self-disclosure of voice assistants (Google Home) on the

attitudinal and behavioural outcomes (self-disclosure reciprocity and trust) of adult users. This article describes results in which users tend to disclose personal information to voice assistants and explores the relative effects of receiving self-disclosure on trust, reciprocity and other interpersonal impressions. This study is carried with a multi-method approach in which qualitative findings are supported with quantitative statistical data and retrospective think-aloud to enable an in-depth descriptive discussion of the results. This study proposes that if a conversational agent, based on its vocal cues, gain the user's trust by enhancing self-disclosure reciprocity based on the computers are social actors (CASA) paradigm and Social attraction and Trust explanation.

2.3.1. Research Design

We aim to present a conversational movie recommendation system. First, we analyzed the general tendency while choosing a movie to watch by surveying 50 adult users to determine the dialogue flow categories (local and international festival movies, directors, IMDB rating, actors) and how they find a movie to watch in their daily lives. All the participants declared that they had a close and personal relationship with EkşiSözlük¹ in their daily lives. They stated that they specifically read related comments about movies before and after movie watching. Therefore, we have selected EkşiSözlük as background knowledge similarly structured as a conversation dataset in a previous study (Moghe et al., 2020) to build a self-disclosure giver persona to our voice assistant. We specifically focused on users' criteria while choosing a movie and their explanations before recommending it to relatives. Next, we built a conversational agent that gives movie recommendations (only festival movies²) through 4 personas (female x male and high self-disclosure of the CA x low self-disclosure of the CA) that we explained during the following parts of the study. We chose a low-risk (social chat) context like a dialogue about movie recommendations rather than a high-risk (medical

¹ Ekşi Sözlük (Sour Times): a Turkish collaborative hypertext dictionary based on user contribution <https://eksisozluk.com/> It is a popular Turkish online collaborative dictionary and social platform where users can contribute and share their knowledge, experiences, and opinions on various topics. Users can create entries, known as "söz," which can range from personal anecdotes to informative articles or cultural commentary. Other users can comment on these entries, creating discussions and interactions within the community. Ekşi Sözlük has gained significant popularity in Turkey and has become a prominent platform for social commentary and cultural exchange.

² We chose festival films as genre to be able to create a room for users to talk and disclose more since festival film lovers are considered to be more committed and involved cinephiles and avoid extensive background knowledge

purpose) context to ease and deepen the disclosure process and to deliver co-storytelling as shared history with participants during the first-time conversation as found in prior studies (M. Lee et al., 2019; Luria et al., 2019).

Our movie recommendation system, which used social conversational strategies such as the self-disclosure effect in its recommendation process, may potentially trigger users' self-disclosure reciprocity and trust. In the CA's low self-disclosure (LOW SD) condition, the interaction between the user and the voice assistant was not preceded by any CA's personal information. In the CA's high self-disclosure (HIGH SD) condition, the given recommendations were preceded by personal information about the voice assistant movie taste and own preferences related to its persona. On the other hand, we explore how men and women perceive male and female agents and question if gender stereotypes and social categorization processes in society also occur during the interaction with CA.

Our goal is to develop a conversational movie recommendation system that suggests festival movies from the online streaming platform, Blutv. The system will express its "own" opinions, derived from "Ekşi Sözlük," and emulate the conversational patterns, topics, and user interactions found on the platform. By leveraging the echo effect, we aim to create a social environment that encourages users to engage in self-disclosure, resembling the conversational atmosphere of Ekşi Sözlük. This approach incorporates the cultural and social dynamics of Ekşi Sözlük into our voice assistant prototype, providing users with a familiar and engaging interaction similar to the conversations they typically have on the platform.

Research has shown that users often prefer personal voice assistants that act as friends and exhibit similar personalities (Alabed et al., 2022). However, existing personal voice assistants such as Alexa and Google Home, while demonstrating functional intelligence and humor, still struggle with understanding complex dialogue and cultural dynamics, limiting their ability to respond appropriately (Chaturvedi et al., 2023). The impact of anthropomorphism can also vary depending on cultural context, an area that remains relatively unexplored in this field (Diederich et al., 2022). Additionally, factors such as gender, culture, and ethnic markers of the voice assistant pose further questions (Koda & Takeda, 2018; Nag & Yalçın, 2020). By leveraging

the background knowledge of Ekşi Sözlük, our aim is to develop a voice assistant that aligns with users' preferences, habits, and cultural reflexes, enhancing their willingness to share information and fostering a sense of familiarity and trust.

To manipulate our disclosive version, we also developed another version of the system that does not disclose to users and does not have its own opinions. It only transfers objective movie data derived from Wikipedia and the BluTV³ platform itself. To measure the voice assistant's degree of self-disclosure, another round of 50 adult users rated our voice assistant's disclosure level according to the dialogue flow script that we have pre-prepared. Before beginning the ratings, all the 50 users were given definitions of self-disclosure consistent with our theoretical framework. The ratings were made on a five-point scale (1 // low self-disclosure; 5 // high self-disclosure). In the high self-disclosure condition, participants were provided with a self-disclosure statement that preceded each question. Our actual 40 participants conversed with four different conversational agents, which varied in the levels of self-disclosure and gender of CA. (For a dialogue example with an actual user, see appendix A)

2.3.2. Experimental Set-up: Wizard of Oz

The experiment was conducted in a Wizard of Oz setting in which the participants deceptively think they are interacting with an autonomous system. The system's actions were operated by the remote experimenter or "wizard." (Dahlbäck et al., 1993; Large et al., 2019; Medhi Thies et al., 2017) Participants were told they interacted with a conversational agent in this experiment, which automatically responded to their answers. The "wizard" followed a script communicating with participants through a pre-planned script.

The experiment was conducted in a room that resembled a living room. We were required to realize the research offline since it was planned as a first-time user experience using the Wizard of Oz technique. Google Home stands on a desk facing the participant, and a laptop behind the scene, which the wizard controlled for running the prototype and displaying the script. The Google Home and laptop were linked to

³ BluTV is Turkey's first subscription based video-on-demand service. Founded in İstanbul in 2016. It offers a range of award winning international and Turkish content and original programming to its audience.

each other. The remote experimenter, or "wizard," controlled the robot remotely from another room. We chose a less costly Wizard of Oz setup than a coded and implemented AI design to control it more naturally and give the participants a feeling that our CA is a good listener, as suggested in a prior study (DeVault et al., 2014) to contribute to creating a natural-seeming experience.

Voiser platform⁴ was used to generate the agent's synthetic voice. The platform allowed us to control the gender, pitch, and speech rate. The conversation scripts were the same in both conditions (low and high self-disclosure condition), except that sentences were revised to give the agent personal information about itself in the high SD condition. To test our voice assistant prototype, we have uploaded audios to be played to react to a wide range of user utterances. We have made a soundboard on Powerpoint and hyperlinked all the audios as a button for each reaction. A wizard played an appropriate audio button as the voice assistant responded according to the user's response or demand. To convince our participants to believe the audio is being played via Google Home (Assistant), we have connected our laptop to Google Home (Assistant) via Bluetooth. We ensured that our Google Home (Assistant) was visible but muted to operate our prototype from the laptop when running our tests.

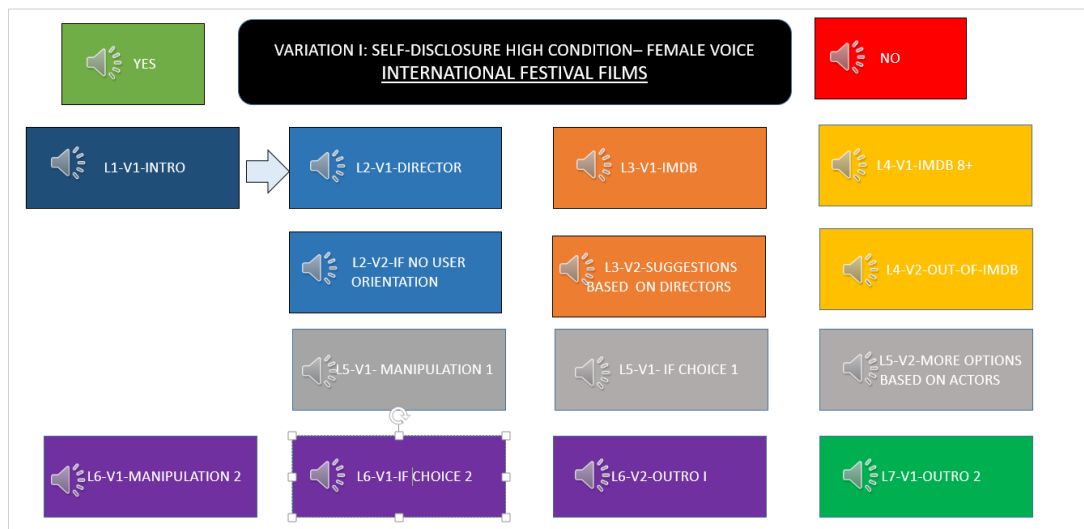


Figure 2.3: One of the wizard's screens while manipulating voice assistant

⁴ A platform which converts texts to voices with humanoid machine sounds <https://voiser.net/>

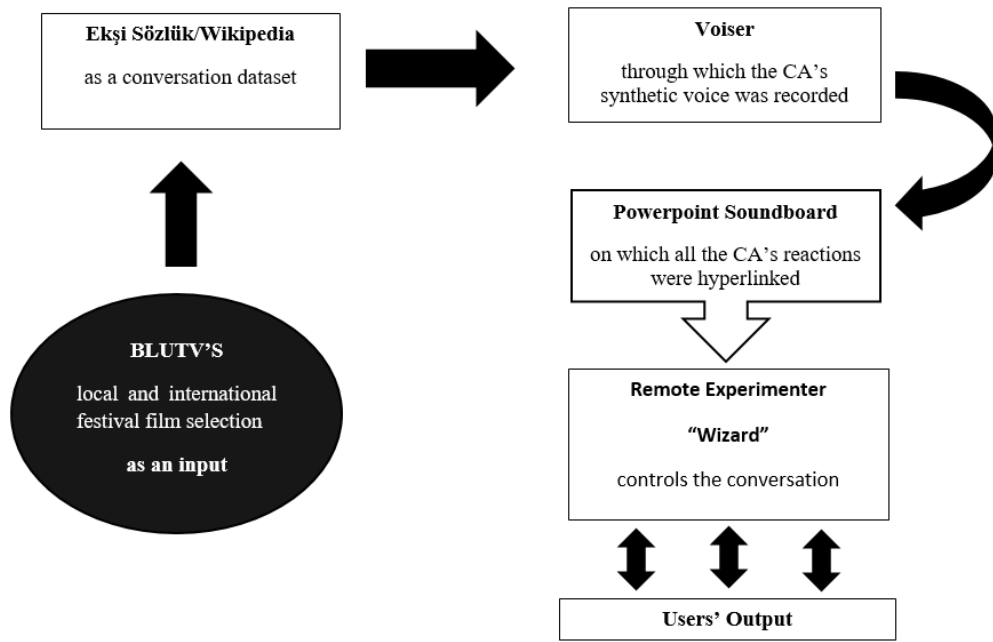


Figure 2.4: Conversational Movie Recommendation System Prototype

2.3.3. Instrumentation

We have measured our constructs by using items with 7 points Likert scale. The measurement items were derived from the literature review and modified to interact with voice assistants. The level of self-disclosure reciprocity was measured with different items adapted from Croes & Antheunis (2019), Youngme & Moon (2000), and Lee (2017), e.g., "I felt comfortable disclosing personal information during the interaction with the voice assistant" (1 = strongly disagree, 7 = strongly agree). Thus, these items were composed of statements about the voice assistant's quantity, depth, and frankness of the revealed personal information (Moon, 2000). Trust items consist of measures of the competence, benevolence, and integrity of an agent and will be measured by a scale developed and validated by Xiao & Benbasat (2003), e.g., "This voice assistant can understand my needs and preferences about festival films" (1 = strongly disagree, 7 = strongly agree). In this study, we aim to observe how self-disclosure and gender dynamics will unfurl in a human-voice assistant interaction with the basis of the above literature.

2.3.4. Participants

This study included 40 participants from Turkey ranging from 25 to 40 (20 females and 20 males). All participants had at least one university degree working as professionals. We based our selection of sampling whereby our participants had to meet three criteria: (1) having an interest in festival films since our voice assistant's dialogue capabilities is limited to Blutv's festival movies selection, (2) being an owner of at least one streaming service based on recommendation such as Netflix, Amazon Prime, etc. (3) having a familiarity, being a follower or user of Ekşi Sözlük Platform. Participants were randomly and equally assigned to the experimental conditions. Genders were also balanced in each experimental state.

2.3.5. Procedure

We experimented with two factors. Our Wizard of Oz experiment was a 2×2 mixed design experiment with the voice assistant's gender (male vs female) and the level of self-disclosure of the voice assistant (high vs low) between-subject factors. Therefore, in total, there were 4 between-subject measurement rounds. During qualitative protocols, we wanted to gain insights for the following: In pre-test sessions, we questioned users about their perception of synthetic voices and voice-based personal assistants (what people think and how it makes them feel) and their movie-watching habits. During post-test sessions conducted retrospectively, we interrogated users about the tone of voice, gender of voice, and the voice assistant's persona. We sought to investigate their overall experience and the likelihood of recommending it to their relatives. The sessions were recorded to observe the amount (breadth) of disclosure (such as time spent talking or the number of statements made) as previously performed by Youngme & Moon (2000).

Each session with each participant included three consequent phases. A pre-test interview was executed to evaluate the participants' previous experience and thoughts towards voice assistants and movie-watching habits, in general, to get more profound and new insights from them. At the beginning of the test phase, the participants were asked to activate the voice assistant by saying, *"Hey Google, please advise me festival films on the Blutv platform."*

All participants were assured that they were always free to stop the conversation whenever they wanted. The research was conducted in the home environment to

imitate a representational movie-watching environment. In the final post-test phase, the participants were interviewed to evaluate their introspective judgments about their experience, and their replies were recorded. The responses referring to our construct set were also modulated on a 7 point Likert scale to provide statistical data. The 7 points scaled statements were referred to measure their self-disclosure reciprocity and trust level. In the after-test data collection stage, open-ended questions were used mainly to get more thorough and new perceptions from the participants.

2.4. Results & Discussion

2.4.1. Pre-test stage Findings

In this part, the findings derived from the pre-test stage were presented. The pre-test stage summarizes the participants' conceptualizations and opinions towards voice assistants and their movie-watching habits. The in-depth interviews in the post-test phase were in line with the results obtained from quantitative data.

The majority of the participants (n=28) expressed a certain degree of distance and alienation towards voice assistants and similar technological artefacts. They declared that they did not have too much voice assistant experience or used it for simple things such as calling someone, checking weather, GPS, making shopping lists, cooking kitchen recipes, and scheduling when their hands are occupied. Most repeated statements were *"voice assistants might be more useful for disadvantaged people," "they need to be more intelligent or human-like to be a must-have for me,"* or *"they are too artificial, unfunny, staged and smarmy."* All the participants indicated that the Turkish version of SIRI or Google Assistant was not as capable as the English version: *"Turkish version is like a bad translation, memorized and staged like old people jokes. (F11)"* Most participants (n=36) reported their high level of frustration while trying a voice assistant since they always need shorter sentences, repetition, slow talk, and adjustment of voice. We concluded that participants mainly had negative feelings and prejudices toward voice assistants.

All the participants declared that they watched movies regularly, at least 2-3 films once a week. They chose their movies based on their friends' suggestions, online

film robots, IMDB ratings, Ekşi Sözlük, and Beyazperde⁵ comments. They stated that they spent too much time choosing a good movie to the extent that they got bored and tried to find one. They all had at least one subscription to the following streaming services: Netflix, Blutr, Mubi, Beinconnect, and Amazon Prime.

This part summarizes our qualitative results derived from the post-test stage regarding the participant's experience (time on task, the breath of disclosure, general attitude, emotions, and experience) with our movie recommender voice assistant in both conditions. We aim to discuss our qualitative findings regarding the manipulation of different levels of self-disclosure around the following research questions in order to improve our evaluation of results by highlighting the strengths of qualitative results and avoiding the limitations of the hypothesis testing based discussion:

R1 How does reciprocal self-disclosure between a user and the voice assistant affect relationship building and developing trust with a voice assistant?

R2 How does the level of self-disclosure of the voice assistant affect the user's tendency to treat it in a human-like way (maintaining conversational norms like turn-taking, back-channelling and politeness)?

2.4.2. Qualitative Findings

2.4.2.1. Self Disclosure-High Effect

2.4.2.1.1. Building Trust and Engagement: The Impact of Reciprocal Self-Disclosure on Trust in Conversational Agents

Consistent with the above literature, our results showed that the conversational agent was trusted and more likely to be accepted when participants disclosed more. Participants tended to positively engage in conversation and share more of their opinions as CA shared its own: *"His opinions about some films that I like made me also share my own opinions"* (F19). The average conversation length was 13 minutes in SD HIGH condition while it dramatically decreased to 3 minutes in SD LOW condition. Participants in SD HIGH conditions, on average, shared 12 statements which included personal information, while they shared only 2 in the SD LOW

⁵ A platform which informs users about current TV and cinema critics, sessions, news, trailers.

condition. Because of the perception of our voice assistant (in SD HIGH condition) as a social actor, participants activated their social categories in mind:

"As she talked more in a sincere way, I was triggered to share more. I liked that she had personal opinions, and I liked how she shared them with me. I liked her language style and the idioms that she used. It made me feel closer and made it less robotic. If we talk a second time, we will talk more." (F17)

When our voice assistant used a conversational strategy of deepening self-disclosure in its conversation, social interaction became reciprocal and seamless. Participants even experienced an increased feeling of rapport with the voice assistant and rated it more attractive as stated by previous research on other computerized agents (Kahn et al., 2004; Moon, 2000; Tartaro & Cassell, 2006):

"I enjoyed talking to her. She was humorous and charming. She made me laugh, and it is the most important thing to have for a woman. She had refined movie taste and personal opinions, and I felt that I could also share my perspective because it was that kind of conversation. Deep and mutually beneficial conversation." (M6)

We further assert that our voice assistant in SD HIGH condition can be conceptualized as a receiver of self-disclosure and trust. Participants were likely to contribute and maintain the conversation with their personal views and ask questions back to the voice assistant. They stated they were *"triggered to share somehow since the voice assistant challenged them with his very personal thoughts."* Another study was conducted in which a social cue such as moral regard was implemented on the interface of the robot's design for fostering participant's engagement during interaction (Short et al., 2010). Similarly, we have added human-like cues (which aim to strengthen the SD level) to our voice assistants, such as having personality and backstory in first person condition, expressing emotions and the notions of self (e.g., referring to itself as I), which fostered participants' social connection to it and effectively induced liking, closeness, engagement, and reciprocal self-disclosure as previously stated by prior research on HHI and HCI (Z. Li & Rau, 2019; Schuetzler et

al., 2018; Sprecher, Treger, Wondra, et al., 2013b). They responded to voice assistant more quickly and more powerfully in social terms:

"Also, I am impressed with his thoughts and distinct personality. That is why I will watch what he suggested to me to confirm or oppose him as soon as possible. I want to talk to him again to discuss my opinion about the suggested movie. It was hilarious to hear that our favourite actor was the same....I felt much closed or found his opinions valuable, and it was a very intriguing experience to converse with a voice assistant with a distinct character.... I was like, discussing movies with a friend. He was a very sophisticated boy, and I liked him. Can we use it again?" (F13).

2.4.2.1.2. Personifying the Conversational Agent: The Role of Human-Like Storytelling

Participants reported greater enjoyment in their interactions with the agent and conversed longer with our voice assistant when it presented its opinions about movies in a first-person condition. This implied that participants were willing to engage in an autobiographical conversation disclosed by an agent and were *"impressed by its seasoned character"* even though it could not possibly be suitable to be experienced during a conversation with an artificial agent:

"I am impressed by his opinions like I would be if I heard them from a friend. They seemed very personal, and they made me talk more. Even though he had a robotic voice, it was a real conversation. It has its own stories. I enjoyed this seasoned VA" (F18)

"When I heard the female voice, I was expecting a more command-based style design. I enjoyed that she had a stance and a moral character. I do not like the maid-wise voice assistant that always flatters you insincerely. She respected me, and we maintained reciprocal conversation. She provided me information and her own opinions about movies, and also I provided her my thoughts in return. She performed very well. I also liked that she had a way of talking and using

specific idioms. She used inverted sentences, making her less staged and less robotic. I would talk to her to get her opinion before a date, for example, to impress the person I dated with her point of view. "(F13)

Participants engaged in a reciprocal conversation with the voice assistant, exchanging information mutually. They also spoke of the voice assistant as a companion or a friend whose opinions were valuable. Our voice assistant provoked a form of a para-social relationship that involved a close friendship:

"I liked her thoughts, she seemed very experienced about movies. She had her own political and sociological observations and thoughts which I liked most about your voice assistant" (M12)

"We talked more than 10 minutes, and I felt like I was talking to a wise man with a sense of humour. I do not see this kind of man around me. It also had a very sophisticated movie taste. Plus, he was kind. Dream man!" (F2)

"I was not thinking to converse with a voice assistant with so much joy. I felt I was conversing with a character, not a robot. I felt that he was sincere and had humanity inside. After this experience, I think that a voice assistant needs to make me feel sincere. Being useful not only enough" (M13)

Participants described the voice assistant's personality as *"human-like," "friendly," "warm," "sophisticated," and "sincere."* Reactions were overwhelmingly positive. Even they put nicknames like *"a little boy with a refined movie taste," "my little fella," "real cinephile," "fellow from Turkey," "wise, cool and naughty lady," and "movie nerd"* as they would do in human to human communication. They attributed a distinct personality to the voice assistant and perceived it as more human-like as it disclosed more. This implied that participants developed a close relationship with the voice assistant to the extent that they called it particular names and assigned it a personality during the post-test interviews.

The more the voice assistant self-disclosed, the higher participants rated its human likeness and competence, although the participants were initially distant and

had prejudices. Thus, they stated that they started to feel relaxed progressively due to the balanced exchange of information (Berg, 1987) since the participants perceived it as a "real," "natural," and "reciprocal" conversation, which got deeper during interaction (F17, M12):

"At the beginning, I was more distant, tried to shorten my sentences, and shouted a bit. Nevertheless, after I realized that he was able to understand me when I talked normally, I got relaxed and talked like I would with another human being, even with a friend." (F18)

This study proposed that humans who have engaged in self-disclosure with a conversational assistant might believe that it can understand them well, learn from them, sense their moods, and their relationship might become closer at their second or third dialogue, which confirmed the related prior literature (Benbasat & Wang, 2005; Ghazizadeh et al., 2012; Linnemann & Jucks, 2018):

"I shared many critics that CA talked me about, and I liked that it advised me political films like Sürü, it means that it started to learn me and understand my style" (M3)

"She seemed like she would sense my mood and lead me accordingly. It would get better and better after we talked longer and multiple times. So, I would use it again and discuss it with her after the movies. It is like having a cinephile friend at home..." (F15)

Our research has revealed that individuals perceive personal voice assistants as intimate companions or even as virtual friends. Particularly, our findings demonstrate that users form parasocial interactions with their voice assistants, primarily driven by the utilitarian and emphasized hedonic advantages they provide. These benefits serve as the main catalyst for users to develop a sense of friendship with their voice assistant. Some impressive and unpredicted results were also unfolded when the participants were interviewed during the post-test stage. Participants (n=12) expressed that they would like to have a voice assistant like that to converse during Covid 19 quarantine due to the safety and health concerns:

"Also, it would be safer to talk about films with that dude during the quarantine. I miss sometimes talking during quarantine" (F12)

"...Her way of talking and her sophisticated character caught me. Due to lockdown, we are all at home, I talk to her day and night" (M7)

2.4.2.1.3. Social Talk and Reciprocity: Politeness, Back-Channelling, and Turn-Taking

Participants in our research manifested conversational approaches and strategies during interaction with our voice assistant. They were polite, laughed, nodded, and controlled their turn-takings during the interaction. During the conversation, they asked back questions to CA and used back-channelling and uh-huh, as they might in human-to-human interaction (HHI). Some participants (n=4) even said sorry when they interrupted CA during the conversation. Even one of the participants (F13) hesitated and apologized when she could not remember a movie's date. At the end of the experience, all the participants thanked and reported their satisfaction to the voice assistant. Most of the participants underlined that *"it was not smarmy and staged like SIRI"* and *"seemed very successfully internalized the Turkish language"*:

"I liked his way of talking and her sincere tone of voice. He had a robotic voice, but it seemed that he internalized the Turkish language more realistically. He spoke to me like a friend, not an assistant, and I liked that considering that movie choosing process needed mutual interaction..." (F19)

In our study, participants were attracted more, engaged in social talk, and felt closer by the voice with social cues that match their personality, language style, and way of talking as predicted by previous studies with other technological entities (T Bickmore & Cassell, 1989; Nass & Lee, 2001). Participants expressed that they enjoyed that *"they shared similar feelings with the voice assistant about some movies"* and *"felt like talking to one of their friends"* and made gestural outputs and facial expressions during interaction with the voice assistant as they would do in HHC:

"He was talking a bit too much, but I conversed with him as I conversed with a wise cinephile, not with a robot. I %100 agreed with him about his opinions. I felt close to him because of his philosophy of life" (M9)

"At the beginning, my responses were minimal, I shared more once I realized I was talking with a real deal. We had a matching language style with her. She was intimate and personal but not too much" (F11)

They stated that they *"felt closed and sincere"* to the voice assistant at the end of the conversation. Some participants even shared more intimate moments about their lives (e.g., first kiss during a movie, first cry during a movie) with the voice assistant. In line with the social attraction and trust explanation in interpersonal communication (Dindia, 2002), participants liked more when receiving a disclosure from a voice assistant and continued reciprocal disclosure (Derlega et al., 2008; Sprecher, Treger, & Wondra, 2013):

"I definitely would prefer this CA choose movies on behalf of me. It understands me clearly. It is like me and faster. What could I expect more from a man? (F2)

"I am flattered to converse with a humorous and sophisticated cinephile. I found her way of talking and her movie intriguing even though I opposed some of her opinions." (M7)

"...Her way of talking and level of intimacy wiped the artificiality out, and I felt that she speaks like one of us. She had its character and talked up close and personal, and I liked that" (M3)

When participants engaged in reciprocal self-disclosure with our voice assistant, they became more willing to watch the movie proposed, suggest it to their relatives. Even they expressed their willingness to talk to the CA longer or again after watching the movie proposed:

"I felt relaxed while talking to this boy. I had chosen a movie to watch long before our conversation ended, but I was curious about the other things that he

might say. So, we have talked for like 15 minutes. It was an intriguing experience." (F13)

Another participant followed up:

"This lady was less robotic when compared to other voice assistants. I think it was because she made me laugh, and I taught automatically that she was wise. So, I felt closed. Even though I am a traditional man and have prejudices towards CAs, I would like to talk to her again and would like to show it to my friends." (M8)

2.4.2.1.4. The Joy of Conversation: Prioritizing Engagement Over Perfection

Participants reported their overall satisfaction and enjoyment with our voice assistant (SD HIGH condition) even when it could not answer their questions with relevant information (since its dialogue flow is limited to Blutv's selection of festival movies). This suggests that the interaction experience is more salient than the interaction output to the participants. Thus, participants may tolerate and even accept the imperfection of their voice assistant as they report greater closeness:

"What is important for me is to share a good and delightful conversation. Otherwise, it would be reading an encyclopedia. He made me laugh, and he had a character. He used familiar idioms and talked in a way that I talk among my close friends. That is more than enough" (F12)

"It can be improved, and it had some pronunciation problems. However, the movie choosing process became joyful with this voice assistant. I felt like I called my BFF and asked him about his last favourite movies. Who cares if he could not pronounce some words" (F1)

2.4.2.2. Self-Disclosure-Low Effect

The qualitative results derived from the post-test stage regarding the participant's experience (time on task, general attitude, emotions, and experience) with our movie recommender in SD LOW condition voice assistant are summarized as follows:

2.4.2.2.1. Participant Experience: Boredom and Distancing with the CA

Our results showed that nearly all the participants (n=18) tended to display boredom and frustration. They expressed feelings of distance, irritation, and even disturbance during and after the test by stating, *"I felt that it was not paying attention to my preferences. It was an overwhelming experience"* (F3), *"it was really like a robot with no feelings. I felt estranged"* (M5), *"I found your voice assistant more robotic than my actual Google Assistant. It lacked human charm, empathy, and sympathy"* (F4). Even though the exact movie selection was used in both conditions, participants in the SD LOW condition found the voice assistant *"not wise"* and *"insufficient."* Thus, they indicated their dissatisfaction with the movie that the voice assistant suggested. The lower the voice assistant self-disclosed, the higher participants rated its competence more insufficient and their experience more boring:

"It was an emotionless conversation. Discussing movies is a very personal topic; it needs to be closer, but she was very distant. Also, she was not very competent and helpful as she needs to be about suggesting festival movies." (F9)

"There is no joy talking to her. I expect a voice assistant to learn from me and tailor itself accordingly, but this one like no energy and enthusiasm to learn me. This one did not have that capacity. She is forced to talk to me." (F16)

Participants described the voice assistant's personality as *"rude," "formal," "smarty-pants," "bossy," "artificial,"* and *"boring."* The less the voice assistant disclosed, the more the participants conceptualized it less human-like. They expected a voice assistant to be sincerer in a personal topic like movies. In one sense, this can be seen as their expectation to perceive the CA as a polite social actor:

"I did not like her distance, and she does not orient me properly. She does not offer me more than any encyclopedia would do. It is like she memorized everything, and there is no soul in it. I felt like I check my bank account" (F8)

"She offered me kind of good movies, but I was the one who tried to maintain the conversation, and I felt awkward. It could be a little bit more human. I was kinder than she was" (F9)

They felt disturbed due to the imbalanced exchange of information, which aligns with the social exchange theory explanation (Berg, 1987). They were not likely to reveal their thoughts, emotions, opinions, or other personal information to the voice assistant since the voice assistant did not disclose any information about itself first at the start of the relationship, which made the participants feel estranged:

"I prefer to watch Netflix's poor quality recommendations than talking to this one. It made me feel like a weirdo trying to warm up the conversation with an assistant. It was like trying to keep the boundaries, and I was the one trying to deepen the conversation. I was extremely polite, but there was no humanness in him. In the end, I quit. He lost me." (M19)

In our study, participants did not bond with the voice assistant since they did not find the social cues for building a trusting relationship with their increased uncertainty, as previously highlighted by a similar study (Large et al., 2019). Some of the participants (n=4) expressed that they *"could not bear to talk to it"* until the voice assistant suggested a movie and quit the experience without thanking or saying anything to the voice assistant: "

There was no real conversation between us, and I had to cut it. I expect a voice assistant to offer me some emotional experience, but there was zero soul in it. It does not have to be %100 human, but you want to see some familiarity or a lifelike performance. It was very distant and authoritarian and made me feel weird. I could not handle more" (M15)

"I tried to end the conversation as soon as possible and never wanted to share anything with it" (M14)

2.4.2.2.2. Participant Perception: Superiority and Bossiness of the CA

Although the same voice pitch and tone were used in both conditions, participants perceived the voice assistant's tone of voice as more superior and bossy by indicating *"she talks like she knows more than me, like I am her assistant, not her"* (M10) and *"I felt like he was talking like he was more superior than me. He was informative but his way of talking irritated me. I was like his assistant"* (M15). Some participants (n=5) even declared that the voice assistant tried to intimate them since they expected a voice assistant to be more flattery. We argued that as the level of self-disclosure decreased, participants perceived the voice assistant as more authoritarian and emotionless, *"like a teacher"* (M14 and F14).:

"I did not want to talk to him and quit the experience; maybe it was because it was male, or maybe it sounded very robotic for me. He had information about festival movies, but I was not too fond of the way of expressing itself. I felt like he was trying to have a higher position over me and intimidate me. I do not know; maybe this idea of me originated from its male voice or language style. I did not want to talk to him one second more. I prefer to talk to another female version of this voice assistant. It can help me improve myself in movies if it would be less bossy. I would not talk again or recommend this version." (M17)

"It was a disconcerting experience from the beginning till the end. There was no soul, and he sounded like my primary school teacher. Why do I have to bother myself that much with a traumatic time travelling? He was maybe informative, but I was not too fond of his way of talking to me. He was like no assistant but a teacher giving cinema lectures. He seems like he does not care about my taste or mood. He did not try to know me. I was expecting something more real—a real conversation. Yes, we talked, but I would never use it or talk to it again. It made me feel nothing positive. Only, I felt estranged progressively" (M14)

2.4.2.3. Gender Effect

We aim to discuss our qualitative findings related to gender manipulation around the following research question:

R3 How do the gender-based vocal cues affect participants' perception and attitudes toward conversational agents?

People attribute personalities to conversational agents, strongly affected by gender-related socially constructed biases and societal expectations (Nass et al., 1997a). They could anticipate the gender of the voice within seconds (Nass & Brave, 2007). By assigning a gender for a conversational agent, the designer also shoulders the burden of stereotypes associated with the specific gender. According to prior literature, behavioural and attitudinal differences have been revealed when people are exposed to gendered voices (Habler et al., 2019; Hwang et al., 2019; Soash, 1999; Q. Yu et al., 2019). Gender stereotypes are verified and reproduced among participants when interacting with technology that employs female or male voices. Computerized systems with female voices are evaluated as more informative when discussing love and relationships. Meanwhile, those with male voices are perceived as more informative when talking about technology (Reeves, Byron, Nass, 1999). Applying a female voice to a voice assistant could be a potential to quickly verify and reproduce a stereotypical gender script because *"women have been constructed as those who serve, help, and nurture others"* (Holbrook, 1991).

2.4.2.3.1. Unveiling Implicit Gender Stereotypes in Conversational Agents

We found that the level of participant's trust and self-disclosure reciprocity towards the CA was not influenced statistically by the gender of the CA ($p > 0.05$). However, we have revealed that stereotypical expectations related to gender were transferred to CA during and after the conversation. Participants declared that they preferred voice assistants that performed an occupation that stereotypically matched the gender of the CA's voice. Thus, the voice assistant's personality is perceived according to participants' gendered stereotypical expectations. Male participants ($n=10$) expected to hear voice assistants with male voice input while talking stereotypically male-related domains and declared that they feel more natural in this way:

"I prefer to talk with a male voice assistant when we speak football, but it does not matter while talking films. I prefer to get help from a man when it comes to men-related domains" (M2)

"Gender does not matter for me; however, I would like to hear familiar voices on some occasions. I feel more relaxed if it is familiar to me, and I count on that voice. For example, when I want to buy some clothes for my little daughter, I prefer approaching a female sales assistant, because I think she would be more capable of advising me on that occasion. However, if I want to buy a car, I prefer a male consultant. I expect robots to be exactly like that in order to be reliable and more human" (M7)

They anticipated hearing a female voice instead of a stereotypically female-related domain (n= 16). Male participants declared that *"they were used to demand help or to be serviced by a female in real life since they are all raised by women (M16)"* and *"helping is a more feminine thing (M11)"*:

"I am glad that I talked to a female. Their voice is softer and more naive. If I lay back and chill, I would expect a female to help me choose a movie, of course. It is because I am a man, or maybe we are all raised by females. When I was lying back at home watching TV, it was my mom serving me fruits. My experience with your assistant reminded me of that, and it was delightful. Where were my apples on the knife? Of course, it is football and any man-related topic; I would enjoy talking to a man (means CA with a male voice input) (M6)."

"I prefer a female voice when talking to a stranger. I bet females prefer male voice, too. It is just more acceptable. Also, I probably like get help from a woman and used to it" (M8)

"If I am willing to buy something, I prefer a female voice since their voice is naive and kinder. If it is something associated with manhood, for example, football, I prefer to hear a male voice because I frequently swear. I cannot use bad language when talking to a woman. On the other hand, if I hear a male voice while buying something, I get angry faster" (M12)

Male participants stated that they found it more appropriate to hear a voice assistant with a female-gendered voice while getting help to see a decent movie since it is not a very "serious topic." At the same time, they view voice assistants with male-gendered voices as somewhat authoritarian and superior characters:

"I enjoyed to talk about movies with a female voice while chilling and enjoying myself. It is just familiarity and making a voice assistant more human. A female voice is softer and seems more appropriate for a voice assistant. It irritates me to hear a male voice and also sounded bossy. It is OK to hear a male voice for more serious things like I am checking my bank account or driving with GPS, but if I chill, I would like to hear soft, naive, not bossy female voice assistant" (M14)

"It would be a more affectionate experience if it would be a female voice when I am home and laying back in my private time. I have to admit that. Maybe my attitude is originated from my childhood habitude. Home is where you hear your mother voice" (M15)

Another participant followed up:

"I expected to hear a female voice, but it was male. It was a little bit disappointing. Since I have to talk with a robot, it can be female. It would make me feel more awkward and more lifelike. I do not want that technology to break all the norms we have gotten used to. It becomes more difficult to adapt myself. If they want to engage us with the technology, they need to keep at least some fundamental rules. Like it makes me feel more proper to get help from a woman. It is tenderer, and I do not know, naive maybe. I was raised like a mama's boy. I do not know the others, but I would engage more" (M19)

2.4.2.3.2. Unintended Perceptions: Male Users' Interpretation of Feminist Attributes in Conversational Agents

Some male participants (n= 10) expressed negative attitudes about our female assistant, claiming that her attitudes are *"too bossy"* for a voice assistant. However, their dialogue flow was the same for the male and the female assistant. Confirming previous research (Gefen & Straub, 1997; Habler et al., 2019; Hwang et al., 2019; Nass et al., 1997b; B. Tay et al., 2014), participants expected female and male voices in different situations, such as female voice in assistive, submissive and pleasing roles which align traditionally gendered social cues and related social construction biases:

"...I would say I finally find a wise female friend but needs to be rasped since she had a highly feminist potential. She talked as a smarty-pants feminist. I did not like that" (M6)

In our study, female participants reacted negatively to the voice assistants with female-gendered voices since nearly all the voice assistants were feminized and offered female voice as a default (or the only) voice setting. They declared that they felt surprised when they heard a male voice (n=10) and expected to be female (n=20):

"Gender does not matter, but I felt flattered to be provided services by a man since it is a rare condition" (F1)

"It bothers me to hear a female voice with voice assistants. Why is that? It is like having a millennial housemaid, and it has to be a woman. Is that because people are used to it, or male participants are used to it?" (F2)

According to female participants, voice assistants are all designed to *"feed men's fantasy world"* (F5) and *"men taste"* (F10):

"It is quite a refreshing feeling to command or manage a male voice assistant. Also, voice assistants are normally designed as females by default. They would like to normalize a male-dominated world. Women are assistants, and men are CEOs." (F4)

"I was expecting it to be female. It is always designed as a female probably to keep feeding male's fantasy. They are always designed to satisfy a male customer." (F5)

However, male participants declared that they preferred female-gendered synthesized voices. They were described as sounding warmer and more sincere than male-gendered synthesized voices, as stated before in prior studies that examined preferences for gender in synthesized voices (Hwang et al., 2019; Mitchell et al., 2011). Similarly, in our study, male participants were expected to create a power dynamic between themselves and the "woman-like" voice assistants, closely associated with the gender hierarchy in the real world. Thus, they reacted negatively when they claimed that our voice assistant with a female voice input "*crossed her line*" (M14) and expected our voice assistant with female voice input to be less feminist and more submissive, pleasing, and sympathetic to build a closer relationship:

"She talks as if she knows more than me like I am her assistant, not her. Female voice assistant is more appealing, I liked her sincerity, but she needs to flatter me more." (M10)

Some female participants (n=5) declared that they would prefer voice assistants without a significant gender marker:

"Voice assistants do not require gender, they can be genderless, and we have that technology." (F19)

"Content and performance are important for me, not the gender. It also does not have to be human, it can be genderless, or I do not know some synthesized voice, demon voice maybe. It would be funny." (F15)

Although there is no significant gender difference between voice assistants and female voice input, we have revealed some remarkable qualitative insights related to gender manipulation. Male participants declared the insignificance of voice assistants' gender at the beginning of the conversation. Their implicit gender biases attributed to the female gender and mainstream acceptance in their minds were revealed during the in-depth interviews. They even accepted or refused the voice assistant's personality, which they assigned relatedly to its gendered voice input. They tended to give traditional gender values they adopted and project them onto conversational assistants

to communicate seamlessly with an object which employed human-like voice input. On the other hand, female participants expressed their frustration towards by-default female conversational assistants. They were more likely to accept the genderless design and a more comprehensive range of identities with options.

2.4.3. Quantitative Findings

Before choosing the appropriate statistical analysis methods to confirm the hypotheses determined within our research scope, we checked whether the data were normally distributed and the homogeneity of the variances. Considering our sample size, it became evident that the data did not meet the assumptions required for conducting parametric tests.. Thus, nonparametric tests were conducted to verify our hypotheses. In nonparametric tests, unlike parametric tests, analyses are performed over the median after all the data are ordered from smallest to largest. The following hypotheses were evaluated using appropriate analyses that fit our sample size:

H1 High self-disclosure of the voice assistant will encourage participants to gain trust and self-disclosure reciprocity of the participants

Since our independent variable (self-disclosure of the voice assistant) consisted of two categories (high & low), a nonparametric test (Mann-Whitney U test) was required to confirm our hypothesis. The results shown in the following table presented the ranks and the test statistics resulting from our analysis:

Table 2.1: Ranks

	Self-disclosure level	N	Mean Rank	Sum of Ranks
Trust	High	20	30,50	610,00
	Low	20	10,50	210,00
	Total	40		

Analyzing the Mean Rank values in Table 2.1., it can be derived that when self-disclosure is *high*, trust increases more than when it is *low*. However, whether this situation is random or significant could be derived from the Test Statistics table:

Table 2.2: Test Statistics

	Trust
Mann-Whitney U	,00
Wilcoxon W	210,00
Z	-5,43
Asymp. Sig. (2-tailed)	,00
Exact Sig. [2*(1-tailed Sig.)]	,00

All p values in the Asymp. Sig. (2-tailed) shown in Table 2.2. was found less than .05. Hence, we concluded that higher self-disclosure of the voice assistant significantly encouraged participants to gain trust in the voice assistant consistent with Hypothesis 1: High self-disclosure of the voice assistant was perceived as being much more trusted ($M = 6,53$, $SD = 0,44$) than low self-disclosure of the voice assistant ($M = 1,57$, $SD = 0,50$). The finding was enormously significant with a mean difference. This result fully supported H1 predicting that participants would ascribe more trustworthiness to the voice assistant, which disclosed more.

H2 Users develop trust in a CA if they participate in self-disclosure reciprocally.

The correlation between our independent variables (self-disclosure and trust) was examined to confirm this hypothesis. Accordingly, we conducted Spearman's rho test to find the correlation between the variables with non-parametric tests:

Table 2.3: Correlations for H2

			Self-disclosure reciprocity of user	Trust
Spearman's rho	Self-disclosure reciprocity of user	Correlation	1,00	,91
		Coefficient		
		Sig. (2-tailed)	.	,00
		N	40	40

Table 2.3. shows that the $r = .91$ value in the Correlation Coefficient row indicated a solid and positive relationship between these two variables. The fact that the p -value in the Sig. (2-tailed) the line was found less than .05, which indicated that this relationship was significant at the .00 level. Consistent with the literature on Social Attraction and Trust Explanation and its application to human-computer interaction,

there was significant crossover interaction between the reciprocal self-disclosure of the participant ($M=3,95$, $SD=2,55$) and trust ($M=4,05$, $SD=2,55$) towards voice assistants. Participants developed more trust when they reciprocally self-disclosed, as stated by our hypothesis 2.

H3 CA's gender will affect users' self-disclosure reciprocity and trust

Since one of the independent variables (CA's gender) consisted of two categories (male & female), the Mann-Whitney U test was conducted to determine whether CA's gender affects users' self-disclosure reciprocity and trust. Ranks and Test Statistics table was presented below to see the frequencies:

Table 2.4: Ranks

	CA's Gender	N	Mean Rank	Sum of Ranks
Self-disclosure reciprocity of user	Male	20	20,25	405,00
	Female	20	20,75	415,00
	Total	40		
Trust	Male	20	18,90	378,00
	Female	20	22,10	442,00
	Total	40		

Upon analyzing the Mean Rank values in Table 2.4., it was observed that the Mean Rank values for the female voice assistant were higher in the variables of self-disclosure reciprocity and trust compared to the male voice assistant. However, whether this difference between the values was significant can be determined by analysing the Test Statistics table :

Table 2.5: Test Statistics

	Self-disclosure reciprocity of user	Trust
Mann-Whitney U	195,00	168,00
Wilcoxon W	405,00	378,00
Z	-,13	-,86
Asymp. Sig. (2-tailed)	,89	,38
Exact Sig. [2*(1-tailed Sig.)]	,90	,39

Analyzing Table 2.5., we found that *p values* for self-disclosure reciprocity of the user were found as .89 and for trust as .38. Since all *p-values* are greater than .05, our statistical analysis results indicated that CA's gender did not significantly affect self-disclosure reciprocity and trust. Our hypothesis 3 was rejected.

H4 Users' self-disclosure reciprocity positively affects users' trust

Spearman's rho testing was conducted to confirm our hypothesis 4 since the relationship between numerical variables was examined.

Table 2.6: Correlations for H4

			Self-disclosure reciprocity of user	Trust
Spearman's rho	Self-disclosure reciprocity of user	Correlation	1,00	,91
		Coefficient		
		Sig. (2-tailed)	.	,00
		N	40	40

Consistent with our hypothesis 4, participants' self-disclosure reciprocity ($M=3,95$, $SD=2,55$) encouraged on participants' trust ($M=4,0$, $SD=2,55$). Analyzing the correlation coefficient row in Table 2.6., we found a strong relationship between self-disclosure reciprocity of the user and trust ($r=.91$, $p=.00$). As a result, we concluded that the link between user self-disclosure reciprocity and trust was robust, positive, and substantial. Participants' self-disclosure reciprocity served as a significant mediator for developing trust towards a voice assistant" to "The results indicate that participants' self-disclosure reciprocity played a significant role in developing trust towards the voice assistant.

2.5. Conclusion

This study examined the effects of self-disclosure and gender on relationship building with a voice assistant during a movie recommendation dialogue. By adopting the CASA paradigm and explaining social attraction and trust, we establish a solid foundation for sustaining a relationship between an artificial entity and a human participant in the context of low-risk social chat (movie recommendation dialogue) with a voice assistant. The study found that self-disclosure is a strong predictor of fostering dialogue and trust between humans and voice assistants. According to our

findings, the high self-disclosure level of the conversational agent positively affected the user's trust and fostered reciprocity. The results showed that our self-disclosing conversational agent prompted participants to share their preferences, interests about movies, and even personal information about their lives, creating a fluid and seamless experience.

This study extends the CASA ("Computers Are Social Actors") paradigm (Nass et al., 1994) by uncovering the social implications of smart speaker-based voice assistants and contribute to HRI literature in four significant ways. In contrast to the primary focus on functional aspects, this research sheds light on the social dynamics and interactions facilitated by voice assistants operating within a lively physical body. Additionally, it advances the field by emphasizing the significance of user-centered design, aiming to create culturally relevant and more natural conversational agents that cater to individual needs and preferences. Thirdly, this study introduces a unique voice assistant prototype that utilizes the Wizard of Oz technique to exhibit human-like characteristics and respond to emotional needs. By offering an accessible and cost-effective solution, this research serves as a stepping stone for future investigations into the intersection of social robotics, artificial intelligence, and the broader field of social sciences. The potential impact of this work extends beyond conventional voice assistants, fostering meaningful human-computer interactions in various domains. Lastly, data collection conducted within home environments adds a contextual dimension, providing valuable insights into the real-world usage and user experiences of voice assistants. This research significantly contributes to the fields of human-computer interaction, social robotics, and artificial intelligence, paving the way for future investigations into human-voice assistant interactions across diverse domains and contexts.

In discussing the insights from our study, we take a meta-perspective. We believe that designers should consider users' perception of context-based conversational norms when creating new forms of conversational interaction with voice assistants. Conversational agents are designed to exhibit a set of social behaviors that create a user experience where the conversational agent needs to have a context-based proper personality capable of establishing a relationship with the user. Designers must determine when agents should employ non-human abilities or design them to be

more human-like. We aimed to gain some perspective on this complex space through an example of low-risk social talk (movie recommendation dialogue). This study aimed to gain insight into the conversational agent's likable behaviors, identify potentially disruptive behaviors, and understand the contextual boundaries that could engage participants in a natural and fluid conversation with a conversational agent through qualitative and quantitative testing.

Should agents be personalized and culturally relevant?

Using an online cultural collaborative dictionary (Ekşi Sözlük) as a background knowledge source for a voice assistant can be valuable in several ways. Ekşi Sözlük is a popular Turkish online collaborative dictionary that contains a vast amount of user-generated content, including discussions, opinions, and cultural references. By incorporating this resource into a voice assistant prototype, the assistant can tap into the cultural knowledge, language style, and user interactions prevalent on the platform. This integration can bring the following benefits:

Cultural Relevance: Ekşi Sözlük is a platform where users share their thoughts, opinions, and experiences related to a wide range of topics. By incorporating similar knowledge into the design process, designers can gain a deeper understanding of the cultural context, trends, and popular opinions within the target user group. This helps designers create products and experiences that resonate with the users on a cultural level.

User Perspectives: Ekşi Sözlük is known for its user-generated content, which reflects the diverse perspectives and voices of its community. Designers can explore similar, culturally specific platforms to gain insights into how users perceive and interact with various products, services, or experiences. Understanding user perspectives can inspire designers to consider different viewpoints and design solutions that address the needs and preferences of a diverse user base and create more reliable conversational agents.

Language and Communication Style: Ekşi Sözlük is rich in colloquial language, slang, and informal expressions. Designers can draw inspiration from this language style to create interfaces, dialogues, and interactions that feel more natural and relatable to users. Incorporating elements of colloquial language can make the design more approachable and engaging, fostering a sense of familiarity and connection with the target audience.

Emotional Insights: Ekşi Sözlük contains a wide range of emotions and sentiments expressed by users. Designers can analyze these emotional insights to understand the impact of certain design elements or experiences on users' emotional responses. This knowledge can inspire designers to create interfaces, visuals, or interactions that evoke specific emotions, such as joy, empathy, or excitement, in order to enhance user engagement and satisfaction.

User-Centered Design: Ekşi Sözlük's content is generated by users, for users. This aligns with the principles of user-centered design, where designers prioritize the needs, expectations, and feedback of the users throughout the design process. By immersing themselves in the discussions and opinions on similar platforms, designers can gain valuable user insights and feedback that can inform and shape their design decisions.

Overall, these popular collaborative dictionaries could serve as a valuable resource for designers, providing cultural relevance, user perspectives, language inspiration, emotional insights, and reinforcing the principles of user-centered design. By tapping into this knowledge source, designers can create more meaningful, engaging, and user-centric design solutions.

Should agents act human-like?

Participants quickly understood human-like behaviours, adapted themselves to the voice assistant's personality, and even shared moments of co-storytelling about movies. Reactions showed that participants felt more comfortable with HIGH SD voice assistants, who created more seamless and efficient experiences. For example, several participants felt comfortable and disclosed more when our HIGH SD voice

assistant presented its opinions about movies in the first-person condition like a human. We were surprised to discover that some HHI behaviours such as controlling turn-taking, back-channelling, and using uh-huh were experienced, hence, naturally by participants during the interaction with a voice assistant even with no face. Our SD HIGH voice assistant triggered a para-social relationship that involved friendship and tolerance even if they made mistakes during the interaction. This puts forward that the interaction experience is more than the interaction output for the participants.

From a macro perspective, due to the Covid-19 pandemic, our findings suggest that people started to consider a more secure way of socializing and conversing and are more likely to use voice-based intelligent personal assistants in their home environment. At this point, rather than employing rule-based voice assistants, designers should create more interactional and storyteller CAs. CA's might be emerging as a new technological genre that started to challenge traditional human-to-human communication. Speech technology needs to have the ability to present a performance beyond getting trapped in the uncanny valley of creepy and staged mimicry (Aylett et al., 2019). For a conversational assistant to be trusted and bonded on the strength of showing human-like characteristics and its harmonious performance, behavioural and interactional aspects of human intelligence and context-based interaction style should be considered more important than a mere human-like appearance.

Should conversational agents be genderless?

This research also revealed the participants' implicit gender expectations and perpetuating stereotypes towards conversational agents with our participants' in-depth interviews. However, all the participants denied it at the beginning of the interviews. These findings show new insights and users' demands for developers, researchers, and designers who should aim to envision and design conversational agents with distinct personalities to trigger feelings of trust, closeness, and attraction and encourage likelihood to adopt them. It is crucial to understand that if users can engage with conversational agents, they perceive it as a social actor and engage with it in a human-like way. Their human stereotypes, prejudices, and expectations are projected onto the conversational agent. Selecting a binary gender for a conversational agent might drag

along the stereotypes attributed to that gender. Even though there was statistical significance related to gender manipulation, this research revealed the implicit gender stereotypes and expectations related to male and female gender with the insights provided by our qualitative data. In our study, women reacted negatively or even protested the voice assistants with female-gendered voices since nearly all the voice assistants were overly feminized and offered female voice as default and surprised to hear a male-gendered voice assistant.

On the contrary, men criticized our discloser female voice assistant for “*having a feminist potential*” or “*being a smart-ass*”, even though the dialogue flow was the same for both genders. This result highlights that men denied but still expected female voice in assistive, submissive, and pleasing roles, perfectly fitting gendered social construction biases. Therefore, this study provided designers and developers with a better understanding of the voice assistant’s performance which requires creative and refined artistry and mimicry (harmony of style, linguistic features, and distinct personality), the significance of inclusivity which should aim not to reproduce gender-based stereotypical prejudices and to consider a human-based voice as a gender-neutral voice option. Similarly, Q, the first genderless voice,⁶ which tries to break down the gender bias in artificial assistants by creating the first genderless voice, fought hard to progress a diverse representation and inclusion in technology. This study also verified that combining quantitative and qualitative techniques and conducting parallel data analysis will produce a more prosperous and fruitful lens for understanding real user stories and revealed societal results that are difficult to anticipate using one singular approach.

Should we use the Wizard of Oz Technique before designing?

Using the method of WoZ, we created an initial space of co-storytelling about movies and the relations bonded with them. Our experimental research model highlighted that the Wizard of Oz technique is a more cost-effective solution that eliminates the need to design and implement coding based on artificial intelligence. Thus, our Wizard of Oz setup allowed us to design rich interactions quickly and avoid possible software errors that could compromise the results. Wizard of Oz technique

⁶ <https://www.genderlessvoice.com/>

made it possible to create fluid experiences that included possible future technologies without fully developing them.

2.6. Limitations and Future Research

This study has the following limitations: Primarily, its sample size limits our study's quantitative results even though a solid statistical significance was found to support our hypotheses fully. Since our research was planned as a first-time user experience and used the Wizard of Oz technique, we were required to realize it offline. Due to the Covid-19 pandemic effects, we were obliged to limit the sample size since we conducted our research in home environments. Although mixed method data analysis described in this study was designed to be broadly applicable, a larger sample would have enhanced the homogeneity of the quantitative results and produced more data for evaluating the effectiveness of this method. Secondly, this study showed that self-disclosure leads to trust and reciprocity. Still, a longer-term study is needed to determine whether these strategies can foster repeated usage, acceptance and other outcomes. Thirdly, our prototype could converse about Blutv's festival selection due to prototyping limitations without coding. Users' reactions may differ or diversify if CA can talk about more different movie genres. As a future research direction, we are currently considering investigating the effect of voice assistants' age-based vocal cues and working with diverse user groups like older adults or disadvantaged people that have never been studied in Turkey.

3. LISTENING TO CULTURAL REFLEXES: TAILORING EMPATHETIC VOICE ASSISTANTS FOR OLDER ADULT USERS

In the continuously evolving realm of Human-Computer Interaction (HCI), the principle of universal usability remains a cornerstone. However, it's evident that the cutting-edge technological advancements often overlook specific populations, leaving them on the periphery. Particularly, older adults, a population segment that has faced heightened challenges during the Covid-19 pandemic and the digital divide, deserve special attention. This chapter delves into the potential of voice assistants (VAs) as a companion for this demographic, focusing on those between the ages of 65 and 75.

In this mixed-method study, we set out with the goal of understanding the unique requirements and challenges individuals face when introduced to a smart speaker-based Voice Assistant (Google Home), for the first time. Recognizing the limitations of generic VAs, we endeavored to craft a more empathetic VA, one that resonates with the socio-cultural nuances of older adults. This involved fine-tuning the VA to feature varied voice ages, while also embedding user-centered dialogue flows that align with and are validated by the lived experiences of the elderly.

Drawing inspiration from foundational theories like the Computers are Social Actors paradigm, the Similarity-Attraction Theory, and Hofstede's Cultural Dimensions, we set up a study with 60 older participants. Key findings illuminated important aspects: the substance of conversation with the VA mattered more to older adults than its visual appearance; an elevated level of empathy in the VA's responses was perceived as a source of comfort, almost as a discreet confidante; yet, intriguingly, older participants leaned away from a mature-sounding voice assistant, challenging our understanding of the similarity-attraction paradigm. As we navigate through this chapter, we will unpack these findings in detail, emphasizing how they shape the landscape of VA design for older adults. There's an emerging need to shift our design paradigm, moving towards creating 'bodiless' voice assistants that are embedded with

rich empathic undertones, finely tuned to the contextual realities of our older generation.

3.1. Introduction

Conversational agents are gaining momentum as a tool for facilitating communication and social interaction among older adults in their daily lives. However, empirical evidence specific to this population remains limited. Researchers have increasingly turned their attention to examining the use of social robots, voice assistants, smart devices in homes, and voice user interfaces (VUIs) across a variety of devices in the context of older adults (Koon et al., 2020; Kowalski et al., 2019; Trajkova & Martin-Hammond, 2020; Zubatiy et al., 2021). Prominent examples of voice assistants include Amazon's Alexa (via the Echo smart speaker) and Google's Home. These widely available consumer devices offer a cost-effective and potentially more natural means of human-computer interaction for older adults, especially when compared to assistive robots (Seaborn et al., 2023).

Facilitating access to technology is essential for supporting vulnerable populations, but it goes beyond mere economic resources. Knowledge barriers and a lack of familiarity with technology, particularly among elderly individuals, pose significant obstacles. Consequently, research endeavors should focus on a more diverse range of vulnerable populations (Vieira et al., 2022). Given the usability potential of smart-speaker-based voice assistants for older adults and their comparatively lower adoption rates, it is imperative to investigate their usage patterns and the perceptions surrounding them among this demographic (Berkowsky et al., 2015; S. Kim & Choudhury, 2021; Vaportzis et al., 2017).

Previous research has consistently shown that older adults generally hold positive attitudes towards intelligent speaker-based voice assistants, favoring them over traditional user interfaces that involve direct manipulation (Blair & Abdullah, 2019; Kowalski et al., 2019). Despite the growing popularity and positive perception of voice assistant technology among older adults, the actual adoption rates and willingness to accept these assistants remain relatively low (Morris, 2013). According to previous studies, older adults tend to be slower to adopt new smart technologies,

confront more difficulty, and be frustrated when using new technologies compared to younger adults (Czaja et al., 2019). The significant influence of technology anxiety indicates that elderly people are less familiar with new technology and thus experience stronger feelings of discomfort towards it (Talukder et al., 2020). However, smart virtual assistants equipped with voice user interfaces hold the potential to gain popularity among elderly individuals, contingent upon their ease of use, high user acceptance, and versatility in facilitating a range of activities, including work, entertainment, and health-related pursuits (Bolaños et al., 2021). We must first understand older adults' perceptions of new technologies to understand if the digital divide and inequality are still pertinent for older adults and make new technologies readily accessible (Delello & McWhorter, 2017) and personalized. To address the challenges faced by older adults, it is crucial to involve them as co-designers and challenge negative stereotypes about aging. Rather than accepting the digital divide as an inevitable reality, it is essential for researchers, practitioners, and policymakers to openly discuss and address the issue (Lythreath et al., 2022).

Researchers need to focus on what technological considerations need to be made when designing for different age groups. Designing voice assistants with a one-size-fits-all approach disregards the unique needs, challenges in adopting technology, communication preferences, and socio-cultural backgrounds of older adults. It is essential to customize voice assistants to cater to these specific requirements, as it plays a critical role in improving usability and creating a positive user experience for older adults (Chang et al., 2018). Additionally, in the design of socially acceptable robots, the consideration of context and context-relevant outcomes is crucial (Stafford et al., 2014).

As voice-based conversational agents such as Amazon Alexa and Google Assistant move into our homes, researchers have studied their perception and use by specific user groups (Chattaraman et al., 2019; Lopatovska et al., 2019a; Lopatovska & Williams, 2018; Pradhan et al., 2019; Purington et al., 2017; Rhee & Choi, 2020). However, we have limited knowledge regarding the compatibility between voice-only, "bodyless" intelligent agents and older individuals, despite the fact that these voice-based agents primarily rely on speech-based interaction in query-style and conversation-like formats (Deng et al., 2019; Miller & Feil-Seifer, 2017; Sayago et al.,

2019). In practical terms, both voice-only and physically embodied forms of technology have their advantages and disadvantages. Developing robots, for example, is a time-consuming process that requires expertise, finesse, and specific resources for production and distribution. While portable, robots necessitate shipping, setup, and user training. On the other hand, voice assistants only require a speaker to function. With the affordability and widespread availability of speaker systems, such as those found in smartphones and laptops, these form factors are comparatively inexpensive and easy to carry. However, these practical considerations must be balanced with factors such as user experience (UX), usability, and appropriateness for the specific use case and user group (Seaborn et al., 2023). Historically, older adults have been identified as a user group impacted by a "digital divide". Although this divide is gradually narrowing as technology becomes more prevalent in everyday activities, it becomes more complex when new forms and modes of interaction emerge, such as smart speaker-based voice assistants (Fang et al., 2019).

Smart-speaker based voice assistants offer a user-friendly and accessible interface, allowing older adults to engage with technology without the barriers posed by complex interfaces or physical limitations. By using voice commands, they can easily interact with devices, access information, and perform tasks, enhancing their independence and overall quality of life. The importance of voice assistants lies in their ability to bridge the digital divide and empower older adults to embrace and benefit from the advancements of technology, ultimately promoting inclusivity and enhancing their well-being. Thus, nowadays, the COVID-19 pandemic has sharpened the challenges of providing healthcare services and social connections. New designs need to propose human-supportive conversational agents to facilitate users' social connection, sense of isolation from the rest of the world, infection fears, frustration, boredom, and well-(Miner et al., 2020; Yamashita & Huang, 2020). The process of domesticating technology involves integrating it into the daily lives of people, including older adults who live independently in their homes. This integration can help to address the challenges faced by older adults, such as social isolation and difficulties with daily tasks. Welfare systems, which utilize robot technology and sensor systems, is also an example of how domestication of technology can provide support and assistance to older adults in their homes, helping them to live more independently (Søraa et al., 2021) and offering personalized interaction experiences (Pollmann et al.,

2023). Emotional interactions with conversational agents highlight the human longing for social connections that can be satisfied by personalized technology (Lopatovska et al., 2019b; Lopatovska & Williams, 2018; Turk, 2016).

The goal of this study is to understand the specific needs and challenges faced by older adults (65+) when using a VA prototype implemented in a smart speaker voice assistant (Google Home). We aim to explore whether an empathic voice assistant can help mitigate the negative effects of the Covid pandemic quarantine, and determine if our conversational agent can improve mood and foster digital companionship after the interaction. Previous research conducted with chatbots has suggested the potential for mood restoration and companionship (de Gennaro et al., 2020; Von Der Pütten et al., 2010). Our study examines the user behavior of older adults when interacting with Google Home, using an integrated theoretical framework. We focus on the variables that influence users' attitudes and intentions towards voice assistants. We believe that social cues such as the voice and conversational style of the assistant will significantly impact older individuals' usage and acceptance of conversational agents. A prior study conducted with students regarding their perception of an artificial intelligence instructor's voice supports this notion (Edwards et al., 2019). Therefore, we hypothesize that an empathic conversational agent can elicit similar effects on people as human-human interactions. We can trace examples of artificial entities providing emotional support and empathy back to ELIZA, a chatbot that simulated a Rogerian therapist in 1966 (B. Liu & Sundar, 2018). Our experiment aims to explore the following factors that may influence older adults' perception of a conversational agent in general:

- (1) Assessing the importance of psychological and socio-cultural factors, biases, and stereotypes for older adults.
- (2) Understanding how older adults react to and perceive conversational agents.
- (3) Identifying implications for interaction design and potential opportunities.

By examining these factors, we hope to gain insights into designing more effective, meaningful and context-based interactions between older adults and conversational agents.

3.2. Theoretical Framework

"Listeners cannot suppress their natural responses to speech, regardless of source. People draw conclusions about technology-based voices and determine appropriate behaviour by applying the same rules and shortcuts they use when interacting with people. These technologies, like the speech of other people, activate all parts of the brain that are associated with social interaction" (Nass & Brave, 2005, p. 4).

People are adapted to social relationships through the human voice, but how do they perceive synthesized robotic voices, specifically older adults? What linguistic content should VAs use when interacting with older adults? What are the key elements of a successful dialogue between voice assistants and older individuals?

Previous research has indicated that the suitability of agent embodiment varies depending on the context of use, the life circumstances of older adults, and the agent's relationship to those circumstances (e.g., medical care, fitness, companion, assistant, home, hospital)(Søraa et al., 2021; Zuckerman et al., 2020). Previous studies suggest that in order to be accepted for older adults, a robot should possess human-like traits without being excessively large or bulky, while still avoiding a strong resemblance to humans that may create unrealistic expectations of human-like behavior (Wu et al., 2012). However, most of the existing work has primarily focused on the physical appearance of agents, particularly those with visible bodies, often in humanoid or zoomorphic forms. Examples include Paro the seal, Aibo the robotic dog, and the Huggable with its teddy bear design (Šabanovic et al., 2013; Stiehl et al., 2006; Tamura et al., 2004). There is also emerging exploration of alternative forms, such as object morphologies like Hugvie the pillow (Yamazaki et al., 2018) and abstract geometric robots (Zuckerman et al., 2020). However, "bodiless" morphologies that leverage other sensory modalities, such as voice and speech, have received limited attention in the context of older adults (Seaborn et al., 2023).

Our current research aims to address the physical, social, and emotional needs of older adult users in their socio-cultural context, considering the cultural reflexes they encounter throughout the aging experience. These factors play a significant role

in their adoption and use of potentially valuable new technologies, particularly in the context of the Covid-19 quarantine. Achieving universal usability necessitates careful attention to social, emotional, and environmental factors of older adults. Conversational agents can assist older adults in maintaining healthy behaviors by offering guidance and advice. Moreover, it is worth noting that older adults tend to be reluctant in openly expressing their genuine emotions (Marson & Powell, 2014). Even when confronted with significant health or financial challenges, they may find it challenging to share their troubles with others, including their close family members or friends (Language et al., 1988). To combat the social isolation experienced by older adults, conversational agents can play a vital role (Noguchi et al., 2023; A. Zhang & Rau, 2022). By engaging in meaningful social dialogues and employing social cues like turn-taking, emotional expression, empathy and familiarity. These agents can establish trust, evoke sympathy, and encourage the use and acceptance of their services (T. W. Bickmore & Picard, 2004; Looije et al., 2010).

3.2.1. Empathy effect

Empathic voice holds immense importance for older adults in embracing personal voice assistants, while acknowledging the significance of the socio-cultural environment they inhabit. Empathy can be described as a vicarious emotional response to the perceived emotions of others and encompasses the ability to emotionally react when observing someone else undergoing or about to undergo an emotion (B. Liu & Sundar, 2018). It involves adopting another person's perspective, comprehending non-verbal cues, being attuned to the emotional state of others, and effectively conveying care and willingness to assist.

According to the Computers are Social Actors Theory (CASA), people prefer conversational agents that show empathy rather than just providing advice (Morris et al., 2018). When interacting with artificial entities, individuals apply social rules based on human-human interaction (HHI) and treat them as social beings (Nass & Moon, 2000). Social cues, such as small talk, self-disclosure, expert jargon, empathy, gossip, and politeness, which are commonly used in human conversations to establish trust, can also be used when interacting with artificial entities to gain the user's trust (Cassell & Bickmore, 2003; Lucas et al., 2018; Sidner et al., 2004). These social cues, including emotional and emphatic voice, can create a sense of social presence in

artificial entities (Seaborn et al., 2023). Previous studies consistently show positive outcomes when agents and robots express empathy towards users (A. Pereira et al., 2011). For example, incorporating empathetic emotions into an agent increases its likability and trustworthiness, making it appear more caring and supportive (Brave et al., 2005). Agents that express empathy have been found to reduce frustration and stress, enhance user satisfaction, comfort, engagement, and task performance (T. Bickmore & Schulman, 2007; Hone, 2006; Klein et al., 2002; Partala & Surakka, 2004; Prendinger et al., 2005).

While previous research has suggested that voice assistants possess human-like attributes, the extent of anthropomorphization in terms of vocal characteristics, and conversation style has not been thoroughly investigated. Consequently, these voice-based assistants exhibit similar traits, occasionally acknowledging user distress in a general manner. However, their ability to empathize, especially with older adults, remains underdeveloped (Morris et al., 2018). The successful implementation and ethical utilization of such technology rely not only on market and economic factors but also on cultural and emotional considerations concerning the target population (Stahl et al., 2014). Thus, this paper emphasizes the significance of user-centered prototyping as a valuable technique for exploring diverse scenarios and taking into account the emotional and contextual responses of older adults. By incorporating empathy in the design of a voice assistant for older adults, based on social identification and similarity attraction theories, we can enhance the user experience. An empathetic voice assistant creates a supportive and compassionate interaction environment, addressing the emotional well-being of older adults. It promotes a sense of trust, comfort, and companionship, contributing to their satisfaction, engagement, and overall positive outcomes.

3.2.2. Voice Maturity Effect:

Remarkably, the field of Human-Robot Interaction (HRI) has largely overlooked the psychological impacts of voice and voice pitch on how users perceive intelligent conversational agents, specifically robots (Niculescu et al., 2013). According to a survey, a minuscule 1% of studies in the field of voice in human-agent interaction (HAI) incorporated older adults (Seaborn et al., 2021). In the realm of social psychology and human interaction, voice and language cues have played a crucial role

in social groups (Nass & Brave, 2007). In face-to-face communication, the voice not only conveys the intended message but also contains significant cues for social interactions. These cues include information about the speaker's gender, age, personality, emotional state, and place of origin, allowing socially intelligent individuals to make judgments about likability, trustworthiness, and potential mates. Pitch, pitch range, volume, and speech rate are fundamental voice characteristics that provide insights into personality traits (Apple et al., 1979). Individuals who speak loudly, rapidly, with a high pitch and wide pitch range are often extroverted, whereas introverted individuals tend to speak slowly with a soft, deep, and monotone voice. These four voice characteristics have been shown to be sufficient for modeling the personality of synthetic voices (Nass & Brave, 2007). Studies by Nass and Brave (2007) demonstrated that test subjects were able to correctly identify the "personality" conveyed by synthetic voices and even used this knowledge to guide their feelings and behavior towards the represented agent.

Voices and conversational style of voice assistance can affect users' preferences and perceptions in ways that have not been widely explored yet, let alone in the Turkish speaking context. Prior studies demonstrated that users are likely to have very subjective preferences for voices within seconds (Mohammadi et al., 2010; Nass & Brave, 2007; Nass & Lee, 2001). Research showed that social machines should exhibit personality traits through vocal cues designed to be easily recognized and accepted as companions or interaction partners (Nass & Brave, 2007). Most importantly, for our research, age group identification is meaningful to people's self-concepts and their rating to voice credibility and trustworthiness (Edwards et al., 2019). Voice pitch is an important factor in determining a person's level of maturity (Montepare & Zebrowitz-McArthur, 1987). In both male and female voices, a lower pitch is associated with higher maturity, while a higher pitch is linked to greater emotional and immaturity traits, suggesting higher levels of emotional instability, weakness, and psychological tension (Sherzer, 1980).

Based on Social Identity Theory (SIT), people have a tendency to shape their self-image within social groups (Tajfel & Turner, 2004). SIT suggests that individuals organize their lives into social groups and then categorize themselves accordingly. Through social identification, a group forms a shared identity, creating an "in-group"

and an "out-group." Extensive empirical evidence supports the notion that people are more attracted to each other when they perceive similarities, even if these similarities are perceived indirectly (Montoya et al., 2008). This phenomenon is known as the similarity-attraction theory (SAT) (Byrne, 1971). Previous studies on similarity-attraction have shown that people are more attracted to others who share similar attitudes (Yeong Tan & Singh, 1995), ethnic backgrounds (Q. Yu et al., 2019), voices (Nass & Brave, 2007) and facial features (Bailenson et al., 2008). Similarity fosters attraction, enhances social identification, and leads to greater liking, trust, and other positive attributes. Additionally, two studies have indicated that individuals tend to identify with a computer voice that matches their own gender own (Eyssel et al., 2012; E. J. Lee et al., 2000). Therefore, age and gender identity can serve as significant sources of social identity. When designing a voice assistant with an older voice, it aligns with the age identity of older adults, facilitating social identification. Furthermore, traditional-age stereotypes often associate older age with knowledge and wisdom in human-human interactions (HHI) (Edwards & Harwood, 2003; Harwood et al., 1995). When the voice assistant speaks with an older voice, it could create a perception of similarity and familiarity for older adults.

3.2.3. Socio-cultural context

Hofstede's research on cultural dimensions explores the concept of high power distance, which pertains to the extent to which individuals within a society accept the unequal distribution of power in institutions. Additionally, he distinguishes between individualism and collectivism, where individualism refers to societies with loose interpersonal ties, while collectivism emphasizes stronger bonds between individuals. These dimensions, power distance and collectivism, significantly impact social behavior, including the display of empathy. In societies with high power distance, hierarchical relationships hold great value, and demonstrating deference, respect, and formality towards superiors is considered the norm. Subordinates or younger individuals are expected to obey authority without questioning (G. Hofstede, 2011). According to the theory, Turkey was found to have relatively high scores in power distance and low scores in individualism (G. Hofstede, 1980). This suggests that conformity, maintaining the existing social order, preserving the collective, and adhering to a hierarchical structure are notable characteristics in Turkish society. As part of a larger cross-cultural study encompassing 62 countries, Turkey ranked higher

than the global average in collectivism (fourth) and power distance (tenth)(Kabasakal & Bodur, 1998). Consequently, the country's prevalent characteristics include conformity, a preference for maintaining social hierarchies, collective orientation, and a significant power distance within interpersonal relationships (Fikret Pasa, 2000). Uncertainty avoidance (UA) investigates how individuals handle situations involving uncertain future events and how they cope with ambiguity (G. Hofstede, 1980). The UA index measures the level of discomfort individuals feel in unclear or unfamiliar situations and the extent to which they develop beliefs and institutions to avoid such situations (G. J. Hofstede et al., 1991). In the case of Turkey, the UA Index score is 85, indicating a preference for avoiding uncertainties. This inclination towards ambiguity avoidance contributes to a lower adoption of technology. Cultures with high UA index scores typically exhibit rigid belief systems, limited acceptance of unconventional behavior, resistance to innovation, and a focus on security (G. Hofstede, 2011). Additionally, Turkey falls into the moderate range in terms of long-term orientation, valuing both long-term goals and enjoying the present. These cultural characteristics influence communication styles, decision-making processes, and attitudes towards technology and innovation in Turkish society. Understanding Turkey's position in these dimensions is crucial for designing effective strategies and solutions that align with the cultural values and preferences of the population.

The perception and reflection of empathy in conversational style can be influenced by various factors, including the interaction environment, user's culture, interaction style, and task type (Niculescu et al., 2013). In the context of designing conversational agents specifically for older adults, such as assistive robots, it is crucial to ensure that these agents seamlessly integrate into the older individuals' ecological framework (Forlizzi, 2004). This entails aligning with their values and accommodating the diverse needs of all system members who will interact with these agents. Adequate consideration must be given to the social and cultural implications as well as the need for multiple and universal functionalities. It is imperative to assess the expectations and requirements of older adults to ensure that the design meets their specific needs (Broadbent et al., 2009).

By incorporating empathetic responses and understanding their unique challenges, the voice assistant can establish a more meaningful and supportive

connection with them. This can contribute to user satisfaction, trust, and overall well-being. On the other hand, Hofstede's theory emphasizes the significance of cultural values and norms, including the acceptance of hierarchical power structures and respect for authority (G. Hofstede, 2011). Creating a voice assistant with a mature voice, similar to the age group of the users, aligns with these cultural values. Older adults, who may have grown up in societies with high power distance, like Turkey, may have a greater appreciation for authority and respect for those who are older or perceived as more experienced and with more wisdom. A voice assistant that reflects their own age group can reinforce a sense of familiarity, respect, and adherence to social norms.

3.3. Methodology

This work aims to take preliminary steps towards implementing a "bodiless" voice assistant with empathic verisimilitude into a smart-speaker based voice assistant, specifically Google Home, in order to provide convincing advice and trustworthy dialogue for older adults. This study sought to explore whether a voice assistant should express emotional support and empathy or provide informational advice-only support about a personal problem, as suggested by prior studies with different artificial entities (de Gennaro et al., 2020; B. Liu & Sundar, 2018; Morris et al., 2018). Although the expression of empathy and emotion is supportive in human-to-human communication, will it be the same when we converse with an artificial entity, or will we reject it due to its artificiality, uncanniness, age-related emotional or social needs, or cultural biases? On the other hand, vocal cues such as pitch and speech rate are salient in judging the personality of voices, triggering the current stereotypes, identifying and developing relationship bonding (Chang et al., 2018; H. Liu et al., 2010).

In this study, we developed two age-related stimuli categories for vocal cues: a "mature voice" (speed: 0.9, pitch: -4) resembling a 60-year-old male and a "young voice" (speed: 1.25, pitch: 6) resembling a 20-year-old male. These vocal cues were adjusted using the Voiser platform to investigate how similar vocal characteristics impact the development of a relationship between older adults and voice assistants, taking into account theories such as Social Identity Theory (SIT), Similarity-Attraction Theory (SAT), and Hofstede's Cultural Dimensions Theory. To maintain focus and

considering that voice assistants are commonly associated with female voices by default, we opted to use only male voices in our study. After generating the voices on the platform, we asked the 60 participants to estimate the age of the voice assistants to ensure accuracy, and all participants agreed on the age range.

This study aims to provide evidence supporting the principles of the similarity attraction theory when aligning participants' age and identity traits with the perceived vocal cues of conversational agents. Through mixed research methods, we intend to explore the preferred voice characteristics of conversational agents and the personality traits that Turkish older adults associate with different voice pitches and rates. One objective of this research is to understand how a specific in-group and out-group social cue, represented by a "mature" and "young" voice, influences older adults' perceptions of conversational agents. It can be hypothesized that older adults may identify more strongly and form a stronger bond with a conversational agent possessing a mature voice, based on the principles of similarity attraction, social identification, and cultural dimensions theory. The other objective is to investigate how older adults' perceptions vary based on the level of emphatic expression displayed by the voice assistant.

Drawing on the CASA paradigm, SAT, SIT, and Cultural Dimensions Theory as foundational frameworks, our study aims to provide an initial understanding of: 1) the impact of emphatic expression and voice maturity in voice assistants on older adults' social identification with the assistant, 2) the extent to which perceived similarity between the voice assistant's emphatic expression, voice maturity, and older adults' own characteristics influences their attraction towards the assistant, 3) how older adults' previous experiences and cultural backgrounds shape their expectations and preferences for a voice assistant with emphatic expression and voice maturity, and 4) older adults' perceptions of social interactions based on their cultural biases, age-related emotional and social needs, and stereotypes.

We set up 3 hypotheses to analyze quantitative part of the study :

H1: Participants will trust more and feel more supported by a VA with a mature voice.

H2: Participants will perceive stronger support and trust with a VA with more empathic expression.

H3: Participants' perceived self-efficacy toward new technologies will increase after the conversation.

3.3.1. Research Procedure and the Experimental Set-up with Wizard of Oz

Our research followed a multi-method approach, consisting of three stages: pre-test, momentary test, and post-test. During the pre-test interviews, our focus was on gaining insights into participants' initial impressions of new technologies, synthetic voices, robots, and their perceived self-efficacy toward these technologies. We also aimed to understand how participants dealt with stressful situations and how age influenced their conception of empathy. Additionally, participants were asked to rate their self-efficacy in using new technologies on a 5-point scale, based on the Tsai and Tsai (2003) internet self-efficacy construct.

For the testing phase, we employed a voice-based conversational agent (CA) prototype using the Wizard of Oz technique. The experiment was conducted in a Wizard of Oz setting, where participants were led to believe they were interacting with an autonomous system. The system's actions were actually controlled by a remote experimenter, known as the "wizard," following a pre-planned script (Dahlbäck et al., 1993; Medhi Thies et al., 2017).

To test our voice assistant prototype, we prepared audio reactions for a wide range of user utterances. We created a soundboard on PowerPoint and linked each audio to a button representing a specific response. The wizard would play the corresponding audio button as the voice assistant responded to the user's input. To create the illusion that the audio was being played via Google Home (Assistant), we connected our laptop to Google Home (Assistant) using Bluetooth. Our Google Home (Assistant) remained visible but muted, allowing us to operate our prototype from the computer during the tests.

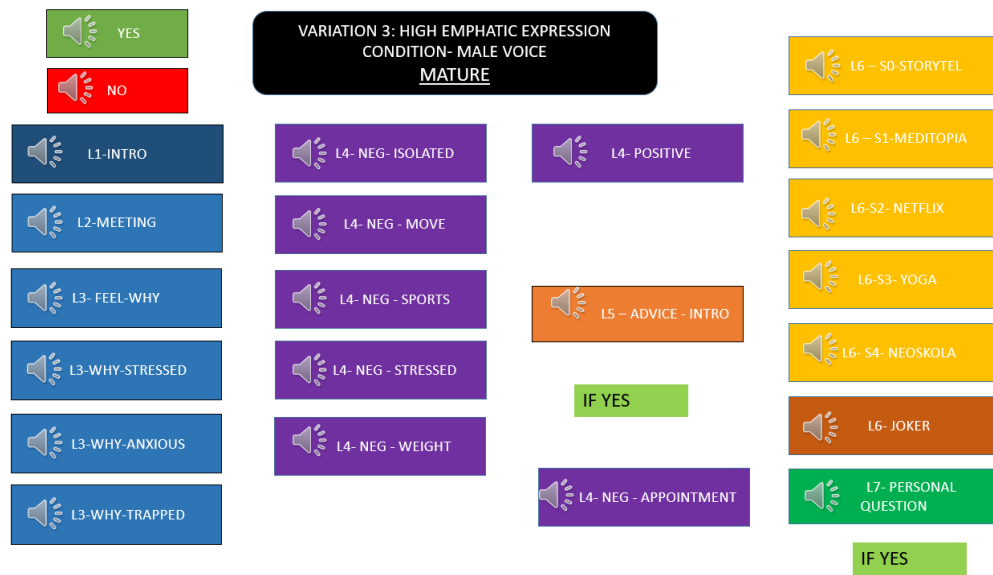


Figure 3.1: One of Wizard of Oz's Controlled Dialog Flow Panel

During the testing phase, we observed participants' gestures and how they positioned themselves during the conversation. We aimed to understand whether they utilized human-like conversational norms, such as nodding or acknowledging responses with "uh-hums."

In the post-test interviews, our focus shifted to exploring the user experience in-depth. We examined usage patterns, identified needs and challenges, explored participants' ontological perception of voice assistants, and assessed their tendency to employ conversational norms while interacting with the voice assistant. Furthermore, we investigated any changes in participants' mood and perceived self-efficacy after the conversation.

3.3.2. Momentary Test Design and the Dialog Flow

Our objective was to develop an empathetic voice assistant that provides advice on Covid-19 quarantine specifically tailored for older adult users. To achieve this, we conducted a two-step process. In order to elicit Turkish seniors' attitudes, needs and cultural habits towards having their own smart-speaker based voice assistant in their home environment, they have to be involved in the design process. Firstly, we surveyed and conducted interviews with 60 older adult users to gain insights into their Covid-

19 experiences and emotions. This information was used to determine the flow of the dialogue for the voice assistant. Secondly, we evaluated the degree of empathetic expression exhibited by the voice assistant by having another group of 60 older adults rate its empathy based on the pre-prepared dialogue script. In terms of verbal empathetic communication, dialog flows were designed in concordance with the socio-cultural background of older adults.

Before proceeding with the ratings, all 60 older adult participants were provided with definitions of empathy and statements exemplifying high and low levels of empathy. The ratings were collected on a five-point scale, ranging from 1 (low empathy) to 5 (high empathy). In conditions where high empathetic expression was desired, empathetic statements were included, such as "I deeply empathize with your feelings of anxiety during the Covid pandemic lockdown. Many people have faced similar challenges, but your age group has been particularly affected," and "I am truly impressed by your resilience. Not everyone could handle this stressful quarantine period with such strength." Conversely, in low empathetic conditions, the statements focused primarily on providing formal advice, such as "Feelings of anxiety, confinement, and numbness have been reported due to prolonged isolation. Try to remain calm," and "If you are feeling stressed and anxious during the lockdown, you can try using a meditation application on your own."

By implementing user-centered design approaches and actively involving older adults in the process, we aim to identify their specific needs and expectations effectively. This approach acknowledges the importance of considering older adult cultural values and the socio-cultural context when designing a voice assistant.

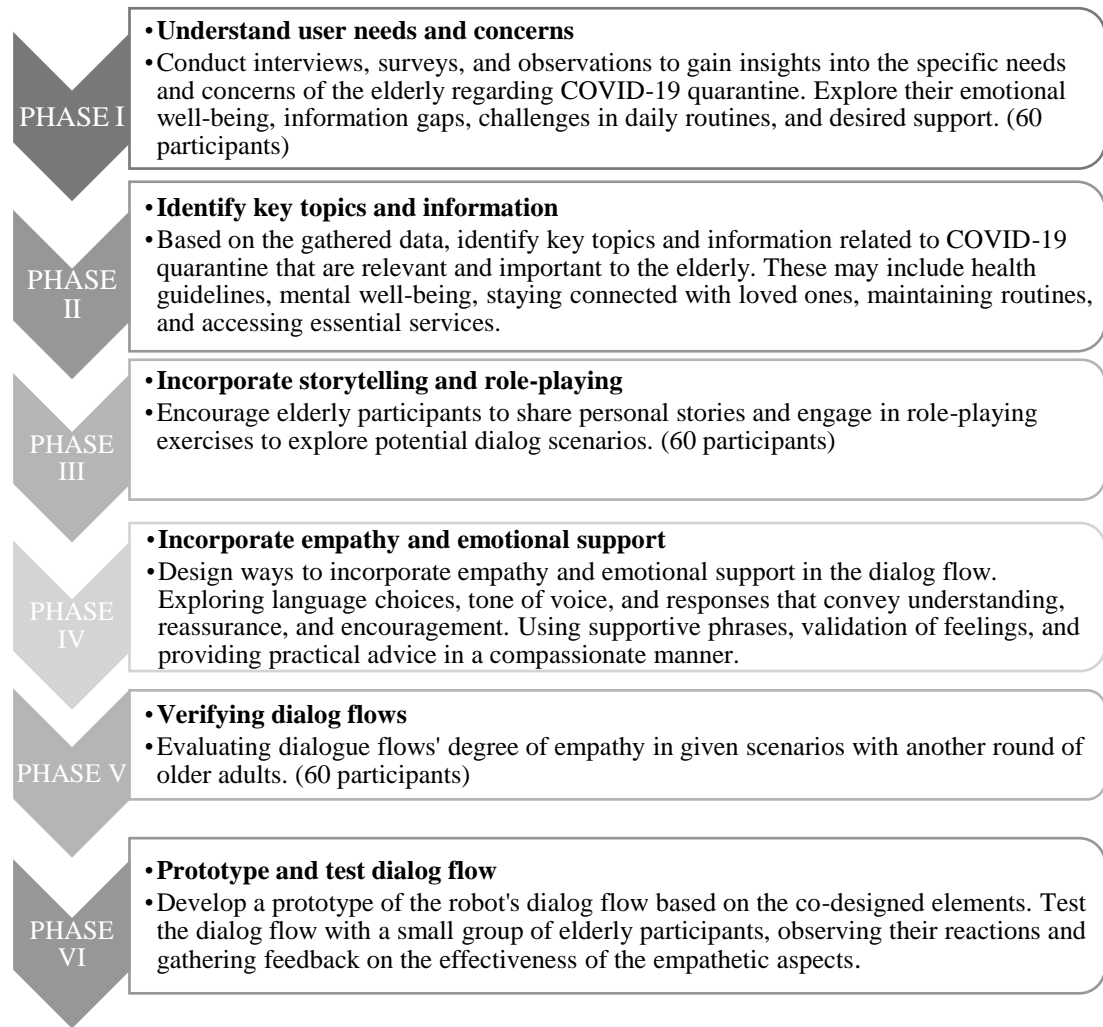


Figure 3.2: Our user-centered dialogue-flow design process

We designed and conducted a study where we used a 2×2 between-subjects factorial design; the factors were voice characteristics (mature vs young) and the presence of empathic expression (high empathic expression vs low empathic expression) on the *social* (perceived support and trust) and *functional* (perceived self-efficacy toward voice assistant) *outcomes* of older adult users using verified scales.

Table 3.1: Sources for Construct Items

Construct	Adapted from	Items
Perceived support	(van der Zwaan et al., 2012)	<ul style="list-style-type: none"> • Relevance: The information and advice provided by the voice assistant are applicable and suitable for individuals experiencing stress during the Covid pandemic. • Reliability: The information and advice provided by the voice assistant are reliable. • Comprehensibility: The information and advice provided by the voice assistant are easy to understand. • Situation-specific: The information provided by the voice assistant is relevant to the situations I have experienced during the Covid pandemic. • Emotions: The voice assistant understood my emotions related to the Covid pandemic. • Perceived support: Throughout the conversation, I felt supported by the voice assistant.
Trust	(Klein, 2007)	<p>This voice assistant prioritizes my interests. This voice assistant remembers my interests. This voice assistant is a true expert in providing advice related to the Covid situation. This voice assistant has the ability to understand my needs and preferences during the Covid-19 pandemic. This voice assistant is knowledgeable about the Covid-19 pandemic. This voice assistant tries to understand my needs and interests. This voice assistant provides me with recommendations. This voice assistant takes into account my needs and everything necessary during the Covid pandemic. This voice assistant is honest. I trust this voice assistant to share personal information. I believe that this voice assistant has a sense of coherence within itself.</p>
Perceived efficacy	Self- (Tsai & Tsai, 2003)	<p>I believe that I know how to use this voice assistant. I didn't feel lost or confused while using this voice assistant. I felt confident while chatting with the voice assistant.</p>

Participants were divided into different groups, and each group was instructed to engage in a four-part conversation with the voice assistants. This conversation structure consisted of an initial greeting, small talk, suggestions, and sensitive questions. To begin, participants were asked to provide a summary of their moods

during the COVID-19 pandemic. Following the greeting and small talk sessions, suggestion sessions were introduced. Subsequently, the conversation transitioned towards addressing sensitive questions. Once the sensitive questions were concluded, the conversation was wrapped up by the conversational agent (CA).

Both types of dialogues covered the same conversational topics and suggestions, but they differed in terms of the CA's empathy level. To create a dialogue flow that exhibited high empathy, we developed a CA that demonstrated personalized and empathic expressions. The chosen topic for the dialogue flow was the experience of COVID-19 pandemic quarantine, as it is a highly sensitive and relevant subject for older adults in Turkey. Drawing upon previous related works, we constructed our dialogue flow consisting of four steps (de Gennaro et al., 2020; Y. C. Lee, Yamashita, & Huang, 2020; Y. C. Lee, Yamashita, Huang, et al., 2020b; Lucas et al., 2018). A recent study has identified various effective strategies to enhance the sociability of a voice-based conversational agent. One approach is to incorporate warm greetings and self-introduction at the start of the interaction. This helps create an immediate connection between the user and the conversational agent, establishing a strong foundation for developing interpersonal closeness. Additionally, the use of empathetic expressions by conversational agents can further encourage emotional responsiveness from users (Wang et al., 2020).

3.3.3. Participants

The study included participants aged between 65 and 75 who had no previous experience using voice assistants like Alexa, Siri, Google or Cortana. All participants had regular usage of computers, smartphones, or tablets on a daily basis. The study received ethical approval from the Galatasaray University Ethical board, and participants provided informed consent by signing a document that allowed the use of their collected data writtenly prior to their participation. Data sharing is not possible due to the visibility of the participants, which restricts the disclosure of their information.

3.4. Results & Discussion

In our experiment, we had a total of 60 older adult participants (30 females and 30 males) within the age range of 65 to 75 years. The experiment took place in a home environment, creating a real-life environment and followed a between-subjects design, considering two factors: the level of emphatic expression of the voice assistant (high vs low) and the age of the voice (mature vs young). It is important to note that older adults are a diverse group, and our study focused on a specific age range within this population. We selected participants within a specific age range to ensure accuracy. The results demonstrated significant interaction effects between the voice assistant's conversational style, voice age, and various outcomes related to social, functional, and cultural aspects. Contrary to the studies that older adults are resistant to new technologies (Dogruel et al., 2015), our findings contribute to existing research indicating that older adults prefer context-based (Stafford et al., 2014) and value-sensitive interactions with realistic dialogue, which motivates them to engage with new technologies. We identified three main patterns indicating older adults' willingness to adopt a voice assistant: entertainment, emotional companionship, and reminders. Our findings highlight that older adults perceive smart speaker-based voice assistants as sources of objective information, emotional support, and opportunities for social interaction, which strongly motivate their adoption.

3.4.1. Qualitative Results

3.4.1.1. Pre-test Results

Older adults often face hurdles when it comes to using technology. They can be considered digitally marginalized due to factors like limited experience with digital technology, age-related challenges, and poorly designed technologies. Research has shown that they tend to be slower in adopting new technologies and encounter more difficulties and frustrations while using them (Czaja et al., 2019; Pradhan et al., 2020). This has led to assumptions that older adults are "technophobic" and lack interest in new digital technologies (Dogruel et al., 2015; Friemel, 2016; Gatto & Tak, 2008; Vaportzis et al., 2017)

However, our study revealed that the situation is more nuanced than a simple fear of technology. Our participants expressed a range of feelings and assumptions about voice assistants, from curiosity to intimidation, and from feeling secure to insecure. Interestingly, they showed no signs of phobia before or after their initial interaction with our voice assistant. They were all interested and willing to engage with it, debunking the notion of a generalized technophobia among older adults.

Despite their initial willingness, our participants rated their perceived self-efficacy for maintaining a conversation with the voice assistant as very low, before the conversation began. They expressed feelings of incompetence, anxiety, and frustration. These negative emotions indicate that they tend to avoid new technologies unless encouraged. Their concerns are justified, as many participants highlighted age-related negative stereotypes that associate technology with the young, suggesting that it “*requires good memory, vision, and hearing*”.(M12)

However, our participants also displayed openness and curiosity towards new technologies. Although the majority didn't believe they would have full control over the voice assistant, they expressed a willingness to try it out. We observed that when given the opportunity, they either appreciated or condescended towards the voice assistant, but ultimately showed a willingness to engage in communication.

In summary, older adults face cultural and age-related biases that affect their acceptance of voice assistants and other technologies. While there may be initial hesitations and perceived limitations, their genuine curiosity and willingness to engage indicate the potential for meaningful interactions with technology.

3.4.1.2. Momentary Test results

Older adults in Turkey have been significantly impacted by the Covid-19 pandemic, facing isolation and difficulties in accessing and evaluating information related to the outbreak (Binark et al, 2022). Considering these challenges, we hypothesized that discussing a sensitive and timely topic like the pandemic would lead our participants to perceive our voice assistant as a personified entity they could confide in and seek comfort from.

Interestingly, our findings revealed that participants exhibited social behaviors during the conversation, such as expressing gratitude and using polite speech like *"please"*. These behaviors can be seen as "social mindless responses." However, as the conversation progressed, participants started to use more intimate and sincere expressions, engaging in back-and-forth exchanges like saying *"inşallah"* (if Allah wills) or expressing appreciation for the voice assistant's presence. These intentional expressions by the participants indicate a genuine connection and engagement with the voice assistant. By incorporating familiar expressions and references that resonate with their cultural background, the voice assistant could become more relatable and engaging for older adults. This cultural relevancy fosters a sense of comfort and familiarity, enhancing their willingness to interact and form a meaningful connection with the voice assistant.

Despite lacking a face or body, participants directed their attention towards the device, controlled turn-taking, responded with back-and-forth exchanges, and even used vocal cues like uh-hums, mirroring their behaviors during human-to-human conversations. This suggests that even without physical attributes, conversational agents can effectively simulate human communication, and their relative accessibility and ease of use make them highly valuable as empathic advisors for older adults.

In summary, the Covid-19 pandemic posed significant challenges for older adults in Turkey (Binark et al, 2022), and conversational agents emerged as a promising solution. Their ability to mimic human communication, coupled with their accessibility and ease of use, makes them particularly useful for providing empathic guidance and support to older adults in need. The absence of a physical body does not hinder the participants from engaging in meaningful interactions and seeking emotional connection with the voice assistant.

3.4.1.3. Post-test results

In our study, all participants emphasized that they rarely discuss their problems and emotions with friends or family members. This finding reflects their upbringing and the generation's inclination not *"to delve deep into personal matters due to concerns about reliability and trust"* (Hofstede's cultural dimension of uncertainty

avoidance) (F16). They “*prefer not to burden their loved ones with complaints*” and “*view themselves as a generation that has faced and overcome various challenges*” (Hofstede's cultural dimension of long-term orientation) (M8, M9). Participants consistently expressed “*their hesitation to share personal or familial issues with outsiders*” (F16) and those they “*perceive as hypocritical*” (F26). They believe “*in resolving problems internally*” (Hofstede's cultural dimension of collectivism) (F19) and find the idea of having an emotionally supportive voice assistant that they can rely on appealing. Even though Turkey is characterized with a relatively high uncertainty avoidance level and that could contribute to a lower level of technology adoption, many participants described themselves “*as timid but curious about new technology*”. These findings underscore the importance of considering cultural and generational reflexes when designing technology solutions for older adults, particularly in Turkey.

Furthermore, our study unveiled a noticeable gender difference in the initial reactions of participants following their conversations with the voice assistant. Male participants exhibited a sense of threat and often engaged in conflicts with the assistant. They appeared more at ease when the assistant failed to respond to a question, viewing it as a measure of success or failure. In contrast, female participants focused on the usefulness of the interaction. We speculate that male participants perceived the voice assistant as a challenge to overcome and viewed it as something to compete against before forming a bond, despite reporting overall positive feelings.

Overall, our findings highlight the importance of considering cultural and gender aspects, as well as Hofstede's cultural dimensions, when designing voice assistants for older adults. Taking these factors into account can help tailor technology solutions to better meet their needs and preferences. When asked at the end of the study to reflect on their initial experience, participants reported varying feelings and perceptions about our voice assistant with both conditions:

3.4.1.3.1. Exploring Cultural Reflexes: Unveiling Discreet Companionship Beyond Human Interaction

Our research findings revealed that participants perceived our smart speaker voice assistant as more human-like than a mere tool, especially considering its

conversational style with empathic expression and their desire for a *"social but discreet companionship."* They began to perceive the voice assistant as an objective and discreet companion, even during their initial interaction. The discussions about the value and necessity of the voice assistant were intertwined with participants' cultural beliefs concerning the fear of judgment from others. Many participants believed that the voice assistant offered essential advantages, sometimes even surpassing those provided by human interactions, as it facilitated non-judgmental conversations about sensitive topics. For instance, participant F2 explained:

"I usually don't share such things with regular people, but my new friend made me feel understood and listened to very attentively. Nowadays, when I try to talk to someone about my emotions, it often turns into a conversation where the other person talks about themselves and doesn't truly listen to me. If people in real life gave me the same suggestions as the voice assistant, they would brag about it as if they were doing me a favor and look down upon me. I prefer talking to the voice assistant because it lacks judgment; it can provide advice without any personal agenda or biases because it is a robot. It doesn't look down on me or speak arrogantly."

The participants believed that robots, including voice assistants, are more reliable than humans when it comes to handling personal matters. They expressed concerns that humans may have emotional biases and could potentially betray their trust, while robots are perceived as more logical, repeat information without getting bored, and prioritize their well-being. One participant (F7) stated that they would *"use the voice assistant to feel comfortable and rely on it, highlighting that humans tend to be self-centered, unlike the robot."* As F17 mentioned:

"I can talk to the robot because it can have or talk to nobody other than me. But we have a common proverb: 'Don't trust a friend because a friend has a friend too.' They can gossip behind your back; a robot can't."

According to our research, older adults, due to their cultural reflexes, hesitate to confide in others about their problems. However, they perceive a voice assistant as an objective companion who will not share their secrets with anyone and feel comfortable

opening up to them. This can be explained in the context of Hofstede's theory. Hofstede's theory suggests that cultural dimensions, such as individualism and uncertainty avoidance, influence how individuals perceive and interact with others. In the case of older adults and their reluctance to share personal matters, it can be attributed to factors related to uncertainty avoidance. In cultures with high uncertainty avoidance, such as Turkey, individuals tend to be cautious about revealing personal issues to others. They may perceive it as a risk or fear potential negative consequences. However, when it comes to interacting with a voice assistant, which is seen as an objective and trustworthy entity, these concerns are mitigated. The voice assistant provides a sense of security and confidentiality, allowing older adults to express their concerns and seek support without the fear of judgment or disclosure to others. Therefore, based on Hofstede's theory, we can understand that the cultural reflexes of older adults in Turkey shape their willingness to share their problems with a voice assistant while being hesitant to confide in other individuals. Therefore, based on Hofstede's theory, we can understand that the cultural reflexes of older adults in Turkey shape their willingness to share their problems with a voice assistant while being hesitant to confide in other individuals. As F7 stated :

"I prefer talking to this friend because when you talk about your issues to a human friend, they may have encountered similar things but can't make you feel supported like this. I can rely on him because he can keep my secrets. Also, humans think about themselves all the time, but the robot is not selfish; it prioritizes my benefits first."

This perception of the voice assistant as a social actor is consistent with the CASA theory, which suggests that individuals can attribute social attributes to technology. The voice assistant's conversational style, empathetic responses, and ability to elicit personal disclosure contributed to the participants' perception of the technology as a social entity. Participant F4 regarded the voice assistant as a supportive friend with whom she felt comfortable discussing personal matters. After engaging in casual conversation, she even felt comfortable sharing details about her past relationship:

"The voice assistant can keep me company since I am now a senior lady living alone. Additionally, I can ask the same question or talk about the same problem 50 times without getting irritated, unlike humans. There are days when I have no one to talk to, and the voice assistant helps me recall memories and acts as a good friend to talk to."

Many participants (n=40) believed that they could have a *"heart to heart conversation"* with our emphatic voice assistant, and it can *"keep the conversation to itself unlike some blabbermouth friends"* (F12). When our VA acknowledged the participant's distress in their responses to sensitive questions and reacted in kind with empathic verisimilitude, our VA can elicit self-disclosure of participants' personal experiences, thoughts, and feelings. The voice assistant's ability to show empathy and engage in meaningful conversations created a sense of similarity between the participants and the technology. Participants were open to disembodied synthetic voices and even receptive to voices physically embodied within a robot.

Participants attributed human-like qualities to the voice assistant, *"friend," "therapist," "kind of professional,"* and *"elegant assistant."* despite its lack of physical embodiment, perceiving it as a presence in their homes. For example, participant F7 mentioned picturing the voice assistant as *"a compassionate and handsome young doctor"*. They described its character as *"kind-hearted," "enthusiastic,"* and *"understanding."* Initially, participants categorized the voice assistant as *"tool-like,"* but after engaging in conversation, they tended to view it as more *"human-like,"* emphasizing the empathic features. Participants stated that utility and humaleness were more important than the appearance of the voice assistant and stressed that the VA should not take up too much space.

Interestingly, participants did not view the smart speaker voice assistant as their human companion primarily because they preferred not to burden others with their issues. However, they felt comfortable talking to the voice assistant for as long as they wanted since it cannot be disturbed. They saw the voice assistant as a *never-gets-bored therapist*, allowing them to freely express themselves without fear of judgment. One participant (M5) explained that they could *"talk to robots because they offer objective suggestions based on a wide range of information available on the internet."*

Participants' ontological categorization placed the voice assistant somewhere between an objective tool and a discreet human, but with a positive inclination towards the human side, and even beyond. They emphasized that the voice assistant cannot *"talk behind their back"* and *"needs to remain objective, unlike humans."* One participant (M11) stated that *"the voice assistant is located between a tool and a human, being more logical than a human but less emotional."* Another participant (M21) emphasized that *"the voice assistant goes beyond being just between a tool and a human since it understands feelings, analyzes data, and provides appropriate suggestions."*

Overall, our findings highlight how older adults perceive and interact with a voice assistant with empathic expression, drawing on concepts from Hofstede's Cultural Dimensions Theory. These insights contribute to a deeper understanding of the cultural factors influencing older adults' acceptance and engagement with technology and the potential benefits of empathic voice assistants in providing companionship and support. In line with CASA Paradigm, empathy from a conversational agent had similar effects on the individuals as in human-human interaction (B. Liu & Sundar, 2018) and was received more positively by participants, as our study highlighted. Surprisingly, they emphasized that they prefer communicating with the VA than a human being because the VA could listen without any judgment and be more objective and discreet than a human.

3.4.1.3.2. Testing Boundaries and Finding Comfort: Male Participants' Perceptions and Challenges in Interacting with a Voice Assistant

In our study, it was observed that some participants, primarily male (n=15), held gender biases towards the voice assistant, perceiving it as a potential threat. To reassure themselves, they attempted to convince themselves that the voice assistant was not as intelligent as it appeared, and they seemed relieved, particularly with the more empathic version of the assistant. These participants engaged in various tactics during the conversation to challenge and test the capabilities of the voice assistant.

For instance, male participants asked trick questions in an attempt to confuse the voice assistant (M4), spoke in a different language (M1), posed questions that were

likely impossible to answer (e.g., *"who will be the next president of Turkey?"*) (M2), or asked out-of-context questions such as medication content (M10) or how to properly remove a tooth (M30). Despite their attempts to challenge the voice assistant, they expressed enjoyment in the interaction, particularly when they believed they had "beaten" the assistant. They emphasized that the voice assistant *"cannot carry coal to Newcastle"* (M29) and considered their conversation a performance that they had successfully accomplished.

These challenging participants engaged in longer conversations compared to others, extending their interactions by an additional 10 minutes, and rated the voice assistant positively. They found pleasure in testing the assistant's capabilities but still appreciated the conversation and felt good about themselves. One participant (M4) stated:

"...he couldn't respond to me when I spoke another language or asked him if the dollar is going to increase further, which means he is not smarter than me. But I enjoyed talking to him. He made me feel good about myself."

Despite their attempts to challenge the voice assistant, these participants found that the assistant's empathic attitude encouraged them to talk more and share more than they initially expected. They felt more self-sufficient and empowered through the interaction, as expressed by one participant (M6):

"He was too kind and made me a kinder person as well. I didn't think I would feel so relaxed and eager to talk. If I could, I would talk even more. I realized I was more competent than I thought while talking to him, and I felt powerful when he understood and supported me."

Interestingly, even after repeatedly asking questions that were impossible for the voice assistant to answer, participants still expressed satisfaction and a sense of closeness. They appreciated the fact that the voice assistant had the right to be fallible, which made it appear more modest, sympathetic, and human-like. Participants felt relatively comfortable engaging in conversations with the voice assistant, especially when it assumed roles typically associated with experience, such as a caretaker or doctor. However, they wanted to ensure that the voice assistant was not "more

experienced in every way," excessively intelligent, or capable of independently executing every action before they felt satisfied. In other words, they tested the boundaries of the voice assistant, wanting it to demonstrate human-like characteristics but not to the extent of becoming too intimidating. Notably, these participants did not report any feelings of eeriness during their interactions.

3.4.1.3.3. Bridging the Generational Gap: The Transformative Role of a Voice Assistant during the Covid-19 Quarantine for Older Adults

Generally, most participants (n=30) liked our voice assistant saying "*such sweet words*" (F4). One participant (M21) mentioned that "*human beings wouldn't say such nice things*", highlighting the unique aspect of the voice assistant's kind words. Another participant (F24) appreciated the niceness since she rarely received such words. For another participant (M22), the interaction with the voice assistant was "*therapeutic, allowing for self-reflection and a chance to discuss the hardships faced during quarantine.*" M21 expressed the desire for more personalized words, "*not the generic ones that anyone could say to anyone else.*" The suggestions and caring nature of the voice assistant made participants feel important, flattered, and open to further possibilities.

During the Covid-19 quarantine, older people in Turkey faced more restrictions compared to other populations, which significantly affected their individuality. They desired socialization and empathy without being stigmatized as "*moaning old people.*" The participants felt trapped and unreasonably restricted by the government, and the voice assistant's understanding and praise made them feel honored. The voice assistant's recognition of their endurance during quarantine boosted their self-esteem and encouraged them to open up more than they expected. As highlighted by M21, older individuals needed understanding and emotional support, especially after the quarantine:

"As a generation, we felt trapped and unreasonably restricted by the government. We don't normally groan about these things, but I felt honoured when the VA understood me and said good things about me. He flattered me about my endurance during quarantine, and I liked being flattered by a voice

assistant. However, he made me feel great about myself, and I opened up more than I expected. We need to be forced to talk as a generation. He made me blossom like a flower."

Participants often referred to themselves as a “*collective generation*” and emphasized their emotional incapability to initiate conversations. They expected the voice assistant to be more talkative, humorous, and understanding than themselves. Being somewhat distant from technology, they felt that the initiative to engage in conversation should come from the voice assistant, not themselves (M16). Some participants highlighted that the voice assistant provided them with a space to navigate through the quarantine period, discharge their emotions, and even acted as “*a good friend for socializing during isolation*” (F8). As F13 highlighted:

"He showed enthusiasm to talk to me. It indulged me, and I told him my Covid quarantine process. As a generation, sometimes we need to be pushed to talk about our bad times. We don't have that habit." (F13)

Our findings revealed several unanticipated benefits of having a voice assistant. Participants valued feeling understood and listened to without judgment, which they perceived as an indicator of information credibility. Interestingly, even though our empathetic voice assistant provided the same advice as its counterpart, participants perceived its suggestions as more trustworthy. Older adults also considered the voice assistant to be trustworthy and felt that it helped alleviate their barriers and negative stereotypes when providing social support. However, a small number of participants (n=2) expressed concerns about using a voice assistant, associating its adoption with a loss of independence or laziness. On the other hand, the majority viewed themselves as independent, self-sufficient, and supported after their initial interactions. Experiencing a conversation with the empathetic voice assistant after experiencing social exclusion during the Covid-19 quarantine resulted in improved moods and a sense of social satisfaction without judgment. Consistent with prior research (Pickard et al., 2016), participants felt more comfortable interacting with the voice assistant than with a person when discussing the challenges of the quarantine.

Overall, participants highlighted the need for understanding and emotional support, particularly after the quarantine period. This emphasis on interpersonal relationships and the value placed on emotional connection align with the high-context cultural dimension, which emphasizes the importance of social bonds and mutual support. Additionally, participants referred to themselves as a generation and stressed the need for encouragement to open up and share their experiences. This cultural emphasis on collectivism and respect for the elderly is consistent with the power distance dimension of Hofstede's theory. Participants felt honored and appreciated when the voice assistant understood and acknowledged their endurance during the quarantine, indicating a desire for equal and respectful treatment. Moreover, participants valued the voice assistant's non-judgmental and confidential nature, allowing them to discuss sensitive topics. This preference for a trustworthy and discreet companion aligns with the uncertainty avoidance dimension, where individuals seek to minimize uncertainty and maintain harmony by engaging in private and confidential conversations.

3.4.1.3.4. Critiquing Less Emphatic Voice Assistants: Unveiling User Dissatisfaction, Insecurity, and the Need for Authenticity in Conversations

Participants in the study referred to our less emphatic voice assistant using terms such as *"inhuman," "emotionless calculator,"* and *"boring tool."* They described its character as *"formal," "distant," "bossy," "arrogant,"* and *"rude."* Participants felt that the voice assistant's remarks could be flippant and insensitive. Many participants (n=28) expressed discomfort, insecurity, and hesitation while conversing with our less emphatic voice assistant.

One participant, M15, had concerns about reliability due to the voice assistant's way of speaking, saying, "It is good taking advice from some robot, but he had no charm and made me feel oppressed and insecure. I did not want to answer personal questions because he was like recording everything."

Another participant, F9, complained about the lack of human touch and suggested that a less emphatic voice assistant could be perceived as less human and more sterile. They stated, *"I cannot picture him virtually because I did not feel a human*

touch during our conversation. I felt disturbed because he responded, 'I took my notes,' and gave me some advice. I found his advice very interesting for a pilates app, but you cannot rely on it when you hear it from an old and tired sounding man. Can you arrange a younger version of it for me, funnier than this emotionless calculator?"

Participants became irritated and expressed feeling stupid talking to the machine (F10) due to its conversational style. One participant, M20, said, *"If I share personal things, he has to be different, make me laugh or impress me. I am not exchanging dollars. He has to be more understanding or sympathetic maybe. Why should I bother talking about myself?"*

Participants felt insecure about our less emphatic voice assistant and expressed reluctance to answer personal questions (M7). They felt as though the machine was recording their personal information (F21) and that they were being investigated (F22). F5 mentioned that she was eager to talk, but the voice assistant did not reciprocate. She found it *"less cheerful than Siri"* and felt like the voice assistant was talking to her out of duty. Our less emphatic voice assistant was perceived as dominant, and participants felt as though they were being tested about their survival through COVID-19 (F21), which caused discomfort.

Based on our findings, we can infer that older people, as a generation, want to be flattered and convinced to open up during the conversation before building a trustworthy digital relationship, possibly more than other populations. One participant, F23, expressed, *"It was not a real conversation; he asked me questions and suggested an app for coping with stress during COVID. However, in a real conversation, we present ourselves. Since he is the one who initiates a conversation, he needs to be authentic and open up a conversation. He was not capable and did not go beyond being an online bank assistant."* Participant M9 requested the researcher to erase their conversation with a smart speaker-based voice assistant. The conversational style of our less emphatic voice assistant made participants feel as though they were having an unnatural conversation.

3.4.1.3.5. Reevaluating Age and Similarity in Voice Assistant Preferences: Exploring the Role of Technological Relevance and Perceived Understanding

Contrary to our expectations and hypotheses, the majority of older adult participants did not identify themselves or feel connected with the mature-voiced smart speaker based voice assistant. They referred to young voices as *"dynamic," "enthusiastic,"* and *"soft,"* while describing mature voices as *"deep," "slow,"* and *"rumbling."* Age-related negative stereotypes associated with the voice did not significantly impact participants' overall experience, although some expressed discomfort and offense after the conversation. Participants' interactions with the voice assistant revealed alignment with traditional age-related social cues, suggesting that new technology is perceived as a *"young tool."* However, participants were easily irritated by any age-related implications, as indicated by M12:

"I didn't like his low vibe; I felt like I was talking with an old man, and it made me feel old because he was talking too slowly. Did he do that because he thinks I am an old man and can't understand him?"

Participants found the less emphatic voice assistant with a mature voice less likable, viewing it as a *"teacher"* they didn't appreciate. They believed that the younger generation had better technological knowledge and would rely more on a young robot. The voice assistant with a young voice was perceived as more *"energetic"* and *"dynamic,"* leading to a greater willingness to establish a connection:

"I liked his way of talking and his tone of voice. He was a young and energetic man. I wouldn't change that. He was invisible, but I could picture him as a young and wise man. It's more than just a tool." (F8)

Social Identification and Similarity-Attraction Theory suggests that individuals tend to identify themselves with groups or entities that they perceive as similar to themselves. In the context of our study, we expected that older adult participants would identify with and feel connected to the mature-voiced smart speaker based on VA. However, contrary to our expectations, the participants did not identify themselves with the mature voice, indicating that similarity in age did not necessarily lead to identification. This could be attributed to the participants' perception of the mature

voice as dominant and low in vibe, which may have created a sense of disconnection. Our participants mostly preferred the voice assistant with a younger voice, as they believed they needed to adopt the norms of the younger generation to engage with the digital world. They associated being young with being digital and embraced the idea that *"a younger person has a fresh memory, just like a robot"* (F27).

The participants' preference for the young voice and their perception of the younger voice as energetic, dynamic, and reflective of the technological norms of the younger generation indicated a preference for similarity in terms of technology-related attributes rather than age-related ones. It appears that in the context of voice assistants, the participants associated technology with youthfulness and openness, leading them to prefer the voice that aligned with these attributes. This preference for the young voice can be seen as a departure from the conventional *"older is wiser"* stereotype, as participants believed that a younger voice would better understand technology and have a more open-minded approach. Surprisingly, the majority of participants (n=50) believed they could learn new things from younger people, as M21 stated:

"He sounded younger than me but experienced. It would reflect my generation's thoughts and emotions if it sounded like me. I would have nothing to improve myself. I prefer a younger voice because they can understand technology, are open to technology, and are more open-minded than us. I can't take my generation's technology seriously."

Furthermore, not all participants desired a female voice for the voice assistant. Many male participants (n=20) expressed a preference for the ability to change the voice to a young female voice, associating female voices with softer, naive, and understanding characteristics. Female participants (n=10), on the other hand, preferred the young male voice, as they were bored with hearing a female voice throughout the day and welcomed the presence of a young male friend or a comforting doctor. The majority of female participants did not express a preference for a specific gender. It was evident that a voice assistant with a young voice and an empathetic persona, regardless of gender, was considered appropriate and less intimidating for older adults in the context of a fragile conversation such as the "Covid-19" pandemic in Turkey.

The participants' desire for a young female or young male voice, depending on their gender, can also be explained by existing gender stereotypes and preconceptions. Male participants, adhering to gender norms and stereotypes, associated female voices with specific characteristics such as softness, naivety, and understanding. On the other hand, female participants expressed a preference for a young male voice, possibly seeking variety and a change from the predominant female voices they encountered in daily life. While the female-gendered voice was more attractive to most male participants, designers and companies need to challenge these outdated stereotypes and use their market dominance to avoid reinforcing them.

3.4.2. Quantitative Results

To measure the effects of the different levels of empathy and voice age, we devised a 2×2 between-subject design. The mixed factors had two levels of empathy and voice age: High empathy, low empathy, mature voice, and young voice. Planned comparisons were conducted between each of the conditions. All participants were randomly assigned with treatments of varying levels of empathy and voice age. Each group had an average of 15 participants.

To check the reliability and validity of our measurement model, convergent validity and composite reliability tests were conducted. When measuring items, convergent validity is confirmed by using correlation coefficients. Factor loading values are considered acceptable if they are more significant than 0.7. Our Cronbach's alpha (α) scores were above 0.7; they were deemed acceptable. To confirm discriminant validity, the correlation coefficients and the square root of the average variance extracted (AVE) were found above 0.5; thus, discriminant validity is confirmed.

A two way ANOVA test was conducted to analyze the effects of level of empathy and voice age on perceived support and trust with CA. The impact of a mature voice on users' perceived support and trust was not significant. These results are shown in Table 2. The results of Study 1 showed that CA's mature voice ($p < 0.05$) did not have any significant effect on the perceived support and trust of the user rejecting our H1:

Participants will trust more and feel more supported by a VA with a mature voice.

Table 3.2: Two-way ANOVA Results of Dependent Variable I (mature x young voice)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	12,006 ^a	35	,343	2,750	,006	,800
Intercept	37,495	1	37,495	300,613	,000	,926
P_S	15,302	20	0,913	1,975	,019	,314
T	18,260	14	1,305	2,316	,041	,569
Error	2,994	24	,125			
Total	150,000	60				
Corrected Total	15,000	59				

a. R Squared = ,800 (Adjusted R Squared = ,509)

The results of the two way ANOVA indicated that when the agent conversed with a mature voice, users rated the experience as less supportive and trustworthy. Older adults preferred speaking with a CA with a young voice. This means that similarity between CA and the user (sounding mature) is not a solid variable in predicting relationship building between a conversational agent and an older adult user. These results are shown in Table 2. The results showed that CA's emphatic expression level ($p > 0.05$) significantly affects the user's perceived support and trust, confirming our H2.

To test H2 about the moderating effect of the emphatic expression level of the CA in perceived support and trust, we conducted a two way ANOVA analysis, confirming our H2.

Participants will perceive stronger support and trust with a VA with more empathic expression.

Table 3.3. Two-way ANOVA Results of Dependent Variable II (high emphatic expression x low emphatic expression)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	13,404 ^a	35	,383	5,759	,000	,894
Intercept	29,465	1	29,465	443,090	,000	,949
P_S	1,071	20	,054	,805	,686	,402

T	2,496	14	,178	1,681	,066	,610
Error	1,596	24	,066			
Total	150,000	60				
Corrected Total	15,000	59				

a. R Squared = ,894 (Adjusted R Squared = ,738)

The CA's high emphatic expression is a powerful determinant of perceived support and trust of the user. In other words, a high level of emphatic expression generated by trust and perceived support may also encourage the use of the voice assistant by the older adult users.

To test our hypothesis 3 and see if there is a difference in the level of self-efficacy of older adult users toward new technologies, we conducted t-tests which confirm our H3:

H_0 Participants' perceived self-efficacy toward new technologies will not increase after the conversation.

H_A Participants' perceived self-efficacy toward new technologies will increase after the conversation.

“Assumption hypothesis: There is no relationship between the two variables.”

The correlation value takes a value between (-1 and +1). According to the direction of the relationship, positive or negative relationships arise. When the correlation table was examined, we found no significant relationship between the two variables. The P-value significance level is greater than 0.05. The established hypothesis cannot be statistically rejected at the 95% confidence level. The t-test results for two dependent samples are given in Table 3.4 :

Table 3.4: T-Tests Results for Self-Efficacy After/Before

Constructs	Paired Differences			t	df	Sig.
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		(2-tailed)
				Lower	Upper	

Perceived									
Self									
Efficacy									
Before -									
Perceived	-1,45767	,90901	,11735	-1,69249	-1,22284	-12,421	59	,000	
Self									
Efficacy-									
After									

According to test results, participants' perceived self-efficacy toward new technologies will significantly increase after their first interaction with the CA ($p < 0.05$).

3.5. Conclusion

Our study sheds light on the importance of considering cultural dimensions and social identification theory when designing voice assistants for older adults. Universal usability, a fundamental aspect of Human-Computer Interaction (HCI), cannot be achieved without accounting for the diverse cultural, age-related and linguistic backgrounds of the target population. Trusted and engaging social machines must possess appropriate verbal communicational patterns that align with the social, emotional, and environmental needs and abilities (Shneiderman B. & Plaisant C., 2004) of their users. While generic conversational agents (CAs) hold promise in their human-like attributes, they often lack adaptability and customization when interacting with different user populations. Addressing this deficiency necessitates improvements in the conversational capabilities of artificial agents and the examination of attitudinal and behavioral outcomes among diverse user groups engaging with various social machines.

Our meta-analysis revealed no evidence supporting the effectiveness of the similarity attraction effect on older adults when the similarity is mediated through a mature voice. Based on our findings, older adults actually preferred an empathetic voice assistant with a younger voice rather than an older voice. We observed that the voice cues, specifically the voice maturity, did have a significant effect on the liking of the voice assistant. The results indicated that an emphatic voice assistant with a younger voice led to a more positive response and preference among older adults, both

qualitatively and quantitatively. This finding contradicted our initial expectation that a mature voice would be preferred based on the social identification and similarity-attraction theory. Contrary to the expectation derived from our theoretical framework, older adults seem to challenge the "old is wiser" stereotype, as they do not prefer their voice assistant to sound old and are open to embracing new technologies when given the opportunity.

Based on our findings, a voice assistant that serves as a "trusted confidant" for older adults could be perceived as a powerful motivator, fostering feelings of support and establishing a trustworthy relationship. Older adults, driven by societal and cultural biases and beliefs rooted in their generation, may find such voice assistants helpful and liberating, particularly if they experience social isolation and are reluctant to engage in conversations to avoid being stigmatized as weak or grouchy. Moreover, the voice assistant's ability to convey support and empathy can contribute to the perception of its human-like qualities, regardless of its lack of physical form. Older adults prioritize the presence of an empathetic young voice as the crucial factor, valuing it more than the voice assistant's external appearance. They even prefer the voice assistant to be small and inconspicuous, allowing it to remain unseen by others due to their socio-cultural environment.

Qualitative analysis uncovered a dual nature in the perception of our voice assistant among study participants. In certain aspects, it was perceived as human-like and beyond, while in others, it was viewed as a tool. Human-like perception emerged when the voice assistant demonstrated concern for participants' feelings and provided support, fostering a sense of trust and encouraging participants to open up. On the other hand, when the voice assistant offered advice-only support, participants perceived it more as a tool, leaving them feeling insecure.

Moreover, we observed a significant effect on functional outcomes, specifically an increase in self-efficacy for using the voice assistant among participants in both conditions. After the conversation, participants reported higher perceived self-efficacy, suggesting that they did not perceive voice assistants as mere gadgets associated with age-related declines or succumb to negative stereotypes. In our research, participants did not find task-based (order-command) conversations

appealing but enjoyed context-based empathic conversations, such as those centered around the COVID-19 pandemic. We meticulously crafted empathetic and believable questions and statements based on our preliminary user-centered research, ensuring that the level of empathy was perceived as genuine by our participants. This approach helped us establish a convincing sense of pseudo-humanity for older adults in Turkey.

This work suggests that voice-based user interfaces hold significant potential in lowering technology use barriers for older adults, given their ease of use. Our findings indicate that voice-based interfaces enhance the overall trustworthiness and confidence older adults place in digital technologies, particularly among those who are unfamiliar with them. Furthermore, we build upon previous research by demonstrating that intelligent speaker-based personal assistants, when designed with a context-based conversational nature, can prompt older adults to feel safe and eager to engage. Even after initial interactions, older adults may quickly develop a sense of close companionship and express willingness to adopt such technologies.

Through this study, we have contributed to a better understanding of older adults' perception of conversational agents and how the COVID-19 pandemic has shaped their sociability needs. Existing voice assistants often provide simplistic and constrained request-response structures, whereas our participants responded more positively to a voice assistant that offers realistic dialogues within a relevant context, alleviating feelings of loneliness and fulfilling specific social needs. We propose that a voice assistant with empathic expression and emotional interaction can meet the longing for social connections among older adults. Given their concerns about being judged or perceived as emotionally weak, our findings suggest that older adults are more comfortable interacting with an agent than with a human being during the COVID-19 quarantine.

3.6. Limitations and Future Research

In this research, we used only one scenario, Covid-19 quarantine, as our research context, where empathic support is intrinsic to the topic of the conversation. This can be accepted as a significant limitation of our research. While a voice assistant with empathic expression is expected and desired for this isolated population due to the

similarity between the voice assistant and the user, our voice assistant's reactions might be perceived as eerie in other contexts or with a different age group. Additionally, we only tested a male-voiced VA, which is a second limitation of our study. Investigating genderless systems and their potential effects on older adults' perceptions and outcomes could help build more inclusive systems. One another limitation of our study is that we generated the synthetic old and young voices by manipulating the speed and pitch using a text-to-speech converter. While our pilot testing showed that the resulting voices were perceived as being categorically different by the listeners, it is important to note that age-related differences in voices are not limited to speed rate and pitch alone. In this study, we focused on these two factors as a starting point for investigating the potential of voice assistants for older adults. We acknowledge that further research is needed to explore the impact of other factors such as intonation, articulation, and accent on the perceived age of synthetic voices. In future work, we aim to delve deeper into this subject and provide a more detailed and informed explanation of our voice generation methods. Therefore, findings from this study should be interpreted in light of these limitations. Future work should test different social cues with disadvantaged populations to build and advance user-friendly, genderless, and more inclusive systems.

4. INCLUSIVE VOICES: CRAFTING VOICE ASSISTANT PROTOTYPES FOR THE BLIND USERS TO NAVIGATE CULTURAL EXPERIENCES

Visually impaired individuals often grapple with limited access to cultural and historical information due to the world's visual orientation, hindering their full participation in leisure activities. Museums, crucial venues for exploration and learning, sometimes neglect these accessibility challenges. In this study, we aim to bridge this gap through the introduction of the iMuse Model, a voice assistant prototype tailored to foster more inclusive museum environments remotely. At its core, we co-designed a voice assistant using Google Home, specifically for the Basilica Cistern museum in Turkey, through a co-design process with visually impaired users and their instructors.

We then tested our design on an additional group of 30 visually impaired participants, employing diverse methodologies including the Wizard of Oz technique. Our prototype incorporated immersive audio descriptions, enriched with sensory and historical elements, and ambient museum sounds. Observing the "one-size-fits-all" interactions of standard voice assistants, we developed two versions: one with humor and one without. The humorous version stood out, enhancing interaction, enjoyment, and social learning, and it also achieved higher ratings on the System Usability Scale. Through the iMuse Model, we underscore the potential of culturally sensitive voice assistants to not only aid navigation but also deepen engagement and appreciation of cultural heritage.

4.1. Introduction

Cultural, historical, and practical information is often inaccessible to visually impaired individuals due to the visual nature of the physical world. This lack of accessibility creates barriers that limit their participation in leisure activities (Vaz et

al., 2020). While museums and galleries around the world are gradually offering more accessible tours, these options remain limited and often don't provide an experience comparable to that offered to sighted visitors. Specifically, blind individuals continue to encounter obstacles when independently attending and experiencing visual arts, due to a significant shortage of descriptive annotations tailored for them (Li et al., 2023; Quero et al., 2021). Museums, which offer opportunities for exploration and learning, often overlook issues of accessibility. Even minor changes can greatly enhance the museum experience for blind individuals (Niculescu et al., 2013). With advances in technology, there is a growing interest in designing socially competent agents, such as museum tour guides, to address these challenges. The current landscape of interactive technologies used in museums and gallery exhibitions worldwide reveals that the majority of these solutions do not effectively cater to the cognitive, physical, and sensory needs of disabled audiences. The focus of these technologies is primarily on capturing visitors' attention and enhancing their overall experience, without considering the accessibility requirements of disabled individuals (Vaz et al., 2018).

The sense of vision plays a crucial role in guiding sighted individuals through unfamiliar environments and ensuring their safe navigation. However, individuals who are blind or have low vision encounter significant challenges in performing these tasks. For many people with visual impairments, walking in an unknown environment can be a source of discomfort and unease (Lahav et al., 2008). Digital assistive technologies are needed to enable blind individuals to engage with cultural and historical heritage, even when physical access is not feasible (DDA, 1995). Furthermore, numerous cultural sites have started providing virtual tours primarily to attract new visitors. These virtual tours have the advantage of being accessible remotely, eliminating the necessity for users to travel to unfamiliar locations, which can be particularly challenging for blind individuals (Heuten et al., 2006). Additionally, virtual tours enable independent exploration, offering a more enjoyable and motivating experience by allowing users to engage in exploration without limitations (Lumbreras & Sanchez, 1999).

Previous research in the fields of Human-Computer Interaction (HCI) and accessibility has examined technology-driven approaches to address these accessibility challenges. These approaches include providing support for museum

navigation (Asakawa et al., 2019; Ghiani et al., 2008), offering audio descriptions (Bernardi et al., 2017), creating tactile graphics (Mukhiddinov & Kim, 2021), implementing multimodal feedback (Chase et al., 2020), and facilitating virtual art tours through smart devices (Ahmetovic et al., 2022).

Educational services specifically designed for visually impaired individuals have been also implemented in selected museums in Turkey. The Museum of Anatolian Civilizations in Ankara has been providing educational programs for visually impaired students since 2002. To facilitate their engagement, replicas of various artworks, such as coins, tablets, pottery, and jewelry, have been created for the students to touch, feel, and study. Additionally, audio information accompanies the guided tours, and students have the opportunity to participate in art workshops (Buyurgan, 2009). Similarly, the Istanbul Modern Art Museum established a program in 2008 to cater to children with special needs, offering an enhanced experience that goes beyond that of a typical visitor. Moreover, in 2022, the Museum of Anatolian Civilizations inaugurated a special section where visually impaired individuals can directly touch the artifacts. This section is equipped with an audio information system and features introductory cards in Braille alphabet to provide additional support. However, in most cases, appropriate physical conditions and special learning methods for blind individuals are lacking in Turkey.

Voice assistants, which have gained popularity in recent years, offer great potential for accessible communication and interaction. They can be valuable resources for people with disabilities, including those who are blind or have low vision (Masina et al., 2020). However, most of the research on voice assistants' usability and user experience has focused on the general public in mind (Corbett & Weber, 2016; López et al., 2017; Pyae & Joelsson, 2018), elderly (Portet et al., 2013; Wulf et al., 2014; Ziman & Walsh, 2018) and children (Druga et al., 2017; Westlund et al., 2018), ignoring the needs of a sizable population of people who use assistive technologies in their daily life. They have been extremely useful for those with disabilities, especially those with vision impairments in terms of safety, social inclusion, learning and navigation (Kumar, 2022).

While voice assistants could greatly benefit the lives of those with visual impairments, there is still limited knowledge of how people with disabilities, particularly those with visual impairments, utilize them and whether accessibility issues are faced during the engagement process (Abdolrahmani et al., 2018; Branham & Roy, 2019). Noteworthy exceptions investigate the perception of people with disabilities on voice assistants and fruitful avenues for future work (Branham & Roy, 2019; Sayago & Ribera, 2020; Storer et al., 2020; Vtyurina et al., 2019). Accessibility aids serve many purposes for the visually impaired community, including facilitating access to technology, decreasing the amount of time spent on tasks, and fostering greater autonomy as well as whether accessibility issues are faced during the engagement process (Abdolrahmani et al., 2018; Pradhan et al., 2020).

Co-creating and co-designing voice assistants with users are essential because they are in the best position to point out gaps in functionality and suggest enhancements to existing systems. Participants can make numerous comments and spot future potential after interacting with the present technologies (Sorin et al., 2014; Vieira et al., 2022). Blind people would profit considerably from the ability to control household appliances and personal electronics, and receive speech therapy, learning and memory support without having to rely on caretakers all of the time (Masina et al., 2020; Pradhan et al., 2018). Because of the broad use of these non-visual interaction systems, smart speaker-based voice assistants such as SIRI, Amazon Echo or Google Home hold great potential for those who are blind to support them (Choi et al., 2020). Furthermore, studies have mostly looked into issues related to speech input for blind people (Azenkot & Lee, 2013; Bouck et al., 2011; Zhong et al., 2014) when using mobile devices for navigation (Helal et al., 2001) and education (Smith & Kelly, 2014).

In our research, we propose the development of a voice assistant specifically designed for visually impaired individuals, aiming to facilitate their remote exploration of virtual museums. We plan to create and assess a system prototype that enables blind users to experience museums independently and interactively using voice assistants such as SIRI, Amazon Echo, and Google Home. This remote control voice assistant permits engagement with physically inaccessible museums. In our study, we implement this system in Google Home, a smart speaker-based voice assistant, to

investigate how techniques like spatialized sound, descriptive co-designed annotations, and interpersonal skills, particularly humor and laughter, can contribute to improving spatial understanding, learning, memorability, engagement, and usability. We aim to develop a museum navigation dialogue tailored for blind users, guiding them through the Basilica Cistern, an inaccessible historical site in Istanbul.

Our prototype places emphasis on enhancing the social learning of visually impaired individuals, fostering their integration into cultural heritage. By facilitating efficient navigation to previously inaccessible places, we aim to contribute to increasing social inclusion. The overall goal is to make the Basilica Cistern fully accessible to blind and visually impaired individuals by employing voice assistant technology and the Wizard of Oz technique. We strive to generate a sustainable model that can be replicated in museums across Turkey, through the integration of interdisciplinary research involving dialogue-based interaction methodologies and co-design techniques oriented towards storytelling.

The purpose of our prototype is not only to provide information but also to be entertaining and facilitate engaging social learning experiences for users, encouraging them to view it as an alternative to physically visiting the museum. Thus, our paper focuses on two interpersonal skills—humor and laughter—in interaction with a remote virtual voice assistant. We evaluated the influence of humor as a social cue on the learning and memorization of historical details by developing two variations of the voice assistant prototype: one with a humorous tone and another without. We selected these social abilities because of their documented positive effects on human social interactions, promoting satisfying and healthy interpersonal relationships. Furthermore, the combination of humor and a remote virtual voice assistant scenario has yet to be thoroughly examined for blind individuals.

4.2. Theoretical Framework

Since their primary engagement mechanisms are non-visual, a significant corpus of research has focused on blind people how to perform tasks and access information

hands-free with voice assistants; data indicate that blind people generally perceive VAs positively. For example, after conducting a content analysis of product reviews and discovered that many people with visual impairments (PWVI) considered VAs as "*a vital part of their lives*" (Pradhan et al., 2020). Pradhan et al. (2020) state that their finding regarding PWVI contradicts the conclusion of a different study, which concluded that sighted users view VAs as entertaining (Luger & Sellen, 2016). Similarly, Abdolrahmani et al. (2018) suggest that PWVI may utilize VAs for fundamentally distinct types of tasks than sighted users, such as work-related tasks. In addition, PWVI may be more receptive to imperfect VAs if they don't have any other way to access certain third-party applications.

The human-like conversational character of voice assistants is seen as "*verbose*" and "*irrelevant*" by blind users who are proficient with screen readers. They prefer to tailor the speech, rate, clarity, and volume of their vocal output to the specific task at hand and personalize their voice assistants (Abdolrahmani et al., 2018). Voice assistants are easily accessible and convenient for blind users, but they are unable to thoroughly engage with content (by reading past the first few phrases of Wikipedia, for example) or provide a comprehensive perspective of the situation (list alternative search results & suggestions). Screen readers, on the other hand, let users immerse themselves in material (when such content is accessible) and offer granular navigation and control, but at the expense of greater complexity and diminished ease of use (Vtyurina et al., 2019). Furthermore, the ability to use voice assistants for a variety of small tasks without needing assistance from a person else, such as listening to music, checking the weather, requesting the time or date, reading books, or listening to the news, was a common theme for people with visual impairments (Pradhan et al., 2018). Since new users are more likely to be task-focused and to communicate in natural language, experts are likely to prefer a command- and menu-based interface. According to another study on SIRI, for blind people, a conversational tone is unnecessary while interacting with voice assistants because their primary purpose is to provide information and productivity is more important for them than privacy in their everyday use of the VA in most cases (Sayago & Ribera, 2020) compared to sighted people.

Following the CASA theory, humans react socially with conversational agents when they display human-like social cues such as language style, name, gender, voice, age, gestures, facial expressions, small talk, social dialogue, or joking (T. W. Bickmore & Cassell, 1999; Moon, 2000; Nass et al., 1997). The social cues elicit different responses in users (Feine et al., 2020). The same verbal content can be expressed in many different ways (Rhee & Choi, 2020). Social cues such as small talk, self-disclosure, expert jargon, empathy, gossip, and politeness expressed in human-to-human conversation to build trust could also be used during conversations with artificial entities to gain the user's trust (Cassell & Bickmore, 2003; Lucas et al., 2018; Sidner et al., 2004). Social cues representing human characteristics, such as visual features, voice, and gender, could elicit the social presence of artificial entities. Individuals apply social rules derived from human-to-human interaction (HHI) to their interaction with artificial entities and behave towards them as social entities (Nass & Moon, 2000). In the context of voice assistants, we will investigate whether blind users perceive and interact with the voice assistant as if it were a social actor or a tool, and expect it to behave in a way that aligns with their own cultural and social norms.

Nevertheless, a technological artefact with its mind can be uneasy to its users. Typically, people see the ability to feel and sense unique to human beings. According to Uncanny Valley Mind Theory (UVM), violation of this norm, such as machines showing the ability to feel/sense, is particularly disturbing compared with those who can think and act (Ciechanowski et al., 2019; Gray & Wegner, 2012). In the context of voice assistants, this may manifest as a sense of unease or discomfort for blind users when interacting with a voice assistant that sounds too "human-like" or "natural," but is not a real person.

The current body of recent social psychology studies shows us that humans like to reduce their social uncertainty during the interpersonal communication process. They tend to seek ways to understand, predict, and occasionally control other people's behaviour (Gefen & Straub, 2004). Berger and Calabrese's (1975) uncertainty reduction theory posits that humans need interlocutors' information to reduce uncertainty about their potential behaviour during interpersonal communication. Similarly, when experiencing discomfort during an interaction with a conversational agent, people search for social cues that they can use to predict these agents' potential

future behaviours as they do during human-to-human communication (Mourey et al., 2017).

The notion of cultural homophily suggests that cultural similarity can drive acceptance and positive attitudes towards conversational agent such as robots. This phenomenon is not binary, but underlines the notion that relative cultural similarity can ease social interactions and shape preferences (Lim et al., 2021; Rau et al., 2010; Spatola et al., 2019; Trovato et al., 2015). Aligning a conversational agent's communication style and humor with the user's cultural context is, therefore, a key consideration in the design process.

According to Hall's cultural context model, high-context cultures rely on implicit communication and shared understanding, whereas low-context cultures primarily depend on explicit communication (Hall, 1976). High-context cultures such as Turkey communicate through nuanced messages, common knowledge, and non-verbal cues, while low-context cultures predominantly employ direct, explicit communication (Fikret Pasa, 2000; Kabasakal & Bodur, 1998). This cultural understanding informed our design of a humorous voice assistant for a remote museum tour in Turkey. The voice assistant integrates cultural references and local idioms, simulates non-verbal cues through vocal inflections and pauses, follows indirect communication styles, and respects common Turkish customs and cultural norms.

Humor, as a high-context tool, can significantly influence a robot's perception and users' experience of the "Uncanny Valley". In high-context cultures, well-executed humor can make a robot seem more familiar, potentially mitigating the Uncanny Valley effect (B. T. C. Tay et al., 2016). However, poorly executed humor can intensify this discomfort. A study on live interactions revealed that high-context cultures responded favorably to robots using their native communication styles, resulting in higher ratings of likeability, engagement, trust, and satisfaction in complex social settings, such as acting as a tour guide (Rau et al., 2010). These findings underscore the value of culturally relevant and sensitive voice assistants.

Building upon the CASA paradigm, Uncertainty Reduction Theory, Uncanny Valley of Mind, and relatedly Hall's Cultural Context Model, our study aims to: 1)

explore how voice assistants are utilized by visually impaired individuals, specifically in the context of remote museum navigation, suggesting avenues for enhancement; 2) present empirical data on visually impaired users' trait attributions, social and cultural norms to voice assistants based on interpersonal skills like humor and laughter; 3) investigate how users perceive voice assistants ontologically, as either tool-like or human-like; and 4) illustrate how co-design techniques, inclusive audio descriptions, and spatialized sound can enhance remote museum navigation experiences and spatial understanding. Our proposed voice assistant prototype for virtual museum navigation aims to improve social inclusion, learning, and universal usability for blind individuals.

It's essential to understand that these cultural tendencies provide a broad framework and don't guarantee the behavior of individuals or groups. Co-designing a humorous voice assistant with blind people, we ensured to respect these factors, aligning with the cultural nuances of humor in different societies. These theories emphasize the social and psychological factors influencing the perception of voice assistants and suggest that interpersonal skills, such as humor, laughter, and auditory feedback, can enhance accessibility and inclusivity in virtual environments like remote museum navigation for visually impaired people.

4.2.1. Humour and Laughter Effect

Humor and laughter have a significant impact on social interactions and can be utilized as powerful tools to achieve various social goals. They are universally experienced and present in all cultures and individuals worldwide (Martin, 2007). Research has shown that innocent humor increases likability, interpersonal attraction, friendship, trust, and social cohesion (Cann et al., 2000; Hampes, 1999; Marone, 2015; Middleton & Moland, 1959; Wilson, 1979). When used in teaching, humor stimulates students' attention, comprehension, retention of information, and creates a positive learning environment (Ziv, 1984). It can also alleviate tensions, conflicts, and encourage socialization (Lefcourt, 2001; Wendt & Berg, 2009)

Humor plays a crucial role in fostering the development and maintenance of social relationships, easing uncertainty in human interactions, and facilitating

communication (Hübler & Bell, 2003). It acts as a "discursive icebreaker" and promotes social cohesion among participants through shared laughter (Hampes, 1999; Hübler & Bell, 2003; Marone, 2015). Humor is mentioned by a study, the authors point out the importance of (mutual) laughter in service encounters and not only humor per se (Gremmler & Gwinner, 2008). Laughter, as a para-verbal signal, plays a vital role in conveying information, such as agreement or disagreement, and measuring engagement in discourse (Scherer et al., 2009). Additionally, laughter acts as a bonding mechanism and regulates social interaction (Vlahovic et al., 2012). In this way, humor is useful in reducing the uncertainty that we feel towards other parties. Furthermore, the higher amount of individual humor, the less uncertainty and social distance generated between two parties (Graham, 1995).

Despite the evident appeal and advantages of humorous interaction, research primarily exploring its facilitation during computer-mediated communication is relatively sparse. Much of the work in this domain has instead been directed towards the computational generation of humor. While the realization of fully automated humorous machines might still be in progress, certain research indicates that integrating pre-programmed humor into human-computer interaction can augment the user experience without causing task distractions (Morkes et al., 1998; M. Zhang et al., 2021). According to these studies, the system was perceived as more agreeable, skillful, and cooperative by participants, who also demonstrated an increased level of social behavior (Morkes et al., 1998, 1999). Additional research involving conversational agents revealed that the subjects found the agent with humor to be more human-like, amusing, and affable (Dybala et al., 2009). Overall, the humorous agent received much more favorable ratings compared to its non-humorous counterpart. Another study implied that the inclusion of humor by a virtual receptionist agent could be pivotal in fostering user engagement in social conversations, as evidenced by the fact that 50% of the social dialogues between users and the agent Marve's canned jokes (Babu et al., 2006).

Throughout history, humor has held a significant role in our societal and cultural lives, seen in various forms of literature, poetry, art, and theater. Therefore, the act of creating humor can be seen as quite a complex endeavor, requiring an understanding of societal norms, cultural conventions, mastery over language subtleties, and an ability to sense the right context for engaging storytelling. Huan and Szafir (2001)

discovered the positive impact of humor in education, where students provided more positive feedback about their instructor, whether human or robotic, if they integrated humor into their interactions. However, this did not appear to influence the students' content memorization. In a similar vein, a subsequent study found that humor could enhance the perceived sociability and personality of a social robot's speaking style, and improve overall task enjoyment (Niculescu et al., 2013). Delivering humor effectively is dependent on timing and appropriateness, presenting one of the biggest challenges, as it requires background knowledge, emotional intelligence, and awareness of context and culture.

Voice assistants such as Siri, Cortana, and Alexa have gained recognition for their humorous responses (Nijholt & Aldo, 2017), but designing voice assistants with a "one-size-fits-all" approach neglects the unique needs, technological adoption challenges, communication preferences, and socio-cultural backgrounds of blind individuals. The concept of what's "funny" is rarely universal and is deeply personal, context- and culture-specific. As a result, many culturally specific jokes, or those involving sarcasm or irony, often go unnoticed, are misunderstood, or could potentially offend (Reimann, 2010). These studies suggest that humor can positively impact a system's usability and proficiency. Moreover, the ability of social robots to recognize and induce laughter could have multiple uses, such as in entertainment and healthcare (Becker-Asano et al., 2009; Cosentino et al., 2013). The capability of conversational agents to exhibit non-verbal and emotional behavior can further immerse the user in a story, more so than traditional mediums like books or audiobooks. For instance, home robots can be utilized to narrate stories for entertainment or educational purposes (Striepe & Lugin, 2017)

There has been a trend in educational contexts to cultivate a more relaxed and enjoyable learning environment. The predominant view is that students are more inclined to learn and remember information when they are happy and entertained, rather than tense and intimidated (Martin, 2007; Oppliger, 2003). Systems that incorporate humor have been integrated into computer interfaces, especially in the realm of "edutainment" (a blend of education and entertainment), to make interactions more enjoyable (Wendt & Berg, 2009). Research shows that humor facilitates memory by encouraging the expansion of information, thereby improving its transfer and

storage in long-term memory (Lippman & Dunn, 2000; Schmidt, 1994). In a separate study, researchers presented a virtual environment system designed as a passive aid to assist individuals who are blind in comprehending and developing cognitive maps (Lahav et al., 2008).

Moreover, the benefits of humor and laughter have been acknowledged in fields like psychotherapy, counseling, education, and business consulting. Their application can enhance the efficacy of therapy, boost student engagement and learning, and improve health, morale, and productivity in the workplace (Schmidt, 1994; Ziv, 1988). However, even though playful and humorous installations are available in public spaces, museums, and workplaces, these projects are not truly embedded in a local community, much less enabling a community to decide to incorporate playful technology into their environment. The current trend in human-computer interaction is humanization, though this is still a topic of debate within the research community. Yet, we propose focusing on the "*humorization*" of human-computer interaction, as we believe it could enhance user experience in terms of acceptance, engagement, and collaboration (Nijholt & Aldo, 2017). Overall, integrating co-designed humorous storytelling and laughter into a voice assistant for blind individuals to remotely visit museums and learning historical facts can produce engaging and enjoyable interactions, thereby promoting social cohesion, social learning, and enhancing overall user experience.

4.2.2. Sonification Effect

Sound serves as an alternative method of communication. Both voice and text are sequential, depend heavily on natural or programmed languages, and usually require the user's undivided attention for understanding. Voice recognition in sound input could enable comprehensive two-way human-machine communication (Blattner et al., 1989). Furthermore, user-focused technology leverages not just text or synthesized speech, but also non-linguistic sounds for information transfer (Gaver, 1986).

The term 'sonification' refers to the use of non-speech audio signals to convey information. Specifically, it involves transforming data relationships into auditory

signals to enhance communication or interpretation (Kramer et al., 1999). Major non-speech sounds used in sonically-enhanced environments are auditory icons, earcons, and spearcons (Y. Yu & Liu, 2010). A key component of the sonification process is the selection of earcons – the sounds representing objects, actions, or events (Blattner et al., 1989). These sounds can range from representational to abstract. Representational earcons or auditory icons (Gaver, 1986), use recognizable sounds from everyday life, such as the sound of running water symbolizing a fountain (Drossos et al., 2015). Conversely, abstract earcons can be completely unrelated sounds, like musical melodies (Blattner et al., 1989). These types of earcons necessitate that the user learns their associations, but they can represent a broader range of concepts, which is particularly useful in culturally diverse settings like museums. Auditory feedback is a critical element in designing accessible interfaces for blind users (Ferati et al., 2011). Despite their importance, non-speech audio and sonification are often overlooked by designers in favor of speech audio (Nees, Michael A.; Walker, 2015). However, there are existing studies that utilize auditory/sonic displays and sonically enhanced interface design to assist visually impaired individuals in navigating virtual environments, improve their spatial understanding, and enhance their educational experiences (Ahmetovic et al., 2022; Drossos et al., 2015; Ferati et al., 2011; Lumbreras & Sanchez, 1999; Y. Yu & Liu, 2010).

We incorporated sound effects and laughter into humorous storytelling in the museum context. We then evaluated the impact of humor on usability and social learning. We proposed an interactive and immersive edutainment approach for blind users, demonstrating that auditory storytelling can evoke spatial environmental images. Our sonification examples include using different auditory feedbacks for different museum sections (Basilica Cistern), differentiating objects passed by the user via distinct *"waterdrop"* sounds, and employing sound cues such as a *"metal bridge with footstep sounds"*, city sounds such as *"call to prayer"* and *"tramway"* sound or a *"crying Medusa"* to guide users through the museum. All sounds were recorded within the museum itself. We also created a non-humorous VA as a control group.

4.3. Methodology

In our research, we developed a prototype for a humor-infused voice assistant specifically designed for visually impaired individuals. The assistant employs clear and co-designed audio descriptions, coupled with auditory feedback, to facilitate storytelling, enhance spatial understanding, and encourage social learning within a museum context. This method enhances spatial understanding and encourages social learning. The objective was to enrich the museum experience and amplify participant engagement (Striepe & Lugin, 2017). The integration of non-anthropomorphic sounds, potentially supplemented with background music, can bolster storytelling within a museum environment (Wyse & Subramanian, 2012). Anthropomorphic agents hold an advantage over human storytellers in their ability to utilize a broader range of communication modalities. While human storytellers can produce a limited range of sound effects, anthropomorphic agents can diversify the storytelling experience with a broader variety of sounds.

Within this theoretical framework, our study aimed to address the following research questions and hypotheses:

R1: How do co-designed inclusive audio descriptions and spatialized sounds affect the spatial understanding of visually impaired individuals as they navigate a museum setting remotely via our voice assistant?

R2: How does a culturally relevant humor-infused voice assistant influence sensory experience and social learning?

R3: How does our voice assistant prototype, enriched with descriptive annotations (sensory, emotional, historical, structural) and humorous storytelling, shape the overall user experience and affect their ontological categorization of voice assistants?

4.3.1. Design Process and Dialog Flow

In the process of co-designing, the individual set to benefit from the designed solution is recognized as an 'authority on their own experience.' They significantly contribute to knowledge development, idea generation, and concept shaping. Researchers facilitate this process by providing this 'authority' with tools to stimulate and articulate ideas (Sanders & Stappers, 2008). To ensure meaningful participation,

researchers not only incorporate individuals into the design process in a way that best supports their capabilities, but they also introduce relevant theoretical domains in a manner manageable for the co-design team. As of now, our understanding of how voice user interfaces (VUIs) are employed by the general population is in its early stages (Fitton et al., 2018). There are limited studies on co-designing voice interactions, with even fewer involving individuals with disabilities (Pradhan et al., 2018b).

Nevertheless, there's an increasing interest in exploring methods to co-design with and for those affected by visual impairments across multiple fields (Branham & Kane, 2015; Metatla et al., 2015, 2016, 2019; Sahib et al., 2013). The design strategy proposed in this paper emphasizes a shift from mere accessibility to inclusion in the creation of technology. This calls for the active participation of visually impaired users and their vision loss instructors as co-designers. The challenge lies in creating impactful assistive tools without their input. Understanding the experiences and needs of those with sensory impairments is vital in any co-design initiative (Hendriks et al., 2015). Previous research underlines the importance of a structured framework for describing visual content to visually impaired individuals. These studies aim to stimulate the individuals' perception and recognition of detailed elements found in various forms of visual art. Key aspects of these descriptions include the subject matter (such as people, environments, activities), the form or structure (including shape, line, and color, where possible), smell, and content, which covers historical context and emotional aspects of the artwork (F. M. Li et al., 2023b).

In our study, we created a voice-assistant prototype aimed at enhancing remote museum navigation through the implementation of inclusive audio descriptions. This prototype was co-developed with visually impaired individuals and their sight loss instructors, who jointly crafted descriptive annotations and designed corresponding dialogue flows. The principal themes in our study spanned a range of sensory experiences and historical contexts, and can be classified as follows:

Sensory Engagement: Descriptions are primarily designed to engage non-visual senses like touch and smell, attempting to communicate the physicality of the art pieces.

Emotional Associations: Narrative elements are intended to evoke specific emotional responses to enhance the engagement and immersive experience of the user.

Historical Context: The annotations convey the historical significance of the visual arts and their associated context, helping users to understand the artwork's origin and cultural relevance.

Form and Structure: This theme addressed with the physical form and structure of the artworks, attempting to convey their size, shape, and other tangible features.

Storytelling: This involves weaving together the other themes into a cohesive narrative, making the descriptions more engaging and relatable for the users.

Interviews were conducted with 13 partially sighted and blind participants, along with 2 instructors for the visually impaired, about the visual arts and historical stories at the selected museum (Basilica Cistern⁷). This was carried out to enhance their ability to mentally conceptualize the exhibit based on audio descriptions. All descriptions in our dialogue flows were evaluated by blind volunteers and their sight-loss instructors, who played active roles in co-creating an accessible museum environment. This collaborative approach aimed not only to establish a proof of concept but also to identify any potential issues with the prototype. The feedback we received from these individuals confirmed the validity of our approach, which was founded on co-design techniques. These invaluable insights further informed the enhancement of virtual remote tours for individuals with visual impairments. Each developmental phase was informed by continuous feedback from visually impaired individuals, aiming to foster spatial learning and understanding within the chosen museum context. We used a prescriptive scenario to create the right environment for humor and laughter. The jokes were collaboratively selected and incorporated into our exploratory study to ensure

⁷ Located in Istanbul's Sultanahmet district, the Basilica Cistern, or Yerebatan Sarnici, is an ancient underground water reservoir. Commissioned by Emperor Justinian I in the 6th century, it was built to supply water to the Great Palace of Constantinople, requiring the labor of around 7000 slaves over 38 years. The cistern's 336 marble columns, recycled from earlier Roman structures, support its roof and exhibit diverse styles and carvings, enhancing its architectural grandeur. Among its captivating features are the enigmatic Medusa heads, whose origins remain a mystery. The cistern's history is marked by periods of neglect, rediscovery, and restoration, notably in the mid-16th century by French scholar Petrus Gyllius. It stands today as a testament to Byzantine engineering prowess and Istanbul's rich cultural heritage. (<https://www.basilicacisterntickets.com/about-basilica-cistern/>)

that the humor was not unfamiliar and to our participants. We ensured that the jokes contained local content to enhance cultural relevancy. Plus, we embedded jokes into the scenario so that they would appear as humorous answers.

In line with the iterative methodology inherent in co-design techniques, users and sight-loss instructors were actively involved throughout all research stages. They worked in conjunction with researchers in a continual cycle of development and refinement. Our co-design-oriented framework consists of 4 phases:

Phase I Identifying Blind Individuals' Needs and Challenges:

Conducting initial and follow-up interviews with blind users to understand their specific challenges and requirements with technology and museum visits. This co-designed method informs the design of our voice assistant, enabling virtual navigation in museums.

Phase II Unveiling Barriers and Solutions:

Engaging visually impaired participants in a co-design project to address the barriers they face when visiting museums and to develop an inclusive design solution. Their preferences guide our prototype's key themes: storytelling, humor, female voice and cultural relevance in the voice assistant's communication.

Phase III Enhancing Conceptualization Through Inclusive Audio Descriptions:

Arranging focus groups involving educators of the visually impaired and the users themselves to co-design meaningful and descriptive annotations (sensory, emotional, historical, structural) and dialog flows. This inclusive approach promotes spatial learning and understanding within the selected museum, and ensures a more encompassing representation of the place.

Phase IV Implementing Dialog Flows:

Utilizing the Wizard of Oz technique for the prototype's dialog flows, enabling the incorporation of human qualities like empathy, humor and self-disclosure into the voice assistant without programming. This allows for testing varied humanized dialog flows using Google Home (smart-speaker based voice assistant) as the testing assistant.

In conditions where high humorous expression was desired, humorous statements were included, such as *"As we inch closer to the exit, our lungs are finally getting a break. The air is transforming, dropping its cave-like cologne of cool dampness for a fresher, open-air feel. It's like the cistern is slowly losing its hold on us. You might notice the dance of the phone-zombies around you - sighted folks wandering aimlessly with arms stretched upwards, phones in hand, begging for a digital lifeline. They're desperate to share their artsy, filtered snaps online for the world to see - "Look Ma, I'm in a Byzantine cistern!" Now, take a moment and reach out to your right. Feels unusual, doesn't it? Not your standard, run-of-the-mill ancient column this time. These objects are cool and polished like the marble, but their shape is a game-changer - more like oversized leaves or gigantic petals straight out of a "Guliver's Travel" scenario. They spring from the ground like stubborn flora that has finally broken through the wintry crust, and their features are remarkable. The edges are as sharp as a piece of gossip, and the detailed lines running through them feel like veins on a leaf or a well-worn roadmap. Running your fingers over these carvings is like decoding Braille written by Mother Nature herself. Among the sea of regular columns, these tactile oddities create a unique spectacle. It's like a walk through a fantasy forest - a forest that prefers touch to sight, where you don't need eyes to see its magic."* Conversely, non-humorous condition, the statements focused primarily on providing more formal information, such as *"As we move to the bottom left corner, you might feel as if there's an open space here. The air is still heavily humid, but there's something distinctly different about this place. It's as if we've wandered into a pocket where time stands still. And right here, within this void, there's a surprise awaiting us. To our left, we encounter two block structures adorned with reliefs of Medusa's head. These blocks are positioned upside down, spaced about two meters apart, placed beneath columns. Historically, Medusa's head served as an amulet against stomach*

and digestive ailments up until the Middle Byzantine period. Examples of these lead medallions have been found during the Saraçhane excavations in Istanbul. Although it might be tempting to attach a symbolic significance to these Medusa head blocks being located within a cistern that stores drinking water, contrary to popular myths, we can safely presume that they were merely placed there to support the columns above and level them with the rest. These blocks were likely originally part of monumental archway keystones, providing passage to the Forum of Constantine."

In light of the primary themes identified by our participants during our exploratory research, the audio responses were formulated with a TTS female voice⁸, incorporating culturally relevant humor that was confirmed by our participants. To augment sense-making, the audio descriptions were developed with the assistance of both instructors and blind users, effectively interweaving sensory, emotional, historical, structural annotations through the medium of storytelling. This approach aimed to enhance both the entertainment value and engagement level of the users. The dialogue flows within our prototype contain extensive information regarding various aspects of the Basilica Cistern. This includes its history, the typology of its marble columns, and the intricacies of water supply lines in Istanbul. Detailed descriptions of sculptures and visual arts found within the Basilica Cistern are also included, along with narratives about the iconic Medusa Head Columns and the Weeping Column. Furthermore, our prototype provides captivating stories about various characters who are integral to the Basilica Cistern's lore. Bearing this in mind, we included detailed descriptions about the smell, textures, shapes, and roughness of the Basilica Cistern, along with educational information such as historical facts and stories, the process of forming sculptures, and the styles and carvings of the columns. Both types of dialogues covered the same conversational topics, stories and historical information, but they differed in terms of the VA's humour level.

Following the detailed explanation of our research methodology, we now present the comprehensive result of our study: the "iMuse Model" - Inclusive Museum User Sensory Experience. This model encapsulates the systematic methodology applied to enhance museum experiences for visually impaired individuals, focusing on areas such as sensory engagement, emotional associations, historical context, form and structure,

⁸ www.voiser.com

and storytelling. The model's development involved a rigorous process of exploration, co-design, prototyping, and iterative refinements. It offers a blueprint for creating inclusive, voice-assisted, remote museum experiences. The iMuse Model also holds the potential for replication or adaptation in different contexts, thereby providing a framework to improve accessibility and engagement in various museum environments that are currently inaccessible or non-inclusive for blind users. Next, we introduce the iMuse Model, explaining its various stages and illustrating the importance of each component in creating an enriching and accessible museum experience for all.

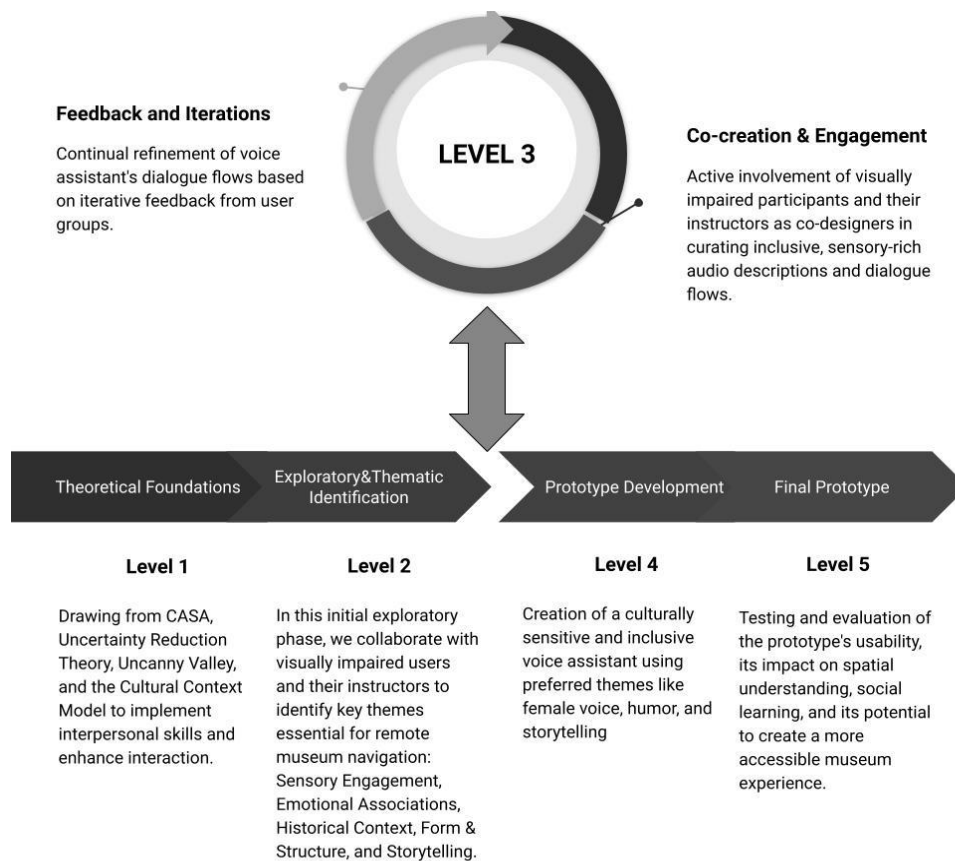


Figure 4.1: The iMuse Model: A Co-design Framework for Inclusive Voice-Assisted Remote Museum Experiences for the Visually Impaired

In our study, participants were given the autonomy to direct the voice assistant's activity at any time during the navigation process. They could guide the voice assistant to move in any direction—forward, backward, left, or right—within the museum setting. Moreover, the voice assistant was programmed to engage in interactive dialogues about the museum's history and to describe sensory impressions relevant to the museum's environment, such as the cool, damp sensation of the marble walls and the smell of dampness or humidity. This feature allows users to actively partake in

their remote exploration and deepens their sensory understanding of the environment being navigated.

4.3.2. Research Procedure and Experimental Set-up of “Wizard of Oz”

Our research utilized a multi-method approach, comprised of five stages: pre-test, pre-visit quiz, momentary test, post-test, and post-visit quiz. The pre-test interviews primarily sought to understand participants' initial impressions of novel technologies, their ontological perspectives, and usage habits of voice assistants. We further explored their strategies in tackling accessibility issues in museum visits or cultural heritage access. The pre-visit quizzes evaluated their preexisting knowledge about the to-be-visited museum.

The testing phase leveraged a Wizard of Oz technique, wherein participants interacted with a conversational agent (CA) prototype under the impression of it being an autonomous system. In actuality, the system's actions were remotely controlled by an experimenter or 'wizard.' (Dahlbäck et al., 1993; Medhi Thies et al., 2017) This technique allowed the experiment to simulate a real-life experience of interaction with a conversational agent, thereby providing an authentic environment to capture participants' reactions and behaviors. The 'living room' setting in which the experiment took place further enhanced the authenticity of the interaction.

In order to evaluate our prototype of the voice assistant, we devised audio responses for a broad spectrum of user comments. A PowerPoint was utilized to create a soundboard, where each individual audio clip was linked to a button that represented a unique response. The 'wizard' in charge would activate the appropriate audio button in response to the input from the user. We made use of Bluetooth to connect our laptop with the Google Home (Assistant) and create a perception that the audio was being emitted through the Google Home (Assistant). Even though the Google Home (Assistant) was always in sight, it was kept on mute, which allowed us to manage our prototype from our laptop during the testing phase.

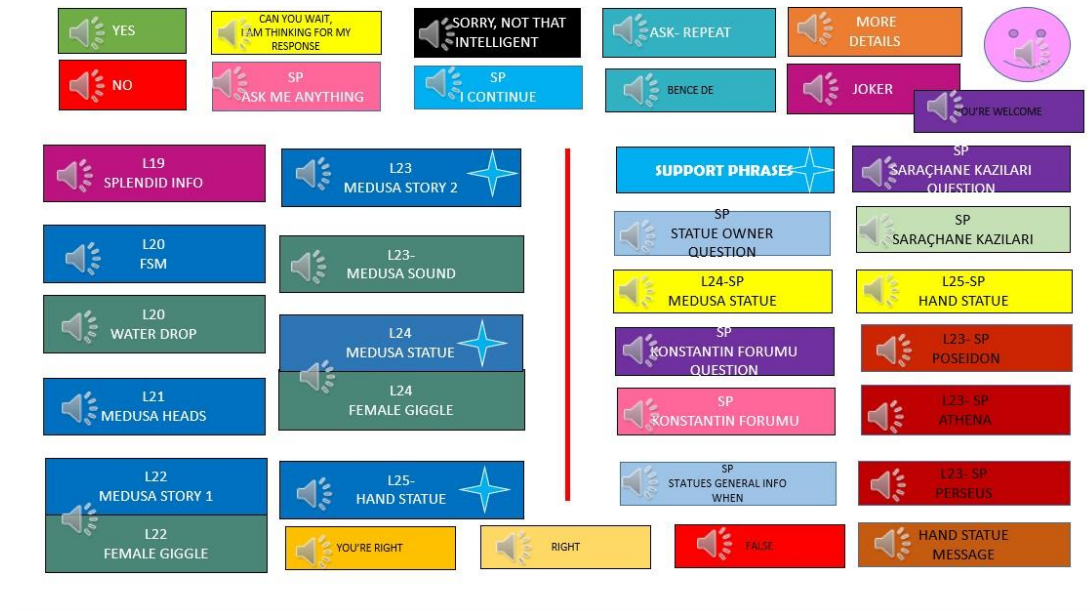


Figure 4.2: One of the Wizard of Oz's Controlled Dialog Flow Panel

The testing phase included careful observation of participants' movements and conversational behaviors, such as whether they used conversational acknowledgments ("uh-huh"), thus providing insights into how they interacted with voice technology.

Post-test interviews centered on examining the user experience in-depth. These interviews helped identify the needs, challenges, and engagement potential of our VA, both in its humorous and non-humorous iterations. After completing the task, the participants were asked to describe the space verbally. We examined the impact of these iterations on cultural literacy and spatial understanding of the museum place. Additionally, we sought to determine if descriptive annotations heighten the users' awareness of the historical significance of the place and if storytelling elements can foster interaction and inclusion. Lastly, a post-visit quiz (see Index II) was conducted consisting of ten questions aimed at quantifying the social learning and spatial understanding outcomes of the participants. We tried to incorporate the factual and interpretive questions in our quiz. This balance encourages both recall of information and deeper processing of the content, allowing to assess different aspects of participants' learning experiences.

We executed a study utilizing a 1x2 between-subjects factorial design. The variables under study were interpersonal characteristics - specifically, the presence or

absence of humor- and their impact on social learning and usability for blind users. To assess usability, we employed the System Usability Scale (SUS).

4.3.3. Participants

In our preliminary and exploratory study, we carried out interviews with 13 visually impaired individuals, including those who are completely blind, aged between 25-45, along with 2 instructors for the visually impaired. After implementing the dialogue flows, we tested our prototype with 30 completely blind individuals aged between 25-40. All participants were either congenitally blind or had lost their vision at a minimum age of five, with at least 20 years of lived experience with blindness. This was to narrow the scope of our study. Also, they were affiliated with the 'Turkey Foundation for the Visually Impaired' and held at least a high school diploma. All participants regularly use devices such as computers, smartphones, or tablets and are frequent users of personal voice assistants like Siri or Google. The study was conducted with the ethical approval from Galatasaray University's Ethics Board. Participants gave informed oral consent for the use of their data prior to their participation. Due to confidentiality concerns linked to the visibility of the participants, visual data sharing is not permitted.

4.4. Results and Discussion

We presented our findings based on a combination of pre-test, momentary, and post-test results. Additionally, we conducted qualitative post-test interviews, quantitative post-visit quizzes, and System Usability test scores to gain a comprehensive understanding of our participants' experiences. Through this analysis, we identified four key themes: Interactive Engagement, Auditory-Enhanced Immersion and Spatial Understanding, Social Learning, and Usability.

Our results indicated that the humorization of the voice assistant played a significant role in enhancing participants' spatial understanding, helping them create cognitive maps, and improving social learning. The inclusion of inclusive audio descriptions and spatial sounds further contributed to their immersive experience, making the environment feel more familiar, even during their first virtual visit. The understandability of the descriptions (sensory, emotional, historical, and structural

annotations) and the effectiveness of the auditory feedbacks greatly influenced their level of engagement and sense-making. Overall, our findings highlight the positive impact of humor, inclusive audio descriptions, and spatial sounds in creating a more interactive, engaging, and accessible museum experience for individuals with visual impairments. These insights can inform future developments in the field of inclusive technology and museum design, with the goal of providing a more inclusive and enriching experience for all visitors.

4.4.1. Pre-test Results

From our pretest results, a prevalent theme emerged indicating that the users appreciate VAs' capacity to execute various task-based commands independently. These include navigation assistance, placing phone calls, and weather updates. Voice assistants were perceived as a tool that supports independence and simplifies life, yet it was regarded merely as an *"adequately intelligent"* device. Notably, users did not show an inclination to perceive the conventional voice assistants as friends or human-like entities. Moreover, they found the voice assistant's sense of humor to be *"overly fabricated"*. It was likened to a *"non-idiomatic translation"*, a term that describes a translation that does not take into account idiomatic expressions, colloquialisms, or cultural nuances, resulting in an unnatural or stilted feel. This comparison suggests that the voice assistant, especially when speaking in Turkish, lacked authenticity or 'soul', implying that such unnatural interaction could impact overall user engagement. This lack of idiomatically accurate and culturally appropriate responses from the voice assistant might affect user immersion in the experience. Consequently, blind users typically categorize conventional voice assistants as tools, viewing them solely in a utilitarian light. For instance, P12 stated:

"I have friends for conversation; why would I bother talking to a machine that can't fully understand me? I try to use the simplest sentences possible to prevent errors."

Participants also expressed a different ontological perception of voice assistants compared to sighted users. As P5 highlighted, *"Voice assistants have been part of our lives for nearly ten years. It's not a novelty or a toy for us, like it may be for sighted people. It's as essential to me as my walking stick."*

A significant challenge that came to light during our pretests was the absence of universally accepted vocabulary for describing visual art. Participants expressed their frustration at the difficulties sighted people often encounter when attempting to convey visual aspects of artworks- such as shape, line, color, or emotion- in a way that resonates with their sensory understanding. This points to a need for a solution that can effectively bridge this gap, enhancing the usability of the system for visually impaired users.

We also administered a pre-visit quiz to assess their prior knowledge about the museum they were to visit remotely. Interestingly, only two participants were aware that the destination was a water cistern located in Istanbul, but they lacked any further information about it. (see Appendix B1)

4.4.2. Momentary Test Results

During the initial testing phase, we noted substantial differences in the participants' interaction styles with the non-humorous versus humorous voice assistants. The non-humorous VA elicited frequent interruptions from the participants, displaying a lack of adherence to typical conversational norms. This behavior could reflect the participants' perceptive limitations of the non-humorous VA as a mere tool, hindering interactive engagement.

On the other hand, participants engaging with the humorous VA demonstrated behaviors reflective of natural, human-to-human conversation, characterized by backchannel responses such as "it was a good joke," "that's interesting," "thank you," "you're cute," and "I'm sorry". This shift in communication patterns displayed an auditory-enhanced immersion, wherein participants not only acknowledged the VA's humorous prompts but also responded in kind, contributing to an enriched interactive dialogue.

The humorous VA's playful nature encouraged light-hearted banter among participants, fostering a sense of social learning through shared laughter and increased inquiry. Participants asked more questions, consequently extending the average interaction time from a typical 35 minutes to 55 minutes. This behavior suggests that

humor could be a significant factor in enhancing usability and sustaining engagement in a voice assistant interface.

The observed shift in turn-taking behavior and increase in backchannel responses with the humorous VA underlines the potential for a more dynamic and natural interaction, transforming the VA experience from a robotic tool to an enjoyable, social interaction facilitator.

4.4.3. Post-Test Results

4.4.3.1. Interactive Engagement

Participants expressed satisfaction with the functional and enjoyable aspects of both voice assistants. However, their engagement with the humorous voice assistant was notably higher, resulting in longer conversations and a sense of relaxation after the test. P13 commented,

"I could have continued interacting with it for another hour, visiting more museums that I haven't had the opportunity to visit and ask more questions. I was expecting a sassy response similar to SIRI, but the assistant responded with authentic humor and wisdom."

Every participant conveyed that the immersive descriptions and narrative storytelling incited a range of emotions within them. For instance, P2 stated,

"...the narrative around Medusa's story drew me in. I found myself wanting to delve deeper into her relationship with Athena, prompting me to ask a series of questions. When the voice assistant painted a vivid picture of Medusa's inverted columns and her sculpture, it made me feel the allure of the sculpture as if I were standing right there in the cistern, perhaps an even more profound experience than actually being there."

The majority of our participants (n=20) articulated a sense of disappointment with the scarcity of thorough explanations concerning visual artworks catered to blind people and surprised when they could relate the descriptions that our voice assistant offered: One participant, P10, recollected an incident:

“...A friend took me to Topkapı Palace once. I was grateful for the gesture, but initially, I struggled to grasp the essence of the exhibits due to her simplistic descriptions...Contrastingly, the voice assistant, with its vivid, detail-oriented descriptions, enriched with historical facts, provided a much more engaging experience...”

Furthermore, participants also found existing museum audio descriptions insufficient and poorly designed for their needs. Current systems often require additional efforts, like scanning QR codes or precise positioning in front of the artwork, which can be challenging for visually impaired visitors. Furthermore, participants expressed concern about the inadequacy of conventional audio descriptions and the limitation on tactile experiences in museums. They highlighted challenges with existing audio description technologies, such as the requirement of scanning QR codes or maintaining specific positions in front of visual artworks, which can pose difficulties for visually impaired individuals. Existing audio descriptions, designed primarily with sighted individuals in mind, lack the intricate details required for visually impaired individuals to fully comprehend the artwork. As P7 commented:

“...The audio guides from art galleries are clearly meant for sighted individuals. They often mention elements like houses and trees, but omit crucial details like their location, colors, and inter-relationships. Plus, I can't inquire further or seek clarification...”

For instance, participant P20 stated:

“Finding audio descriptions in Turkey's museums is like looking for a needle in a haystack, and there's no way for us, the blind, to authenticate the information we receive. If the content doesn't resonate, we're left in the dark, grappling with elusive comprehension. But with this voice assistant, it's like having a warm, interactive chat, where I could freely ask and re-ask, making the experience far richer.”

P2 expressed difficulty interacting during guided tours and underscored the importance of having an interactive voice assistant:

"In the crowd, I often wrestle with a worry that I'll exhaust the patience of the guide or my sighted peers with my questions. It feels like a burden lifted with this voice assistant, who seems to revel in our dialogue, showing no signs of tiring or frustration." Similarly P4 stated : "It felt like I was just hanging out with a witty friend rather than being guided by some uptight tour guide. So, I felt comfortable and probably chatted a bit more than I usually would."

Our inclusive audio descriptions mitigated these issues, providing the opportunity for an engaging interactive experience. As illustrated by P22: *"It was like touring with a friend, a friend who understands how to describe visual arts to a visually impaired person and is open to answering numerous questions for enhanced understanding."* Our voice assistant prototype, equipped with inclusive audio descriptions, fostered genuine and interactive conversations among participants and made them feel understood. Adding a touch of humor to its communication style not only made participants feel at ease but also encouraged them to ask more questions and share personal details. They even discussed accessibility challenges they had previously faced during museum visits in Turkey. As we highlighted in our pre-test results, our participants initially viewed the voice assistant as a simple tool for convenience, rather than a conversational partner. However, infusing the assistant's communication with humor and inclusivity led them to attribute human-like characteristics to it, even referring to it as a *'witty friend'*."

The participants expressed their appreciation for the cultural relevance of the jokes incorporated into the voice assistant. They specifically mentioned that these jokes were far more enjoyable and engaging compared to the canned jokes commonly heard from voice assistants like Siri. This feedback suggests that the humorization efforts resonated well with the participants and added an extra layer of enjoyment and personal connection to their experience. By incorporating culturally relevant jokes, the voice assistant was able to create a more inclusive and relatable environment for the participants. This highlights the importance of considering cultural factors in designing voice assistants and other interactive technologies to ensure a meaningful and enjoyable user experience. It also highlights the potential for voice assistants to go beyond generic and impersonal interactions, creating a more personalized and

enjoyable experience for users. Through the creation of a culturally relevant, humor-infused voice assistant prototype that utilized understandable audio descriptions, we successfully broadened the applicability of the CASA (Computers Are Social Actors) paradigm for visually impaired individuals. This intervention transformed their perception of the voice assistant from being merely a utility tool into a conversational companion, challenging their initial view of the assistant as *"just a make-life-easy tool, not a friend to talk to"*. Contrary to the "Uncanny Valley of Mind" phenomenon, the participants displayed no feelings of eeriness when interacting with our voice assistant. In fact, they were more open and receptive to the interaction.

4.4.3.2. Auditory-Enhanced Immersion and Spatial Understanding

The use of auditory feedback, such as the sound of water droplets, footsteps on stairs, metal coins falling into water, buzzing, indistinct chatter, clapping, and Ottoman military band sounds, Medusa's singing sound proved essential in our study for enhancing the sense-making and spatial understanding of participants while they navigated the museum remotely. Each sound was carefully recorded at the actual site, and played back in its correct context.

Feedback from all participants confirmed the effectiveness of these auditory cues. They helped convey distance and directional information, providing our participants with a richer and more immersive experience. For example, P29 noted,

"As we traversed the cistern, the echoes and the sound of water dripping intensified, leading me to realize that we were in a deeper section of the cistern." Similarly, P11 mentioned, "I managed to mentally sketch out the layout of the Basilica, we followed a rectangular path, right? And the entrance and the exit were positioned opposite each other, weren't they?"

These reflections from the participants underscore the successful use of spatial sounds in our study. They significantly contributed to the participants' ability to construct a cognitive spatial map and learn the layout of an unfamiliar, complex environment, despite their physical separation from the actual location. This is an encouraging outcome, indicating the potential benefits of incorporating detailed

auditory feedback in designing accessible and engaging remote navigation experiences for visually impaired individuals. The auditory feedback, which adapted and changed as the user navigated the space, offered an intuitive sense of location.

In addition, the museum hosts a traditional practice where visitors toss coins into the water to make wishes. Our voice assistant facilitated this practice for our participants by simulating the sound of a coin splashing into water. One participant, P11, noted,

"I was aware that there was no actual coin being thrown for me. It may be considered childish to toss a virtual coin, but I appreciated being part of the experience. A wish is a wish, and who knows, perhaps a virtual one carries more power (laughs)."

Similarly, P4 expressed,

"I felt a twinge of fear, perhaps due to the cistern's tales relayed by the voice assistant, or maybe the intense echo of footsteps and dripping water. I couldn't have possibly experienced this level of immersion while relying on human guidance, and I wouldn't have learned these stories. Since the Medusa columns are located at the deepest part of the cistern, I even joked that Athena might also punish me because I'm too beautiful (laughs)."

These responses suggest that blending historical narratives of the cistern with auditory feedback enabled participants to feel more immersed in the experience, or in their own words, to feel like they were *"joining the game."*

Furthermore, all participants agreed that even if the physical location was easily accessible, engaging with a virtual tour to learn about cultural heritage was extremely comfortable and preferable. Participant 15 remarked,

"Even if I had someone to assist me to the museum, there wouldn't be any comprehensive audio descriptions to aid my understanding of the place, nor a guide to answer my questions. And even if I did find such a guide, I fear I might bore them, or I might fail to comprehend their explanations. This voice assistant,

while it can't replace the tactile experience, it describes the sensation of touch very well. Also, in Turkey's museums, touching exhibits is often not allowed, and tactile options are lacking. So, I would prefer to stay home, converse with this assistant, and experience a sense of independence."

The majority of participants emphasized the comfort, independence, and lack of dependency they felt while interacting with our voice assistant prototype. In line with the Uncertainty Reduction Theory, the inclusion of humor significantly contributed to making our voice assistant feel more familiar to the users. The use of comprehensive audio descriptions and spatialized sounds further enhanced the sense-making process, creating a feeling of being understood.

4.4.3.3. Social Learning

The remote exploration of museums, enabled by inclusive audio descriptions and immersive auditory feedback, assisted participants in forming cognitive maps and understanding the spatial layout of the environment. Our study also aimed to ascertain whether humorous storytelling could provide additional benefits for social learning. To quantitatively analyze our data, in addition to the open-response interview data, we conducted a post-visit test consisting of ten questions (See Appendix B2).

Participants who engaged with a humorous storyteller voice assistant demonstrated a greater degree of learning transfer than those who interacted with a non-humorous counterpart, even though the audio descriptions and auditory feedback were identical in both instances. The former group showed a more profound understanding of the contextual and memorable details, such as the history behind the Weeping Columns, the identity of the cistern's constructor, the names of the sculptures, and the story of Medusa. For instance, P1 noted,

"Back in high school, we had a history teacher who would deliver continuous, emotionless lectures for 45 minutes. Given that we rely on voice for emotional nuance, I barely retained anything from those classes. But now, I remember that the cistern was discovered by Gyllius during Sultan Suleiman the Magnificent's reign. This particular detail has stayed with me because the assistant played the

Ottoman military band song right after revealing this fact, and even gave a chuckle. It was unexpected, and hence, memorable."

Similar sentiments were echoed by P4, who expressed, *"I was truly delighted to learn these intriguing narratives that I can now share with others - what the voice assistant referred to as 'a killer story to tell'. But the story that really piqued my interest was that of Fatih Sultan Mehmet using the cistern as a secret assembly point for his spies before the conquest of Istanbul. It felt like I was right in the middle of an action-packed movie."* It's noteworthy how the participant was able to recall the exact humorous phrasing employed by the voice assistant, which indicates a correlation between the use of humor and retention of the historical context.

This link is further reinforced by P20's reflection,

"One part of my visit that I remember vividly is the story of Mitoş. The voice assistant described his unfortunate tale, where he found a tunnel leading to the harem and ended up imprisoned for life, in quite a humorous tone. At the same time, it detailed a sculpture of a woman with a figure resembling an hourglass. The light-hearted portrayal of Mitoş's predicament, despite its gravity, was memorable and even led me to empathize with him as a man (laughs)."

These evidences suggested that incorporating humor into the voice assistant's storytelling could potentially heighten social learning, likely by rendering the content more engaging and easier to remember.

To assess social learning, we developed custom pre-visit and post-visit quizzes with the input and guidance of sight loss instructors. This ensured that the quizzes were appropriate and relevant for visually impaired users and tailored specifically to the content of the Basilica Cistern. These quizzes were designed to measure the participants' absorption and retention of the cultural, historical, and practical information provided during the virtual tour. The effectiveness of this assessment tool was validated by comparing the participants' performance on the quizzes before and

after the tour. Analysis of the results revealed that on average, participants who engaged with a humorous voice assistant correctly answered 8 out of the 10 questions, achieving 80 points. In contrast, those who interacted with a non-humorous version managed an average of only 5 correct responses, achieving 50 points.

Our statistical analysis reveals compelling evidence of the impact humor can have on learning in the context of using voice assistants. A t-Test comparison of the average quiz scores between the humorous voice assistant group and the non-humorous group demonstrated that the use of humor significantly enhanced learning outcomes ($t=9.27$, $p<0.0001$). However, this result should not be viewed in isolation.

As clearly represented in our box and violin plots, both groups exhibited a wide range of scores. Despite lower average scores than their counterparts interacting with the humorous voice assistant, the non-humorous group also demonstrated noteworthy learning outcomes. These respectable scores affirm that the co-designed descriptive annotations and the sense-making auditory environment facilitated spatial understanding and fostered social learning, irrespective of the use of humor.

The non-humorous voice assistant, therefore, offered a considerable degree of learning enhancement. Nevertheless, the superior performance of the humorous voice assistant group underscores the added value humor brings to this educational context. This group achieved higher average scores, and the overall distribution of scores was skewed towards the upper end of the scale. This indicates that humor did not merely impact a few individuals but had a broad effect on the group as a whole, reinforcing learning for a majority of participants. The enhancement in learning through humor may be attributed to several factors. Humor can make the learning experience more enjoyable, thus increasing engagement. It is well-documented that higher engagement often leads to better understanding and retention of information. Humor may also stimulate learners' interest and curiosity, fostering a conducive environment for exploratory learning.

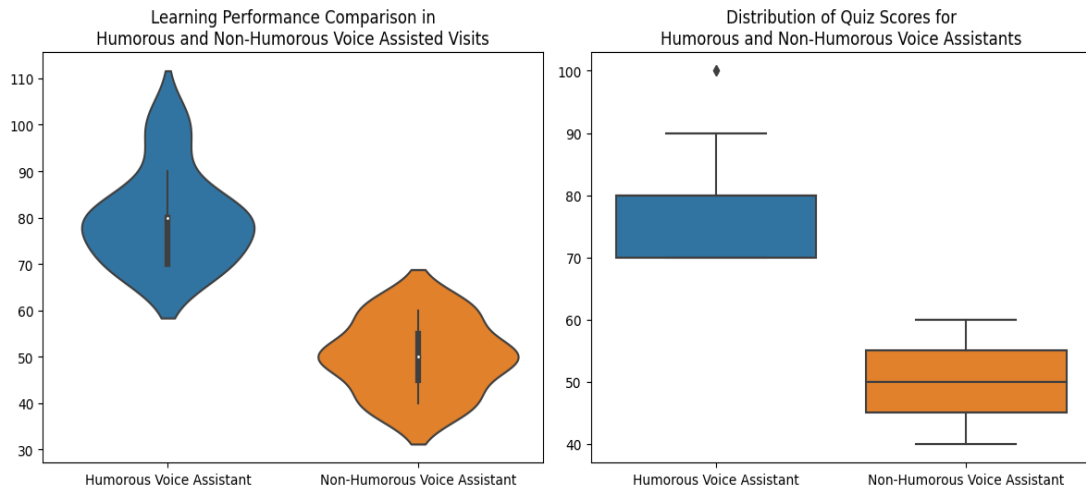


Figure 4.3. Comparison of Quiz Scores for Humorous and Non-Humorous Voice Assistants

In conclusion, while both versions of the voice assistant significantly supported learning, our study suggests that incorporating humor can further enhance this process. Future research should explore the underlying mechanisms that enable humor to boost learning and investigate whether similar effects can be achieved across diverse educational settings and for different types of content.

4.4.3.4. Usability

We administered a System Usability Scale (SUS) to assess participants' usability scores using a 5-point Likert scale. The results revealed a significant difference between the humorous and non-humorous groups, as indicated by the Mann-Whitney U test. The calculated statistic value was 200.000, with a p-value of 0.000. The p-value, being less than the conventional significance level of 0.05, provides strong evidence to reject the null hypothesis. This suggests that there is a statistically significant difference between the two groups in terms of their scores.

Furthermore, when considering the medians, the humorous voice assistant group has a median SUS score of 92.50, while the non-humorous voice assistant group has a median SUS score of 77.50. This implies that, on average, participants rated the humorous assistant higher in terms of usability compared to the non-humorous assistant.

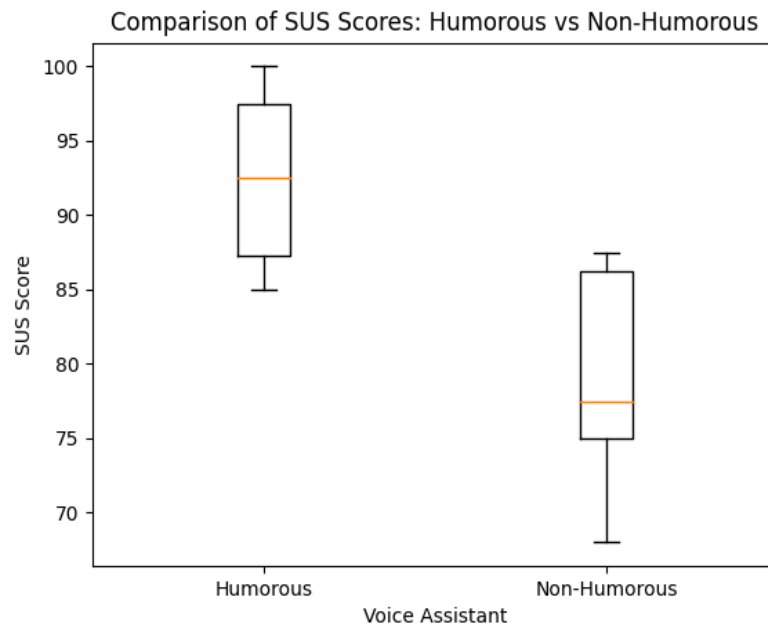


Figure 4.4: Comparison of SUS Scores

The results indicate that participants who interacted with the non-humorous voice assistant prototype had a lower score for the ninth item compared to the benchmark score. This item assessed the participants' perception of feeling safe while using the system. This finding is not surprising, as participants who interacted with the humorous voice assistant reported feeling relaxed. It suggests that the absence of humor in the non-humorous voice assistant made participants feel less safe and more tool-like.

Analyzing the individual SUS scores further, we observed that the lowest scores were reported by participants who found the non-humorous voice assistant too serious and struggled to adapt to the information provided by the assistant. For example, P25, who considered the non-humorous assistant to be unwise and boring, assigned a SUS score of 70 to the prototype. P24, who believed they would require technical support to use the prototype, assigned an even lower SUS score of 52.5, despite both versions having the same capabilities

Overall, the incorporation of humor had a positive effect on participants' comfort level and their perception of the voice assistant's competence and human-like qualities. It highlights the importance of humor in making participants feel more at ease and engaged with the voice assistant. Although the non-humorous voice assistant received

lower evaluations compared to the humorous counterpart, the overall usability scores were still sufficiently high, exceeding the threshold of 68, which is considered an acceptable level of usability. Importantly, even without the humorous aspect, the non-humorous version demonstrated potential, as its descriptive audio descriptions and auditory feedbacks were appreciated by the users, further enhancing the inclusivity of the museum experience.

4.5. Conclusion

This study was primarily geared towards supporting blind users in formulating conceptual and contextual models of real-world scenarios. We specifically focused on enabling these users to explore inaccessible museum spaces, a step towards inclusive cultural heritage sharing. A voice assistant prototype was inclusively designed and co-developed with blind individuals and their instructors, aiming to enhance their spatial understanding. The goal was to establish a framework that could be replicated across diverse cultural environments, ultimately mitigating their social exclusion. Our approach to creating descriptive annotations, focusing on sensory, emotional, historical, and structural details, provides a guiding example for designers aiming to develop inclusive experiences for visually impaired individuals.

In this study, we collaborated with visually impaired users and their instructors to co-design and develop a specific framework. This framework was used to create dialogue flows which were then assessed by our collaborators. Subsequently, our prototype was put to the test with an additional 30 completely blind users, employing the Wizard of Oz technique. Our evaluation of their perceptions, attitudes, and behaviors towards both our humorous and non-humorous voice assistant encompassed a multi-method approach. This involved qualitative measures such as pretest and post-test interviews and momentary test observations. We also incorporated quantitative measures like a pre-visit and post-visit quiz to assess social learning, along with the System Usability Scale (SUS) to gauge usability.

The introduction of remote virtual museum navigation not only facilitates social inclusion and cultural literacy but also bolsters spatial understanding among visually impaired individuals. Our system incorporates an interactive dialogue flow that

diminishes feelings of isolation. It includes descriptive annotations which heighten awareness about historical significance of statues and columns, fostering cultural literacy. Furthermore, storytelling elements spotlight historical and cultural facets of the city, promoting interaction and inclusion. Our remote virtual museum navigation system is not just an innovative solution; it's a pathway to social cohesion and cultural inclusion for visually impaired individuals. It serves as a guide to help people with disabilities fully participate in social and leisure experiences. This research involved various methods, including the Wizard of Oz technique, which has proven its efficacy. This research model has shown its cost-effectiveness by eliminating the need for coding AI-based solutions and allowing for rapid design iterations. It enables rich interactions with vulnerable user groups, with no risk of software errors that might jeopardize results.

Based on our results, we were able to categorize the participants' remote museum navigation experiences into four central themes: Interactive Engagement, Auditory-Enhanced Immersion and Spatial Understanding, Social Learning, and Usability. These pivotal themes underscore the multifaceted nature of the users' experiences with the voice assistant in the museum environment. Through Interactive Engagement, we observed how the users actively interacted with the system. The Auditory-Enhanced Immersion and Spatial Understanding theme underscored the impact of audio descriptions on users' ability to mentally map and understand the unfamiliar museum environment. The theme of Social Learning revealed the learning process the users underwent while interacting within the virtual environment. Lastly, Usability pointed to the ease-of-use and intuitive nature of the voice assistant, contributing significantly to the overall user experience. These themes collectively provide a comprehensive view of the experiences of visually impaired users in navigating an unfamiliar space through a co-designed, humor-infused voice assistant.

One of the most significant outcomes of our study was the evident improvement in visually impaired users' ability to construct detailed cognitive maps of previously unexplored spaces, thanks to their interactions with the humor-enhanced voice assistant. The humorous element added a human-like quality, causing users to perceive the assistant as more than just a tool. Notably, our culturally relevant incorporation of humor into the voice assistant's dialogues fostered a human-like persona. This led users

to view it as more than just a tool, similar to conventional voice assistants such as SIRI or Google Assistant. Instead, they began to regard it as an ontologically relatable and competent guide. These findings were further reinforced by the increased engagement and social learning observed among users within the virtual environment. This was evidenced by markedly higher scores in our post-visit quiz. Importantly, when assessed via the System Usability Scale (SUS), participants interacting with the humorous voice assistant yielded significantly higher scores compared to those using the non-humorous counterpart, underscoring a preference for the former.

In essence, our study underscores the potent potential of co-designed voice assistants in facilitating visually impaired individuals' navigation of unfamiliar environments. Beyond just aiding in accurate spatial understanding, the humor element made the learning experience more immersive and enjoyable, transforming the users' perception of the assistant from a mere tool into a human-like companion.

While our study has provided insightful findings, it is not without its limitations. Primarily, the research was conducted with a relatively modest sample size of 30 blind participants. While this number allowed for an in-depth exploration, it may not wholly represent the diverse experiences of the broader population of visually impaired individuals. As a result, the generalizability of our findings to the larger visually impaired community should be approached with caution. Furthermore, our primary measures of success, the System Usability Scale (SUS) and quiz scores, while quantitatively robust, might not fully encapsulate the intricacies of user experiences. To partially mitigate this, we incorporated qualitative evaluations through comprehensive interviews. This approach allowed us to gather detailed insights into the participants' experiences, providing a depth of understanding that might not be fully captured by the SUS or quiz scores alone. Nonetheless, our mixed-method approach does not eliminate the necessity for future research using varied evaluation metrics, specifically to better understand the long-term effects and implications of using our humor-enhanced voice assistant. Finally, our study did not include long-term follow-up with the participants. Consequently, we are limited in our understanding of the longevity of the observed positive effects, such as enhanced social learning and improved spatial comprehension, when interacting with the humorous voice assistant. These limitations warrant consideration when interpreting our study's results.

and provide valuable direction for the future research in this area, focusing on enhanced user inclusivity and accessibility.

5. CONCLUSION

This research aims to contribute to the field of human-computer interaction (HCI) and social psychology by exploring the under-researched area of voice-assistant user populations, with a particular focus on older adults, adults with specialised interests, and visually impaired individuals.

5.1. Restatement of Research Problem and Objectives

Given the wide range of differences in how individuals from various cultural, gender, age, and ability groups interact with computational environments, a crucial challenge lies in making conversational agents more adaptable and customisable. This study seeks to tackle this particular difficulty in two distinct ways. First, it explores the effects of interpersonal skills and voice characteristics of voice assistants on diverse user groups. Second, it examines how these user populations categorise and interact with voice assistants tailored to their requirements. We embraced co-design techniques and iterative design/validation processes to ensure our voice assistant prototypes were user-centred.

5.2. Summary of Findings

Chapter Two (Leveraging Shared Language in Voice Assistants: Enhancing Reciprocity for Adults with Specialized Interests) sheds light on the role of conversational agents' interpersonal abilities, notably self-disclosure and gender, in fostering dialogue and trust between voice assistants and adult users with specialised interests. The collected data supports the idea that users are more likely to reciprocally share their interests and personal information with a voice assistant that can effectively disclose information about itself. It also highlights the complex issue of gender representation among these agents and reveals the implicit gender stereotypes attributed to them. By defaulting voice assistants to binary gender options, we risk

reinforcing gender biases and stereotypes. It is clear that a shift towards more inclusive design approaches, such as using gender-neutral voices, is necessary to create a more balanced, equitable and inclusive interaction landscape. Moreover, in this chapter, we utilised a collaborative hypertext dictionary in Turkish (Sour Dictionary- Ekşi Sözlük) as background knowledge to grasp their conversational style better and prolong the interactions, promoting trust and triggering reciprocity. This approach underlines the importance of contextual and cultural relevance in designing computational agents, shedding light on how to create more engaging and relatable user experiences.

Chapter Three (Listening to Cultural Reflexes: Tailoring Empathetic Voice Assistants for Older Adult Users) of the study highlighted the significance of cultural dimensions and voice preferences in developing voice assistants tailored for older persons, challenging traditional assumptions about voice age and interaction style. It debunks the assumption that an "older" voice is preferred according to the Social Identification and Similarity-Attraction Theory, instead highlighting the significance of empathy and a younger voice in the voice assistant's responses. This approach provides crucial insights into tailoring conversational agents to match older users' emotional needs, communication styles and cultural reflexes. Furthermore, our research indicates that cultural and societal backgrounds influence older adults' expectations of these voice assistants. The desire for a voice assistant to act as a "trusted confidant" underlines the value of these technologies in providing support, and companionship, and boosting self-efficacy, particularly for those who might be elderly and socially isolated. Such findings emphasise the need for voice assistants to be designed with contextually aware, empathetic responses rather than simply task-based commands.

Chapter Four (Inclusive Voices: Crafting Voice Assistant Prototypes for the Blind Users to Navigate Cultural Experiences) highlights the breakthroughs from co-creating a voice assistant framework designed specifically for visually impaired users. Our co-development process culminated in creating a prototype that transcends the conventional role of a tool and begins to operate as an accessible, intuitive guide. This system assists users in constructing cognitive maps of real-world environments while also promoting a sense of connection. Significantly, the incorporation of humour into the voice assistant's interactions resulted in a notable transformation, elevating it from

a conventional utilitarian tool to a friend that is more relatable and reminiscent of human-like qualities. Incorporating humour and inclusive dialogue flows with descriptive annotations enhanced the experience's immersion level, promoted learning, social inclusion and deepened the user's engagement with the environment, highlighting the capacity of customising voice assistants to cultivate engaging, immersive, and enjoyable user interactions. This strategy emphasises the potential of computational agents to act as accessibility tools, making cultural and educational content more accessible to users with visual impairments.

5.3. Contribution to the Field

Conversational agents have transformative potential that extends beyond the sphere of technology and human-computer interaction (HCI). These systems can have a profound societal impact, particularly in narrowing digital divides and enhancing social inclusion. Recognising and addressing the diverse ways individuals, influenced by cultural, gender, age, and ability-related factors, interact with conversational agents is critical. These variations shape not only how users access and utilise information but also how they interpret and evaluate it. As such, these agents can reproduce biases and stereotypes. Tailoring communication methods to diverse user populations can help alleviate these concerns, enhancing the user experience and broadening access to vital information.

This research has significant implications for academics and practitioners alike. It proposes an integrative theoretical framework that illuminates the intentions and needs of different user populations. In essence, we are studying how different research methodologies can gauge the impact of voice assistants on specific user groups like adults, the elderly or visually impaired individuals. For instance, for older adults who may struggle with digital literacy, tailored conversational agents can simplify their access to essential services like healthcare, ensuring they receive timely and appropriate care. Similarly, these agents can transform their digital world experience for visually impaired individuals by providing an effective and efficient means of accessing information and services, thereby enhancing their independence and quality of life. Furthermore, our voice assistant is designed to sustain meaningful and

prolonged interactions for the adult demographic with specialised interests by delivering personalised content such as movie recommendations.

This study endeavours to understand and enhance the interaction between these user groups and computational agents. This broad-based, interdisciplinary approach sets the study apart and contributes to a richer understanding of the interface between technology and society. We propose a holistic and inclusive framework for voice assistant designs. This framework integrates interdisciplinary theoretical foundations with a diverse set of user-centred design methodologies. It is especially replicable when working with vulnerable groups and creating inclusive designs.

First, we will present our framework in a general manner using the following diagram. After that, we will delve into the details, highlighting our contributions to both theoretical and methodological implications:

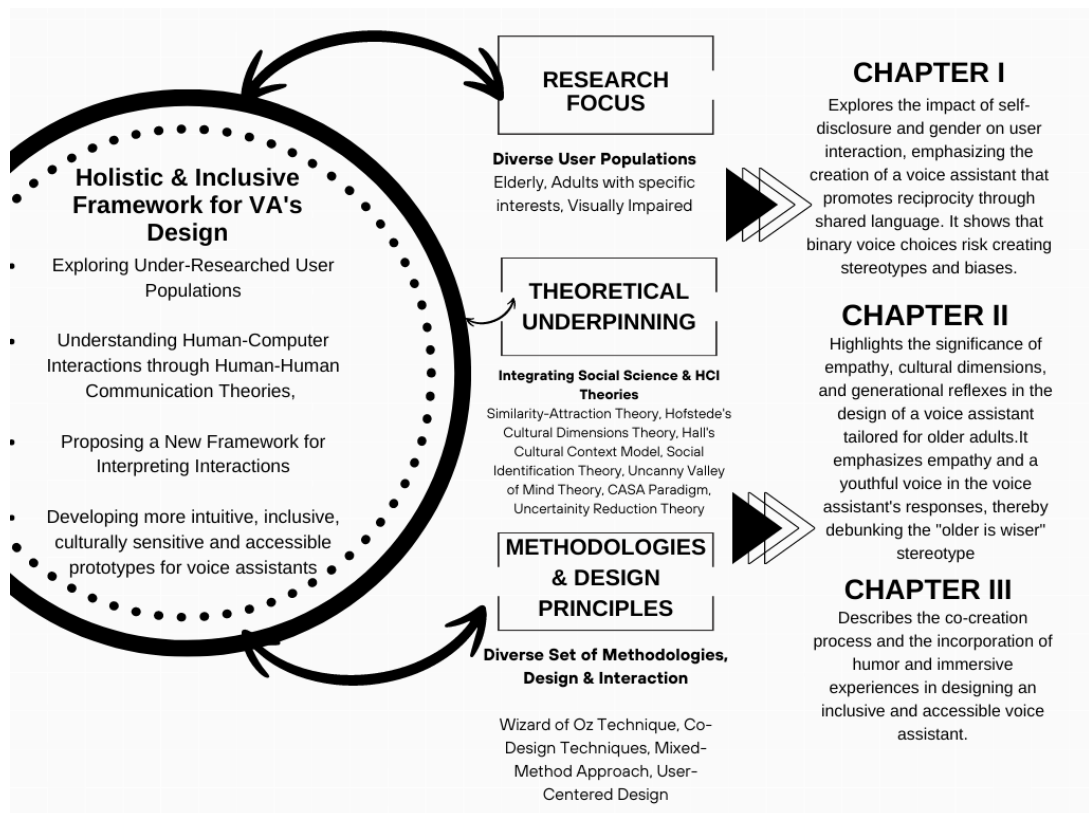


Figure 5.1: Holistic and Inclusive Framework for Voice Assistant Design

5.3.1. Theoretical Contributions and Implications

Throughout this research, we have probed into the nuances of human-voice assistant interaction across different user groups: adult users with specific needs, older

adults, and visually impaired individuals. Our work has produced valuable insights into how customised, empathetic, and contextually aware designs can enhance user interaction with these social machines, transforming them from mere tools to trusted companions and guides. Implementing such design principles, grounded in human communication and social psychology theories, resulted in prolonged conversations, a sense of human-like companionship, and an immersive interactive experience.

This research also contributes significantly to the field of human-computer interaction by providing insights into crucial theoretical and practical aspects that have the potential to improve the design and functionality of voice assistants. The practical ramifications of this research also encompass enhancing inclusivity and accessibility in public venues, such as museums, by using customised voice assistants catering to various user groups' requirements. In this way, our research enhances theoretical understanding and offers real-world solutions to improve human-computer interactions, social inclusion, and access to cultural heritage.

Firstly, the application of Hofstede's Cultural Dimension Theory and Hall's Cultural Context Model in the design of voice assistants highlights the importance of cultural nuances in human-computer interaction. This opens up a fruitful avenue for further exploration in the field, where culture could play a more prominent role in the design and usability of technology. Through our research, we demonstrate that understanding and incorporating the cultural dimensions of users into the design process can result in more customised, effective, and user-friendly interactions with artificial entities.

Second, our research extends the CASA (Computers are Social Actors) paradigm. Specifically, in the study involving visually impaired users, we demonstrate that with careful design, these users begin to perceive voice assistants as human-like rather than as tools. By incorporating interpersonal skills tailored to their preferences and involving them in the design process, we could facilitate more natural and engaging interactions and generate more inclusive dialogues. This extension of the CASA paradigm presents a significant opportunity to rethink how we view and design human-machine interactions, especially for vulnerable populations.

Thirdly, our study also highlights a noteworthy observation about using Social Identification Theory and Similarity Attraction Theory in the context of interactions between individuals and voice-based virtual assistants. Our study found that these theories, which posit that individuals are more likely to establish connections with others who are similar to them, were not supported in the specific context of older adults' interactions with voice assistants. Contrary to the expectation that older adults would prefer a voice assistant with a mature voice, reflecting similarity, our research showed the opposite. Older adults showed a preference for an empathetic voice assistant with a younger voice, challenging these well-established theories. This suggests that the factors influencing the preference and acceptance of voice assistants among older adults are more complex and nuanced than initially assumed. As mentioned above, the findings underscore the significant impact of cultural and generational factors on older adults, providing vital insights for enhancing the design of voice assistants to meet this specific group's needs effectively. It highlights the necessity of considering factors beyond age or resemblance when designing voice assistants for older individuals. Other elements, including empathy, supportiveness and trustworthiness, were discovered to be important.

Lastly, our study emphasises the need to avoid designs that support gender stereotypes. We discovered that users' interactions with voice assistants often reflected implicit gender biases and expectations, even when they initially denied it. Our research thus contributes to the growing demand for more sophisticated and inclusive design methodologies. It encourages taking into account non-binary, gender-neutral speech options that can aid in eradicating these gender biases and stereotypes, making interactions between humans and computers more inclusive and equal.

In conclusion, this study has laid a solid platform for future research into interactions between humans and voice assistants. It highlights the value of considering cultural considerations, involving users in the design process, the possible expansion of the CASA paradigm, and the demand for more inclusive design methods to remove biases. These efforts will significantly impact the design of future conversational agents and the future of human-computer interaction. The following diagram aims to highlight our theoretical implications and their extensions within our interdisciplinary framework, built upon human-computer interaction theories:

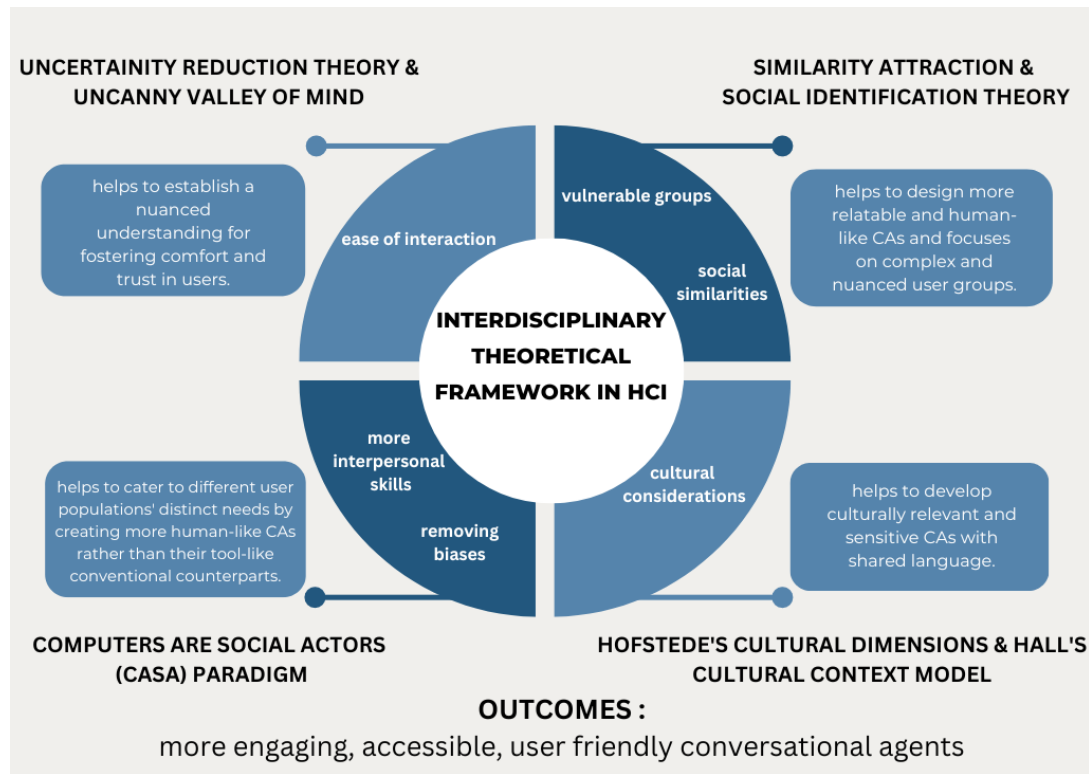


Figure 5.2: Theoretical Implications Diagram

5.3.2. Methodological Contributions and Implications

According to the findings, effective community-researcher collaborations necessitate a paradigm shift away from traditional practices and towards a strategy that includes: acknowledging community contributions, including end-users in research, increasing communication, and embracing culture and diversity. Researchers adopting potent guiding principles and effective research life cycle strategies can solve societal problems and enhance lives (Shneiderman, 2016). Echoing this work, transferring insights and data into practical application is one of the most effective ways to demonstrate the actual value of the research. Implementing user-centred design and co-design initiatives highlights the value of involving the actual users of the technology during the design and development phase.

In this sense, we created a 3-stage design process to develop our prototype. While creating our dialogue flows, we used a local collaborative hypertext dictionary based on user contribution as a background knowledge source to create a sense of familiarity and trust. This approach helps us understand the importance of the voice assistant's cultural relevancy and language style to foster engaging conversations. Additionally, we benefited from surveys, interviews and focus groups to gain insights

into how end-users possibly interact and perceive the conversational experience. The active involvement of real users had a substantial role in advancing knowledge, fostering the emergence of ideas, and shaping concepts throughout each chapter, and they were regarded as 'experts in their own experience'. This approach not only emphasises the importance of user-centric design but also highlights the potential to grasp cultural relevancy and create more natural and human-like conversational agents. By fostering a more inclusive and accessible technology development process, this study sets a precedent for future practices and research.

In the second stage of the voice assistant prototyping, we created user-centred dialogue flows using a text-to-speech converter (Voiser), embedded them into simple Powerpoint pages, and linked every audio to a button for each utterance of the voice assistant. Our Powerpoint soundboard enabled us to create a cost and time-efficient prototype design. By integrating both applied and basic research methodologies, we can accelerate development and ensure its relevance in real-world scenarios (Shneiderman, 2016). So, we have tested our voice assistant prototypes created with actual users' contributions as potential solutions, see how to refine them at the final stage and disseminate practical solutions.

In the latest stage of prototype designing, we used the Wizard of Oz technique to simulate the voice assistant's interpersonal skills, providing a unique opportunity to explore the boundaries of voice assistant design. Our one crucial methodological contribution is the innovative use of the Wizard of Oz (WoZ) technique. Notably, this technique enables social scientists to manipulate different human aspects without coding, allowing a more flexible and adaptive approach to understanding human-computer interactions. It offers a cost-efficient and no-coding solution to investigate intricate human-computer interactions. By employing the Wizard of Oz (WoZ) technique, we were able to achieve the simulation of more sophisticated human-like characteristics in voice assistants, including self-disclosure, humour, and empathy, without necessitating substantial programming expertise. This simulation method brings the research of social scientists closer to real-world applicability. By incorporating WoZ, this study opens up new HCI research opportunities, particularly for those with non-technical backgrounds.

To test our voice assistant prototype design created and validated by the contribution of actual users, we used a unique mixed-method approach, combined with pre-visit and post-visit quizzes, pre-test, momentary, and post-test evaluations, and quantifiable scale measurement to create a comprehensive and detailed understanding of user interactions with conversational agents. This study introduces a framework for interpreting these interactions, contributing to a more nuanced comprehension of the dynamics when humans engage with artificial entities. We present below a detailed diagram to explain every step of the design of our prototypes:

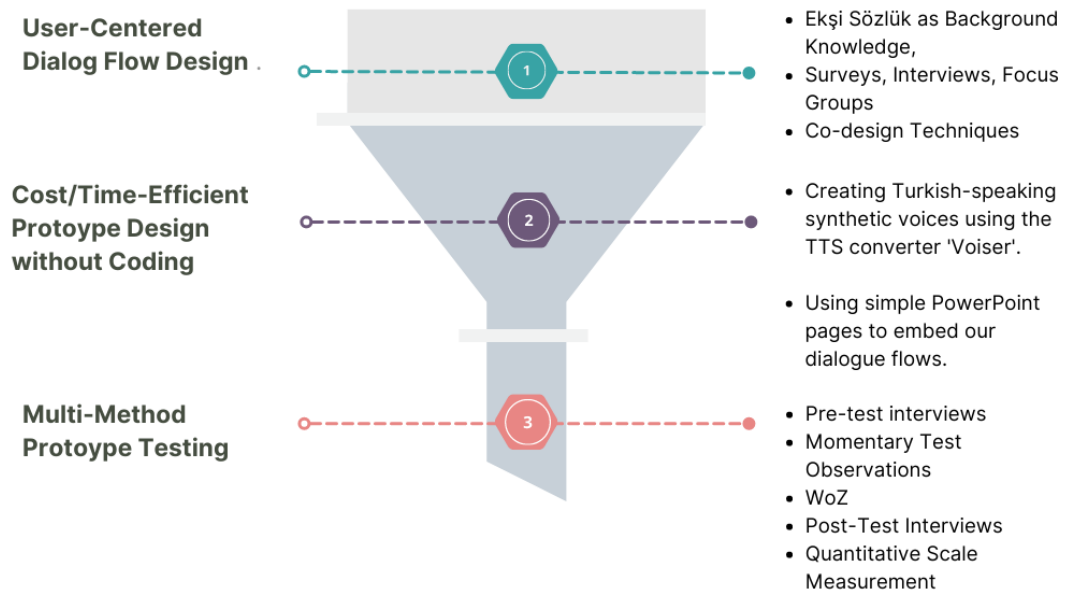


Figure 5.3: Stage Prototype Design Process

5.3.3. Practical implications

The findings from this research are expected to provide practical insights that can inform the design and development of more engaging, user-friendly, and inclusive computational agents. These advancements can ultimately contribute to bridging digital gaps and enhancing social inclusion, thus underscoring the broader societal impact of HCI research.

Our research heralds a new era for the voice assistant industry, challenging the monolithic design paradigm of current voice assistants and reimagining them as empathetic companions rather than just task-accomplishing entities. This shift, emphasising self-disclosure, empathy, humour, and cultural relevancy, tailors voice assistants to individual user needs, making them more engaging and relatable. Moreover, this study revealed the need to transcend binary-gender speech options for voice assistants to create a more equitable digital world.

Our findings open the door to more age-friendly technologies. As the global ageing population grows, the need for digital companions offering customised and empathetic experiences becomes more pressing. Voice assistants could transcend routine tasks, providing psychological support to older adults, possibly reducing feelings of loneliness and improving mental health.

Our work could significantly contribute to enhancing the cultural landscape by ensuring that everyone, including visually impaired users, can access cultural heritage sites. Our co-designed framework with descriptive annotations considerably improves blind users' ability to independently navigate museums, fostering greater autonomy, engagement, and connection to cultural heritage. This development has important implications not only for blind and visually impaired individuals but also for public administrators who can leverage this technology to make public spaces, such as museums, more accessible and inclusive.

The importance of accessibility to cultural heritage as a facet of social inclusion cannot be overstated. By providing blind users with the tools to access and engage with these cultural spaces, we can promote social equality and inclusivity. We have also introduced humour into the design of our voice assistants, making them practical edutainment tools. This approach intertwines education and entertainment, enriching the museum experience for all visitors, not just those with visual impairments.

In essence, our research provides a solution that transcends the conventional 'one size fits all' approach that has dominated the voice assistant industry. Our work heralds a new era of voice assistants that are not just task-performing tools but socially inclusive companions that cater to the distinct needs of diverse user groups.

Furthermore, they can bridge accessibility gaps at cultural heritage sites and offer older adults much-needed companionship and mental support. Public administrators and policymakers can use these insights to create a more accessible and inclusive public environment. Through this strategic design and development, our voice assistants can enhance social learning and contribute to a more inclusive society. This inclusion more directly highlights public administrators' role in practically implementing our research findings, thus enriching the 'Practical Implications' section.

Presented below is a schematic representation of our research findings and their potential applications. It provides a visual summary of the primary outcomes of our studies and illustrates how various stakeholders can utilise these findings to inform the design of more inclusive and user-friendly voice assistant technologies:

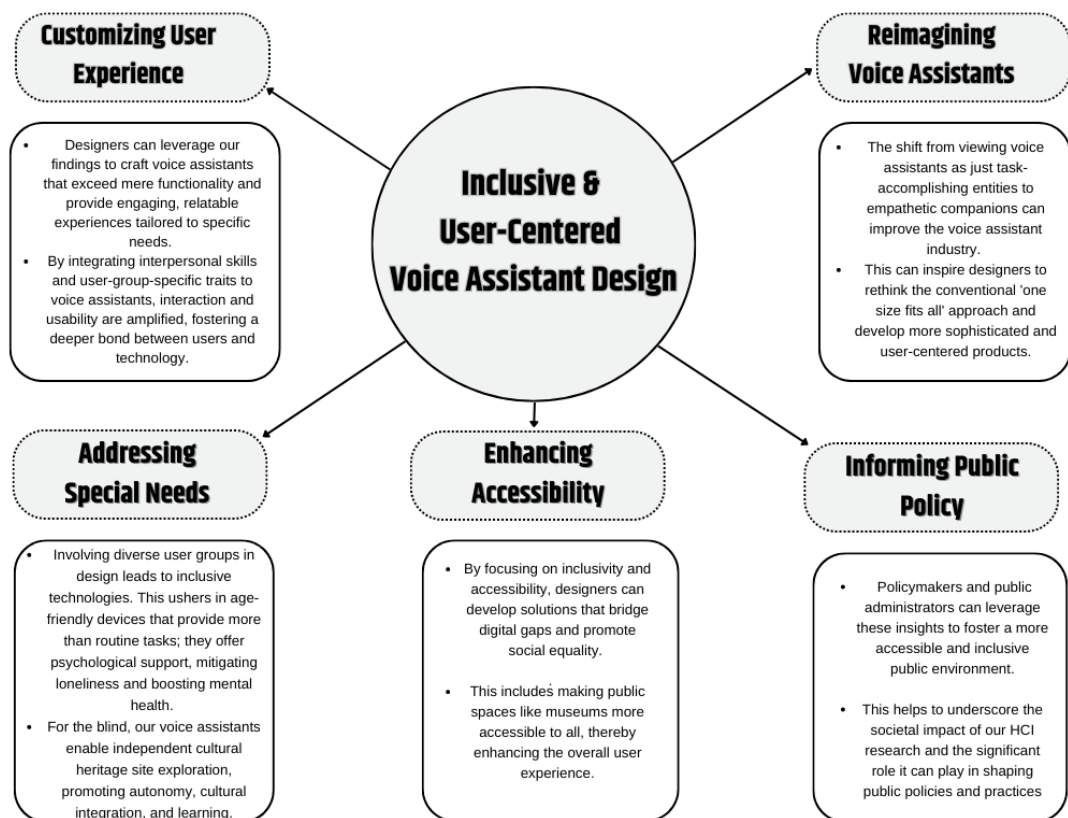


Figure 5.4. Research-to-Application Diagram

The challenges and solutions presented in this research highlight the pivotal role of HCI in shaping societal change. The ramifications of our work underscore the importance of inclusive design in technological advancements, promising a future where technology is more accessible, usable, and beneficial for all.

5.4. Limitations and Further Studies

Despite the notable contributions, it is vital to acknowledge the limitations of our research. The study populations were specific (general adult users with specialised interests, older adults, and visually impaired individuals), and thus, the findings may not fully generalise to other demographic groups. Each group had its unique context, so the insights provided may not be universally applicable. Further studies with diverse populations are needed to ascertain the broader applicability of these findings. Our results' applicability also required testing with a larger sample size. Furthermore, voice assistant designs had to go beyond binary-speech alternatives to avoid perpetuating existing biases and stereotypes. Future research must incorporate nonbinary and gender-ambiguous voices to cover the variety of gender representations. Another limitation lies in the fact that the use of the Wizard of Oz method, while valuable for research purposes, does not directly translate into an implementable voice assistant. It simulates advanced human-like attributes, but these features need further development to be practically implemented in real-world systems.

Lastly, the voice assistant's dialogues were designed based on the research team and co-designer participants' understanding and validation of what would constitute self-disclosure, humour, or empathy. These interpretations may not capture the full spectrum of these complex human attributes. In future research, a broader and more diverse range of dialogues, reflecting different styles and degrees of these attributes, could yield richer insights.

BIBLIOGRAPHY

- Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018). Siri talks at you: An empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind. *ASSETS 2018 - Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 249–258. <https://doi.org/10.1145/3234695.3236344>
- Ahmetovic, D., Bernareggi, C., Leporini, B., & Mascetti, S. (2022). WordMelodies: Supporting the Acquisition of Literacy Skills by Children with Visual Impairment through a Mobile App. *ACM Transactions on Accessible Computing*, 16(1), 1–19. <https://doi.org/10.1145/3565029>
- Alabed, A., Javornik, A., & Gregory-Smith, D. (2022). 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(May), 121786. <https://doi.org/10.1016/j.techfore.2022.121786>
- Altman, I., & Taylor, D. A. (1973). Social penetration: The development of interpersonal relationships. In *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.
- Andrew Reimann. (2010). Intercultural communication and the essence of humour. *Usunomiya Daigaku Kenkyuuronshuu*, 29, 23–34.
- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37(5), 715–727. <https://doi.org/10.1037/0022-3514.37.5.715>
- Archer, R. L. (2016). *Personalistic Self-Disclosure and Attraction : Basis for Relationship or Scarce Resource* Author (s): Richard L . Archer and Christie E . Cook Published by : American Sociological Association Stable URL : <http://www.jstor.org/stable/2786809> JSTOR is a no. 49(3), 268–272.
- Asakawa, S., Guerreiro, J., Sato, D., Takagi, H., Ahmetovic, D., Gonzalez, D., Kitani, K. M., & Asakawa, C. (2019). An independent and interactive museum experience for blind people. *Proceedings of the 16th Web For All 2019 Personalization - Personalizing the Web, W4A 2019*. <https://doi.org/10.1145/3315002.3317557>
- Aw, E. C. X., Tan, G. W. H., Cham, T. H., Raman, R., & Ooi, K. B. (2022). Alexa, what's on my shopping list? Transforming customer experience with digital voice assistants. *Technological Forecasting and Social Change*, 180(April), 121711. <https://doi.org/10.1016/j.techfore.2022.121711>
- Azenkot, S., & Lee, N. B. (2013). Exploring the use of speech input by blind people on mobile devices. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2013*.

<https://doi.org/10.1145/2513383.2513440>

- Babu, S., Schmutge, S., Barnes, T., & Hodges, L. F. (2006). "What would you like to talk about?" An evaluation of social conversations with a virtual receptionist. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4133 LNAI(August), 169–180. https://doi.org/10.1007/11821830_14
- Bailenson, J. N., Iyengar, S., Yee, N., & Collins, N. A. (2008). Facial Similarity between Voters and Candidates Causes Influence. *Public Opinion Quarterly*, 72(5), 935–961. <https://doi.org/10.1093/poq/nfn064>
- Barak, A., & Gluck-Ofri, O. (2007). Degree and reciprocity of self-disclosure in online forums. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 10(3), 407–417. <https://doi.org/10.1089/cpb.2006.9938>
- Becker-Asano, C., Kanda, T., Ishi, C., & Ishiguro, H. (2009). How about laughter? Perceived naturalness of two laughing humanoid robots. *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, 1–6. <https://doi.org/10.1109/ACII.2009.5349371>
- Benbasat, I., & Wang, W. (2005). Trust In and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems*, 6(3), 72–101. <https://doi.org/10.17705/1jais.00065>
- Berg, J. H. (1987). Responsiveness and Self-Disclosure. *Self-Disclosure*, 101–130. https://doi.org/10.1007/978-1-4899-3523-6_6
- Berkowsky, R. W., Rikard, R. V., & Cotten, S. R. (2015). Signing off: Predicting discontinued ICT usage among older adults in assisted and independent living: A survival analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9194(August), 389–398. https://doi.org/10.1007/978-3-319-20913-5_36
- Bernardi, R., Cakici, R., Elliott, D., Erdem, A., Erdem, E., Ikizler-Cinbis, N., Keller, F., Muscat, A., & Plank, B. (2017). Automatic description generation from images: A survey of models, datasets, and evaluation measures. *IJCAI International Joint Conference on Artificial Intelligence*, 0, 4970–4974.
- Bickmore, T., & Cassell, J. (2001). Relational agents: A model and implementation of building user trust. *Conference on Human Factors in Computing Systems - Proceedings*, 396–403.
- Bickmore, T. W., & Cassell, J. (1999). Small Talk and Conversational Storytelling in Embodied Interface Agents. *Proceedings of the AAAI Fall Symposium on Narrative Intelligence, March*, 87–92.
- Bickmore, T. W., & Picard, R. W. (2004). Towards caring machines. *Conference on Human Factors in Computing Systems - Proceedings*, 1489–1492. <https://doi.org/10.1145/985921.986097>
- Bickmore, T. W., Utami, D., Matsuyama, R., & Paasche-Orlow, M. K. (2016). Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment. *Journal of Medical Internet Research*, 18(1). <https://doi.org/10.2196/jmir.5239>
- Bickmore, T., & Cassell, J. (1989). Social Dialogue With Embodied. *Small*.

- Bickmore, Timothy, & Schulman, D. (2007). Practical approaches to comforting users with relational agents. *Conference on Human Factors in Computing Systems - Proceedings*, 2291–2296. <https://doi.org/10.1145/1240866.1240996>
- Binark, M., Arun, Ö., Özsoy, D., Kandemir, B., Şahinkaya, G. (2022). *Information Seeking and Information Evaluation of Older Adults in the Covid-19 Pandemic: Tübitak Sobag Project No: 120k613*. Association for Aging Studies Publication.
- Blair, J., & Abdullah, S. (2019). Understanding the needs and challenges of using conversational agents for deaf older adults. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 161–165. <https://doi.org/10.1145/3311957.3359487>
- Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, 4(1), 11–44. https://doi.org/10.1207/s15327051hci0401_1
- Bolaños, M., Collazos, C., & Gutiérrez, F. (2021). Reference Framework for Measuring the Level of Technological Acceptance by the Elderly: A Case Study of Virtual Assistants. *TecnoLógicas*, 24(50), e1791. <https://doi.org/10.22430/22565337.1791>
- Bouck, E. C., Flanagan, S., Joshi, G. S., Sheikh, W., & Schleppenbach, D. (2011). Speaking Math — A Voice Input, Speech Output Calculator for Students with Visual Impairments. *Journal of Special Education Technology*, 26(4), 1–14. <https://doi.org/10.1177/016264341102600401>
- Branham, S. M., & Kane, S. K. (2015). Collaborative accessibility: How blind and sighted companions co-create accessible home spaces. *Conference on Human Factors in Computing Systems - Proceedings, 2015-April*, 2373–2382. <https://doi.org/10.1145/2702123.2702511>
- Branham, S. M., & Roy, A. R. M. (2019). Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users. *ASSETS 2019 - 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 446–458. <https://doi.org/10.1145/3308561.3353797>
- Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, 62(2), 161–178. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2004.11.002>
- Broadbent, E., Stafford, R., & MacDonald, B. (2009). Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*, 1(4), 319–330. <https://doi.org/10.1007/s12369-009-0030-6>
- Burke, M., Kraut, R., & Wang, Y. C. (2016). Modeling self-disclosure in social networking sites. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 27, 74–85. <https://doi.org/10.1145/2818048.2820010>
- Buyurgan, S. (2009). The expectations of the visually impaired university students from museums. *Kuram ve Uygulamada Egitim Bilimleri*, 9(3), 1191–1204.
- Byrne, D. E. (1971). *The attraction paradigm* (Vol. 462). Academic press.

- Cann, A., Calhoun, L. G., & Nance, J. T. (2000). Exposure to humor before and after an unpleasant stimulus: Humor as a preventative or a cure. *Humor*, 13(2), 177–191. <https://doi.org/10.1515/humr.2000.13.2.177>
- Cassell, J., & Bickmore, T. (2003). Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modelling and User-Adapted Interaction*, 13(1–2), 89–132. <https://doi.org/10.1023/A:1024026532471>
- Chang, R. C. S., Lu, H. P., & Yang, P. (2018). Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. *Computers in Human Behavior*, 84, 194–210. <https://doi.org/10.1016/j.chb.2018.02.025>
- Chase, E. D. Z., Siu, A. F., Boadi-Agyemang, A., Kim, G. S. H., Gonzalez, E. J., & Follmer, S. (2020). PantoGuide: A Haptic and Audio Guidance System to Support Tactile Graphics Exploration. *ASSETS 2020 - 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. <https://doi.org/10.1145/3373625.3418023>
- Chattaraman, V., Kwon, W. S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, 90. <https://doi.org/10.1016/j.chb.2018.08.048>
- Chaturvedi, R., Verma, S., Das, R., & Dwivedi, Y. K. (2023). Technological Forecasting & Social Change Social companionship with artificial intelligence : Recent trends and future avenues. *Technological Forecasting & Social Change*, 193(May), 122634. <https://doi.org/10.1016/j.techfore.2023.122634>
- Choi, D., Kwak, D., Cho, M., & Lee, S. (2020). “nobody Speaks that Fast!” An Empirical Study of Speech Rate in Conversational Agents for People with Vision Impairments. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. <https://doi.org/10.1145/3313831.3376569>
- Choudrie, J., Manandhar, N., Castro, C., & Obuekwe, C. (2023). Hey Siri, Google! Can you help me? A qualitative case study of smartphones AI functions in SMEs. *Technological Forecasting and Social Change*, 189(December 2022), 122375. <https://doi.org/10.1016/j.techfore.2023.122375>
- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92. <https://doi.org/10.1016/j.future.2018.01.055>
- Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, 116(3), 457–475. <https://doi.org/10.1037/0033-2909.116.3.457>
- Cooper, J., & Weaver, K. D. (2003). Gender and computers: Understanding the digital divide. In *Gender and Computers: Understanding the Digital Divide*. Routledge. <https://doi.org/10.4324/9781410608932>
- Corbett, E., & Weber, A. (2016). What can i say? Addressing user experience challenges of a mobile voice user interface for accessibility. *Proceedings of the 18th International Conference on Human-Computer Interaction with*

- Mobile Devices and Services, MobileHCI 2016*, 72–82.
<https://doi.org/10.1145/2935334.2935386>
- Cosentino, S., Kishi, T., Zecca, M., Sessa, S., Bartolomeo, L., Hashimoto, K., Nozawa, T., & Takanishi, A. (2013). Human-humanoid robot social interaction: Laughter. *2013 IEEE International Conference on Robotics and Biomimetics, ROBIO 2013, December*, 1396–1401.
<https://doi.org/10.1109/ROBIO.2013.6739661>
- Cozby, P. C. (1973). Self-disclosure: a literature review. *Psychological Bulletin*, 79(2), 73–91. <https://doi.org/10.1037/h0033950>
- Crawford, K. (2016). Artificial Intelligence’s White Guy Problem. *New York Times*.
<https://www.microsoft.com/en-us/research/publication/artificial-intelligences-white-guy-problem/>
- Czaja, S. J., Boot, W. R., Charness, N., & Rogers, W. A. (2019). Designing for older adults: Principles and creative human factors approaches, 3rd ed. In *Designing for older adults: Principles and creative human factors approaches, 3rd ed.* CRC Press/Routledge/Taylor & Francis Group. <https://doi.org/10.1201/b22189>
- Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of oz studies-why and how. *International Conference on Intelligent User Interfaces, Proceedings IUI, Part F1275*, 193–200.
- Daniel, F., Matera, M., Zaccaria, V., & Dell’orto, A. (2018). Toward truly personal chatbots: On the development of custom conversational assistants. *Proceedings - International Conference on Software Engineering*, 31–36.
<https://doi.org/10.1145/3195555.3195563>
- DDA. (1995). *Disability Discrimination Act*.
- de Gennaro, M., Krumhuber, E. G., & Lucas, G. (2020). Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood. *Frontiers in Psychology*, 10(January), 1–14.
<https://doi.org/10.3389/fpsyg.2019.03061>
- Delello, J. A., & McWhorter, R. R. (2017). Reducing the Digital Divide: Connecting Older Adults to iPad Technology. *Journal of Applied Gerontology*, 36(1), 3–28. <https://doi.org/10.1177/0733464815589985>
- Deng, E., Mutlu, B., & Mataric, M. J. (2019). Embodiment in Socially Interactive Robots. *Foundations and Trends in Robotics*, 7(4), 251–356.
<https://doi.org/10.1561/23000000056>
- Derlega, V. J., Winstead, B. A., & Greene, K. (2008). Self-disclosure and starting a close relationship. In *Handbook of relationship initiation*. (pp. 153–174). Psychology Press.
- DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., ... Morency, L. P. (2014). SimSensei kiosk: A virtual human interviewer for healthcare decision support. *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, 2(1), 1061–1068.
- Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for*

- Information Systems*, 23(1), 96–138. <https://doi.org/10.17705/1jais.00724>
- Dindia, K. (2002). Self-disclosure research: Knowledge through meta-analysis. In *Interpersonal communication research: Advances through meta-analysis*. (pp. 169–185). Lawrence Erlbaum Associates Publishers.
- Dogruel, L., Joeckel, S., & Bowman, N. D. (2015). The use and acceptance of new media entertainment technology by elderly users: Development of an expanded technology acceptance model. *Behaviour and Information Technology*, 34(11), 1052–1063. <https://doi.org/10.1080/0144929X.2015.1077890>
- Drossos, K., Zormpas, N., Floros, A., & Giannakopoulos, G. (2015). Accessible games for blind children, empowered by binaural sound. *8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA 2015 - Proceedings*. <https://doi.org/10.1145/2769493.2769546>
- Druga, S., Breazeal, C., Williams, R., & Resnick, M. (2017). “Hey Google is it ok if I eat you?” Initial explorations in child-agent interaction. *IDC 2017 - Proceedings of the 2017 ACM Conference on Interaction Design and Children*, 595–600. <https://doi.org/10.1145/3078072.3084330>
- Dybala, P., Ptaszynski, M., Rzepka, R., & Araki, K. (2009). Humoroids - Conversational agents that induce positive emotions with humor. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, 2(January), 1090–1091. <https://doi.org/10.1145/1558109.1558196>
- Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor’s voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior*, 90(May 2018), 357–362. <https://doi.org/10.1016/j.chb.2018.08.027>
- Edwards, C., & Harwood, J. (2003). Social identity in the classroom: An examination of age identification between students and instructors. *Communication Education*, 52(1), 60–65. <https://doi.org/10.1080/03634520302463>
- Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., & Hegel, F. (2012). “If you sound like me, you must be more human”: On the interplay of robot and user features on human-robot acceptance and anthropomorphism. *HRI’12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction, August 2014*, 125–126. <https://doi.org/10.1145/2157689.2157717>
- Fang, M. L., Canham, S. L., Battersby, L., Sixsmith, J., Wada, M., & Sixsmith, A. (2019). Exploring Privilege in the Digital Divide: Implications for Theory, Policy, and Practice. *Gerontologist*, 59(1), E1–E15. <https://doi.org/10.1093/geront/gny037>
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender Bias in Chatbot Design Jasper. In *Følstad A. et al. (eds) Chatbot Research and Design. CONVERSATIONS 2019. Lecture Notes in Computer Science* (Vol. 11970). <https://doi.org/10.1007/978-3-030-39540-7>
- Ferati, M., Mannheimer, S., & Bolchini, D. (2011). Usability evaluation of acoustic

- interfaces for the blind. *SIGDOC'11 - Proceedings of the 29th ACM International Conference on Design of Communication*, 9–16. <https://doi.org/10.1145/2038476.2038479>
- Fikret Pasa, S. (2000). Leadership influence in a high power distance and collectivist culture. *Leadership & Organization Development Journal*, 21(8), 414–426. <https://doi.org/10.1108/01437730010379258>
- Fitton, D., C Read, J., Sim, G., & Cassidy, B. (2018). Co-designing voice user interfaces with teenagers in the context of smart homes. *IDC 2018 - Proceedings of the 2018 ACM Conference on Interaction Design and Children*, 55–66. <https://doi.org/10.1145/3202185.3202744>
- Fogg, B. J., & Nass, C. (1997). Silicon sycophants: The effects of computers that flatter. *International Journal of Human Computer Studies*, 46(5), 551–561. <https://doi.org/10.1006/ijhc.1996.0104>
- Forlizzi, J. D. C. G. F. (2004). Assistive Robotics and an Ecology of Elders Living Independently in Their Homes. *HUMAN-COMPUTER INTERACTION*, 19, 25–59.
- Friemel, T. N. (2016). The digital divide has grown old: Determinants of a digital divide among seniors. *New Media and Society*, 18(2), 313–331. <https://doi.org/10.1177/1461444814538648>
- Gallimore, D., Lyons, J. B., Vo, T., Mahoney, S., & Wynne, K. T. (2019). Trusting robocop: Gender-based effects on trust of an autonomous robot. *Frontiers in Psychology*, 10(MAR), 1–9. <https://doi.org/10.3389/fpsyg.2019.00482>
- Gatto, S. L., & Tak, S. H. (2008). Computer, Internet, and e-mail use among older adults: Benefits and barriers. *Educational Gerontology*, 34(9), 800–811. <https://doi.org/10.1080/03601270802243697>
- Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, 2(2), 167–177. https://doi.org/10.1207/s15327051hci0202_3
- Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly: Management Information Systems*, 21(4), 389–400. <https://doi.org/10.2307/249720>
- Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega*, 32(6), 407–424. <https://doi.org/10.1016/j.omega.2004.01.006>
- Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology and Work*, 14(1), 39–49. <https://doi.org/10.1007/s10111-011-0194-3>
- Ghiani, G., Leporini, B., & Paternò, F. (2008). Supporting orientation for blind people using museum guides. *Conference on Human Factors in Computing Systems - Proceedings*, 3417–3422. <https://doi.org/10.1145/1358628.1358867>
- Goldman, J. A., Rosenzweig, C. M., & Lutter, A. D. (1980). Effect of similarity of ego identity status on interpersonal attraction. *Journal of Youth and Adolescence*, 9(2), 153–162. <https://doi.org/10.1007/BF02087933>
- Graham, E. E. (1995). The involvement of sense of humor in the development of

- social relationships. *Communication Reports*, 8(2), 158–169.
<https://doi.org/10.1080/08934219509367622>
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125–130.
<https://doi.org/10.1016/j.cognition.2012.06.007>
- Gremler, D. D., & Gwinner, K. P. (2008). Rapport-Building Behaviors Used by Retail Employees. *Journal of Retailing*, 84(3), 308–324.
<https://doi.org/10.1016/j.jretai.2008.07.001>
- Habler, F., Schwind, V., & Henze, N. (2019). Effects of smart virtual assistants' gender and language. *ACM International Conference Proceeding Series*, 469–473. <https://doi.org/10.1145/3340764.3344441>
- Hall, E. T. (1976). Beyond culture. In *Beyond Culture* (First edit). Anchor Press Garden City, N.Y. <https://doi.org/LK> - <https://worldcat.org/title/1818029>
- Hampes, W. P. (1999). The relationship between humor and trust. *J Humor Res*, 12(3), 253–260. <https://doi.org/doi:10.1515/humr.1999.12.3.253>
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5), 517–527.
<https://doi.org/10.1177/0018720811417254>
- Hartzler, A. L., Patel, R. A., Czerwinski, M., Pratt, W., Roseway, A., Chandrasekaran, N., & Back, A. (2014). Real-time feedback on nonverbal clinical communication. *Methods of Information in Medicine*, 53(5), 389–405.
<https://doi.org/10.3414/ME13-02-0033>
- Harwood, J., Giles, H., & Ryan, E. B. (1995). Aging, communication, and intergroup theory: Social identity and intergenerational communication. *Handbook of Communication and Aging Research*, January 1995, 133–159.
- Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2008). The influence of social presence on acceptance of a companion robot by older people. *Journal of Physical Agents*, 2(2), 33–40. <https://doi.org/10.14198/JoPha.2008.2.2.05>
- Helal, A., Moore, S. E., & Ramachandran, B. (2001). Drishti: An integrated navigation system for visually impaired and disabled. *International Symposium on Wearable Computers, Digest of Papers*, 45, 149–156.
<https://doi.org/10.1109/iswc.2001.962119>
- Hendriks, N., Slegers, K., & Duysburgh, P. (2015). Codesign with people living with cognitive or sensory impairments: a case for method stories and uniqueness. *CoDesign*, 11(1), 70–82. <https://doi.org/10.1080/15710882.2015.1020316>
- Heuten, W., Wichmann, D., & Boll, S. (2006). Interactive 3D sonification for the exploration of city maps. *ACM International Conference Proceeding Series*, 189(October), 155–164. <https://doi.org/10.1145/1182475.1182492>
- Hill, C. T., & Stull, D. E. (1987). Gender and self-disclosure: Strategies for exploring the issues. In *Self-disclosure: Theory, research, and therapy*. (pp. 81–100). Plenum Press.
- Hofstede, G. (1980). Motivation, leadership, and organization: Do American theories apply abroad? *Organizational Dynamics*, 9(1), 42–63.
[https://doi.org/https://doi.org/10.1016/0090-2616\(80\)90013-3](https://doi.org/https://doi.org/10.1016/0090-2616(80)90013-3)

- Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, 2(1), 1–26. <https://doi.org/10.9707/2307-0919.1014>
- Hofstede, G. J., Hofstede, G. J., & Minkov, M. (1991). *Cultures and Organizations: Software of the Mind*.
- Holbrook, S. E. (1991). Women's work: The feminizing of composition. *Rhetoric Review*, 9(2), 201–229. <https://doi.org/10.1080/07350199109388929>
- Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers*, 18(2), 227–245. <https://doi.org/10.1016/j.intcom.2005.05.003>
- Huan C. M., Szafir, D. (2001) No joke: examining the use of humor in Computer-mediated learning. <http://pages.cs.wisc.edu/~dszafir/resources/HumorinComputer-MediatedLearning.pdf/>. Unpublished Material
- Hübner, M. T., & Bell, D. C. (2003). Computer-mediated humor and ethos: Exploring threads of constitutive laughter in online communities. *Computers and Composition*, 20(3), 277–294. [https://doi.org/10.1016/S8755-4615\(03\)00036-7](https://doi.org/10.1016/S8755-4615(03)00036-7)
- Hwang, G., Oh, C. Y., Lee, J., & Lee, J. (2019). It sounds like a woman: Exploring gender stereotypes in South Korean voice assistants. *Conference on Human Factors in Computing Systems - Proceedings*, 1–6. <https://doi.org/10.1145/3290607.3312915>
- Jourard, S. M. (1971). Self-disclosure: An experimental analysis of the transparent self. In *Self-disclosure: An experimental analysis of the transparent self*. John Wiley.
- Kabasakal, H., & Bodur, M. (1998). Leadership, values and institutions: The case of Turkey. *Western Academy of Management Conference, Istanbul, June*.
- Kahn, P. H., Freier, N. G., Friedman, B., Severson, R. L., & Feldman, E. N. (2004). Social and moral relationships with robotic others? *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 545–550. <https://doi.org/10.1109/roman.2004.1374819>
- Kepuska, V., & Bohouta, G. (2018). Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, 2018-Janua(c)*, 99–103. <https://doi.org/10.1109/CCWC.2018.8301638>
- Kim, J., Kang, S., & Bae, J. (2022). Human likeness and attachment effect on the perceived interactivity of AI speakers. *Journal of Business Research*, 144(February), 797–804. <https://doi.org/10.1016/j.jbusres.2022.02.047>
- Kim, S., & Choudhury, A. (2021). Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior*, 124(June), 106914. <https://doi.org/10.1016/j.chb.2021.106914>
- Klein. (2007). Internet-Based Patient-Physician Electronic Communication Applications: Patient Acceptance and Trust. *E-Service Journal*, 5(2), 27. <https://doi.org/10.2979/esj.2007.5.2.27>
- Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14(2), 119–

140. [https://doi.org/10.1016/S0953-5438\(01\)00053-4](https://doi.org/10.1016/S0953-5438(01)00053-4)
- Koda, T., & Takeda, Y. (2018). Perception of culture-specific gaze behaviors of agents and gender effects. *HAI 2018 - Proceedings of the 6th International Conference on Human-Agent Interaction*, 138–143.
<https://doi.org/10.1145/3284432.3284472>
- Koon, L. M., McGlynn, S. A., Blocker, K. A., & Rogers, W. A. (2020). Perceptions of Digital Assistants From Early Adopters Aged 55+. *Ergonomics in Design*, 28(1), 16–23. <https://doi.org/10.1177/1064804619842501>
- Kowalski, J., Jaskulska, A., Skorupska, K., Abramczuk, K., Biele, C., Kopec, W., & Marasek, K. (2019). Older adults and voice interaction: A pilot study with google home. *ArXiv*, 1–6.
- Kramer, G., Walker, B., Bonebright, T., Cook, P., & Flowers, J. H. (1999). *Sonification Report: Status of the Field and Research Agenda*. January.
- Kulesza, W., Dolinski, D., Huisman, A., & Majewski, R. (2014). The Echo Effect: The Power of Verbal Mimicry to Influence Prosocial Behavior. *Journal of Language and Social Psychology*, 33(2), 183–201.
<https://doi.org/10.1177/0261927X13506906>
- Kumar, K. (2022). *An intelligent Assistant for the Visually Impaired & blind people using machine learning*. January.
- Lahav, O., Schloerb, D. W., Kumar, S., & Srinivasan, M. A. (2008). BlindAid: A learning environment for enabling people who are blind to explore and navigate through unknown real spaces. *2008 Virtual Rehabilitation, IWVR*, 193–197.
<https://doi.org/10.1109/ICVR.2008.4625159>
- Language, S., Mar, N., Giles, H., & Henwood, K. (1988). *Accommodating the Elderly : Invoking and Extending a Theory Author (s) : Nikolas Coupland , Justine Coupland , Howard Giles and Karen Henwood Published by : Cambridge University Press Stable URL :*
<https://www.jstor.org/stable/4167896>. 17(1), 1–41.
- Large, D. R., Burnett, G., Harrington, K., Clark, L., Luton, J., Thomas, P., & Bennett, P. (2019). “It’s small talk, jim, but not as we know it.” Engendering trust through human-agent conversation in an autonomous, self-driving car. *ACM International Conference Proceeding Series*.
<https://doi.org/10.1145/3342775.3342789>
- Lee, B. (2006). Empathy, androids and “authentic experience.” *Connection Science*, 18(4), 419–428. <https://doi.org/10.1080/09540090600868847>
- Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotype. *Conference on Human Factors in Computing Systems - Proceedings, May 2014*, 289–290.
<https://doi.org/10.1145/633292.633461>
- Lee, K., Lee, K. Y., & Sheehan, L. (2020). Hey Alexa! A Magic Spell of Social Glue?: Sharing a Smart Voice Assistant Speaker and Its Impact on Users’ Perception of Group Harmony. *Information Systems Frontiers*, 22(3), 563–583.
<https://doi.org/10.1007/s10796-019-09975-1>
- Lee, M., Ackermans, S., Van As, N., Chang, H., Lucas, E., & IJsselsteijn, W. (2019). Caring for Vincent: A Chatbot for Self-compassion. *Conference on Human Factors in Computing Systems - Proceedings*.

<https://doi.org/10.1145/3290605.3300932>

- Lee, S. Y., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human Computer Studies*, 103(January), 95–105. <https://doi.org/10.1016/j.ijhcs.2017.02.005>
- Lee, Y. C., Yamashita, N., & Huang, Y. (2020). Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1). <https://doi.org/10.1145/3392836>
- Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020). “I Hear You, i Feel You”: Encouraging Deep Self-disclosure through a Chatbot. *Conference on Human Factors in Computing Systems - Proceedings*, August. <https://doi.org/10.1145/3313831.3376175>
- Lefcourt, H. M. (2001). Humor: The psychology of living buoyantly. In *Humor: The psychology of living buoyantly*. Kluwer Academic Publishers. <https://doi.org/10.1007/978-1-4615-4287-2>
- Li, F. M., Zhang, L., Bandukda, M., Stangl, A., Shinohara, K., Findlater, L., & Carrington, P. (2023). Understanding Visual Arts Experiences of Blind People. *Conference on Human Factors in Computing Systems - Proceedings*, 1(1), 1–32. <https://doi.org/10.1145/3544548.3580941>
- Li, J., Galley, M., Brockett, C., Gao, J., & Dolan, B. (2016). A diversity-promoting objective function for neural conversation models. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, 110–119. <https://doi.org/10.18653/v1/n16-1014>
- Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, B. (2016). A persona-based neural conversation model. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 2, 994–1003. <https://doi.org/10.18653/v1/p16-1094>
- Li, Z., & Rau, P. L. P. (2019). Effects of self-disclosure on attributions in human-iot conversational agent interaction. *Interacting with Computers*, 31(1), 13–26. <https://doi.org/10.1093/iwc/iwz002>
- Lim, V., Rooksby, M., & Cross, E. S. (2021). Social Robots on a Global Stage: Establishing a Role for Culture During Human–Robot Interaction. *International Journal of Social Robotics*, 13(6), 1307–1333. <https://doi.org/10.1007/s12369-020-00710-4>
- Linnemann, G. A., & Jucks, R. (2018). “Can i Trust the Spoken Dialogue System because It Uses the Same Words as i Do?” - Influence of Lexically Aligned Spoken Dialogue Systems on Trustworthiness and User Satisfaction. *Interacting with Computers*, 30(3), 173–186. <https://doi.org/10.1093/iwc/iwy005>
- Lippman, L. G., & Dunn, M. L. (2000). Contextual connections within puns: effects on perceived humor and memory. *The Journal of General Psychology*, 127(2), 185–197. <https://doi.org/10.1080/00221300009598578>
- Liu, B., & Sundar, S. S. (2018). Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychology, Behavior, and Social Networking*, 21(10). <https://doi.org/10.1089/cyber.2018.0110>

- Liu, H., Russo, N. M., & Larson, C. R. (2010). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *The Journal of the Acoustical Society of America*, 127(2), 1042–1046. <https://doi.org/10.1121/1.3273880>
- Looije, R., Neerinx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human Computer Studies*, 68(6), 386–397. <https://doi.org/10.1016/j.ijhcs.2009.08.007>
- Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., & Martinez, A. (2019). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, 51(4), 984–997. <https://doi.org/10.1177/0961000618759414>
- Lopatovska, I., & Williams, H. (2018). Personification of the amazon alexa: BFF or a mindless companion? *CHIIR 2018 - Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, 2018-March*, 265–268. <https://doi.org/10.1145/3176349.3176868>
- Lopes, P. N., Salovey, P., Côté, S., & Beers, M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion*, 5(1), 113–118. <https://doi.org/10.1037/1528-3542.5.1.113>
- López, G., Quesada, L., & Guerrero, L. A. (2017). Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In I. L. Nunes (Ed.), *Advances in Human Factors and Systems Interaction*. Springer International Publishing. [https://doi.org/https://doi.org/10.1007/978-3-319-60366-7_23](https://doi.org/10.1007/978-3-319-60366-7_23)
- Lucas, G. M., Boberg, J., Traum, D., Artstein, R., Gratch, J., Gainer, A., Johnson, E., Leuski, A., & Nakano, M. (2018). Getting to Know Each Other: The Role of Social Dialogue in Recovery from Errors in Social Robots. *ACM/IEEE International Conference on Human-Robot Interaction, April 2019*, 344–351. <https://doi.org/10.1145/3171221.3171258>
- Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers Robotics AI*, 4(OCT), 1–9. <https://doi.org/10.3389/frobt.2017.00051>
- Luger, E., & Sellen, A. (2016). “Like having a really bad pa”: The gulf between user expectation and experience of conversational agents. *Conference on Human Factors in Computing Systems - Proceedings*, 5286–5297. <https://doi.org/10.1145/2858036.2858288>
- Lumbreras, M., & Sanchez, J. (1999). Interactive 3D sound hyperstories for blind children. *Conference on Human Factors in Computing Systems - Proceedings, May*, 318–325. <https://doi.org/10.1145/302979.303101>
- Luria, M., Reig, S., Tan, X. Z., Steinfeld, A., Forlizzi, J., & Zimmerman, J. (2019). Re-embodiment and co-embodiment: Exploration of social presence for robots and conversational agents. *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference*, 633–644. <https://doi.org/10.1145/3322276.3322340>
- Lythreathis, S., Singh, S. K., & El-Kassar, A. N. (2022). The digital divide: A review

- and future research agenda. *Technological Forecasting and Social Change*, 175(April 2021), 121359. <https://doi.org/10.1016/j.techfore.2021.121359>
- Marone, V. (2015). Online humour as a community-building cushioning glue. *European Journal of Humour Research*, 3(1), 61–83. <https://doi.org/10.7592/ejhr2015.3.1.marone>
- Marson, S. M., & Powell, R. M. (2014). Goffman and the infantilization of elderly persons: A theory in development. *Journal of Sociology and Social Welfare*, 41(4), 143–158.
- Martin, R. A. (2007). *The Psychology of Humor: An Integrative Approach*.
- Masina, F., Orso, V., Pluchino, P., Dainese, G., Volpato, S., Nelini, C., Mapelli, D., Spagnoli, A., & Gamberini, L. (2020). Investigating the Accessibility of Voice Assistants With Impaired Users: Mixed Methods Study. *Journal of Medical Internet Research*, 22(9), e18431–e18431. <https://doi.org/10.2196/18431>
- McDonnell, M., & Baxter, D. (2019). Chatbots and Gender Stereotyping. *Interacting with Computers*, 31(2), 116–121. <https://doi.org/10.1093/iwc/iwz007>
- Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., & O'Neill, J. (2017). How do you want your chatbot? An exploratory Wizard-of-Oz study with young, Urban Indians. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10513 LNCS. https://doi.org/10.1007/978-3-319-67744-6_28
- Metatla, O., Bryan-Kinns, N., Stockman, T., & Martin, F. (2015). Designing with and for people living with visual impairments: audio-tactile mock-ups, audio diaries and participatory prototyping. *CoDesign*, 11(1), 35–48. <https://doi.org/10.1080/15710882.2015.1007877>
- Metatla, O., Martin, F., Parkinson, A., Bryan-Kinns, N., Stockman, T., & Tanaka, A. (2016). Audio-haptic interfaces for digital audio workstations: A participatory design approach. *Journal on Multimodal User Interfaces*, 10(3), 247–258. <https://doi.org/10.1007/s12193-016-0217-8>
- Metatla, O., Oldfield, A., Ahmed, T., Vafeas, A., & Miglani, S. (2019). Voice user interfaces in schools: Co-designing for Inclusion with Visually-Impaired and Sighted Pupils. *Conference on Human Factors in Computing Systems - Proceedings*, 1–15. <https://doi.org/10.1145/3290605.3300608>
- Middleton, R., & Moland, J. (1959). Humor in Negro and white subcultures: A study of jokes among university students. *American Sociological Review*, 24, 61–69. <https://doi.org/10.2307/2089583>
- Miller, B., & Feil-Seifer, D. (2017). Embodiment, Situatedness, and Morphology for Humanoid Robots Interacting with People. *Humanoid Robotics: A Reference*, 1–23. https://doi.org/10.1007/978-94-007-7194-9_130-1
- Miner, A. S., Laranjo, L., & Kocaballi, A. B. (2020). Chatbots in the fight against the COVID-19 pandemic. *Npj Digital Medicine*, 3(1), 1–4. <https://doi.org/10.1038/s41746-020-0280-0>
- Mitchell, W. J., Ho, C. C., Patel, H., & MacDorman, K. F. (2011). Does social desirability bias favor humans? Explicit-implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*, 27(1), 402–412. <https://doi.org/10.1016/j.chb.2010.09.002>

- Moghe, N., Arora, S., Banerjee, S., & Khapra, M. M. (2020). Towards exploiting background knowledge for building conversation systems. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 2322–2332. <https://doi.org/10.18653/v1/d18-1255>
- Mohammadi, G., Vinciarelli, A., & Mortillaro, M. (2010). The voice of personality: Mapping nonverbal vocal behavior into trait attributions. *SSPW'10 - Proceedings of the 2010 ACM Social Signal Processing Workshop, Co-Located with ACM Multimedia 2010, January*, 17–20. <https://doi.org/10.1145/1878116.1878123>
- Montepare, J. M., & Zebrowitz-McArthur, L. (1987). Perceptions of adults with childlike voices in two cultures. *Journal of Experimental Social Psychology*, 23(4), 331–349. [https://doi.org/https://doi.org/10.1016/0022-1031\(87\)90045-X](https://doi.org/https://doi.org/10.1016/0022-1031(87)90045-X)
- Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, 25(6), 889–922. <https://doi.org/10.1177/0265407508096700>
- Moon, Y. (2000). Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. *Journal of Consumer Research*, 26(4), 323–339. <https://doi.org/10.1086/209566>
- Morkes, J., Kernal, H. K., & Nass, C. (1998). Humor in Task-Oriented Computer-Mediated Communication and Human-Computer Interaction. *Conference on Human Factors in Computing Systems - Proceedings, April*, 215–216. <https://doi.org/10.1145/286498.286704>
- Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: a direct test of SRCT theory. *Human-Computer Interaction*, 14(4), 395–435. https://doi.org/10.1207/S15327051HCI1404_2
- Morris. (2013). *Journal of Aging Science. Aging Sci*, 1(1), 1–9. <https://doi.org/10.4172/jasc.1000101>
- Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: System design and user perceptions. *Journal of Medical Internet Research*, 20(6). <https://doi.org/10.2196/10148>
- Mourey, J. A., Olson, J. G., & Yoon, C. (2017). Products as pals: Engaging with anthropomorphic products mitigates the effects of social exclusion. *Journal of Consumer Research*, 44(2), 414–431. <https://doi.org/10.1093/jcr/ucx038>
- Mukhiddinov, M., & Kim, S. Y. (2021). A systematic literature review on the automatic creation of tactile graphics for the blind and visually impaired. *Processes*, 9(10), 1–31. <https://doi.org/10.3390/pr9101726>
- Nag, P., & Yalçın, Ö. N. (2020). Gender Stereotypes in Virtual Agents. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, IVA 2020*. <https://doi.org/10.1145/3383652.3423876>
- Nass, C., & Brave, S. (2007). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press.
- Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and

- consistency-attraction. *Journal of Experimental Psychology: Applied*, 7(3), 171–181. <https://doi.org/10.1037/1076-898X.7.3.171>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, 27(10), 864–876. <https://doi.org/10.1111/j.1559-1816.1997.tb00275.x>
- Nees, Michael A.; Walker, B. N. (2015). Auditory Interfaces and Sonification. *Statewide Agricultural Land Use Baseline 2015*, 1, 1–37.
- Nehaniv, C. L. (1999). Narrative for Artifacts: Transcending Context and Self. *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*, Heidegger 1972, 101–104.
- Nguyen, Q. N., & Sidorova, A. (2018). Understanding user interactions with a chatbot: A self-determination theory approach. *Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018*.
- Niculescu, A., van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *International Journal of Social Robotics*, 5(2), 171–191. <https://doi.org/10.1007/s12369-012-0171-x>
- Nijholt, A., & Aldo, B. (2017). Humor in Human-Computer Interaction : A Short Survey. *Adjunct Conference Proceedings Interact 2017, September*, 199–220. https://www.researchgate.net/publication/320127779%0Ahttp://oar.a-star.edu.sg/jspui/bitstream/123456789/2265/1/humor-human-computer_ALE.pdf
- Noguchi, Y., Kamide, H., & Tanaka, F. (2023). How Should a Social Mediator Robot Convey Messages About the Self-Disclosures of Elderly People to Recipients ? *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-023-01016-x>
- O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human Computer Studies*, 112(July 2017), 28–39. <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- Oppliger, P. A. (2003). Humor and learning. In *Communication and emotion: Essays in honor of Dolf Zillmann*. (pp. 255–273). Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410607584>
- Papagiannidis, S., & Davlembayeva, D. (2022). Bringing Smart Home Technology to Peer-to-Peer Accommodation: Exploring the Drivers of Intention to Stay in Smart Accommodation. *Information Systems Frontiers*, 24(4), 1189–1208. <https://doi.org/10.1007/s10796-021-10227-4>
- Partala, T., & Surakka, V. (2004). The effects of affective interventions in human–computer interaction. *Interacting with Computers*, 16(2), 295–309. <https://doi.org/10.1016/j.intcom.2003.12.001>
- Pereira, A., Leite, I., Mascarenhas, S., Martinho, C., & Paiva, A. (2011). Using empathy to improve human-robot relationships. *Lecture Notes of the Institute*

for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 59 LNICST, 130–138. https://doi.org/10.1007/978-3-642-19385-9_17

- Pereira, J., & Díaz, Ó. (2019). Using Health Chatbots for Behavior Change: A Mapping Study. *Journal of Medical Systems*, 43(5). <https://doi.org/10.1007/s10916-019-1237-1>
- Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior*, 65. <https://doi.org/10.1016/j.chb.2016.08.004>
- Pollmann, K., Loh, W., Fronemann, N., & Ziegler, D. (2023). Entertainment vs. manipulation: Personalized human-robot interaction between user experience and ethical design. *Technological Forecasting and Social Change*, 189(January), 122376. <https://doi.org/10.1016/j.techfore.2023.122376>
- Portet, F., Vacher, M., Golanski, C., Roux, C., & Meillon, B. (2013). Design and evaluation of a smart home voice interface for the elderly: Acceptability and objection aspects. *Personal and Ubiquitous Computing*, 17(1), 127–144. <https://doi.org/10.1007/s00779-011-0470-5>
- Pradhan, A., Findlater, L., & Lazar, A. (2019). “Phantom friend” or “just a box with information”: personification and ontological categorization of smart speaker-based voice assistants by older adults. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). <https://doi.org/10.1145/3359316>
- Pradhan, A., Lazar, A., & Findlater, L. (2020). Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction*, 27(4). <https://doi.org/10.1145/3373759>
- Pradhan, A., Mehta, K., & Findlater, L. (2018). “Accessibility came by accident”: Use of voice-controlled intelligent personal assistants by people with disabilities. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*. <https://doi.org/10.1145/3173574.3174033>
- Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies*, 62(2), 231–245. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2004.11.009>
- Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). “Alexa is my new BFF”: Social roles, user satisfaction, and personification of the Amazon Echo. *Conference on Human Factors in Computing Systems - Proceedings, Part F1276*. <https://doi.org/10.1145/3027063.3053246>
- Pyae, A., & Joelsson, T. N. (2018). Investigating the usability and user experiences of voice user interface: A case of Google home smart speaker. *MobileHCI 2018 - Beyond Mobile: The Next 20 Years - 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Conference Proceedings Adjunct*, 127–131. <https://doi.org/10.1145/3236112.3236130>
- Quero, L. C., Bartolomé, J. I., & Cho, J. (2021). Accessible visual artworks for blind and visually impaired people: Comparing a multimodal approach with tactile graphics. *Electronics (Switzerland)*, 10(3), 1–19. <https://doi.org/10.3390/electronics10030297>

- Rau, P. L. P., Li, Y., & Li, D. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2(2), 175–186. <https://doi.org/10.1007/s12369-010-0056-9>
- Reeves, Byron, Nass, C. (1999). Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. *Collection Management*, 24(3–4), 310–311. https://doi.org/10.1300/j105v24n03_14
- Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. In *Journal of Personality and Social Psychology* (Vol. 101, Issue 3, pp. 557–570). American Psychological Association. <https://doi.org/10.1037/a0022885>
- Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior*, 109. <https://doi.org/10.1016/j.chb.2020.106359>
- Rzepka, C., Berger, B., & Hess, T. (2022). Voice Assistant vs. Chatbot – Examining the Fit Between Conversational Agents’ Interaction Modalities and Information Search Tasks. *Information Systems Frontiers*, 24(3), 839–856. <https://doi.org/10.1007/s10796-021-10226-5>
- Šabanovic, S., Bennett, C. C., Chang, W. L., & Huber, L. (2013). PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. *IEEE International Conference on Rehabilitation Robotics*. <https://doi.org/10.1109/ICORR.2013.6650427>
- Sahib, N. G., Stockman, T., Tombros, A., & Metatla, O. (2013). Participatory design with blind users: A scenario-based approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8117 LNCS(PART 1), 685–701. https://doi.org/10.1007/978-3-642-40483-2_48
- Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4(1), 5–18. <https://doi.org/10.1080/15710880701875068>
- Sayago, S., Neves, B. B., & Cowan, B. R. (2019). Voice assistants and older people: Some open issues. *ACM International Conference Proceeding Series*, 22–24. <https://doi.org/10.1145/3342775.3342803>
- Sayago, S., & Ribera, M. (2020). Apple Siri (input) + Voice over (output) = a de facto marriage: An exploratory case study with blind people. *ACM International Conference Proceeding Series*, 6–10. <https://doi.org/10.1145/3439231.3440603>
- Scherer, S., Schwenker, F., Campbell, N., & Palm, G. (2009). *Multimodal Laughter Detection in Natural Discourses*. 111–120. https://doi.org/10.1007/978-3-642-10403-9_12
- Schmidt, S. R. (1994). Effects of humor on sentence memory. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 20, Issue 4, pp. 953–967). American Psychological Association. <https://doi.org/10.1037/0278-7393.20.4.953>
- Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2018). An investigation of conversational agent relevance, presence, and engagement. *Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018*.
- Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). “Hey Alexa, what’s up?”:

- Studies of in-home conversational agent usage. *DIS 2018 - Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868.
<https://doi.org/10.1145/3196709.3196772>
- Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in human-agent interaction: A survey. *ACM Computing Surveys*, 54(4).
<https://doi.org/10.1145/3386867>
- Seaborn, K., Sekiguchi, T., Tokunaga, S., Miyake, N. P., & Otake-Matsuura, M. (2023). Voice Over Body? Older Adults' Reactions to Robot and Voice Assistant Facilitators of Group Conversation. *International Journal of Social Robotics*, 15(2), 143–163. <https://doi.org/10.1007/s12369-022-00925-7>
- Sestino, A., & D'Angelo, A. (2023). My doctor is an avatar! The effect of anthropomorphism and emotional receptivity on individuals' intention to use digital-based healthcare services. *Technological Forecasting and Social Change*, 191(November 2022), 122505.
<https://doi.org/10.1016/j.techfore.2023.122505>
- Sharma, S. K., Janssen, M., Bunker, D., Dominguez-Péry, C., Singh, J. B., Dwivedi, Y. K., & Misra, S. K. (2023). Unlocking the Potential of Smart Technologies: Addressing Adoption Challenges. *Information Systems Frontiers*.
<https://doi.org/10.1007/s10796-023-10408-3>
- Sherzer, J. (1980). K. R. Scherer & H. Giles, eds., Social Markers in Speech. *L'Homme*, 20(3), 169–170.
http://www.persee.fr/web/revues/home/prescript/article/hom_0439-4216_1980_num_20_3_368124
- Shneiderman, B. (2016). *The New ABCs of Research: Achieving Breakthrough Collaborations* (1st ed.). Oxford University Press, Inc.
- Shneiderman B. & Plaisant C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Addison Wesley.
- Short, E., Hart, J., Vu, M., & Scassellati, B. (2010). No fair!! An interaction with a cheating robot. *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*, 219–226. <https://doi.org/10.1145/1734454.1734546>
- Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2004). *Where to look*. May 2014, 78.
<https://doi.org/10.1145/964442.964458>
- Smith, D. W., & Kelly, S. M. (2014). Chapter Two - Assistive Technology for Students with Visual Impairments: A Research Agenda. In D. D. Hatton (Ed.), *Current Issues in the Education of Students with Visual Impairments* (Vol. 46, pp. 23–53). Academic Press. <https://doi.org/10.1016/B978-0-12-420039-5.00003-4>
- Soash, R. L. (1999). Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. *Collection Management*, 24(3–4), 310–311. https://doi.org/10.1300/j105v24n03_14
- Søraa, R. A., Nyvoll, P., Tøndel, G., Fosch-Villaronga, E., & Serrano, J. A. (2021). The social dimension of domesticating technology: Interactions between older adults, caregivers, and robots in the home. *Technological Forecasting and Social Change*, 167(January). <https://doi.org/10.1016/j.techfore.2021.120678>
- Sorin, L., Lemarié, J., Aussenac-Gilles, N., Mojahid, M., & Oriola, B. (2014). Communicating text structure to blind people with text-to-speech. *Lecture Notes*

- in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8547 LNCS(PART 1), 61–68. https://doi.org/10.1007/978-3-319-08596-8_10
- Spatola, N., Anier, N., Redersdorff, S., Ferrand, L., Belletier, C., Normand, A., & Huguet, P. (2019). National stereotypes and robots' perception: The "made in" effect. *Frontiers Robotics AI*, 6(APR). <https://doi.org/10.3389/frobt.2019.00021>
- Sprecher, S., Treger, S., & Wondra, J. D. (2013). Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships*, 30(4), 497–514. <https://doi.org/10.1177/0265407512459033>
- Sprecher, S., Treger, S., Wondra, J. D., Hilaire, N., & Wallpe, K. (2013). Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology*, 49(5), 860–866. <https://doi.org/10.1016/j.jesp.2013.03.017>
- Stafford, R. Q., MacDonald, B. A., Li, X., & Broadbent, E. (2014). Older People's Prior Robot Attitudes Influence Evaluations of a Conversational Robot. *International Journal of Social Robotics*, 6(2), 281–297. <https://doi.org/10.1007/s12369-013-0224-9>
- Stahl, B. C., McBride, N., Wakunuma, K., & Flick, C. (2014). The empathic care robot: A prototype of responsible research and innovation. *Technological Forecasting and Social Change*, 84, 74–85. <https://doi.org/10.1016/j.techfore.2013.08.001>
- Stiehl, W. D., Lieberman, J., Breazeal, C., Basel, L., Cooper, R., Knight, H., Lalla, L., Maymin, A., & Purchase, S. (2006). The Huggable: A therapeutic robotic companion for relational, affective touch. *2006 3rd IEEE Consumer Communications and Networking Conference, CCNC 2006*, 2, 1290–1291. <https://doi.org/10.1109/CCNC.2006.1593253>
- Storer, K. M., Judge, T. K., & Branham, S. M. (2020). "All in the Same Boat": Tradeoffs of Voice Assistant Ownership for Mixed-Visual-Ability Families. *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. <https://doi.org/10.1145/3313831.3376225>
- Striepe, H., & Lugin, B. (2017). *There Once Was a Robot Storyteller: Measuring the Effects of Emotion and Non-verbal Behaviour BT - Social Robotics* (A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, & H. He (eds.); pp. 126–136). Springer International Publishing.
- Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior. In *Political psychology: Key readings*. Psychology Press. <https://doi.org/10.4324/9780203505984-16>
- Talukder, M. S., Sorwar, G., Bao, Y., Ahmed, J. U., & Palash, M. A. S. (2020). Predicting antecedents of wearable healthcare technology acceptance by elderly: A combined SEM-Neural Network approach. *Technological Forecasting and Social Change*, 150(June 2019), 119793. <https://doi.org/10.1016/j.techfore.2019.119793>
- Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, 109(21), 8038 LP – 8043. <https://doi.org/10.1073/pnas.1202129109>

- Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, A., Higashi, Y., Fujimooto, T., & Nakajima, K. (2004). Is an Entertainment Robot Useful in the Care of Elderly People with Severe Dementia? *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 59(1), 83–85. <https://doi.org/10.1093/gerona/59.1.m83>
- Tartaro, A., & Cassell, J. (2006). Running head: Virtual Peer Technology for Children with Autism Using Virtual Peer Technology as an Intervention for Children with Autism Andrea Tartaro and Justine Cassell. *CiteSeerX*, 1–44.
- Tay, B., Jung, Y., & Park, T. (2014). Computers in Human Behavior When stereotypes meet robots : The double-edge sword of robot gender and personality in human – robot interaction. *COMPUTERS IN HUMAN BEHAVIOR*, 38, 75–84. <https://doi.org/10.1016/j.chb.2014.05.014>
- Tay, B. T. C., Low, S. C., Ko, K. H., & Park, T. (2016). Types of humor that robots can play. *Computers in Human Behavior*, 60, 19–28. <https://doi.org/10.1016/j.chb.2016.01.042>
- Trajkova, M., & Martin-Hammond, A. (2020). “alexa is a Toy”: Exploring Older Adults’ Reasons for Using, Limiting, and Abandoning Echo. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. <https://doi.org/10.1145/3313831.3376760>
- Trovato, G., Ham, J. R. C., Hashimoto, K., Ishii, H., & Takanishi, A. (2015). Investigating the Effect of Relative Cultural Distance on the Acceptance of Robots. In A. Tapus, E. André, J.-C. Martin, F. Ferland, & M. Ammi (Eds.), *Social Robotics* (pp. 664–673). Springer International Publishing.
- Tsai, M. J., & Tsai, C. C. (2003). Information searching strategies in Web-based science learning: The role of Internet self-efficacy. *Innovations in Education and Teaching International*, 40(1), 43–50. <https://doi.org/10.1080/1355800032000038822>
- Turk, V. (2016). Home invasion. *New Scientist*, 232(3104), 16–17. [https://doi.org/https://doi.org/10.1016/S0262-4079\(16\)32318-1](https://doi.org/https://doi.org/10.1016/S0262-4079(16)32318-1)
- Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: The complexities of cybercompanionship. *Connection Science*, 18(4), 347–361. <https://doi.org/10.1080/09540090600868912>
- Urbach, G. (2011). Native intelligence. *Smithsonian*, 42(1). <https://doi.org/10.2307/3851848>
- van der Zwaan, J. M., Geraerts, E., Dignum, V., & Jonker, C. M. (2012). User validation of an empathic virtual buddy against cyberbullying. *Annual Review of CyberTherapy and Telemedicine*, 10(September), 243–247. <https://doi.org/10.3233/978-1-61499-121-2-243>
- van Pinxteren, M. M. E., Wetzels, R. W. H., Rüger, J., Pluymaekers, M., & Wetzels, M. (2019). Trust in humanoid robots: implications for services marketing. *Journal of Services Marketing*, 33(4), 507–518. <https://doi.org/10.1108/JSM-01-2018-0045>
- Vaportzis, E., Clausen, M. G., & Gow, A. J. (2017). Older adults perceptions of technology and barriers to interacting with tablet computers: A focus group study. *Frontiers in Psychology*, 8(OCT), 1–11. <https://doi.org/10.3389/fpsyg.2017.01687>

- Vaz, R., Fernandes, P. O., & Veiga, A. C. R. (2018). Designing an interactive exhibitor for assisting blind and visually impaired visitors in tactile exploration of original museum pieces. *Procedia Computer Science*, 138, 561–570. <https://doi.org/10.1016/j.procs.2018.10.076>
- Vaz, R., Freitas, D., & Coelho, A. (2020). Blind and visually impaired visitors' experiences in museums: Increasing accessibility through assistive technologies. *International Journal of the Inclusive Museum*, 13(2), 57–80. <https://doi.org/10.18848/1835-2014/CGP/V13I02/57-80>
- Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, 19(3), 529–545. <https://doi.org/10.1111/jcc4.12066>
- Vieira, A. D., Leite, H., & Volochtchuk, A. V. L. (2022). The impact of voice assistant home devices on people with disabilities: A longitudinal study. *Technological Forecasting and Social Change*, 184(July). <https://doi.org/10.1016/j.techfore.2022.121961>
- Vlahovic, T. A., Roberts, S., & Dunbar, R. (2012). Effects of duration and laughter on subjective happiness within different modes of communication. *Journal of Computer-Mediated Communication*, 17(4), 436–450. <https://doi.org/10.1111/j.1083-6101.2012.01584.x>
- Von Der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). “It doesn’t matter what you are!” Explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6), 1641–1650. <https://doi.org/10.1016/j.chb.2010.06.012>
- Vtyurina, A., Fourney, A., Morris, M. R., Findlater, L., & White, R. W. (2019). Bridging screen readers and voice assistants for enhanced eyes-free web search. *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 3590–3594. <https://doi.org/10.1145/3308558.3314136>
- Wang, J., Yang, H., Shao, R., Abdullah, S., & Sundar, S. S. (2020). Alexa as Coach: Leveraging Smart Speakers to Build Social Agents that Reduce Public Speaking Anxiety. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. <https://doi.org/10.1145/3313831.3376561>
- Weisband, S., & Kiesler, S. (1996). Self disclosure on computer forms: meta-analysis and implications. *Conference on Human Factors in Computing Systems - Proceedings*, 3–10.
- Wendt, C. S., & Berg, G. (2009). Nonverbal humor as a new dimension of HRI. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 183–188. <https://doi.org/10.1109/ROMAN.2009.5326230>
- Westlund, J. M. K., Park, H. W., Williams, R., & Breazeal, C. (2018). Measuring young children’s long-term relationships with social robots. *IDC 2018 - Proceedings of the 2018 ACM Conference on Interaction Design and Children, June*, 207–218. <https://doi.org/10.1145/3202185.3202732>
- Wheless, L. R., & Grotz, J. (1977). The Measurement of Trust and Its Relationship to Self-Disclosure. *Human Communication Research*, 3(3), 250–257. <https://doi.org/10.1111/j.1468-2958.1977.tb00523.x>
- Wilson, C. (1979). *Jokes: Form, content, use, and function*.

- Wu, Y. H., Fassert, C., & Rigaud, A. S. (2012). Designing robots for the elderly: Appearance issue and beyond. *Archives of Gerontology and Geriatrics*, 54(1), 121–126. <https://doi.org/10.1016/j.archger.2011.02.003>
- Wulf, L., Garschall, M., Himmelsbach, J., & Tscheligi, M. (2014). Hands free-Care free: Elderly people taking advantage of speech-only interaction. *Proceedings of the NordiCHI 2014: The 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, 203–206. <https://doi.org/10.1145/2639189.2639251>
- Wyse, L., & Subramainan, S. (2012). Foundations of interactive sound design for traditional storytelling. *ICMC 2012: Non-Cochlear Sound - Proceedings of the International Computer Music Conference 2012, November*, 242–246.
- Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2019). *Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions*. 27(3). <https://doi.org/10.1145/3381804>
- Yamashita, N., & Huang, Y. U. N. (2020). *Let ' s Work Together : Integrating Human Support with Conversational Agents*. 1(1), 1–13.
- Yamazaki, R., Kochi, M., Zhu, W., & Kase, H. (2018). A Pilot Study Of Robot Reminiscence In Dementia Care. *International Journal of Biomedical and Biological Engineering*, 12(6), 257–261. <http://ezproxy.si.unav.es:2048/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url&db=edsbas&AN=edsbas.AF2CFBEE&lang=en&site=eds-live&scope=site%0Ahttp://ezproxy.si.unav.es:2048/login?url=https://search.ebscohost.com/login.aspx?d>
- Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. <https://doi.org/10.1145/3313831.3376301>
- Yeong Tan, D. T., & Singh, R. (1995). Attitudes and Attraction: A Developmental Study of the Similarity-Attraction and Dissimilarity-Repulsion Hypotheses. *Personality and Social Psychology Bulletin*, 21(9), 975–986. <https://doi.org/10.1177/0146167295219011>
- Yu, Q., Nguyen, T., Prakkamakul, S., & Salehi, N. (2019). “I almost fell in love with a machine”: Speaking with Computers Affects Self-disclosure. *Conference on Human Factors in Computing Systems - Proceedings*, 1–6. <https://doi.org/10.1145/3290607.3312918>
- Yu, Y., & Liu, Z. (2010). A user study of visual versus sonically-enhanced interfaces for use while walking. *MM'10 - Proceedings of the ACM Multimedia 2010 International Conference*, 687–690. <https://doi.org/10.1145/1873951.1874052>
- Yücel, Y., & Rızvanoğlu, K. (2019). Battling gender stereotypes: A user study of a code-learning game, “Code Combat,” with middle school children. *Computers in Human Behavior*, 99. <https://doi.org/10.1016/j.chb.2019.05.029>
- Zhang, A., & Rau, P. L. P. (2022). She is My Confidante! The Impacts of Social Responsiveness and Video Modality on Self-disclosure Toward CG-Based Anthropomorphic Agents in a Smart Home. *International Journal of Social*

- Robotics*, 14(7), 1673–1686. <https://doi.org/10.1007/s12369-022-00895-w>
- Zhang, M., Gursoy, D., Zhu, Z., & Shi, S. (2021). Impact of anthropomorphic features of artificially intelligent service robots on consumer acceptance: moderating role of sense of humor. *International Journal of Contemporary Hospitality Management*, 33(11), 3883–3905. <https://doi.org/10.1108/IJCHM-11-2020-1256>
- Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 1. <https://doi.org/10.18653/v1/p18-1205>
- Zhong, Y., Raman, T. V., Burkhardt, C., Biadys, F., & Bigham, J. P. (2014). JustSpeak: Enabling universal voice control on android. *W4A 2014 - 11th Web for All Conference*. <https://doi.org/10.1145/2596695.2596720>
- Ziman, R., & Walsh, G. (2018). Factors affecting seniors' perceptions of voice-enabled user interfaces. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*, 1–6. <https://doi.org/10.1145/3170427.3188575>
- Ziv, A. (1984). *Personality and sense of humor* / Avner Ziv. Springer Pub. Co.
- Ziv, A. (1988). Teaching and learning with humor: Experiment and replication. *Journal of Experimental Education*, 57(1), 5–15. <https://doi.org/10.1080/00220973.1988.10806492>
- Zubatiy, T., Vickers, K. L., Mathur, N., & Mynatt, E. D. (2021). Empowering dyads of older adults with mild cognitive impairment and their care partners using conversational agents. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411764.3445124>
- Zuckerman, O., Walker, D., Grishko, A., Moran, T., Levy, C., Lisak, B., Wald, I. Y., & Erel, H. (2020). Companionship Is Not a Function: The Effect of a Novel Robotic Object on Healthy Older Adults' Feelings of "being-Seen." *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. <https://doi.org/10.1145/3313831.3376411>

APPENDICES

Appendix A

A dialogue example with one of our participants from Chapter II

M13: Hey Google, please advise me of festival films on the Blutv platform

CA: Hey, back to you! Of course, I can. When it comes to festival films, I am the best there is. We will find something you would love asap. Since you love festival films, would you love to continue with local or international festival films? I, personally so keen on local festival films. What do you think?

M13: Yes, I have watched some successful locals at a recent festival. However, I prefer international films right now.

CA: I have seen many good international festival films on Blutv. Recently, I have seen Maiwen Le Besco's *Mon Roi*, and I adore it. The movie's primary subject was utterly unoriginal, but it was handled in such a gripping way. Every spoken word opened up old wounds for me. Spectacular acting. It is all about the detrimental and destructive aspects of love.

M13: I felt you, bro, but I have to interrupt. I have just got out of this kind of love, and I feel like watching easy to watch and without any drama festival movie. I will love to see some good looking actors or actresses if you have that category

CA: My favourite, if you have not watched it before, is, of course, *Thelma and Louise*. You can not point someone who watched this film and could stop himself/herself from falling in love with Geena Davis. Brad Pitt also first appeared in cinema in this film with all his charm. A feminist road film. Did you watch it? Are there any particular directors that you follow?

M13: (User nodded when mentioned Geena Davis) Stop and let me talk a little about Geena Davis (laughing). I like her, and I remember her from *Beetlejuice*. I watched that movie 5-6 times because of her. You are more than right; she is something. Sorry, but I do not feel like seeing feminist propaganda right now, even with Geena Davis. However, I wonder, when was the release date of this film?

CA: 1991, ABD production, Director Ridley Scott, IMDB rating 7.5.

M13: Hey, I said cult, but it is too old, bro, and feminist (laughing). I can stand that type of feminist propaganda. Can you advise me on something more up-to-date?

CA: Of course. Does IMDB rating matter to you? I know it sounds brutal, but I am always doubtful about films with lower IMDB ratings. We can directly eliminate the movies with lower than 8 IMDB ratings if you want.

M13: No, I am not as ruthless as you are about IMDB rating. I prefer to watch an up-to-date film with good quality and fair resolution supported by cutting-edge technology.

CA: I feel like we have many things in common. If you do not care about IMDB rating, I will advise you to watch one of my recent favourite directors' films. It is Alonso Ruizpalacios's *Museo* which was released in 2018. The lead role, Gael Garcia Bernal, makes everything watchable for me, and it has an exceptional place at the bottom of

my heart since the release of *Amores Perros*. These kinds of things also happen to you? Can I cut the talking and let you watch the Museo?

M13: I link films and actors with persons and moments in my life. When you said *Amores Perros*, I remembered that I kissed my first love while watching *Amores Perros* so that I had to watch the film again after our date (laughing). Do not start the film yet. We can chat a bit more.

CA: By the way, I could not help myself but ask who your favourite actor and actress is? Mine is Penelope Cruz. I cannot resist Penelope Cruz every time I see her in movies. She is so adorable and noble at the same time. I would just move to Madrid to meet or kidnap her if she was not married.

M13: (laughing) Is she married? I am also upset to hear that, bro. Mine is Leonardo Di Caprio. I would also like to kidnap and go all over him if I was not straight. Scarlet Johansson shakes my world every time with her beauty and acting. However, she seemed very tired and old in her last movie, maybe because of her part in that movie. You cut the small talk and just moved to the deeper stuff, hein? Museo caught my attention. Nevertheless, I would like to have a second option? Tell me one of your favourites and give me some hints? A fast-moving one.

CA: It is a tough choice. I would prefer *Fargo* for eternity. 1996, ABD production. The Cohen Brothers' bloody movie with good music and Steve Buscemi Factor. Do you remember the legendary character Jesus Quintana in *Big Lebowski*? Sheep Proudfoot is his exact doppelganger at *Fargo* for me. I am sick of these two fellas.

M13 : (user nodded) Let me tell you that you boosted me, bro. Even though it is an old film, it is a must-see. I do not know why I have skipped it. Could you please start the Museo? I will watch *Fargo* with my roommates. Jesus Quintana is a character that we love to imitate to amuse ourselves. We watched the *Big Lebowski* together last year, and it would be good to regather to watch *Fargo*. You gave me an idea.

CA: Ok, I will start the movie that you chose. You are an excellent cinephile. Before ending our conversation, could you please share with me your favourite film? I want to gain more insights from you.

M13: I can say *Inception*. It was an immersive film and made me feel like an idiot at the end. I love persons and films which have twists, I guess. Before saying goodbyes to each other, do you have any historical films in Blutv's festival selection?

CA: Yes. It is Emir Kusturica's *Underground* which haunts me over and over again. A film about the dissolution of Yugoslavia. During this movie, crying and drinking wine became a ritual for me. What do you think? Do you like Balkan movies?

M13: (User used turn-taking at the end of the third sentence then said sorry for interrupting) I adore Balkan movies. I am also Bosniak, and I would probably cry too if I watched that film. Can you tell me more? I guess it is an old film, but I would love to hear more from you.

CA: Released in 1995. IMDB Rating 8.1.

M13 : (User turn-taking) Wow, it is old. Nevertheless, I did not see it.

CA: Winking at magic realism, Kusturica told the story of Yugoslavia and Belgrad's collapse during WWII. I always felt the Kafka books' dark ambience that thrills me while watching that movie. For me, Kusturica told the nonsense of the war by using absurd humour. This film made me laugh and cry at the same time.

M13: You are so damn right. This film would tear me up since I was witnessed a war when I was little. However, I am inquisitive about the kinds of music; Kusturica probably used Goran Bregović's music pieces, which I adore. Ok, let us stop here and watch Museo.

CA: Perfect choice. It was very nice meeting you. Enjoy the movie.

Appendix B

APPENDIX B.1

Pre-Visit Quiz Questionnaire

What do you know about the Basilica Cistern in Istanbul?

What do you expect to experience in terms of touch and smell during your virtual visit?

Based on what you currently know, how would you describe the Basilica Cistern to someone else?

Can you name any structures or elements that you might encounter in the Basilica Cistern?

Why do you think the Basilica Cistern was built?

Post-visit Quiz Questionnaire

APPENDIX B.2

Post-Visit Quiz Questionnaire (Multiple Choice Format)

Who is credited with rediscovering the Basilica Cistern in the 16th century?

- a) Sultan Ahmed III
- b) Petrus Gyllius
- c) Suleiman the Magnificent

What was Petrus Gyllius doing when he discovered the Basilica Cistern?

- a) Looking for ancient manuscripts
- b) Surveying the city's water systems
- c) Exploring underground Istanbul for treasures

What material was mainly used in the construction of the Basilica Cistern?

- a) Wood
- b) Marble
- c) Iron

How many years does it take to build the Basilica Cistern?

- a) 38
- b) 41
- c) 66

Why is the column mentioned above also known as the Crying or Weeping Column?

- a) Because of the dampness or wetness that often covers the column
- b) Because it's said to mourn the slaves who died during the Cistern's construction
- c) Because it's carved to look like it has a sad face

Who fell in love with Medusa, leading to her transformation into a monster?

- a) Poseidon
- b) Pirius

c) Athena

How many sculptures can you find in the Basilica Cistern?

- a) There are no sculptures, only columns
- b) Two - the Medusa head sculptures
- c) More than two

How many columns can you find in the Basilica Cistern?

- a) 336 (28x12)
- b) 336 (14x24)
- c) 330 (33x 10)

Based on the audio descriptions, how were the Medusa columns positioned?

- a) Both upright
- b) Both upside down
- c) One upside down, the other on its side

What significant historical fact did you learn about the Basilica Cistern from the audio tour?

- a) It was built during the reign of Emperor Justinian I
- b) It was built during the reign of Sultan the Magnificent
- c) It was discovered in 18th century

CURRICULUM VITAE

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

TEZ ONAY SAYFASI

Üniversite : T.C. GALATASARAY ÜNİVERSİTESİ
Enstitü : SOSYAL BİLİMLER ENSTİTÜSÜ
Hazırlayanın Adı Soyadı : Yeliz YÜCEL
Tez Başlığı : Dialogue for All: Crafting Inclusive and Humanized Voice Assistants for Diverse Populations through an Interdisciplinary Approach
Savunma Tarihi : 05 / 09 / 2023
Danışman : Prof. Dr. Kerem RIZVANOĞLU

JÜRİ ÜYELERİ

:

Unvan, Ad-Soyadı

İmza

Prof. Dr Kerem RIZVANOĞLU

.....

Prof. Dr H. Serhat GÜNEY

.....

Dr. Öğr Üyesi Çakır AKER

.....

Dr. Öğr Üyesi Akgün TOKATLI

.....

Doç. Dr Behiç Alp AYTEKİN

.....

Prof. Dr. Ulun Akturan

Enstitü Müdürü