TURUN YLIOPISTO
UNIVERSITY OF TURKU

Deepfake User Personas for Good

From Maleficent to Helping Users

Ilkka Kaate

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. E OSA – TOM. 138 | OECONOMICA | TURKU 2025

TURUN YLIOPISTO
UNIVERSITY OF TURKU

DEEPFAKE USER PERSONAS FOR GOOD

From Maleficent to Helping Users

Ilkka Kaate

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. E OSA – TOM. 138 | OECONOMICA | TURKU 2025

University of Turku

Turku School of Economics
Department of Marketing and International Business
Marketing
Doctoral Programme of Turku School of Economics

Supervised by

Professor Rami Olkkonen
Turku School of Economics
University of Turku
Finland

Associate Professor Joni Salminen
University of Vaasa
Finland

Reviewed by

Professor Jürgen Ziegler
Universität Duisburg-Essen
Germany

Associate Professor Cristina Gena
Università degli Studi di Torino
Italy

Opponent

Professor Jürgen Ziegler
Universität Duisburg-Essen
Germany

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Cover Image: Kokotewan / Adobe Cloud

ISBN 978-952-02-0428-0 (PRINT)
ISBN 978-952-02-0429-7 (PDF)
ISSN 2343-3159 (Print)
ISSN 2343-3167 (Online)
Painosalama, Turku, Finland 2025

UNIVERSITY OF TURKU
Turku School of Economics
Department of Marketing and International Business
Marketing
ILKKA KAATE: Deepfake User Personas for Good – From Maleficent to Helping Users
Doctoral Dissertation, 156 pp.
Doctoral Program of Turku School of Economics
October 2025

ABSTRACT

This dissertation explores the emerging role of deepfake personas in human–computer interaction and marketing, investigating their usability, user perceptions, and ethical considerations in design tasks. While deepfake technology is often associated with malicious use, this study shifts the focus toward deepfakes' constructive potential, particularly in the creation of deepfake user personas – AI-generated video representations that embody persona attributes such as goals, needs, and behavioral traits. These deepfake personas offer a novel way to enhance user-centered design by providing potentially more immersive, emotionally engaging, and humanlike representations of target users. However, challenges such as the uncanny valley effect, emotional and demographic biases, and technical inconsistencies raise questions about the credibility and inclusiveness of deepfake personas. The dissertation consists of five empirical studies addressing key research gaps, including comparisons between deepfake, human, and conventional persona modalities, the effects of deepfake distortions, glitches, on user perceptions of and behavior with deepfakes, and the demographic and emotional diversity of commercially available deepfake avatars. Applying both qualitative and quantitative methods, the dissertation provides insights into how users perceive deepfake personas, how deepfake persona perceptions affect users' task performance, and how design and marketing professionals might responsibly implement deepfake personas. The findings reveal that while DFs can improve empathy and engagement, their effectiveness depends on technical quality, ethical transparency, and alignment with user expectations. By investigating deepfake personas across experimental and observational contexts, this dissertation contributes to the academic understanding of persona use in digital environments and proposes practical guidelines for ethically and effectively integrating deepfake technology into design workflows.

KEYWORDS: human-computer interaction, user experience, deepfake, deepfake persona

TURUN YLIOPISTO
Turun Kauppakorkeakoulu
Markkinoinnin ja kansainvälisen liiketoiminnan laitos
Markkinointi
ILKKA KAATE: Deepfake-käyttäjäpersoonat hyvän asialla – pahantahtoisuudesta käyttäjien avuksi
Väitöskirja, 156 s.
Turun kauppakorkeakoulun tohtoriohjelma
Lokakuu 2025

TIIVISTELMÄ

Tämä väitöskirja tarkastelee deepfake-persoonien kasvavaa roolia ihmisen ja tietokoneen välisessä vuorovaikutuksessa sekä markkinoinnissa. Tutkimus selvittää deepfake-persoonien käytettävyyttä, käyttäjäkokemuksia ja eettisiä näkökulmia suunnittelutehtävissä. Vaikka deepfake-teknologia yhdistetään usein haitalliseen käyttöön, tässä tutkimuksessa huomio on deepfake-teknologian rakentavassa potentiaalissa, erityisesti deepfake-käyttäjäpersoonien luomisessa. Deepfake-käyttäjäpersoonilla tarkoitetaan tekoälyn luomia videohahmoja, jotka ilmentävät käyttäjäpersoonien ominaisuuksia, kuten tavoitteita, tarpeita ja käyttäytymismalleja. Deepfake-käyttäjäpersoonat tarjoavat uuden tavan vahvistaa käyttäjäkeskeistä suunnittelua tarjoamalla käyttäjistä mahdollisesti immersiivisempiä, tunteellisesti puhuttelevampia ja inhimillisempiä esityksiä. Samalla deepfake-käyttäjäpersooniin liittyy kuitenkin haasteita, kuten outo laakso -ilmiö, tunteelliset ja demografiset vinoumat sekä tekniset epäjohdonmukaisuudet, jotka herättävät kysymyksiä deepfake-käyttäjäpersoonien uskottavuudesta ja inklusiivisuudesta. Väitöskirja koostuu viidestä empiirisestä tutkimuksesta, jotka vastaavat keskeisiin tutkimus-aukkoihin. Näihin kuuluvat muun muassa vertailu deepfake-, ihmispohjaisten ja perinteisten persoonamuotojen välillä, deepfake-häiriöiden ja virheiden vaikutukset käyttäjien havaintoihin ja toimintaan sekä kaupallisten deepfake-avatarien demografinen ja emotionaalinen monimuotoisuus. Laadullisia ja määrällisiä menetelmiä hyödyntäen väitöskirja tarjoaa uutta tietoa siitä, miten käyttäjät kokevat deepfake-käyttäjäpersoonat, miten nämä kokemukset vaikuttavat tehtävistä suoriutumiseen ja kuinka suunnittelu- ja markkinointialojen ammattilaiset voivat ottaa deepfake-käyttäjäpersoonat käyttöön vastuullisesti. Tulokset osoittavat, että deepfake-käyttäjäpersoonat voivat parantaa empatiaa ja sitoutumista, mutta deepfake-käyttäjäpersoonien vaikuttavuus riippuu teknisestä laadusta, eettisestä läpinäkyvyydestä ja käyttäjien odotusten täyttymisestä. Tarkastelemalla deepfake-käyttäjäpersoonia sekä kokeellisissa että havainnoivissa konteksteissa väitöskirja syventää akateemista ymmärrystä deepfake-käyttäjäpersoonien käytöstä digitaalisissa ympäristöissä ja esittää käytännön suosituksia deepfake-teknologian eettisestä ja tehokkaasta hyödyntämisestä suunnitteluprosesseissa.

AVAINSANAT: ihmisen ja tietokoneen vuorovaikutus, käyttäjäkokemus, deepfake, deepfake-persoona

Acknowledgements

In the late 1990's we were driving by the University of Turku campus with my mother, sister, and my grandfather's friend Kaarina. Kaarina told us that one day me and my sister would be on that hill, referring to the University campus hill. I did not understand the manifestation at that moment but now looking back, it is quite easy to see. Kaarina saw something I could not see, and she planted a seed in my and my sister's mind. Cheers!

I want to thank my opponent Professor Jürgen Ziegler for his expertise and time. I also want to thank Associate Professor Cristina Gena for being the pre-examiner for my dissertation. Most valuable contribution, much obliged. I also want to thank Associate Professor Joni Salminen for taking me in his project and serving as my supervisor. I have learnt a lot. I am also grateful, and I want to thank Professor Rami Olkkonen for being my supervisor and offering academic, and non-academic, wisdom and stories. Displacement is what keeps the boat afloat.

I studied and graduated from geography (on that very same university hill) and during my studies in geography, I saw a lot of people preparing their doctoral thesis in geography and wondering that maybe that could be something for me as well. One major contributor to my thinking and an example of scientific work motivation was my boss at the geography department, Niina Käyhkö. My close friend from geography, Juuso Suomi, defended his dissertation in 2014 which gave me more latent motivation for my own coming thesis work, although not in geography. As a struck of destiny, I ended up studying marketing at the Turku School of Economics (TSE) in 2015. During my studies at TSE, I was employed in an ad agency, working with my professional mentor and friend Tommi Salenius. Tommi taught me a lot about digital marketing and channeled information from his own professional mentor Joni Salminen with whom I got acquainted during the years I worked with Tommi. After a few years, I ended up working on Joni's project, one fruit of which is my doctoral dissertation. The trip has been the most educational yet fast. Joni introduced me to a research group that included a master coder Soon-Gyo Jung and all-around master Bernard J. (Jim) Jansen. Both these gentlemen I owe my gratitude. I also want to thank my colleagues at the TSE marketing department and my co-authors for constant support, discussions, and mood-lifting.

During the past few years, I have found relaxation and comfort from the arts, yet I still struggle with the definition of art. Music has played a big role in my life for a long time now and more push to my music enthusiasm was given by our geography band Icefingers. I thank Jouni, Harri, Joni, Matti, and Lasse (Lasse, maybe we will meet one day…) for the well-needed mix of feelings and sense of success and belonging. In 2019, I joined the TSE choir, Cantus Mercurialis (CM). The people in CM have given such joy to my life that it is hard to describe. I have found friends in CM, I have learnt a lot about music in CM, and I have developed skills in reading music and how to listen to music. The same with the University of Turku Choir and also with the male choir Naskalit. Positive energy, positive experiences, and good friends to support life and, moreover, my work on the dissertation.

In 2021, I found theater. From theater I found Suvi, as much as Suvi found me, and we became partners in life. Thank you, Suvi, for everything. You are giving me the gift of parenthood in the form of what we now call Zelda. I also want to thank Olavi for taking me, as a novice, in his stage production. We never guess the positive implications of our actions. It takes faith in one to see what others themselves cannot see. Theater also boosted my enthusiasm and opportunities in music and from that side of theater I want to thank especially Aleksi, Antti, Elina (Tee), Elina (Tuo), and Petteri. Definitely not least, I thank my family. Mother, father, and sister. Their constant support, although not necessarily always outspoken but sensed, has driven me to the point in life where I am now. The learnings from my childhood and from my youth have taught me things that cannot be taught. Things that can only be observed, absorbed, and refined.

October 2025
Ilkka Kaate

Table of Contents

Figures

Tables

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Kaate, I., Salminen, J., Santos, J., Jung, S.-G., Olkkonen, R., Jansen, B. The Realness of Fakes: Primary Evidence of the Effect of Deepfake Personas on User Perceptions in a Design Task. *International Journal of Human – Computer Studie*, 2023; 178: 1–21. https://doi.org/10.1016/j.ijhcs.2023.103096

II Kaate, I., Salminen, J., Jung, S.-G., Almerekhi, H.., Jansen, B. How Do Users Perceive Deepfake Personas? Investigating the Deepfake User Perception and Its Implications for Human-Computer Interaction. *CHItaly '23: Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter*, 2023; 15: 1–12. https://doi.org/10.1145/3605390.3605397

III Kaate, I., Salminen, J., Santos, J., Jung, S.-G., Almerekhi, H., Jansen, B. ”There Is something Rotten in Denmark”: Investigating the Deepfake Persona Perceptions and Their Implications for Human-centered AI. *Computers in Human Behavior: Artificial Humans*, 2024; 2: 1–14. https://doi.org/10.1016/j.chbah.2023.100031

IV Kaate, I., Salminen, J., Al Tamine, R., Jung, S.-G., Jansen, B. Is Deepfake Diversity Real? Analyzing the Diversity of Deepfake Avatars. *Expert Systems With Applications*, 2025; 269: 1–12. https://doi.org/10.1016/j.eswa.2025.126382

V Kaate, I., Salminen, J., Jung, S.-G., Rizun, N., Revina, A., Jansen, B. Getting Emotional Enough: Analyzing Emotional Diversity in Deepfake Avatars. *NordiCHI '24: Proceedings of the 13th Nordic Conference on Human-Computer Interaction*, 2024; 13: 1–12. https://doi.org/10.1145/3679318.3685398

1 Introduction

1.1 Background and Aims of the Study

Deepfakes, referring to images, videos, or audio created or modified with artificial intelligence (AI) tools, capable of depicting both real and fictional humans, carry the reputation of being used for malicious purposes (Ahmed, 2021; Ahmed et al., 2023; Appel & Prietzel, 2022; Bode, 2021; Bray et al., 2022; Chadha et al., 2021; Fallis, 2021). However, although deepfake technology has been used for malicious purposes, deepfakes could be used for good purposes. One example of deepfake use for good is user personas (later *personas*). Personas are fictional characters that represent real user groups first presented in goal-directed design to help designers understand the end users better (Cooper, 1999; Nielsen, 2019; Rönkkö et al., 2004). I.e., to bring end users closer to designers. User can refer to (a) the people using personas, hence, persona user, or (b) to the user whom the persona represents, hence, the end user. Personas have been conventionally presented as flat-file personas (conventional personas), such as PDFs and posters, but at the emergence of AI and machine learning (ML) into various domains in human life, the use of deepfake technology has become possible also with personas introducing deepfake personas (DFs) which are deepfakes carrying the properties conventionally found in personas such as persona goals, needs, and wants (Cooper, 1999). On the other hand, a deepfake avatar (DA) is a digital entity, the outer shell of a deepfake, with a human-like appearance, generated by computer software to engage, inform, and interact with users across information systems, applications, and services. DAs are used as the facial and bodily models for image and video deepfakes (Aljaroodi et al., 2019; Kaate, Salminen, Jung, Rizun, et al., 2024; Kaate, Salminen, Al Tamime, et al., 2025).

AI and ML have significantly transformed human-computer interaction (HCI) and marketing, among others, in the past few years (Duvvuri, 2021; Kaate, Salminen, Jung, Santos, et al., 2025; Kaate, Salminen, Jung, Xuan, et al., 2025; Salminen, Mustak, et al., 2022). Among the advancements in AI and ML, deepfake technology has emerged as a powerful tool with the potential to also revolutionize how personas are created, utilized, and perceived. Personas themselves have been under scientific interest since the late 1990's when (Cooper, 1999) published his work *The Inmates*

Are Running the Asylum although personas as such had been used before, for example in marketing (Caballero et al., 2014), but Cooper formulated properly the outline of persona creation and persona use in design for the first time in the field of HCI as design tool. DFs hold promise for enhancing user experience (UX) and providing more personalized and engaging interactions in marketing and HCI.

The increasing sophistication of deepfake technology has opened new avenues for persona development and persona research, particularly regarding users' perceptions of and interactions with personas (Kaate, Salminen, Jung, Xuan, et al., 2025; Salminen, Jung, et al., 2021). The significance of deepfake technology lies in deepfake technology's ability to create personas that can mimic real humans (Bode et al., 2021). Deepfake technology's capability extends beyond mere visual replication since it holds, optimally, the subtle expressions and behaviors of a human being that make human interaction more natural and relatable, potentially making DFs a prominent tool for design (Renier et al., 2024). In marketing, DFs can enhance customer engagement by providing tailored experiences resonating with customers on a personal level (Whittaker et al., 2021). This emotional depth can enhance UX in various applications, from customer service to educational tools. In HCI, DFs can facilitate better user-centered design (UCD) by offering realistic and empathetic user representations (Danry et al., 2022). However, the use of deepfakes has also raised ethical concerns and practical challenges that must be carefully addressed (Ahmed, 2021; Fallis, 2021). There is also the question about how realistic deepfakes really are and how users perceive deepfakes (Barari et al., 2021; Preu et al., 2022; Wagner & Blewer, 2019), a question most famously raised already in the 1970's by (Mori, 1970).

There are challenges associated with the use of deepfake technology for good, one concern being the deepfakes potentially being perceived as eerie or unsettling, a phenomenon known as the uncanny valley effect (Mori, 1970). The uncanny valley effect occurs when a digital representation, such as deepfake, is perceived as being very close to human but has slight imperfections that make it appear very strange or unnatural to users. DFs can display a range of emotions and behaviors that make interactions with them more natural and relatable, yet imperfections like uncanny valley effect can occur. In the context of deepfakes, the effects of uncanny valley have remained largely unexplored yet addressing the challenge of uncanniness is crucial for ensuring that DFs are effective and well-received by DF users.

Deepfake technology has the potential to transform how personas are created and deployed. Conventional personas, persona profiles, while useful, often lack the dynamism and emotional depth that can truly engage users, customers, and designers (Bonnardel et al., 2016; Kaate, Salminen, Jung, Santos, et al., 2025). DFs, by contrast, could offer a more immersive and realistic approach to personas, making it easier for designers and other persona users to empathize with and understand the needs, goals, and behaviors of different user groups, which is the primary purpose of personas (Cooper, 1999). However, the realism of DFs can be misleading leaving room for users' different perceptions about the realism and humanlikeness of the DFs and DF perceptions are, still, quite unexplored grounds in persona research.

The integration of deepfake technology into HCI, design, and marketing processes can significantly enhance the effectiveness of personas. By creating more humanlike and engaging persona representations, DFs can help bridge the gap between user needs and design solutions. For instance, in a design task, DFs could convey user emotions and reactions more vividly, helping designers create products and services that are more aligned with user expectations and preferences (Danry et al., 2022). In the marketing context, also, DFs can serve as useful tools to make it easier for marketers to empathize with potential customers thus making it easier for the marketers to create marketing that resonates with the customers (Whittaker et al., 2021). Affecting the usability of deepfakes are also the emotional and demographic properties of the deepfakes and DAs. AI-generated imagery, such as deepfakes, have been found to include emotional and demographic biases possibly affecting the usability of DFs. However, what remains unclear are the current extent and effect of DAs' and DFs' current emotional and demographic capabilities on the usability of DAs and DFs.

From the marketing point of view, personas and DFs are tools for market segmentation (Chandra et al., 2022; Revella, 2015; Whittaker et al., 2021). Market segmentation is the practice of dividing the consumer or business market into smaller groups of customers based on shared characteristics (An, Kwak, Jung, Salminen, & Jansen, 2018). This process is seen as crucial for developing targeted marketing strategies that resonate with specific customer segments, i.e., customer groups. DFs have a role in enhancing marketing segmentation by providing more detailed and realistic representations of these customer segments. For example, DFs can be used to create customized marketing messages that appeal to different customer groups (Campbell, Plangger, Sands, & Kietzmann, 2022; Campbell, Plangger, Sands, Kietzmann, et al., 2022). By, possibly, reflecting the diverse emotions, behaviors, and preferences of various customer segments better than conventional personas, DFs can help marketers develop more effective and personalized marketing campaigns. This personalized approach can lead to higher engagement with the products and services of the company and better customer satisfaction, as the marketing messages are more likely to resonate with the target customers (Casaca & Miguel, 2024).

DAs themselves could have a potential effect on the users' DF perceptions through emotional and demographic properties of the DAs (Aljaroodi et al., 2019; Becker et al., 2024). DAs, as they are trying to reflect the wide variety of humans, vary in their facial emotions, age, gender, and ethnic group making it worthwhile to ask whether these characteristics affect DF perceptions and do deepfake service providers have gaps in, for example, demographics of DAs compared to the world-wide user base of the deepfakes. Creating deepfakes is, at the moment, possible through commercial deepfake service providers which offer a selection of DAs for deepfake creation with limited emotional and demographic properties, possibly affecting users’ persona perceptions.

So, DFs have potential helping HCI designers and marketers to better reach the goals, needs, and wants of the targeted end users and audiences but despite the potential of DFs, *user perceptions of DFs in design tasks, i.e., for good purposes, has not been examined in depth*. This dissertation aims to explore DFs’ use for good, how DFs are perceived by DF users, and what challenges are associated in using DFs in design tasks. By examining the theoretical frameworks and practical implications of using deepfake technology especially in HCI, this dissertation seeks to provide a comprehensive understanding of how DFs can be effectively and ethically integrated into use for good purposes.

This dissertation addresses these challenges by conducting five empirical studies that together provide a comprehensive view of DFs in design and marketing contexts. Articles I-III examine DF user perceptions, to demonstrate both the DFs’ potential to increase empathy and engagement in users and the vulnerabilities associated with technical glitches and uncanny realism. Articles IV-V shift the focus to the broader ecosystem of DA generation, exposing demographic and emotional biases in commercial DA platforms. Synthesizing these findings, the dissertation demonstrates that DFs can be valuable tools for UCD if DFs are developed with both perceptual authenticity and systemic inclusivity in mind.

1.2 Research Gap and Objectives

How users perceive DFs is influenced by multiple factors (Becker et al., 2024; Gupta et al., 2020; Preu et al., 2022), and the effects of DF perceptions on DF user behavior are also diverse (Bode, 2021; Weisman & Peña, 2021). While there has been research on viewing patterns and perceptions of deepfake videos and audio (M. Li et al., 2021; Wang et al., 2022), there is a gap in comprehensive empirical studies focusing on users' perceptions of DFs in experimental settings. Studying DF perceptions is crucial because these perceptions can significantly influence the usability and effectiveness of DF applications, such as how DFs can be used in design tasks. Studies exploring how users respond to and perceive DFs are scarce. Thus, studies on DF perceptions’ impact on DF use are needed.

The research gap of this dissertation consists of the research gaps of the individual articles forming the dissertation (Table 1). Firstly, there is a lack of knowledge in how DFs are perceived compared to narrative (text) and classic persona (conventional flat-file persona profile) modalities. Personas come in different modalities, and the user perceptions of classic (Salminen, Nielsen, et al., 2018) and narrative personas (Kunur et al., 2016; Nielsen, 2004) have been studied and factors affecting the persona usability of classic and narrative persona modalities have been explored yet user perceptions of DFs' effect on user behavior in a design task are scarce. Prior studies have found that making deepfakes as humanlike as possible has a positive effect on deepfake perceptions (Chadha et al., 2021) but this also introduces the *uncanny valley* problem (Mori, 1970) when nearly reaching a complete humanlikeness but not quite makes the deepfake eerie. Having a deepfake almost humanlike has the potential of making the deepfake disturbing for the users affecting deepfake perception (Weisman & Peña, 2021).

Secondly, user perceptions of deepfakes have been studied (Becker et al., 2024; Preu et al., 2022) as well as how users perceive humans (Bertoni et al., 2022). However, the user perception differences of DFs and human personas (personas acted by humans in videos) in an experimental setting have not been studied in depth. Knowing the perception differences between DFs and human personas is crucial because it helps identify the unique attributes and potential shortcomings of DFs compared to human personas and reveals the possible factors affecting DF and human persona perceptions.

Thirdly, deepfakes are known to occasionally have sudden irregularities commonly called glitches, i.e., irregularities in the deepfakes that are perceived as weird or unnatural by the users (Appel & Prietzel, 2022; Bode, 2021; Hasan & Salah, 2019) that could affect the DF perception. However, the effects of such glitches on deepfake and human persona perceptions and on user behavior in an experimental setting have not been studied in depth. Glitches come in different forms such as eeriness of the eyes (Appel & Prietzel, 2022), ears (Agarwal & Farid, 2021), and voice (Müller et al., 2021) and the effects of these glitches, when observed by users, on persona perception and users' task performance have not been studied empirically. Furthermore, developing strategies to minimize these glitches will be crucial for improving the realism and acceptance of deepfake technology.

Fourthly, commercial deepfake service providers have penetrated the deepfake creation market rapidly during the past few years offering easy access to deepfake generation around the world (Colossyan, 2022; Elai, 2022; Synthesia, 2022). However, demographic similarity between a deepfake and its user has been found important for users identification, a user's feeling of similarity, with the deepfake (Aljaroodi et al., 2019; Wagner & Blewer, 2019) thus making the demographic coverage of deepfakes and DAs, on which the visual appearance of the deepfake is based, important. However, in the commercial deepfake services, there are hundreds of DAs available for deepfake creation (Colossyan, 2022; Elai, 2022; Synthesia, 2022), but the diversity of DAs' demographic properties compared to global demographics has not been studied in depth.

Finally, the deepfakes' effectiveness in capturing and reflecting the wide range of human emotions has not been studied in depth. In prior studies of the role of displaying human emotions in artificial characters, such as deepfakes, with a wide range of human-like facial expressions increases deepfakes' perceived humanlikeness (Bhargave et al., 2013; Chen & Jack, 2017). Yet, the current state of deepfake service providers' DAs' emotional facial expression status is unclear. Potential emotional bias in DAs can lead to a lack of humanlikeness and authenticity in human-deepfake interactions where a wider range of emotions would be necessary (Becker et al., 2024).

Table 1. Research gaps for each dissertation article.

Research Gap	Methodology	Article
Lack of scientific knowledge on how users perceive deepfake user personas compared to narrative and classic persona modalities and how persona perception differences affect design task performance.	Empirical experiment with deepfake, narrative, and classic persona modalities and statistical analysis of survey answers from 91 participants.	Article I: The Realness of Fakes: Primary Evidence of the Effect of Deepfake Personas on User Perceptions in a Design Task
Lack of scientific knowledge on how users perceive DFs versus human personas.	Empirical experiment with deepfake and human personas with qualitative thematic analysis of answers and observations from 46 participants.	Article II: How Do Users Perceive Deepfake Personas? Investigating the Deepfake User Perception and Its Implications for Human-Computer Interaction
Lack of scientific knowledge on the effects of deepfake glitches on persona perception and user behavior.	Empirical experiment with deepfake and human personas with statistical analysis of survey answers and behavioral data from 46 participants.	Article III: "There Is something Rotten in Denmark": Investigating the Deepfake Persona Perceptions and Their Implications for Human-centered AI
Lack of scientific knowledge on how demographic diversity is represented in commercial deepfake services.	Quantitative observation and statistical analysis.	Article IV: Is Deepfake Diversity Real? Analyzing the Diversity of Deepfake Avatars
Lack of scientific knowledge on the effectiveness of deepfakes in capturing and reflecting human emotions.	Quantitative observation and statistical analysis.	Article V: Getting Emotional Enough: Analyzing Emotional Diversity in Deepfake Avatars

1.3 Research Questions

To address the identified research gaps (Table 1), five research questions (RQs) were formulated which were further divided into individual RQs in the research articles (Table 2). Based on the research gaps presented in Table 2, this dissertation explores the impacts of DFs on user perceptions, DF perceptions' implications for user behavior, and the emotional and demographic diversity of DAs. The RQs presented in the five articles comprising this dissertation investigate these aspects of DFs and DAs.

Table 2. Original research question from the articles.

Dissertation Research Question	Research Questions in the Research Articles	Article
RQ1: To what degree do deepfake personas improve or worsen designer persona perceptions and design task performance?	How does persona modality affect designers' perceptions of the persona?	Article I: The Realness of Fakes: Primary Evidence of the Effect of Deepfake Personas on User Perceptions in a Design Task
	How does persona modality affect design task performance?	
	Is there an uncanny valley effect associated with deepfake personas?	
RQ2: How do users perceive deepfake personas for a design task?	How do people perceive deepfakes for a design task?	Article II: How Do Users Perceive Deepfake Personas? Investigating the Deepfake User Perception and Its Implications for Human-Computer Interaction
RQ3: How do users observe deepfake persona distortions affect users' perception of the deepfake persona?	What type of glitches do users observe in deepfake personas?	Article III: "There Is something Rotten in Denmark": Investigating the Deepfake Persona Perceptions and Their Implications for Human-centered AI
	How do users' (a) perceptions of the persona and (b) behavior vary between deepfake and human personas?	
	How does users' glitch perception affect their (a) perceptions of the persona and (b) behavior?	
RQ4: How inclusively do deepfake avatars represent the global population?	What are the central demographic gaps in deepfake avatars?	Article IV: Is Deepfake Diversity Real? Analyzing the Diversity of Deepfake Avatars
	How well do deepfake avatars represent the global population?	
RQ5: How accurately do deepfake avatars express human facial expressions?	What are the key emotional expressions that define the diversity in emotional expressions exhibited by DAs?	Article V: Getting Emotional Enough: Analyzing Emotional Diversity in Deepfake Avatars
	In what ways, if any, do DAs exhibit bias in capturing and reflecting human emotional expressions?	
	Are there significant differences in the way DAs express emotions based on age, gender, and ethnic group?	

RQ1 (*To what degree do deepfake personas improve or worsen designer perceptions and design task performance?*) explores the different aspects of users' DF perceptions and DF perceptions' potential effect on design task performance. Deepfakes themselves and persona perceptions' effect on design task performance has been studied before yet there is a knowledge gap in what ways DFs affect users' design task performance. Deepfake technology's ability to create realistic depictions of human-like characters carries potential for immersive personas for design use but scientific research on the potential of such realistic depictions remains scarce. DF properties such as possible uncanniness (Mori, 1970) i.e., an artificial character being close to humanlikeness making the artificial character disturbing, and the uncanniness' effect on DF usability in design tasks has not been studied. Also, personas come in different forms or modalities, such as conventional 2D personas (paper, PDF) and narrative personas (textual persona narrative) (Kunur et al., 2016; Salminen, Nielsen, et al., 2018). The effects of persona modality (conventional persona or narrative persona, and also chat personas) on persona perception has been studied (Kaate, Salminen, Jung, Santos, et al., 2025; Salminen, Liu, et al., 2020) but not the effects of DFs on persona perception. In addition, the effects of DFs on design task performance have remained an unexplored avenue.

RQ2 (*How do users perceive deepfake personas for a design task?*) explores how users perceive DFs and the subsequent effects on user behavior. While much of the existing literature focuses on the detection of deepfakes and deepfakes' negative implications, such as misinformation and manipulation (Mustak et al., 2023; Vaccari & Chadwick, 2020), there has been less emphasis on understanding how users perceive and interact with deepfakes in a *positive way*. RQ2 aims to fill that knowledge gap by investigating the perceptions users have when engaging with deepfakes in a controlled experiment. Understanding how users perceive deepfakes is important, for example, for the development of usable and effective information systems. If deepfakes are perceived negatively, such as being untrustworthy or creepy, deepfakes' integration into information systems could be hindered. Conversely, positive deepfake perceptions could benefit user experience with deepfakes and increase acceptance of the deepfake technology and increase DFs' usefulness in design.

RQ3 (*How do users observe deepfake persona distortions affect users' perception of the deepfake persona?*) aims to reveal how users perceive DFs if and when users encounter glitches, visual or auditory distortions, in the DFs. DF perceptions are vital for the usability of the DFs as the perception directly impacts, for example, the user's trust and overall acceptance of deepfakes (Seymour et al., 2021). Also, persona realness' effect on the persona perceptions and user behavior is a less explored area. DFs may carry glitches in deepfakes' voice and movement possibly lowering the perceived realness of the DF (Appel & Prietzel, 2022; Bode et al., 2021; Hasan & Salah, 2019).

RQ4 (*How inclusively do deepfake avatars represent the global population?*) explores the disparities in the representation of different demographic groups in deepfakes. Demographic disparities, e.g., a bias towards some demographics while neglecting other demographics, could have an effect on the quality of interaction between the deepfake and its user. Understanding the potential demographic gaps is the first step in addressing the gaps, which is essential for creating more inclusive and equitable deepfakes. Deepfakes are increasingly used in customer service, education, and other applications where deepfakes interact with a diverse user base (Bach et al., 2022; Vassilakopoulou & Hustad, 2021). If deepfakes predominantly represent only certain demographics, deepfakes may not resonate with or be relatable to all deepfake users (Renier et al., 2024). This can affect the quality of the deepfake user experience, especially for those users from demographic groups that are demographically underrepresented in current deepfakes. By identifying and addressing these demographic gaps, deepfakes can be created to better serve a diverse global user base, potentially improving user experience with deepfakes.

RQ5 (*How accurately do deepfake avatars express human facial expressions?*) emphasizes the significance of facial expressions in interpersonal communication. Deepfakes, designed to be *anthropomorphic*, aim to simulate human-like interactions (Nass & Moon, 2000). Understanding the range and diversity of facial expressions in DAs is important for improving user satisfaction and engagement, among others, in various applications. Identifying and addressing potential biases in the facial expressions of deepfakes based on gender, ethnicity, and age is also important for user satisfaction and engagement. Emotional biases in DAs can affect the perceived authenticity and effectiveness of interactions with the deepfakes in users from diverse backgrounds, potentially reinforcing stereotypes or failing to accurately represent the emotional complexity of the interaction between the user and the deepfake (Buolamwini & Gebru, 2018; Zhou et al., 2024). Previous studies have found the importance of facial emotional expressiveness to user engagement (Jack & Schyns, 2015; McDuff & Czerwinski, 2018; Miao et al., 2022) but as such the effects of deepfakes' emotional expressiveness, or the lack of it, has not been thoroughly addressed in prior studies.

1.4 Dissertation Structure

This dissertation is divided into seven chapters. This first chapter is the introductory chapter presenting the background for the study, research gaps, and research questions. The second chapter puts forth the theoretical background for the dissertation by discussing previous studies and scientific literature giving the conceptual framework for the dissertation. The third chapter presents the research methodology of the dissertation as well as the research philosophy of the dissertation.

The fourth chapter presents the key findings of the five research articles. The fifth chapter discusses the theoretical contributions and practical implications of the five articles, and limitations of the research. The sixth chapter is the conclusive chapter and finally, the seventh chapter of the dissertation includes the five research articles also presented in Table 2.

2 Theoretical Background

2.1 The Concept, Role, and History of Personas

2.1.1 The Concept of Personas

The functionality of personas in design and HCI is influentially grounded in the idea of goal-directed design (Cooper, 1999), where personas represent archetypal user groups that help guide product and service features toward user goals. In the fast-evolving landscape of product and service development, one of the most significant challenges companies face is ensuring effective communication about end users among diverse cross-functional teams (Howard, 2015a). This issue is often exacerbated in large organizations where specialists from various disciplines, such as software engineering, marketing, project management, and design, struggle to align their goals and methodologies. To bridge this gap, the concept of *personas* emerged in the late 1990s (Cooper, 1999) as a tool providing a common language for development teams that helps the teams focus on the needs, goals, and wants of the end users of products or services.

Personas are detailed and realistic representations of end users or customers who interact or will interact with a product or service (Cooper, 1999; Howard, 2015a). Personas have been conventionally presented as two-dimensional presentations, persona profiles, on paper, PDFs, etc. (Figure 1). Persona profiles have, among other possible information, a persona facial image, demographic details of the persona, and a collection of persona's goals, needs, and wants (Cooper, 1999). Personas are not merely fictional. Personas are rather constructed from rigorous data collection methods including ethnographic studies (White & Devitt, 2021), surveys (Kaate, Salminen, Jung, Almerekhi, et al., 2023; Kaate, Salminen, Jung, Santos, et al., 2024; Kaate, Salminen, Santos, et al., 2024; Kaate, Salminen, Jung, Santos, et al., 2025), or from website analytics data (B. J. Jansen et al., 2020; Salminen, Jansen, et al., 2018). The purpose of creating personas is to distill complex end user data into archetypal representations of end-user groups, encapsulating end-users' goals, motivations, and needs (Cooper, 1999) for design or planning. For example, a persona might represent "*Sam, a 35-year-old marketing manager who relies on productivity tools to manage his team's campaigns and who lives in Philadelphia* *with his wife and daughter.*" By personifying user data in this way, personas help persona users empathize with the end user and make informed design decisions that cater better to the end users' needs, goals, and wants (Cooper, 1999). Personas are an important component in the toolkit in modern HCI and business trying to create user-centered products and services (Anvari et al., 2015). By providing a psychologically compelling means to communicate user data, personas facilitate better teamwork and more informed design decisions, ultimately leading to products that resonate better with their end users (Anvari et al., 2017; Anvari & Richards, 2016).

Figure 1. An example of a conventional persona profile with persona facial image and demographic information on the left, biographic, frustrations, and goals in the center, and personality traits and motivations on the right (Freepik, 2025).

Personas provide several benefits for the design process. First, personas enhance communication. Personas serve as a touchstone for end-user discussions in and between teams, ensuring that every team member, regardless of their specialty, can understand and contribute to user-focused design decisions (Massanari, 2010). Personas create a unified vision of the end user, which can be referenced through the development process. Second, personas enable UCD. By integrating personas into UCD practices, teams can better align their design choices with real end user needs. This alignment is meant to result in products or services that are more intuitive and satisfying for the end-users to use, thus driving user satisfaction and loyalty (Howard, 2015b). Third, personas offer informed decision-making. Personas carry a wealth of user information, providing actionable knowledge that guides design and development. Personas help prioritize features and design elements that will most benefit the target users.

2.1.2 Role and History of Personas

2.1.2.1 Personas in Marketing

Personas was introduced in the realm of marketing to help companies understand their customers better by segmenting the market into homogeneous groups based on common characteristics (Ketamo et al., 2010). Conventional personas, as used in early marketing practices, focused on qualitative data and narrative profiles (Cooper, 1999; J. S. Pruitt & Adlin, 2006). These developments set the stage for this dissertation's focus on DFs, which represent a novel modality within personas' historical development. Marketing researchers and practitioners used personas to create detailed persona profiles representing typical users within these segments, thereby informing product development and marketing strategies to better meet customer needs and preferences (Caballero et al., 2014).

Personas in marketing have evolved significantly since the 1990s, becoming a tool for understanding and effectively targeting marketing messages to customer segments (Caballero et al., 2014; Revella, 2015). Personas provide marketers a finely grained understanding of customers', or potential customers', needs, goals, and motivations. As such, personas can be seen as manifestations of *market orientation* directing the company interest towards the market by giving more detailed and vivid depictions of the company's customers or potential customers directing the offering of the company to a more client-friendly direction. Market orientation is a business philosophy that prioritizes understanding and responding to customer needs, competitor actions, and market trends to create value (Kohli & Jaworski, 1990). Market orientation involves a company-wide commitment to gathering market intelligence, sharing insights across company's departments, and adapting strategies to enhance customer satisfaction and competitive advantage making personas a valuable tool for market orientation.

The purpose and role of personas in marketing is divided into several aspects. Firstly, personas enable marketers to adopt a *customer-centric approach* by providing a deeper understanding of the target audience (Fenton et al., 2022). This approach ensures that marketing efforts are relevant and resonate with the target audience, potentially leading to higher engagement and conversion rates, whatever the primary goal for the marketing. Conventionally the process of creating personas in marketing begins with ideation, where marketers brainstorm potential characteristics and needs of the target audience (Fenton et al., 2022). This stage involves documenting initial assumptions and views about the persona in a qualitative manner. Presently, data-driven, quantitative, persona creation methods have largely replaced manual persona generation methods. With advancements in online analytics and overall user and client data gathering, data-driven personas (DDPs) have emerged (Salminen, Jansen, et al., 2018). DDPs leverage large-scale data and computational techniques, offering advantages such as automated data collection, behavioral accuracy and scalability. DDPs address many challenges of manual persona methods, providing a more accurate and dynamic representation of customer segments (B. J. Jansen et al., 2020; McGinn & Kotamraju, 2008; Zhang et al., 2016).

Secondly, by *segmenting* the market into distinct personas, i.e., customer segments, marketers can tailor their messages and strategies to meet the needs and preferences of different customer segments (Afwan et al., 2024). Segmentation leads to more effective targeting of products and services and better positioning, differentiating the company offering from its competitors, of products and services. Thirdly, personas inform product or service development by highlighting the needs and preferences of different client or consumer groups (Caballero et al., 2014). Using personas in development processes ensures that products and services are designed with the end user in mind, increasing the likelihood of market acceptance. Fourthly, understanding the preferred channels and content types of different personas helps in developing effective communication strategies (Salminen, Jansen, et al., 2018). This knowledge allows marketers to craft messages that are more likely to persuade and convert the target audience. Knowing the preferred media and content types of different personas helps in developing a more effective content strategy (Salminen, Jansen, et al., 2018). Finally, personas are continuously refined and improved to better reflect current market situation. Market is constantly developing and to keep up with the market, marketers must evolve personas accordingly (Fenton et al., 2022).

2.1.2.2 Personas in HCI

The concept of personas has been a revolutionary tool in HCI and UCD, evolving since personas' inception in the late 1990s. The persona technique has found a critical place in HCI and software development, adapting over time to meet the changing needs of these fields. Personas were formally introduced to HCI by Alan Cooper in his book ”The Inmates Are Running the Asylum” (Cooper, 1999), where Cooper described personas as tools for goal-directed design. Cooper's persona approach emphasized the creation of detailed, fictional characters based on user research to guide design decisions. These fictional characters, personas, are not real people but rather synthesized profiles created from observations of real users, aiming to ensure that the design process remains user-focused (Goh et al., 2017).

Personas are often contrasted with other user-research techniques such as user archetypes, mental models, and scenarios. While user archetypes provide broad generalizations of user groups, personas offer more specific, contextualized representations that integrate behavioral and attitudinal data (Matthews et al., 2012). Mental models, on the other hand, focus on how users conceptualize a system and its functions (Gentner & Stevens, 2014), whereas personas emphasize user needs and motivations within specific contexts bringing end users closer to designers. Similarly, scenarios describe hypothetical user interactions with a product but do not provide the rich personal details and motivations that personas do (Carroll, 2000). The persona technique is grounded in theories of cognitive psychology and human decision-making, particularly in the way narratives and empathy influence perception and design choices (Nielsen, 2019). By embodying user attributes in a coherent and memorable way, personas facilitate empathy, which aids designers in making user-centered decisions (Salminen, Jung, et al., 2020).

In the years after Cooper's book, the use of personas in HCI has been refined and expanded. Researchers and practitioners have developed various approaches to persona creation and utilization, often adapting personas to fit different contexts and methodologies. Cooper's original concept of goal-oriented personas remains a foundational approach in HCI. This method involves creating personas with specific goals and motivations that drive their interaction with a product. The primary aim is to focus on what users want to achieve and how users use the product or service to meet their goals (Goh et al., 2017). Subsequent developments introduced role-based personas, which categorize users based on their roles and tasks within a particular context (J. S. Pruitt & Adlin, 2006). This method considers the lifecycle of personas, ensuring personas remain relevant and updated throughout the design process.

Research has shown that personas improve UCD practices by enhancing empathy, communication, and decision-making among design teams. For instance, (Friess, 2012) found that design teams using personas engaged in more user-focused discussions compared to teams relying solely on raw user data. Similarly, (Long, 2009) demonstrated that personas improve designers' alignment by creating shared mental models of the user base, reducing misinterpretations and design conflicts. Personas also serve as cognitive artifacts, aiding memory retention and recall of user needs during the design process (Kozar & Miaskiewicz, 2009). Furthermore, personas can positively influence design choices when integrated with real-world data, as shown by (Chapman & Milham, 2006), who found that personas based on real user data reduce the risk of biased assumptions in user modeling. However, the effectiveness of personas depends on their validity and how well they represent actual user behavior, a factor that is sometimes overlooked in persona development.

The integration of personas into agile development practices has been a significant evolution in persona applications. Agile development practices, characterized by iterative development and continuous feedback, require flexible tools that can adapt to changing requirements and user feedback (Caballero et al., 2014). In an agile development context, personas are often used at the beginning of the design or development project to refine initial development cycles and personas are continually refined based on user feedback and evolving requirements. This kind of iterative approach ensures that the personas remain relevant and accurately represent the user base throughout the project lifecycle. For instance, (Chamberlain et al., 2006) advocate for using personas during the exploratory phase to establish a clear understanding of user needs, goals, and behavior before any development or design begins.

Despite their widespread adoption, personas are not without criticism. One major concern is the risk of stereotyping and oversimplification. When not based on rigorous user research, personas can reinforce biases rather than mitigating biases (Hill et al., 2017). This issue is particularly evident when personas are created based on intuition rather than empirical data, leading to inaccurate representations of user diversity (Rönkkö et al., 2004). Additionally, some studies question whether personas truly influence design outcomes or if personas serve more as rhetorical tools that designers use to justify existing assumptions (Nielsen et al., 2015). Another limitation is the adoption challenge: research by (Blomquist & Arvola, 2002) found that design and development teams sometimes struggle to integrate personas into agile workflows, particularly in fast-paced environments where up to date personas are difficult to maintain. Furthermore, conventional persona profiles are sometimes perceived as static, whereas real users exhibit dynamic and evolving behaviors, making it difficult to ensure personas remain relevant throughout the design process (McGinn & Kotamraju, 2008). Addressing these limitations is important for better persona usability requiring a more dynamic, data-driven approach to persona development, incorporating real-time user analytics and iterative refinements as well as new modalities of personas, such as deepfakes.

Progress in AI and data analytics have led to the emergence of AI-generated and DDPs, which seek to overcome some of the traditional limitations of manually created personas. AI-generated personas leverage ML techniques to synthesize large-scale user data, creating dynamic personas that evolve based on real-time interactions (Salminen, Jansen, et al., 2018). These approaches allow for greater scalability and personalization, making it possible to generate personas that are continuously updated with new behavioral insights (Salminen, Jansen, et al., 2018). The integration of real-time analytics, adaptive persona modeling, and ethical frameworks to ensure that personas remain inclusive, accurate, and useful in dynamic HCI contexts are factors affecting persona use and research in the future.

2.2 Different Forms of Personas

2.2.1 Quantitative Personas

While widely used in HCI and marketing, personas have faced criticism and challenges. One primary concern is the risk of creating personas that may not accurately and engagingly represent real users (B. Jansen et al., 2021). This issue can arise if personas are based on insufficient or biased data, leading to designs that do not effectively meet end user needs. The data for persona creation can be from surveys, interviews, or numerical data from website analytics and user data for example from mobile apps (Howard, 2015b; Salminen, Jung, et al., 2020) or the data for persona creation can be sales teams' or development teams' hunch about the potential users. Additionally, the time and resources required to create and maintain personas can be substantial, potentially conflicting with the fast-paced nature of agile design and development (Goh et al., 2017).

To address the need for speed and accuracy in persona creation, DDPs have been evolving during the past years (B. J. Jansen et al., 2020). DDPs represent an advancement in the field of HCI, combining the, often, qualitative creation process of conventional personas with quantitative data analytics (Adlin & Pruitt, 2009; An, Kwak, Jung, Salminen, Admad, et al., 2018; B. Jansen et al., 2021; B. J. Jansen et al., 2020; Park & Kang, 2022; Zhang et al., 2016). Unlike conventional personas, which often rely on limited qualitative data, DDPs leverage extensive datasets obtained from digital interactions, such as website data, mobile application user data, and social media, resulting in more accurate and dynamic representations of users as the date keeps updating in the data source (Adlin & Pruitt, 2009; An, Kwak, Jung, Salminen, Admad, et al., 2018; B. Jansen et al., 2021; B. J. Jansen et al., 2020; Park & Kang, 2022; Zhang et al., 2016). DDPs offer several key advantages: DDPs are seen as accurate and reliable due to their grounding in large datasets. Dynamic updating capabilities keep DDPs relevant over time and scalability allows DDPs to represent large and diverse user populations while actionable insights facilitate informed decision-making (B. Jansen et al., 2019). DDP approach involves collecting and structuring the data, using methods such as clustering and natural language processing to identify distinct user *segments*, and transforming these segments into comprehensive personas (B. Jansen et al., 2019). As any persona, DDPs typically include demographic and behavioral information, enriched with qualitative data to provide a holistic view of the users (Adlin & Pruitt, 2009; B. Jansen et al., 2021).

The primary advantage recognized with DDPs is their enhanced accuracy and reliability compared to conventional, ”manual” personas which often utilize qualitative data (Park & Kang, 2022). Conventional personas can suffer from biases and may not reflect the broader user population accurately (Park & Kang, 2022). Being created from, usually, large datasets, DDPs are statistically robust and credible among persona users. DDPs offer a cost-effective and efficient alternative to the time-consuming and expensive process of developing conventional personas, making DDPs accessible to organizations with limited resources (S.-G. Jung et al., 2017). Moreover, DDPs can be updated dynamically to reflect changes in user behavior, ensuring their continued relevance and utility for the design and development teams (Salminen, Jung, et al., 2020).

The transition in the field of personas from conventional personas to DDPs represents a shift towards more evidence-based and dynamic user representations (Bonnardel et al., 2016). From an HCI perspective, DDPs provide a bridge between the empathy-driven approach of traditional personas and the precision of analytics (B. J. Jansen et al., 2020; Nielsen et al., 2015; Salminen, Jung, et al., 2020). DDPs enable designers and developers to make informed, user-centered decisions backed by data of the very same target groups for the target groups. This integration carries a deeper understanding of user needs, goals, and behaviors, leading to more effective and engaging user interfaces, experiences, products, and services as the personas evolve when the users evolve creating an iterative loop for persona development. Furthermore, DDPs can serve as interactive interfaces themselves, offering stakeholders access to underlying analytics and user data, thereby enhancing personas' usability and applicability in real-time decision-making processes (Nielsen et al., 2015).

DDPs are utilized in various domains, including system design, digital marketing, and content creation (Guan et al., 2021). In system design, DDPs help designers understand diverse user needs and preferences, leading to the development of more intuitive and user-friendly user interfaces. As the system under development evolves and user data is collected, the personas based on the user data evolve. In digital marketing, DDPs enable marketers to craft personalized and effective strategies by understanding the specific characteristics and behaviors of different user segments, resulting potentially in higher engagement and better conversion outcomes (Chaisuwan et al., 2019; Kaate, Salminen, Jung, Olkkonen, et al., 2023). Marketing content creators benefit from personas by gaining insights into what types of content resonate with different customer segments, allowing for the creation of more engaging and relevant content (Clarke, 2015; Revella, 2015).

Despite their advantages, implementing DDPs presents challenges, such as ensuring data quality, addressing algorithmic biases, and navigating ethical considerations (McGinn & Kotamraju, 2008; Zhang et al., 2016). The accuracy and quality of DDPs heavily depends on the quality of the data based on which the personas are generated. Inaccurate or incomplete data can lead to misleading personas and poor decision-making. Algorithms used in DDP creation can introduce biases, particularly if the input data is biased, requiring careful monitoring and mitigation (Salminen, Jung, et al., 2020).

The future of DDPs in HCI looks promising, with advancements in data analytics, ML, and user interface design enhancing DDPs' capabilities. Integrating AI and ML better in DDP creation can further automate and refine the creation and updating of DDPs. Developing real-time personas that adapt dynamically as new data is collected will provide even more accurate and timely user representations. Enhancing interactivity in personas can allow persona users to explore and manipulate underlying data, gaining a more detailed understanding of the personas. Expanding persona use to other fields such as healthcare, education, and public services can unlock new opportunities for improving UX (Jansen et al., 2019).

DDPs represent a significant evolution in HCI design and development practices, offering a more accurate, dynamic, and scalable approach to understanding the users and designing for the users (B. Jansen et al., 2021; Park & Kang, 2022). DDPs combine the empathy of conventional personas with the analytical power of data science, providing valuable insights that enable better decision-making in, for example, design and enhance UX across various domains. As technology continues to evolve, the potential for DDPs to impact HCI and beyond will presumably grow, making DDPs an essential tool for UCD and development practices.

2.2.2 Qualitative Personas

Qualitative personas are research-based user representations developed through qualitative methods such as interviews, ethnographic studies, and contextual inquiries (Misztal-Radecka & Indurkhya, 2020). Qualitative personas are rooted in narrative construction and empathy-building, designed to communicate a rich, human-centered understanding of users (Matthews et al., 2012; Nielsen, 2019). Qualitative personas focus on narratives to capture users' motivations, behaviors, and pain points in depth. Qualitative personas provide detailed, empathy-driven insights that help design teams make user-centered decisions. While qualitative personas offer deep contextual understanding, they are usually resource-intensive and may lack broad generalizability due, usually, to small sample sizes (Mulder & Yaar, 2006).

Developing qualitative personas involves a structured and iterative process that aims to capture the depth of user experiences through, for example, interviews or ethnographic studies. This process typically involves defining the objectives, aligning persona creation with specific design or business goals (B. J. Jansen et al., 2022). The objectives determine the scope of user research and guide the selection of appropriate methods for persona data collection. Researchers must identify key user segments whose needs and behaviors will shape the eventual personas. This phase in the qualitative persona generation process is critical in establishing a clear focus for persona generation and ensuring that the personas remain relevant to their intended application.

The primary data collection phase relies on qualitative research techniques such as in-depth interviews, contextual inquiries, and ethnographic field studies (Misztal-Radecka & Indurkhya, 2020). Contextual inquiry, which involves observing users in their natural environments, provides additional insights into how individuals interact with products and services in real-world scenarios. This method is particularly useful in uncovering implicit needs and workarounds that users may not articulate in interviews. Ethnographic research, which involves immersive studies in users' environments over an extended period, offers a deeper understanding of cultural and social factors influencing behavior (Misztal-Radecka & Indurkhya, 2020).

After collecting data, the next phase involves systematic data analysis to identify common patterns and behavioral themes from the data. Thematic analysis is commonly employed, wherein transcripts from interviews and observations are coded to extract key themes related to user goals, challenges, and behaviors (B. J. Jansen et al., 2022), eventually forming personas. This process involves categorizing qualitative data into meaningful clusters, allowing researchers to uncover insights that transcend individuals and represent broader user archetypes. The iterative nature of qualitative data analysis ensures that personas are constructed based on empirical evidence rather than anecdotal or subjective interpretations.

Once behavioral patterns and archetypes emerge, researchers synthesize the findings into persona profiles that reflect distinct user segments (B. J. Jansen et al., 2022). Qualitative persona profiles typically include similar components as do quantitative personas, a combination of demographic details, behavioral tendencies, goals, and pain points (Salminen, Guan, et al., 2020). While demographic information such as age, occupation, and background provides context, the primary emphasis of qualitative personas is on users' motivations, behaviors, and attitudes. To ensure validity, researchers may conduct persona user workshops where design team members review and refine the personas, refining persona alignment with the collected data and resonate with design and business objectives (B. J. Jansen et al., 2022; Korsgaard et al., 2020).

Refinement and validation play a crucial role in ensuring that personas remain useful throughout the design process. Some studies suggest validating personas through user feedback, where actual users or domain experts assess whether the personas accurately represent their lived experiences (Ramos et al., 2021). Additionally, integrating usability testing and iterative user research can further refine qualitative persona characteristics over time, making qualitative personas adaptable to evolving user needs.

2.3 Deepfakes

2.3.1 Deepfakes in General

Deepfakes, a term that first gained prominence in 2017, refer to synthetic media where faces, voices, and other human elements are convincingly altered or generated using advanced AI techniques (Gamage et al., 2022). The first attempt for a deepfake video happened already in 1997 in a program called *Video Rewrite* (Bregler et al., 1997). The program utilized vocal and visual mimicking of real humans using original footage of historical figures as the base for altered video production. The Video Rewrite program worked by analyzing the phonemes (speech sounds) in the original video along with the associated facial movements, then rearranging and seamlessly blending mouth images to align with the phoneme sequence of the new audio, i.e., the audio the new video creator wanted the character to speak (Bregler et al., 1997). The main intended application for the technology at the time was in film dubbing.

After Video Rewrite, the wider development of deepfakes can be traced to the use of generative adversarial networks (GANs) (Goodfellow et al., 2014; Karras et al., 2019), a class of ML frameworks where two neural networks, the generator and the discriminator, are pitted against each other to produce increasingly realistic outputs; That is, deploying GAN is an iterative process where two computer software bounce data between each other making the data, such as a video, more realistic (Jacobsen, 2024). This technology allows the creation of hyper-realistic videos, audio, and images that can mimic real individuals or generate fictional humans.

Deepfakes are generated through a process that involves extensive training of GANs on large datasets of human images and videos (Masood et al., 2023). The generator network creates a deepfake video, while the discriminator network evaluates the deepfake video's authenticity against real human videos. Through iterative refinement, the generator improves its ability to produce convincing deepfake videos that the discriminator cannot distinguish from real human videos (Masood et al., 2023). This process requires vast amounts of computational power and high-quality training data to achieve realistic results. The sophistication of deepfake technology has rapidly advanced, making it increasingly difficult for even trained human eyes to discern deepfakes from real humans.

From the film dubbing purposes in the beginning (Bregler et al., 1997), the use of deepfakes has spanned various domains, both benign and malicious. On one hand, deepfakes are employed in entertainment, such as in movies and video games, to create special effects, resurrect deceased or old actors, or just to make the fake world of a movie or a game more realistic (Murphy et al., 2023). On the other hand, deepfakes are notorious for their potential misuse in creating fake news, political disinformation, and non-consensual pornography (Barari et al., 2021; Vaccari & Chadwick, 2020). These malicious applications have raised significant ethical and legal concerns, as they can deteriorate trust in the media, manipulate public opinion, and cause severe personal harm (I. Sharma et al., 2023).

Deepfakes are used in various contexts, including social media, politics, and journalism and the production of deepfakes has become increasingly easier with commercial deepfake software (Colossyan, 2022; Elai, 2022; Synthesia, 2022). In social media, deepfakes can spread misinformation rapidly, reaching a wide audience before the content can be verified (P. Sharma et al., 2024). Politically, deepfakes pose a threat to the integrity of elections and political discourse by fabricating statements or actions of public figures (I. Sharma et al., 2023; Vaccari & Chadwick, 2020). In journalism, the challenge with deepfakes lies in maintaining the credibility of information when deepfake technology can produce convincing fake content (Shin & Lee, 2022). The pervasive nature of deepfakes necessitates the development of advanced detection methods to identify and mitigate their impact.

The evolution of deepfake technology has also spurred a corresponding development in detection techniques. ML algorithms are being developed to identify subtle inconsistencies and artifacts that differentiate deepfakes from genuine media. However, this has led to an ongoing ”arms race” between deepfake creation and detection technologies (P. Sharma et al., 2024). The detection methods, while improving, still face significant challenges in keeping pace with the rapidly advancing deepfake generation techniques.

However, despite all the malicious use of deepfakes, deepfakes can also be used for good purposes. In the *entertainment industry*, deepfakes have been utilized to create realistic visual effects (Usukhbayar & Homer, 2020). For instance, the Dalí Lives exhibition at the Dalí Museum in Florida employed deepfake technology to animate the artist Salvador Dalí, providing visitors with an interactive and immersive experience (Mihailova, 2021). In *education*, deepfakes have been used to enhance learning experiences. By animating historical figures or creating realistic simulations, educators can offer students engaging and interactive content that deepens their understanding of complex subjects (Erduran, 2024). In *medicine*, deepfake technology has been explored for its potential to improve patient care and medical training. For example, deepfakes can simulate realistic patient scenarios, providing medical students with valuable hands-on experience in a controlled environment (Hamilton, 2024). These positive applications highlight the versatility of deepfake technology when utilized responsibly. However, before this dissertation, the use of deepfakes as personas in a design context has not been studied.

2.3.2 Anthropomorphism

Anthropomorphism, the tendency to attribute human characteristics, emotions, and intentions to non-human entities, plays a central role in the perception and reception of deepfakes (Huang, 2024). The human inclination to perceive agency in artificial creations, such as deepfakes, has long been studied in psychology and cognitive science, where it has been linked to evolutionary survival mechanisms that favor the recognition of social agents in ambiguous stimuli (Epley et al., 2007). This inherent tendency in humans to anthropomorphize extends to digital avatars, robots, and AI-generated content, including deepfakes.

One significant aspect of deepfakes is their high degree of anthropomorphism, or the extent to which deepfakes are perceived as humanlike (Waytz et al., 2010). Research in HCI and social robotics suggests that the more humanlike a digital entity appears, the more people are likely to interact with it as if it were a real person (Smith et al., 2021). The application of anthropomorphism in the context of deepfakes is particularly relevant due to the increasing realism of deepfakes. As deepfakes replicate human facial expressions, voice patterns, and even microexpressions with precision, deepfakes elicit natural responses from viewers, who instinctively perceive deepfakes as human entities (Korshunov & Marcel, 2018; Nass & Moon, 2000). Research in HCI has shown that people tend to engage with and respond to digital human representations in ways similar to how they interact with real humans (Kim, 2024). This phenomenon, sometimes referred to as the ”computers as social actors” (CASA) paradigm, a concept introduced by (Nass et al., 1994), suggests that when deepfakes are perceived as convincingly human-like, deepfakes may evoke trust, empathy, or even deception, depending on the context in which the deepfakes are deployed.

Anthropomorphism in deepfakes has significant implications for both the ethical considerations and the effectiveness of deepfake application. In entertainment and education, the more human-like and engaging a synthetic character appears, the more likely it is to create a compelling and immersive experience (Waytz et al., 2010). For example, when historical figures are animated using deepfake technology, the sense of authenticity and relatability can enhance engagement and retention of historical knowledge (Pataranutaporn et al., 2021). However, this same mechanism raises concerns in the context of misinformation. When users attribute human-like credibility to deepfakes, users may be more susceptible to believing manipulated content, even when presented with evidence of its artificial origin (Chesney & Citron, 2019).

The degree of anthropomorphism in deepfakes also intersects with the uncanny valley effect (Mori, 1970). Deepfakes that fall into this perceptual uncanny valley may trigger distrust or cognitive dissonance among viewers, which could either limit deepfakes’ effectiveness in positive applications or serve as a natural barrier to deception in harmful applications (MacDorman & Chattopadhyay, 2016). As deepfake technology continues to advance, mitigating the uncanny valley effect while maintaining ethical considerations will be crucial in determining how deepfakes are received in different contexts.

In the context of this dissertation, anthropomorphism is central to understanding how DFs function as tools for design. The effectiveness of DFs in HCI is contingent on the extent to which deepfakes elicit anthropomorphic responses and influence trust. Anthropomorphism helps explain why users may emotionally engage with or attribute intentionality to DFs. While the field of HCI has explored AI-generated characters and virtual agents in various capacities (Aljaroodi et al., 2019; Do et al., 2024; Huang, 2024), the specific application of deepfakes as *personas* remains an underexplored area. Understanding how anthropomorphism shapes the perception of DFs provides knowledge into deepfakes' potential for usability and credibility, as well as the ethical challenges deepfakes pose.

2.3.3 Demographic Diversity in Deepfakes

The demographic diversity of personas has been found to play a crucial role in shaping user engagement, trust, and inclusivity in design (Salminen, Jung, et al., 2022). In persona-based design, where personas are used to guide the development of products and services, the demographic attributes of the personas, including ethnicity, gender, age, and socio-cultural background, impact how users relate to and interpret these representations (Renier et al., 2024). With the emergence of AI-generated personas (S.-G. Jung et al., 2025) and deepfake technology, the construction of personas can move beyond traditional persona-generation methods toward synthetic, realistic personas that can be dynamically adapted for different design contexts and for different persona users. This shift in persona creation raises new questions about the implications of demographic diversity in DFs and the potential biases embedded in DF creation.

Personas serve as a tool to help designers understand and empathize with end-users, yet personas' effectiveness is tied to how well they represent diverse persona user groups (Cooper, 1999; Salminen, Jung, et al., 2022). Social identity theory (Tajfel & Turner, 1986) suggests that people naturally group themselves and others into categories, such as race, gender, or age, based on shared traits, helping them understand who they are in relation to others.

In the context of personas, demographic diversity of the personas influences whether persona users feel represented or excluded from a design process and demographic diversity can also increase societal understanding (Hall et al., 2013). Research has demonstrated that the users of virtual characters, including deepfakes, engage more with the virtual characters that reflect the users' own demographic backgrounds (Do et al., 2024). Conversely, the lack of demographic diversity in personas can lead to stereotyping, exclusion, and design biases, particularly when personas fail to account for intersectional user identities.

The data in DFs is, or can be, just the same as for conventional personas or for DDPs meaning that the persona information (demographics, goals, wants, etc.) is based on some data but for DFs, the deepfake itself is generated based on the persona data. Bias risks are imminent from imbalanced training datasets of the DF systems (Zhou et al., 2024). For example, the DF system might not be able to produce realistic African deepfakes but the deepfake generation system might be able to produce realistic Caucasian deepfakes due to insufficient training data for deepfake generation. Studies on AI bias have shown that AI models often overrepresent certain demographic groups while underrepresenting others, leading to distorted or homogenized persona profiles (Buolamwini & Gebru, 2018). In the context of persona-based design, these biases may reinforce stereotypical or inaccurate end-user representations, affecting the inclusivity of designed products and services.

Bias in DFs extends beyond demographic underrepresentation to cultural and linguistic biases, which may influence how personas are designed and perceived (Bender et al., 2021). Studies in algorithmic fairness have demonstrated that AI-generated representations tend to replicate biases present in the AI training data, meaning that DFs, as AI-generated representations, may unintentionally reinforce gendered or racially skewed portrayals (Mitchell et al., 2019). Demographic bias and gaps in deepfake systems' can hinder DFs' use for good potentially affecting how users interpret and relate to DFs and addressing such biases and gaps requires intentional AI training dataset diversification and fairness-aware AI training to ensure that DFs accurately represent global populations.

2.4 Persona Perceptions

Persona perceptions of conventional personas are shaped by how well the persona representations align with users' expectations, the extent to which they facilitate empathy, and personas' perceived utility in the design process. While personas have long been valued for helping designers conceptualize user needs and behaviors (Cooper, 1999), personas' effectiveness depends on how they are constructed, presented, and utilized (Chapman & Milham, 2006; Matthews et al., 2012). Research suggests that persona perceptions vary significantly depending on, for example, the role of the persona user, the level of engagement with personas, and the degree to which personas are seen as credible and representative (Matthews et al., 2012).

One of the persona user perceptions of conventional personas is the personas' ability to convey user empathy among persona users who may have limited direct interaction with the actual users (Harley, 2015). By humanizing data, personas help bridge the gap between raw statistics and the real-world experiences of users. When personas are well-researched and detailed, they can create strong mental models for the persona users about the end users, leading to design decisions that are more aligned with end-user needs (Nielsen, 2019). Empirical studies suggest that personas can serve as a cognitive tool, improving designers' ability to recall user needs, which enhances engagement with persona-driven design practices (Salminen, Jung, et al., 2021).

However, critical perceptions of conventional personas highlight several limitations that can hinder personas' effectiveness. A common criticism is that personas can feel artificial or disconnected from reality, particularly if they lack clear ties to empirical data (Matthews et al., 2012). If personas are based on limited research, subjective assumptions, or generic user stereotypes, they may be perceived as over-simplified caricatures rather than valuable design tools (Turner & Turner, 2011). The degree to which personas feel authentic plays a major role in personas' perceived usefulness. When personas are too generic, they may fail to resonate with design teams, leading to disengagement from the end-users (Miaskiewicz & Kozar, 2011; Rönkkö et al., 2004).

Conventional personas often provide fixed, snapshot-like descriptions of end-users, which may not fully capture the evolving nature of user behavior and preferences over time (Rönkkö et al., 2004). In industries where end-user needs change rapidly, such as technology and digital services, conventional personas may be seen as outdated, reducing their perceived relevance (S. Jung et al., 2019). Perceiving personas as outdated has led some practitioners to view conventional, static personas as a one-time exercise rather than an ongoing and updating tool in the design process (Friess, 2012).

Another challenge in persona perception is the level of trust persona users place in the persona's authenticity. If persona users suspect that personas are based on designer intuition rather than real end-user data, persona users may be viewed with skepticism (Rönkkö et al., 2004). This skepticism is particularly common in organizations where there is a strong reliance on data-driven decision-making, as personas without clearly documented sources may appear less rigorous compared to direct end-user insights (Salminen, Guan, et al., 2021).

The effectiveness of personas, conventional, data-driven and AI-generated personas, depends on various perceptual factors that shape persona user engagement and interaction with the persona (Marsden & Pröbster, 2019; Matthews et al., 2012). Several key attributes influence how persona users perceive and respond to personas, including realism, distractions, empathy, credibility, completeness, usefulness, clarity, willingness to use, immersion, and sense of control (Salminen, Santos, et al., 2020). Studies on these persona perceptions are abundant, including a study developing persona perception scale (PPS) by (Salminen, Santos, et al., 2020).

Realism plays a crucial role in the acceptance of and engagement with personas. Studies have demonstrated that higher visual fidelity leads to stronger empathic responses and increased persona user engagement (Salminen, Liu, et al., 2020). When personas exhibit realistic visual and behavioral characteristics, they are more likely to evoke trust and emotional connection. *Distractions*, such as unnatural behaviors, visual inconsistencies, or awkward animation, are properties found in DFs, and these properties can significantly reduce user immersion and disrupt the persona experience (Appel & Prietzel, 2022; Bode et al., 2021; Hasan & Salah, 2019). Prior research has shown that external distractions can alter perceived realism and narrative transportation of any digital media (Busselle & Bilandzic, 2008). Minimizing these distractions in AI-generated personas is critical to maintaining an engaging and coherent interaction between the AI-generated persona and its user.

The ability of a persona to evoke *empathy* is another essential factor influencing user perceptions of the persona. Empathy toward personas is strongly linked to personas' relatability (B. J. Jansen et al., 2020; Nielsen, 2019). A systematic review of extended reality environments suggests that users develop stronger empathic connections with realistic and well-designed digital entities (Martingano et al., 2021). This implies that personas should be designed for emotional engagement by incorporating human-like properties in the persona. *Credibility* is a central attribute affecting trust and willingness to engage with personas. Prior research has found that a persona's attractiveness and perceived authenticity correlate with its credibility (Fogg, 2003). Persona users are more likely to trust and rely on personas that exhibit coherent and consistent characteristics, reinforcing personas' perceived credibility.

Personas' *completeness*, referring to the depth and coherence of the information the persona conveys, also shapes user perceptions. Prior studies have identified completeness as a key factor in how users evaluate personas (Adlin & Pruitt, 2009; Marsden & Pröbster, 2019). When personas provide comprehensive and well-structured details, they are perceived as more useful and trustworthy (Salminen, Santos, et al., 2020). *Clarity* in the presentation of a persona attributes ensures that persona users can easily understand and relate to the personas. Existing research suggests that clear and concise information presentation significantly enhances user comprehension and engagement of personas (Park & Kang, 2022). Designing personas with transparent and accessible information is vital for effective communication of persona information.

Immersion refers to the extent to which users feel absorbed in an experience involving personas. Studies on immersive media suggest that higher immersion levels enhance empathy and presence, leading to more engaging interactions (Calvert & Hume, 2022; Jennett et al., 2008; Smith et al., 2021) making immersion especially important to DFs.

3 Research Methodology

3.1 Ontological and Epistemological Perspectives

Behind every research, there are some basic assumptions about the world and phenomena they study. How the world is seen and how the phenomena are studied are dependent on the very nature of the world and the phenomena under inspection but also dependent on how the world and the phenomena are thought to be.

The ontological and epistemological foundations of this dissertation stem from the intersection of HCI, AI, and user experience research, with a particular focus on DFs. Ontology in this dissertation concerns the nature of reality in which DFs exist and what they are, while epistemology addresses how knowledge about DFs is constructed and understood.

The ontological assumptions of this dissertation are rooted in critical realism and pragmatism. Critical realism posits that an objective reality exists independent of human perception, but our understanding of the objective reality is mediated by social and technological constructs (Bhaskar, 2008). In the context of DFs, this means acknowledging that DFs, while artificially generated, are perceived and interacted with as if they possess social and psychological characteristics akin to real human entities (Dourish, 2001; Latour, 2007). This ontological stance aligns with research in HCI and AI ethics, which recognizes that AI-generated personas can shape user perceptions, decision-making, and behavior despite their artificial origins (Floridi & Cowls, 2019; Mittelstadt et al., 2016).

In the context of this study, pragmatism also offers a suitable philosophical lens for examining DFs, emphasizing the practical consequences and real-world applications of deepfake technology (Gualeni, 2015). Rather than seeking universal truths or ontological certainties, a pragmatic approach foregrounds the utility, experiential impact, and context-specific effectiveness of DFs in design and user research. This aligns with the dissertation's focus on how users perceive, interact with, and derive value from DFs across varying design tasks. Pragmatism also supports iterative inquiry and reflexivity, allowing the research to remain open to evolving technological, ethical, and social dynamics surrounding deepfakes, while prioritizing actionable knowledge that can inform responsible implementation in human–computer interaction and design practice.

DFs function as simulacra (Baudrillard, 2019), meaning they are representations that, while not tied to real individuals, acquire a form of reality within HCI. In this sense, DFs are not ontologically independent beings but sociotechnical constructs that gain meaning through interaction, perception, and engagement (Brey, 2018). This aligns with post-phenomenological perspectives, which argue that technologies do not simply reflect reality but actively mediate human experiences, shaping how users perceive digital artifacts (Ihde, 2010; Verbeek, 2011). While DFs themselves might lack agency, their perceived realism influences user cognition, trust, and emotional engagement, making DFs' ontological status both virtual and socially real (Huang, 2024).

DFs do not inherently possess human qualities, yet users can interact with DFs as if the DFs were real. DFs' perceived realism is seen depending on cognitive, social, and cultural frameworks, rather than on an intrinsic property of the technology itself (Jasanoff, 2004). Empirical findings in the dissertation articles support this view, demonstrating that users respond to DFs in ways that mirror real-world interactions, even when these DFs lack biological substance (Kaate, Salminen, Jung, Almerekhi, et al., 2023). The uncanny valley effect, referring to the uneasy or disturbed feeling users experience when they see digital figures or robots that look nearly human, but not quite realistic enough, creating a sense of something being "off" or unsettling, further reinforces the idea that DFs exist in a relational ontological space where their reality is contingent on how well they approximate human authenticity (Kätsyri et al., 2015; Mori, 1970).

The epistemological foundation of this dissertation aligns with constructivism, which holds that knowledge is socially and contextually constructed rather than objectively discovered (Berger & Luckmann, 2011). This perspective acknowledges that user perceptions of DFs are not purely technological phenomena but are shaped by human cognition, sociocultural expectations, and technological design (Jasanoff, 2004; Suchman, 2007). As a result, DFs are not just digital artifacts but epistemic objects. Entities that generate knowledge through user interaction and perception (Cetina, 2009).

This study approach aligns with pragmatist-empirical epistemology (Morgan, 2007), which values methodological pluralism and, in the context of this dissertation, recognizes that knowledge about DFs is best understood through diverse scientific approaches. Controlled experiments measure user perceptions through surveys and behavioral data, while qualitative methods, such as think-aloud recordings and transcripts, capture the detailed ways users make sense of DFs. These methods ensure a rigorous empirical foundation for understanding how DFs influence cognitive and emotional engagement in HCI contexts. Moreover, this dissertation engages with critical epistemology, particularly in examining the biases embedded in deepfakes (Miller & Taddeo, 2017). The study of demographic gaps in deepfakes highlights the epistemic limitations of current AI models, which tend to reflect biases in the AI training data rather than offering a neutral representation of human diversity (Binns, 2017; Birhane, 2021).

By adopting the ontology and epistemology presented earlier, this dissertation positions DFs as both technological artifacts and socio-cognitive constructs. While DFs are artificially generated, their effects on users are real and measurable, warranting both empirical investigation and theoretical reflection. The mixed-methods approach used in this dissertation offers a comprehensive understanding of how DFs are perceived, interpreted, and integrated into HCI. These ontological and epistemological framing bridges the gap between technological realism and social perception, reinforcing that DFs are not merely digital artifacts but entities that acquire ontological significance through human interaction and mediated representation.

3.2 Methodology in the Articles

3.2.1 Data Collection and Data Analysis

Analyzing DF perceptions, deepfakes' facial expressions, and deepfakes' demographic properties on deepfake provider services were analyzed with various methods (Table 3).

Table 3. Research approaches in the research articles I-V.

Article	**Research approach**	**Reasoning for choosing the approach**
Article I: The Realness of Fakes: Primary Evidence of the Effect of Deepfake Personas on User Perceptions in a Design Task (Kaate, Salminen, Santos, Jung, et al., 2023)	**General approach**: Quantitative.	To study the perception differences between different persona modalities.
	Data collection: User experiment, survey.	To understand users' deepfake perceptions and the effect of deepfakes in a real-life design task.
	Data analysis: ANOVA	To analyze the differences in user perceptions between different persona modalities and between different persona user groups.
Article II: How Do Users Perceive Deepfake Personas? Investigating the Deepfake User Perception and Its Implications for Human-	**General approach**: Qualitative	To understand the underlying reasons for users' varying DF perceptions.
	Data collection: User experiment, survey open-ended answers, session	To understand real users' DF perceptions in a real design task.

Article	**Research approach**	**Reasoning for choosing the approach**
Computer Interaction (Kaate, Salminen, Jung, Almerekhi, et al., 2023)	recordings and transcripts, study notes.	
	Data analysis: Thematic analysis	To study the open-ended answers on users' DF perceptions.
Article III: "There Is something Rotten in Denmark": Investigating the Deepfake Persona Perceptions and Their Implications for Human-centered AI (Kaate, Salminen, Santos, et al., 2024)	**General approach**: Quantitative and qualitative	To understand users' DF perceptions and its implications on persona user experience.
	Data collection: User experiment, survey, session recordings and transcripts	To understand real users' DF perceptions in a real design task.
	Data analysis: Multivariate analysis of covariance (MANCOVA) and thematic analysis of open-ended answers	To analyze the effects of users' gender, age, and deepfake familiarity on DF perceptions.
Article IV: Is Deepfake Diversity Real? Analyzing the Diversity of Deepfake Avatars (Kaate, Salminen, Al Tamime, et al., 2025)	**General approach**: Quantitative.	To understand the demographic variations in deepfake avatars in deepfake avatar services.
	Data collection: Collecting data on deepfake avatars' demographic variation in deepfake avatar services. Visual and automated tagging of deepfake avatar emotional expressions.	To gain first-hand knowledge of the demographic variations in deepfake avatars in deepfake avatar services.
	Data analysis: Descriptive metrics on demographic variables in deepfake avatars. Chi-squared test of independence.	To analyze the variation of deepfake avatars' demographic variation in deepfake avatar services.
Article V: Getting Emotional Enough: Analyzing Emotional Diversity in Deepfake Avatars (Kaate, Salminen, Jung, Rizun, et al., 2024)	**General approach:** Quantitative.	To understand the facial expressions variations in deepfake avatars in deepfake avatar services.
	Data collection: Collecting data on deepfake avatars' emotional expressions' variation in deepfake avatar services. Visual and automated tagging of deepfake avatar emotional expressions.	To gain first-hand knowledge of the facial expressions variation in deepfake avatars in deepfake avatar services.
	Data analysis: Comparative metrics of deepfake avatars' emotional expressions between deepfake avatar services. Mann-Whitney U test, Kruskal-Wallis test, and Wilcoxon test.	To analyze the variation of deepfake avatars' facial expression variation in deepfake avatar services.

In Article I, data was collected online with participants recruited from Prolific which is an online research platform (Palan & Schitter, 2018). The study adopted a within-subjects experimental design. This design involved observing the same participants under different conditions to evaluate their perceptions of different persona modalities, including classic, narrative, and DFs.

A total of 90 participants (N=90) took part in the study. Three personas were generated using a commercial deepfake service. To examine the effect of persona modality on users' perceptions and task performance, each participant interacted with all three modalities. The presentation order was randomized to control for order effects and ensure balanced, unbiased data collection.

The survey data collection utilized Qualtrics, an online survey platform. The data analysis in Article I employed ANOVA to analyze differences in user perceptions across three persona modalities (deepfake, classic, and narrative). ANOVA was chosen as the data analysis method because each participant interacted with all persona types, allowing the control for individual differences and reducing within-subject variability (Field, 2012). This method enabled a more precise estimation of the effects of persona modality on dependent variables such as empathy, credibility, and immersion. Additionally, post hoc pairwise comparisons were conducted to determine specific differences between the modalities while controlling for multiple comparisons, ensuring statistical robustness. Since the study involved multiple dependent variables. This analytical approach ensured that any detected effects were not due to chance but reflected genuine differences in user perception.

Article II employed a between-subjects experimental design with 46 participants. Participants were divided into groups assigned to different study conditions, specifically guided and non-guided regarding the possibility of glitches (distortions) in the deepfakes. That is, some of the participants received the information that the personas might include glitches, some of the participants did not receive this information. The think-aloud method was chosen because it allowed the capturing of real-time cognitive and emotional reactions of the study participants without the distortions associated with retrospective reporting (Eccles & Arsal, 2017). Think-aloud recordings were transcribed. The data analysis employed thematic analysis of the transcribed think-aloud data to investigate how users perceive DFs. Thematic analysis of the think-aloud data was applied to identify recurring patterns in participants' responses (Braun & Clarke, 2006). The thematic analysis followed an inductive approach, meaning that themes were derived from the data rather than imposed beforehand, ensuring a more grounded understanding of users' persona perceptions.

Article III utilized a mixed-method approach in a between-subjects experimental design. A total of 46 participants interacted with both a DF and a human persona in a within-subjects experimental setting allowing for a comparative analysis of user perceptions between the DFs and human personas. Data collection involved both qualitative and quantitative measures, including transcribed think-aloud recordings, survey data collection with Qualtrics, observational notes by the study supervisors, and behavioral data collection to capture a comprehensive knowledge of users' DF perceptions. Survey data were analyzed using repeated measures MANCOVA. The use of mixed methods was justified because it allowed to complement the quantitative findings with qualitative findings, thereby providing a more comprehensive understanding of users' persona perceptions (Creswell & Plano Clark, 2018). The repeated measures MANCOVA was employed to analyze multiple dependent variables (e.g., trust, credibility, and empathy) while controlling for covariates such as users' prior experience with deepfake technology. MANCOVA was selected because it reduced the risk of Type I errors (false positive) that would arise from conducting multiple separate ANOVAs (Tabachnick & Fidell, 2013). Additionally, pairwise comparisons with Bonferroni correction were conducted to identify specific differences between the experimental conditions. The think-aloud data was thematically coded, which helped contextualize the quantitative results by capturing users' subjective experiences with the DFs. The integration of these methods provided a more detailed analysis of how users perceive DFs.

Article IV employed a data collection involving the selection of deepfake service providers and categorizing DAs based on their demographic variables. Data on 202 DAs was manually collected from three deepfake service providers to gain an understanding of the differences and gaps of and in deepfake service providers' demographics. This approach enabled the identification of demographic biases and provided insights into the demographic diversity of DAs. A systematic analysis of 202 DAs from the deepfake providers was conducted. Quantitative content analysis was used to categorize DAs based on key demographic attributes such as age, gender, ethnicity, and linguistic representation.

Quantitative content analysis was chosen because it allowed for a systematic, objective analysis of representational patterns (Coe & Scacco, 2017). Chi-squared tests of independence were applied to determine whether certain demographic groups were statistically over or underrepresented in the DAs. This approach was justified because chi-squared analysis is well-suited for categorical data and enables the detection of significant disparities in representation (Barceló, 2018). Furthermore, descriptive statistics were used to report frequencies and proportions of DAs, providing an overview of the extent of demographic biases. By employing a structured coding scheme and inter-coder reliability checks, it was ensured that the demographic classification of the DAs was consistent and replicable.

Article V employed facial emotion detection of 202 DAs using Amazon AWS Rekognition, a computer system designed for facial emotion recognition, supplemented with human inspection of 202 DAs, and non-parametric statistical tests. AWS Rekognition was chosen for its ability to efficiently analyze a large dataset of 202 DAs, providing objective measures of emotional expression frequencies as it has been used in similar tasks in previous studies (Joodi et al., 2023; Mane & Shah, 2019; Rafael et al., 2020). However, given the well-documented biases in algorithmic emotion detection (McDuff & Czerwinski, 2018), human rater was included to code the same data of 202 DAs and to verify the results of AWS, ensuring greater reliability. The study applied nonparametric statistical tests, Mann-Whitney U, Kruskal-Wallis, and Wilcoxon tests, because the data on emotional facial expressions were not normally distributed, making parametric tests inappropriate (Field, 2012). Mann-Whitney U test was used to compare gender-based differences in emotional expression, while Kruskal-Wallis test was used to test differences across ethnic groups. Post-hoc Wilcoxon pairwise comparisons were conducted when significant group differences were found in the Kruskal-Wallis test.

3.2.2 Validity and Reliability

Reliability pertains to the consistency of measurements over time, while validity relates to the accuracy and appropriateness of the measures used to investigate a particular phenomenon (Drost, 2011). In qualitative research, reliability and validity focus on ensuring the rigor and trustworthiness of the research process, whereas in quantitative research, reliability and validity pertain to the statistical evaluation of measurement instruments (Lincoln & Guba, 1985).

In quantitative research, validity is evaluated through three main types: content validity, criterion validity, and construct validity (Ping, 2004). Content validity assesses how well the measurement items capture the theoretical construct. Criterion validity examines whether the construct corresponds with other established measures of the same concept. Construct validity determines the extent to which the measured construct accurately represents the intended concept (Ping, 2004). Construct validity is further analyzed through convergent and discriminant validity. Convergent validity is demonstrated when different measures of the same construct show a strong correlation, while discriminant validity ensures that distinct constructs remain separate (Ping, 2004).

For Article I, the reliability of the study was ensured using a repeated-measures design, where each participant was exposed to multiple persona modalities, reducing individual biases. The study's validity was strengthened through the use of established and validated metrics, such as the PPS (Salminen, Santos, et al., 2020) and the Immersive Experience Questionnaire (IEQ) (Jennett et al., 2008), ensuring that constructs like empathy, credibility, and immersion were measured robustly. However, the study acknowledged limitations in external validity due to its reliance on a controlled experimental setting, which may not fully capture real-world persona usage contexts.

For Article II, reliability was addressed through rigorous qualitative coding procedures in the think-aloud analysis, with multiple researchers independently coding transcripts and resolving discrepancies through discussion. I.e., interrater agreement was formed before the final decision on the themes found in the data. However, the study noted potential limitations in ecological validity, as the experimental setup may not fully represent how users interact with DFs in real-world settings. Additionally, the relatively small sample size (n=46) may affect the generalizability of the study, though it falls in the range generally acceptable for user studies in HCI.

For Article III, reliability was ensured using standardized experimental procedures and structured data collection. Interrater reliability was applied to qualitative data analysis to ensure consistency in identifying themes. The study's validity was strengthened through its mixed-methods approach, which allowed for cross-validation between qualitative think-aloud data and quantitative survey responses. Quantitative data was collected using established and validated measures, PPS (Salminen, Santos, et al., 2020), and behavioral metrics measured directly on the user study platform. However, construct validity could be affected by the subjective nature of glitch perception, and external validity may be constrained due to the nature of the experimental design.

For Article IV, reliability was reinforced through systematic content analysis with predefined coding frameworks applied consistently across 202 DAs. The study ensured interrater reliability by having multiple coders analyze the demographic attributes of the DAs, with discrepancies resolved through discussion. Validity was enhanced by comparing deepfake demographic distributions to real-world population data, providing an objective benchmark. However, limitations in validity were noted, as demographic diversity may not fully capture the details of how users perceive inclusivity in DAs.

For Article V, reliability was supported by using both automated facial emotional expression detection with AWS Rekognition (Joodi et al., 2023) and human assessment and validation of the results to cross-check accuracy in the DAs' emotional facial expressions tagging. Interrater agreement was calculated to assess the consistency between automated and human evaluations of the emotional facial expressions. Validity was strengthened using a large dataset (202 DAs) and statistical analysis to detect biases in emotional expression. However, the study acknowledged potential biases in facial emotion recognition algorithms in AWS, which may not equally capture emotional expressions across different demographic groups, impacting validity.

4 Summary of Findings in the Research Articles

4.1 Article I: The Realness of Fakes: Primary Evidence of the Effect of Deepfake Personas on User Perceptions in a Design Task

The study investigates how DFs influence user perceptions in a design task compared to classic and narrative personas. The study identifies several key findings. First, DFs trigger a higher sense of abnormality and discomfort, aligning with the uncanny valley effect, more than classic and narrative personas. Users rated DFs significantly lower in empathy, credibility, completeness, clarity, willingness to use, and sense of control. Interestingly, DFs did not enhance immersion or extend dwell time in the design task, contrary to initial expectations. Users spent comparable amounts of time with DFs as they did with other persona modalities. The current deepfake technology, with its inherent flaws, reduces its effectiveness as a tool in design tasks, indicating the need for technological improvements in deepfakes. Despite these limitations, deepfakes have potential for future persona applications, particularly due to their realistic human mimicry and dynamic, immersive formats for personas, given the future technological improvements. The study emphasizes the necessity for further research to refine deepfake technology and optimize persona designs, considering the diverse user reactions to different persona modalities. These findings suggest that while DFs are not yet optimal for design tasks, their future potential justifies continued exploration and enhancement of the deepfake technology.

4.2 Article II: How Do Users Perceive Deepfake Personas? Investigating the Deepfake User Perception and Its Implications for Human-Computer Interaction

The study investigates user perceptions of DFs in design tasks, focusing on DF perceptions' impact on HCI. The study identified key themes affecting user perception, including human-likeness, distractions, and trust. User perception of deepfakes is heavily influenced by perceived realness and human-likeness of the deepfakes, affected by deepfake factors like manners, speech, trustworthiness, emotional expressions, and the character-user connection. Users found unnatural eye movements and other glitches in DFs significantly lowering users' perception of authenticity and trust in the DFs. The study suggests that to improve DF UX, DFs must be integrated ethically and transparently, ensuring user awareness when deepfakes are used. The research also highlights that while deepfake present opportunities to enhance UX, these benefits can only be realized by addressing current technological limitations and ensuring responsible use of the deepfake technology. This emphasizes the importance of ongoing improvements in deepfake technology and the ethical considerations necessary for deepfakes' deployment in HCI contexts.

4.3 Article III: "There Is something Rotten in Denmark": Investigating the Deepfake Persona Perceptions and Their Implications for Human-centered AI

The study examines user perceptions of DFs in design tasks and their implications for human-centered AI. An experimental study involving 46 participants revealed that glitches in DFs, such as unnatural eye, voice, and body movements, significantly impact user experience with deepfakes and users' task performance. Users detected these glitches, leading to perceptions of reduced authenticity, trust, and emotional connection. Despite users spending similar amounts of time with DFs as with the human persona, the detected glitches caused a decrease in users' confidence in the task and an increase in perceived strangeness of the DF. The findings suggest that the current state of deepfake technology is not yet optimal for deepfake use in design tasks. The study highlights the need for significant technological improvements to minimize such glitches and enhance user acceptance and trust in DFs. By addressing these issues, DFs can be more effectively integrated into human-centered AI applications, improving deepfake utility and deepfake UX.

4.4 Article IV: Is Deepfake Diversity Real? Analyzing the Diversity of Deepfake Avatars

The study examines the demographic diversity of 202 DAs from three deepfake service providers. Key findings reveal significant demographic disparities in deepfakes, with a strong bias towards the young white demographic, and an underrepresentation of older demographics, Asians, Middle Easterners, and various occupations. Although gender representation is nearly balanced (99 male and 103 female DAs), this does not fully mitigate the overall lack of demographic diversity in the DAs. The study highlights a pronounced overrepresentation of younger age groups (Millennials and Generation X) and Caucasians, while Indigenous peoples, Asians, and Middle Eastern individuals are significantly underrepresented. Additionally, the research identifies a limited range of occupations among the DAs, covering only seven distinct job roles, and highlights the insufficient linguistic diversity of DAs, with the service providers covering only approximately 1% of the world's languages. The study recommends that deepfake service providers prioritize demographic inclusivity, create DAs that represent a wider range of racial backgrounds and age groups, and expand linguistic capabilities to better reflect global demographic diversity. These steps are essential for creating more inclusive and equitable deepfake technologies that effectively serve a diverse, global deepfake user base.

4.5 Article V: They Are "Real" but Are They Real? Analyzing Emotional Diversity in Deepfake Avatars

The study investigates the emotional diversity in DAs and effectiveness of DAs in conveying a range of emotions. By analyzing 202 DAs using Amazon AWS Rekognition and human examiner, the study identified a significant "positivity bias," with avatars predominantly displaying smile and calm facial expressions. Gender differences were notable, with male DAs scoring higher in "smile" and "calm" facial emotions, while female DAs scored higher in "surprised", "fear," and "happy" facial emotions. Ethnic differences were also significant; European and Hispanic DAs exhibited a broader range of "smile", "happy", and "calm" facial emotions compared to other ethnic groups. No significant facial emotional bias in the DAs was found based on age. The study concludes that the emotional representation among DAs is limited, necessitating more balanced emotional diversity to enhance realism and effectiveness in DAs. The findings emphasize the importance of improving deepfake technology for a wider and more authentic range of emotional facial expressions, thus better serving diverse deepfake applications and user groups.

5 Discussion

5.1 Theoretical Contributions

The theoretical contributions of this dissertation are based on the theoretical contributions of Articles I-V (Table 4).

5.1.1 Theoretical Contributions Article I

Article I presents several theoretical contributions to the field of HCI and persona design by answering *RQ1: To what degree do deepfake personas improve or worsen designer persona perceptions and design task performance?* Firstly, Article I examines the impact of DFs on designers' perceptions of personas, revealing that DFs elicit lower empathy, credibility, and immersion compared to classic and narrative personas. The study highlights the significant effect of the "uncanny valley", where DFs are perceived as eerie and less trustworthy compared to other studied persona modalities negatively influencing DFs' acceptance and effectiveness in the design task (Mori, 1970). The findings refine existing theories of trust and authenticity in digital interactions, showing that DFs, despite their realism, do not inherently enhance credibility or engagement. Moreover, the study challenges the assumption that realism increases immersion, revealing that DFs do not significantly improve user involvement or dwell time in design tasks (Jennett et al., 2008). The study also contributes to theories of technological adoption by illustrating the gap between deepfake technological maturity and user acceptance of deepfakes, emphasizing that current DFs are not yet seamlessly integrated into user workflows. Finally, the study advances the discourse on ethical AI and responsible design, emphasizing the need for transparency and user-centered development to mitigate the unsettling effects of DFs. By addressing these dimensions, the study provides a deeper theoretical understanding of how realism, credibility, and usability intersect in DFs, paving the way for future advancements in persona-based design and human-AI interaction. Beyond prior persona research that has largely emphasized textual or narrative representations (Cooper, 1999; J. Pruitt & Adlin, 2010), Article I introduces DFs as a new persona modality, thereby expanding theoretical discussions on how persona modality shapes persona users' empathy and engagement.

5.1.2 Theoretical Contributions Article II

Article II answers *RQ2: How do users perceive deepfake personas for a design task?* advances theoretical discussions on HCI, trust in AI, and DF realism by examining how users perceive DFs in a design task. The study expands existing models of digital trust and authenticity by showing that perceived realness in DFs is influenced not just by deepfake appearance but also by factors such as mannerisms, speech, and emotional expression (Glikson & Woolley, 2020; Jacovi et al., 2021; Kietzmann et al., 2021). Article II highlights how minor glitches, such as unnatural eye movement, significantly deteriorate trust towards the DF, reinforcing the role of subtle behavioral cues in user acceptance of DFs (Y. Li et al., 2018). Furthermore, the study refines theories on human-likeness of deepfakes and cognitive engagement by demonstrating that DFs, despite their advanced realism, can become distractions rather than enhancers of UX. The study findings also contribute to ethical AI design by emphasizing the necessity of transparency in deepfake applications, as users must be aware of the artificial nature of the DFs to maintain trust. By integrating insights from HCI, cognitive psychology, and AI ethics, the study deepens the theoretical understanding of how users cognitively process and emotionally respond to DFs (Ahmed, 2021), shaping future approaches to DF design. While previous HCI research on persona perception has focused on trust and credibility in relation to design fidelity (Friess, 2012; Matthews et al., 2012), Article II extends the existing literature by showing how persona users articulate their perceived persona authenticity and discomfort when interacting with DFs.

5.1.3 Theoretical Contributions Article III

Article III answers *RQ3: How do users observe deepfake persona distortions affect users' perception of the DF?*, contributes to the theoretical understanding of DFs in HCI by revealing how users perceive and respond to distortions, glitches, i.e., deepfake glitch perception, in DFs. The study extends theories of trust and authenticity in deepfakes by demonstrating that even minor inconsistencies in DFs, such as unnatural eye movements, voice issues, and rigid facial expressions, can significantly erode user confidence in DFs and reduce perceived credibility. The study also refines the uncanny valley theory (Mori, 1970) implicating that DFs, despite their realism, are not inherently more engaging or immersive, as detected distortions disrupt the illusion of deepfakes' human-likeness. Additionally, the study advances human-centered AI theories by emphasizing the need for technological refinement and user-informed design to improve acceptance and usability. The findings support the growing discourse on ethical AI by highlighting the importance of transparency, as users may react negatively when they sense artificiality without explicit disclosure, eroding trust and empathy towards the deepfake (Y. Li et al., 2018). By integrating insights from cognitive psychology, AI ethics, and HCI, the study deepens the theoretical understanding of how technological imperfections shape user experiences, reinforcing the necessity of refining deepfake technology to align with human expectations of deepfakes to better integrate deepfakes into design processes. Building on the uncanny valley theory (MacDorman & Chattopadhyay, 2016; Mori, 1970), Article III contributes by empirically demonstrating how subtle perceptual glitches in DFs erode users' trust and confidence in the persona.

5.1.4 Theoretical Contributions Article IV

Article IV, by answering *RQ4: How inclusively do deepfake avatars represent the global population?*, advances theoretical discussions on digital representation, AI bias, and inclusivity in HCI by critically examining the demographic diversity of DAs. The study contributes to theories of algorithmic bias (Kordzadeh & Ghasemaghaei, 2022) by demonstrating that deepfake technology disproportionately represents young, white individuals while significantly underrepresenting older demographics, Asians, Middle Eastern populations, and various occupational identities. The study refines existing frameworks on digital identity and fairness in AI (Mir et al., 2022) by revealing how DAs replicate systemic demographic and linguistic imbalances, reinforcing concerns about bias in ML training datasets. Furthermore, it extends theories of human-likeness and social presence by showing that the limited diversity of DAs constrains their ability to serve as effective, relatable digital personas for the global user base. The findings also contribute to ethical AI and responsible design principles, emphasizing the necessity for more inclusive dataset curation and improved generative AI models that reflect broader demographic and linguistic diversity. The study integrates insights from algorithmic fairness, HCI, and media representation, deepening the theoretical understanding of how DAs shape inclusivity and digital identity in AI-driven communication. Whereas earlier critiques of algorithmic bias have mainly centered on facial recognition and image (Buolamwini & Gebru, 2018; Crawford, 2021), Article IV extends the debate into the design domain by showing how demographic imbalances in commercial DA services constrain the inclusivity of DFs.

5.1.5 Theoretical Contributions Article V

Article V, by answering *RQ5: How accurately do deepfake avatars express human facial expressions?*, contributes to theoretical discussions on emotional expressiveness in DFs (López-Gil et al., 2022) by analyzing how DAs convey emotions across gender and ethnic groups. The study extends theories of digital realism and anthropomorphism (Shah et al., 2025; Stojnić, 2015) by demonstrating that DAs exhibit a significant ”positivity bias”, with emotions like smiling and calmness being overrepresented, while negative or complex emotions appear less frequently (Raji & Buolamwini, 2019; Stumpf et al., 2024). Article V refines existing frameworks on emotional diversity in artificial agents by revealing systematic differences in how male and female avatars, as well as different ethnic group avatars, express emotions, indicating that current DAs do not accurately capture the full spectrum of human facial expression. The study emphasizes the need for emotionally more inclusive and diverse training datasets to create DAs that are more representative by their emotional expressiveness. The study also points to the necessity for future research to address cultural variations in emotional expression and the ethical implications of deploying deepfake technology responsibly. Additionally, the study contributes to theories of representation and bias in AI (Leavy, 2018; Shahbazi et al., 2023) by showing that the emotional portrayal of DAs does not reflect real-world emotional diversity, which may influence deepfake user perceptions and trust (Caci et al., 2024). The findings have implications for the design of human-like deepfakes, emphasizing the need for greater balance in emotional representation to enhance authenticity and relatability (Jack & Schyns, 2015; McDuff & Czerwinski, 2018; Miao et al., 2022). While emotional expressiveness has been studied in avatars and embodied conversational agents (López-Gil et al., 2022; Nass & Moon, 2000), Article V provides one of the first systematic analyses of positivity bias in DAs, highlighting how limited emotional repertoires restrict DAs realism and potential for nuanced design use.

Table 4. Responses to research questions: theoretical contributions and managerial implications of Articles I-V.

Article	Research Question	Theoretical Contribution	Managerial Implications
Article I	RQ1: To what degree do deepfake personas improve or worsen designer persona perceptions and design task performance?	DFs trigger the uncanny valley effect, reducing trust and immersion. The study challenges assumptions that realism improves user engagement and highlights the need for user-centered AI design.	Managers should use DFs with caution, ensuring high technical quality, ethical safeguards, and ongoing evaluation of the deepfakes, while recognizing that conventional persona profiles may still be effective in certain use cases.

Article	Research Question	Theoretical Contribution	Managerial Implications
Article II	RQ2: How do users perceive deepfake personas for a design task?	Subtle behavioral cues in deepfakes shape deepfake perception, with glitches reducing trust. The study refines digital trust theories and emphasizes transparency for improving user acceptance of DFs.	Managers should ensure that DFs are realistic, ethically transparent, user-centered, and continuously refined through usability testing, as users' trust and engagement depend on minimizing technical flaws in deepfakes.
Article III	RQ3: How do user observed deepfake persona distortions affect users' perception of the deepfake persona?	Deepfake glitches erode user confidence, refining theories of trust. The study highlights the need for technological improvements and transparency in deepfakes.	Managers should integrate DFs ethically and transparently, invest in deepfake detection tools, and prioritize user feedback to enhance realism and trust in deepfakes, ensuring deepfakes improve UX without compromising credibility.
Article IV	RQ4: How inclusively do deepfake avatars represent the global population?	Deepfake avatars lack demographic and occupational diversity, reinforcing algorithmic bias. The study highlights the need for inclusive AI development to ensure fairness and better digital demographic representation.	Managers should prioritize demographic diversity in deepfake avatars by using inclusive deepfake creation models, adopting a multi-provider approach, and supporting diverse training data for deepfake creation to ensure equitable user engagement, avoid demographic bias, and enhance the effectiveness of deepfake applications.
Article V	RQ5: How accurately do deepfake avatars express human facial expressions?	Deepfake avatars exhibit emotional bias, overrepresenting positive emotions while lacking emotional diversity. The study calls for more balanced emotional AI training datasets to enhance deepfake avatar authenticity.	Managers should enhance the emotional and demographic diversity of deepfake avatars by investing in inclusive training data for deepfake creation ensuring emotionally authentic, relatable, and context-appropriate interactions with deepfakes that improve deepfake UX across diverse applications.

5.2 Managerial Implications

5.2.1 Managerial Implications for Article I

The managerial implications of the article highlight several key aspects regarding the use of DFs in design tasks. Managers should recognize that while DFs have potential in HCI for enhancing user empathy and engagement, DFs currently face challenges, particularly concerning their credibility, usability, and the uncanny valley effect. The uncanny valley effect impacts the acceptance and effectiveness of DFs making their seamless adoption in design workflows difficult. Managers should be cautious about relying solely on DFs for critical design decisions, as the study found that traditional personas might still be more effective in certain contexts. The realism offered by DFs can enhance user immersion, but only if the technical quality of the deepfakes is high enough to avoid the uncanny valley effect. Therefore, investment in state-of-the-art deepfake technology and continuous quality improvement of deepfakes is crucial. Moreover, ethical considerations are paramount. Deepfake technologies can be misused, leading to potential biases and ethical dilemmas. Managers must implement vetting processes for the creation and use of deepfakes to mitigate these risks. Additionally, educating DF users about the strengths and limitations of DFs is necessary to set realistic expectations and nurture informed decision-making. Finally, as the field of deepfake technology evolves, managers should stay informed about the latest developments and be prepared to adapt their strategies accordingly. Regularly updating and testing DFs against other persona modalities can help ensure DFs are used effectively and ethically in enhancing UCD processes.

5.2.2 Managerial Implications for Article II

Article II highlights key managerial implications for organizations integrating DFs into different processes. First, user perception of DFs is critical to the success of DF applications. Managers must ensure that DFs are designed to align with user expectations of human-likeness, realism, and trustworthiness. Given that DF users are highly sensitive to deepfake distortions or glitches, including unnatural eye movements and voice inconsistencies, managers must invest in refining deepfake technology to minimize these distortions before deploying DFs. Second, ethical and transparent implementation of DFs is essential for user trust and adoption. Organizations should clearly disclose when deepfakes are being used and provide users with control over their interactions with the deepfakes. Without transparency, deepfake applications risk being perceived as deceptive, which could hinder user engagement and trust. Third, deepfakes should be designed with a user-centered approach to enhance deepfake UX. Simply making deepfakes more realistic is not enough. Organizations must ensure that deepfakes meaningfully improve UX rather than becoming distractions. This involves careful attention to details like manners, speech patterns, and emotional expressiveness of the deepfakes. Managers should prioritize deepfake usability testing, including eye-tracking studies and user feedback, to assess how deepfakes influence engagement and task performance to which the deepfakes are used. Fourth, long-term success in deepfake adoption depends on continuous technological improvements and user adaptation. Since users can detect deepfake abnormalities, failing to meet expectations lowers perceived credibility of the deepfakes. Companies investing in deepfake applications must focus on developing deepfakes that better replicate human-like details and human behavioral cues lowering the chance of glitches. By addressing these challenges proactively, deepfake technology can evolve into a valuable tool for enhancing interaction in digital environments, rather than remaining an experimental technology.

5.2.3 Managerial Implications for Article III

For the managerial implications of Article III, managers should ensure that the integration of deepfakes into company processes is ethical and transparent, informing users when and how deepfakes are utilized. This transparency helps in maintaining user trust towards deepfakes and preventing the misuse of deepfake technology for fraudulent activities. Additionally, managers need to prioritize the development of robust detection and verification tools to tackle the malicious use of deepfakes. It is crucial to create deepfakes that enhance the UX by appearing authentic and reliable, thus promoting user engagement and trust with the deepfakes. Managers should also consider user feedback on the realism and human-likeness of deepfakes to further improve the deepfake technology. The successful implementation of DFs can potentially increase user interaction quality and provide a more immersive UX in various applications, from virtual customer service to educational tools.

5.2.4 Managerial Implications for Article IV

For the managerial implications of Article IV, the lack of demographic diversity in DAs poses significant challenges for organizations utilizing DFs. Failure to represent diverse populations in DAs can alienate users, reinforce stereotypes, and widen the digital divide, making deepfake technology less inclusive and less effective in engaging global audiences. Organizations deploying deepfakes should prioritize demographic inclusivity by selecting or developing deepfake models that reflect a broad spectrum of ethnicities, ages, and occupations to avoid marginalizing and alienating user groups and limiting accessibility. To ensure a more consistent and inclusive deepfake UX, companies should adopt a multi-provider approach when using DAs, as different deepfake service providers exhibit varying levels of demographic representations of DAs, possibly minimizing demographic bias. Furthermore, deepfake service providers must advocate for more diverse training datasets in DA development to improve the demographic balance of DAs. Investing in more advanced deepfake techniques and ML models will help create more representative and contextually relevant DAs, which are essential for improving user engagement, trust, and effectiveness in professional and social applications of deepfakes.

5.2.5 Managerial Implications for Article V

For the managerial implications of Article V, organizations utilizing DAs must prioritize emotional diversity to enhance realism and engagement of the deepfakes. The DAs studied in Article V exhibit a strong positivity bias with an overrepresentation of smiles and calm expressions while lacking a broader range of facial emotions. Expanding the emotional expressiveness of DAs is critical for applications in customer service, education, and mental health support, where detailed and contextually appropriate emotional facial expressions of deepfakes improve deepfake UX. Deepfake developers should invest in more inclusive emotional training datasets to ensure that DAs represent diverse emotional facial expressions across gender, ethnicity, and age groups. The lack of emotional variety in certain demographics may stem from their underrepresentation in deepfake training datasets, highlighting the need for balanced demographic inclusion in deepfake development. Organizations can improve both demographic and emotional diversity of deepfakes to create more authentic and relatable interactions with deepfakes. Additionally, ethical and policy considerations should guide DA implementation, especially in sensitive contexts such as healthcare and customer interactions. Policymakers and industry leaders should establish guidelines that ensure the responsible deployment of emotionally expressive DAs, preventing misrepresentation or potential biases when DAs are used and deepfakes are created. Increasing emotional diversity in DAs is not only an ethical necessity but also a practical step toward creating DFs that effectively serve a diverse base of DF use cases.

5.3 Limitations and Future Research Directions

While this dissertation provides novel insights into DFs and their applications especially in HCI, limitations must be acknowledged. These limitations highlight areas where further research is needed to refine the understanding of DFs, DF user perceptions, and DFs' impact on design and decision-making processes. One of the primary limitations is the generalizability of the findings. The studies presented in this dissertation rely on controlled user experiments and data collected from specific participants. While these methods allow for controlled and replicable research, the methods may not fully capture the diversity of real-world contexts in which DFs might be deployed. Future research should expand the scope of empirical studies to include diverse user groups, cultural contexts, and professional settings to enhance external validity. Another limitation is the technological maturity of DFs at the time of the research. Deepfake technology has been and is evolving rapidly, and the studies in this dissertation are based on DAs and DFs generated using commercial-grade deepfake services available at the time of data collection. The quality of DAs, particularly their ability to express detailed facial emotions and reduce the uncanny valley effect, is likely to improve in the future. Subsequent studies should reassess DF perceptions as the deepfake technology advances to determine whether improvements in realism mitigate the negative effects observed in this research.

The uncanny valley effect remains a significant obstacle in the use of DFs. While this dissertation explores how DF glitches and distortions impact user trust and engagement towards DFs, future research should examine how different approaches to DF design, such as stylization or partial anthropomorphism, could minimize the perceived uncanniness of the deepfakes. Studies incorporating eye-tracking data, physiological response to deepfakes, and longitudinal studies could provide a deeper understanding of how users cognitively and emotionally respond to DFs over extended interactions. This dissertation also highlights concerns about bias in DFs, particularly in terms of demographic representation and facial expressions. The analysis of commercial DA platforms reveals that most available DAs disproportionately represent young, white individuals, while other demographics remain underrepresented. This lack of diversity raises ethical concerns and potential usability issues, as users from underrepresented backgrounds may struggle to relate to or engage with DFs. Future research should explore methods for ensuring more inclusive deepfake generation training datasets and investigate how demographic alignment between DFs and their users influences user engagement and trust towards the DFs. Furthermore, the emotional diversity of DAs was found to be limited, with an overrepresentation of positive emotions such as smiling and calmness. This limitation restricts the applicability of DFs in scenarios where more detailed or context-specific emotional expressions are required, such as customer service, therapy, or educational settings. Future studies should investigate techniques for enhancing the emotional realism of DFs and assess how users respond to a broader range of emotional expressions in deepfakes. In addition, the long-term effects of DF exposure remain an open research question. While this dissertation provides evidence of immediate user reactions to DFs, it does not examine how user perceptions evolve over repeated interactions with DFs. Future research should explore whether prolonged exposure to DFs leads to increased acceptance, desensitization, or changes in trust towards deepfakes over time. Longitudinal studies could provide knowledge into whether users adapt better to DFs or continue to experience discomfort due to, for example, the uncanny valley effect or the glitches detected in the DFs.

Finally, the ethical and legal considerations of DFs warrant further exploration. As deepfake technology becomes more accessible and widely adopted, concerns about misuse, deception, and privacy must be addressed. Future research should examine how transparency, disclosure, and user control mechanisms influence trust and acceptance of DFs in real-world applications. Additionally, policy-oriented research could explore guidelines for responsible deepfake use in business, education, and public communication to ensure ethical deployment.

In conclusion, while this dissertation advances the understanding of DFs and their implications for HCI and marketing, critical areas remain for future research. Addressing the limitations identified here will be essential for refining DF technology, enhancing its usability, and ensuring that it serves diverse user groups in ethical and effective ways.

6 Conclusion

This dissertation has examined the role and implications of DFs in HCI and marketing, focusing on how users perceive, respond to, and engage with DFs in design tasks. Based on the five empirical studies, the findings offer valuable insights into both the strengths and the constraints of using DFs as tools in UCD. The results highlight that while DFs offer new possibilities for enhancing empathy and immersion in design processes, DFs' effectiveness is currently constrained by technical imperfections and perceptual challenges. One key issue is the uncanny valley effect, when a persona appears nearly but not quite human, causing discomfort and reducing the perceived realism, credibility, and emotional resonance of the deepfake. Users were also found to be highly sensitive to visual and auditory distortions, glitches, in DFs, such as unnatural eye movements, which can disrupt trust and negatively affect DF UX. These findings highlight the need for continued technological refinement to improve the stability and realism of deepfakes used in design contexts. Furthermore, the studies revealed significant demographic and emotional biases in commercially available DAs. Emotional portrayals skewed toward positivity, with male avatars exhibiting predominantly calm and smiling expressions, and female avatars more frequently showing emotions like happiness, surprise, and fear. This imbalance raises worries regarding inclusivity and the risk of perpetuating stereotypes, especially in applications that rely on authenticity and representational fairness. To ensure that DFs serve their intended function in a responsible and impactful way, developers and designers must focus on improving technical quality, ensuring demographic and emotional diversity, and promoting transparency in DF use. Ethical implementation, user-centered evaluation, and inclusive design principles are crucial for realizing the potential of DFs as tools for good in HCI and marketing. Future DFs hold promise for their ethical use in HCI and marketing given that the DF creators use DFs ethically. While DFs technical quality improves, the general public need better media literacy in a world where information is presented by real humans and DFs. The fine line between fake and real gets thinner.

Looking further ahead, the rapid development of generative AI presents several promising directions for future use and research of DFs. One avenue lies in the evolution from video-based DFs to **interactive, conversational personas.** These interactive DFs could embody not only the visual and vocal features of personas, but also respond dynamically to questions, simulate decision-making, and adapt their communication based on context. Such DFs may further enhance empathy and engagement in design and marketing; however, they also raise new challenges related to transparency, agency, and data provenance. Another promising field involves the use of **user simulations**, where synthetic personas interact autonomously within simulated environments to test system prototypes, conduct usability evaluations, or model user behavior. These simulations, if grounded in realistic behavioral data, could reduce the need for early-stage user recruitment while supporting iterative, data-driven design. As AI-generated personas become more common in design and marketing, ensuring their ethical and practical implementation will become an increasingly important responsibility for researchers and practitioners alike.

References

Adlin, T., & Pruitt, J. (2009). Putting personas to work: Using data-driven personas to focus product planning, design, and development. *Human-Computer Interaction: Development Process*, 95.

Afwan, A. S., Pradina, D. A., Kurniawati, Y., & Heikal, J. (2024). Persona-Based Segmentation: Customer Data Mapping as the Foundation of Corporate Marketing Strategy (Case Study on Scarlett Whitening). *JEMSI (Jurnal Ekonomi, Manajemen, Dan Akuntansi)*, *10*(1), 750–757. https://doi.org/10.35870/jemsi.v10i1.2018

Agarwal, S., & Farid, H. (2021). Detecting Deep-Fake Videos from Aural and Oral Dynamics. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 981–989. https://doi.org/10.1109/CVPRW53098.2021.00109

Ahmed, S. (2021). Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes. *Personality and Individual Differences*, *182*, 111074.

Ahmed, S., Ng, S. W. T., & Wei Ting, A. B. (2023). Understanding the role of fear of missing out and deficient self-regulation in sharing of deepfakes on social media: Evidence from eight countries. *Frontiers in Psychology*, *14*, 609.

Aljaroodi, H. M., Adam, M. T. P., Chiong, R., & Teubner, T. (2019). Avatars and Embodied Agents in Experimental Information Systems Research: A Systematic Review and Conceptual Framework. *Australasian Journal of Information Systems*, *23*. https://doi.org/10.3127/ajis.v23i0.1841

An, J., Kwak, H., Jung, S., Salminen, J., Admad, M., & Jansen, B. (2018). Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. *ACM Transactions on the Web*, *12*(4), 1–26. https://doi.org/10.1145/3265986

An, J., Kwak, H., Jung, S., Salminen, J., & Jansen, B. J. (2018). Customer segmentation using online platforms: Isolating behavioral and demographic segments for persona creation via aggregated user data. *Social Network Analysis and Mining*, *8*(1), 54. https://doi.org/10.1007/s13278-018-0531-0

Anvari, F., & Richards, D. (2016). A Method to Identify Talented Aspiring Designers in Use of Personas with Personality. In L. A. Maciaszek & J. Filipe (Eds.), *Evaluation of Novel Approaches to Software Engineering* (Vol. 599, pp. 40–61). Springer International Publishing. https://doi.org/10.1007/978-3-319-30243-0_3

Anvari, F., Richards, D., Hitchens, M., & Babar, M. A. (2015). Effectiveness of Persona with Personality Traits on Conceptual Design. *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, *2*, 263–272. https://doi.org/10.1109/ICSE.2015.155

Anvari, F., Richards, D., Hitchens, M., Babar, M. A., Tran, H. M. T., & Busch, P. (2017). An empirical investigation of the influence of persona with personality traits on conceptual design. *Journal of Systems and Software*, *134*, 324–339. https://doi.org/10.1016/j.jss.2017.09.020

Appel, M., & Prietzel, F. (2022). The detection of political deepfakes. *Journal of Computer-Mediated Communication*, *27*(4), zmac008. https://doi.org/10.1093/jcmc/zmac008

Bach, M. P., Zoroja, J., & Vukšić, V. B. (2022). Review of corporate digital divide research: A decadal analysis (2003-2012). *International Journal of Information Systems and Project Management*, *1*(4), 41–55. https://doi.org/10.12821/ijispm010403

Barari, S., Lucas, C., & Munger, K. (2021). *Political deepfakes are as credible as other fake media and (sometimes) real media*.

Barceló, J. A. (2018). Chi-Square Analysis. In S. L. López Varela (Ed.), *The Encyclopedia of Archaeological Sciences* (1st ed., pp. 1–5). Wiley. https://doi.org/10.1002/9781119188230.saseas0090

Baudrillard, J. (2019). *Simulacra and simulation* (S. F. Glaser, Trans.). Univ. of Michigan Press.

Becker, C., Conduit, R., Chouinard, P. A., & Laycock, R. (2024). Can deepfakes be used to study emotion perception? A comparison of dynamic face stimuli. *Behavior Research Methods*, *56*(7), 7674–7690. https://doi.org/10.3758/s13428-024-02443-y

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Berger, P. L., & Luckmann, T. (2011). *The social construction of reality: A treatise in the sociology of knowledge*. Open Road Media Integrated Media.

Bertoni, L., Kreiss, S., & Alahi, A. (2022). Perceiving Humans: From Monocular 3D Localization to Social Distancing. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 7401–7418. https://doi.org/10.1109/TITS.2021.3069376

Bhargave, A., Jadhav, N., Joshi, A., Oke, P., & Lahane, S. (2013). Digital ordering system for restaurant using Android. *International Journal of Scientific and Research Publications*, *3*(4), 1–7.

Bhaskar, R. (2008). *A realist theory of science*. Routledge.

Binns, R. (2017). *Fairness in Machine Learning: Lessons from Political Philosophy*. https://doi.org/10.48550/ARXIV.1712.03586

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Blomquist, A., & Arvola, M. (2002). Personas in action: Ethnography in an interaction design team. *Proceedings of the Second Nordic Conference on Human-Computer Interaction*, 197–200.

Bode, L. (2021). Deepfaking Keanu: YouTube deepfakes, platform visual effects, and the complexity of reception. *Convergence: The International Journal of Research into New Media Technologies*, *27*(4), 919–934. https://doi.org/10.1177/13548565211030454

Bode, L., Lees, D., & Golding, D. (2021). The Digital Face and Deepfakes on Screen. *Convergence: The International Journal of Research into New Media Technologies*, *27*(4), 849–854. https://doi.org/10.1177/13548565211034044

Bonnardel, N., Forens, M., & Lefevre, M. (2016). Enhancing Collective Creative Design: An Exploratory Study on the Influence of Static and Dynamic Personas in a Virtual Environment. *The Design Journal*, *19*(2), 221–235. https://doi.org/10.1080/14606925.2016.1129145

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2022). Testing Human Ability To Detect Deepfake Images of Human Faces. *arXiv Preprint arXiv:2212.05056*.

Bregler, C., Covell, M., & Slaney, M. (1997). Video Rewrite: Driving visual speech with audio. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '97*, 353–360. https://doi.org/10.1145/258734.258880

Brey, P. (2018). The strategic role of technology in a good society. *Technology in Society*, *52*, 39–45. https://doi.org/10.1016/j.techsoc.2017.02.002

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (Vol. 81, pp. 77–91). PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html

Busselle, R., & Bilandzic, H. (2008). Fictionality and Perceived Realism in Experiencing Stories: A Model of Narrative Comprehension and Engagement. *Communication Theory*, *18*(2), 255–280. https://doi.org/10.1111/j.1468-2885.2008.00322.x

Caballero, L., Moreno, A. M., & Seffah, A. (2014). Persona as a Tool to Involving Human in Agile Methods: Contributions from HCI and Marketing. In S. Sauer, C. Bogdan, P. Forbrig, R.

Bernhaupt, & M. Winckler (Eds.), *Human-Centered Software Engineering* (Vol. 8742, pp. 283–290). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44811-3_20

Caci, B., Giordano, G., Alesi, M., Gentile, A., Agnello, C., Lo Presti, L., La Cascia, M., Ingoglia, S., Inguglia, C., Volpes, A., & Monzani, D. (2024). The public mental representations of deepfake technology: An in-depth qualitative exploration through Quora text data analysis. *PLOS ONE*, *19*(12), e0313605. https://doi.org/10.1371/journal.pone.0313605

Calvert, J., & Hume, M. (2022). Immersing learners in stories: A systematic literature review of educational narratives in virtual reality. *Australasian Journal of Educational Technology*, 55–71. https://doi.org/10.14742/ajet.7032

Campbell, C., Plangger, K., Sands, S., & Kietzmann, J. (2022). Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising. *Journal of Advertising*, *51*(1), 22–38. https://doi.org/10.1080/00913367.2021.1909515

Campbell, C., Plangger, K., Sands, S., Kietzmann, J., & Bates, K. (2022). How Deepfakes and Artificial Intelligence Could Reshape the Advertising Industry: The Coming Reality of AI Fakes and Their Potential Impact on Consumer Behavior. *Journal of Advertising Research*, *62*(3), 241–251. https://doi.org/10.2501/JAR-2022-017

Carroll, J. M. (2000). *Making use: Scenario-based design of human-computer interactions*. MIT Press.

Casaca, J. A., & Miguel, L. P. (2024). The Influence of Personalization on Consumer Satisfaction: Trends and Challenges. In A. T. Rosário, R. N. Cruz, & L. B. Moniz (Eds.), *Advances in Marketing, Customer Relationship Management, and E-Services* (pp. 256–292). IGI Global. https://doi.org/10.4018/979-8-3693-3455-3.ch010

Cetina, K. K. (2009). *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press. https://doi.org/10.2307/j.ctvxw3q7f

Chadha, A., Kumar, V., Kashyap, S., & Gupta, M. (2021). Deepfake: An Overview. *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020*, 557–566.

Chaisuwan, B., Chantamas, M., & Prasongsukarn, K. (2019). Consumer Persona: Segmentation Scheme Based on Buddhist Temperaments for Digital Marketing. *Multifaceted Explorations of Consumer Culture and Its Impact on Individuals and Society*, 173–195. https://doi.org/10.4018/978-1-5225-6120-0.ch010

Chamberlain, S., Sharp, H., & Maiden, N. (2006). Towards a Framework for Integrating Agile Development and User-Centred Design. In P. Abrahamsson, M. Marchesi, & G. Succi (Eds.), *Extreme Programming and Agile Processes in Software Engineering* (Vol. 4044, pp. 143–153). Springer Berlin Heidelberg. https://doi.org/10.1007/11774129_15

Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. *Psychology & Marketing*, *39*(8), 1529–1562. https://doi.org/10.1002/mar.21670

Chapman, C. N., & Milham, R. P. (2006). The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(5), 634–636. https://doi.org/10.1177/154193120605000503

Chen, C., & Jack, R. E. (2017). Discovering cultural differences (and similarities) in facial expressions of emotion. *Current Opinion in Psychology*, *17*, 61–66. https://doi.org/10.1016/j.copsyc.2017.06.010

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753.

Clarke, M. F. (2015). The Work of Mad Men that Makes the Methods of Math Men Work: Practically Occasioned Segment Design. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3275–3284. https://doi.org/10.1145/2702123.2702493

Coe, K., & Scacco, J. M. (2017). Content Analysis, Quantitative. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (1st ed., pp. 1–11). Wiley. https://doi.org/10.1002/9781118901731.iecrm0045

Colossyan. (2022). *Choose your language*. https://www.colossyan.com/choose-your-language

Cooper, A. (1999). The Inmates are Running the Asylum. In U. Arend, E. Eberleh, K. Pitschke, U. Arend, E. Eberleh, & K. Pitschke (Eds.), *Software-Ergonomie '99* (Vol. 53, pp. 17–17). http://link.springer.com/10.1007/978-3-322-99786-9_1

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third edition). Sage.

Danry, V., Leong, J., Pataranutaporn, P., Tandon, P., Liu, Y., Shilkrot, R., Punpongsanon, P., Weissman, T., Maes, P., & Sra, M. (2022). AI-Generated Characters: Putting Deepfakes to Good Use. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–5. https://doi.org/10.1145/3491101.3503736

Do, T. D., Benjamin, J., Protko, C. I., & McMahan, R. P. (2024). *Cultural Reflections in Virtual Reality: The Effects of User Ethnicity in Avatar Matching Experiences on Sense of Embodiment*. https://doi.org/10.48550/ARXIV.2407.10412

Dourish, P. (2001). *Where the Action Is: The Foundations of Embodied Interaction*. The MIT Press. https://doi.org/10.7551/mitpress/7221.001.0001

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, *38*(1), 105–123.

Duvvuri, V. (2021). Expanding Digital Marketing Campaigns With Machine Learning Built Lookalike Audiences Having Varying Prior User Characteristics. In Á. Rocha, J. L. Reis, M. K. Peter, R. Cayolla, S. Loureiro, & Z. Bogdanović (Eds.), *Marketing and Smart Technologies* (pp. 201–212). Springer. https://doi.org/10.1007/978-981-33-4183-8_17

Eccles, D. W., & Arsal, G. (2017). The think aloud method: What is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, *9*(4), 514–531. https://doi.org/10.1080/2159676X.2017.1331501

Elai. (2022). *List of Supported Languages—Elai.io. Text-to-video AI*. https://elai.io/elai-languages

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Erduran, S. (2024). Deepfakes and students' deep learning: A harmonious pair in science? *Science*, *385*(6712), eadr8354. https://doi.org/10.1126/science.adr8354

Fallis, D. (2021). The Epistemic Threat of Deepfakes. *Philosophy & Technology*, *34*(4), 623–643. https://doi.org/10.1007/s13347-020-00419-2

Fenton, A., Heinze, A., Osborne, M., & Ahmed, W. (2022). How to Use the Six-Step Digital Ethnography Framework to Develop Buyer Personas: The Case of Fan Fit. *JMIR Formative Research*, *6*(11), e41489. https://doi.org/10.2196/41489

Field, A. (2012). *Discovering statistics using SPSS* (3. ed., repr). Sage.

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann Publishers.

Freepik. (2025). *Free Vector | Buyer persona infographics with photo*. Freepik. https://www.freepik.com/free-vector/buyer-persona-infographics-with-photo_9925542.htm#fromView=keyword&page=1&position=1&uuid=49d7d1d8-97dd-4815-ad41-91f5122fdcdf&query=Editable+Persona+Template

Friess, E. (2012). Personas and decision making in the design process: An ethnographic case study. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1209–1218. https://doi.org/10.1145/2207676.2208572

Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022). Are Deepfakes Concerning? Analyzing Conversations of Deepfakes on Reddit and Exploring Societal Implications. *CHI Conference on Human Factors in Computing Systems*, 1–19.

Gentner, D., & Stevens, A. L. (Eds.). (2014). *Mental models*. Psychology Press, Taylor & Francis Group.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goh, C. H., Kulathuramaiyer, N., & Zaman, T. (2017). Riding Waves of Change: A Review of Personas Research Landscape Based on the Three Waves of HCI. In J. Choudrie, M. S. Islam, F. Wahid, J. M. Bass, & J. E. Priyatma (Eds.), *Information and Communication Technologies for Development* (Vol. 504, pp. 605–616). Springer International Publishing. https://doi.org/10.1007/978-3-319-59111-7_49

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 27* (pp. 2672–2680). Curran Associates, Inc. http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

Gualeni, S. (2015). *Virtual worlds as philosophical tools: How to philosophize with a digital hammer*. Palgrave Macmillan.

Guan, K. W., Salminen, J., Nielsen, L., Jung, S.-G., & Jansen, B. J. (2021). Information Design for Personas in Four Professional Domains of User Experience Design, Healthcare, Market Research, and Social Media Strategy. *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS'21)*, 10. https://doi.org/10.24251/HICSS.2021.540

Gupta, P., Chugh, K., Dhall, A., & Subramanian, R. (2020). The eyes know it: FakeET- An Eye-tracking Database to Understand Deepfake Perception. *Proceedings of the 2020 International Conference on Multimodal Interaction*, 519–527. https://doi.org/10.1145/3382507.3418857

Hall, S., Evans, J., & Nixon, S. (Eds.). (2013). *Representation* (Second edition). Sage ; The Open University.

Hamilton, A. (2024). Artificial Intelligence and Healthcare Simulation: The Shifting Landscape of Medical Education. *Cureus*. https://doi.org/10.7759/cureus.59747

Harley, A. (2015). Personas Make Users Memorable for Product Team Members. *Nielsen Norman Group*.

Hasan, H. R., & Salah, K. (2019). Combating Deepfake Videos Using Blockchain and Smart Contracts. *IEEE Access*, *7*, 41596–41606. https://doi.org/10.1109/ACCESS.2019.2905689

Hill, C. G., Haag, M., Oleson, A., Mendez, C., Marsden, N., Sarma, A., & Burnett, M. (2017). Gender-Inclusiveness Personas vs. Stereotyping: Can We Have it Both Ways? *Proceedings of the 2017 CHI Conference*, 6658–6671. https://doi.org/10.1145/3025453.3025609

Howard, T. W. (2015a). Are personas really usable? *Communication Design Quarterly*, *3*(2), 20–26. https://doi.org/10.1145/2752853.2752856

Howard, T. W. (2015b). Are Personas Really Usable? *Communication Design Quarterly Review*, *3*(2), 20–26. https://doi.org/10.1145/2752853.2752856

Huang, J. (2024). *Understanding individuals' experience with virtual characters from the perspective of perceived authenticity* [Nanyang Technological University]. https://doi.org/10.32657/10356/175025

Ihde, D. (Ed.). (2010). *Technology and the lifeworld: From garden to earth*. Indiana University Press.

Jack, R. E., & Schyns, P. G. (2015). The Human Face as a Dynamic Tool for Social Communication. *Current Biology*, *25*(14), R621–R634. https://doi.org/10.1016/j.cub.2015.05.052

Jacobsen, B. N. (2024). Deepfakes and the promise of algorithmic detectability. *European Journal of Cultural Studies*, 13675494241240028. https://doi.org/10.1177/13675494241240028

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635. https://doi.org/10.1145/3442188.3445923

Jansen, B. J., Jung, S.-G., Nielsen, L., Guan, K. W., & Salminen, J. (2022). How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment. *Pacific Asia Journal of the Association for Information Systems*, *14*, 1–28. https://doi.org/10.17705/1pais.14301

Jansen, B. J., Salminen, J. O., & Jung, S.-G. (2020). Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics. *Data and Information Management*, *4*(1), 1–17. https://doi.org/10.2478/dim-2020-0005

Jansen, B., Jung, S., & Salminen, J. (2019). Creating Manageable Persona Sets from Large User Populations. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3313006

Jansen, B., Salminen, J., Jung, S., & Guan, K. (2021). *Data-Driven Personas* (1st ed., Vol. 14). Morgan & Claypool Publishers.

Jasanoff, S. (Ed.). (2004). *States of Knowledge* (0 ed.). Routledge. https://doi.org/10.4324/9780203413845

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

Joodi, M. A., Saleh, M. H., & Khadhim, D. J. (2023). Proposed Face Detection Classification Model Based on Amazon Web Services Cloud (AWS). *Journal of Engineering*, *29*(4), 176–206. https://doi.org/10.31026/j.eng.2023.04.12

Jung, S., Salminen, J., & Jansen, B. J. (2019). Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, 1–6. https://doi.org/10.1145/3290607.3312955

Jung, S.-G., An, J., Kwak, H., Ahmad, M., Nielsen, L., & Jansen, B. J. (2017). Persona Generation from Aggregated Social Media Data. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1748–1755. https://doi.org/10.1145/3027063.3053120

Jung, S.-G., Salminen, J., Aldous, K. K., & Jansen, B. J. (2025). PersonaCraft: Leveraging language models for data-driven persona development. *International Journal of Human-Computer Studies*, *197*, 103445. https://doi.org/10.1016/j.ijhcs.2025.103445

Kaate, I., Salminen, J., Al Tamime, R., Jung, S., & Jansen, B. J. (2025). Is Deepfake Diversity Real? Analyzing the Diversity of Deepfake Avatars. *Expert Systems with Applications*, *269*, 126382. https://doi.org/10.1016/j.eswa.2025.126382

Kaate, I., Salminen, J., Jung, S.-G., Almerekhi, H., & Jansen, B. J. (2023). How Do Users Perceive Deepfake Personas? Investigating the Deepfake User Perception and Its Implications for Human-Computer Interaction. *Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter*, 1–12. https://doi.org/10.1145/3605390.3605397

Kaate, I., Salminen, J., Jung, S.-G., Olkkonen, R., & Jansen, B. J. (2023). How Can Intelligent Persona Features Support Online Advertising Work? *ACHI 2023 : The Sixteenth International Conference on Advances in Computer-Human Interactions*, 65–67. https://www.thinkmind.org/download_full.php?instance=ACHI+2023

Kaate, I., Salminen, J., Jung, S.-G., Rizun, N., Revina, A., & Jansen, B. J. (2024). Getting Emotional Enough: Analyzing Emotional Diversity in Deepfake Avatars. *Nordic Conference on Human-Computer Interaction*, 1–12. https://doi.org/10.1145/3679318.3685398

Kaate, I., Salminen, J., Jung, S.-G., Santos, J. M., Häyhänen, E., Xuan, T., Azem, J., & Jansen, B. J. (2024). Modeling the New Modalities of Personas: How Do Users' Attributes Influence Their Perceptions and Use of Interactive Personas? *Adjunct Proceedings of the 32nd ACM Conference* *on User Modeling, Adaptation and Personalization*, 164–169. https://doi.org/10.1145/3631700.3664882

Kaate, I., Salminen, J., Jung, S.-G., Santos, J. M., Häyhänen, E., Xuan, T., Azem, J. Y., & Jansen, B. (2025). The 'fourth wall' and other usability issues in AI-generated personas: Comparing chat-based and profile personas. *Behaviour & Information Technology*, 1–17. https://doi.org/10.1080/0144929X.2025.2469659

Kaate, I., Salminen, J., Jung, S.-G., Xuan, T. T. T., Häyhänen, E., Azem, J. Y., & Jansen, B. J. (2025). "You Always Get an Answer": Analyzing Users' Interaction with AI-Generated Personas Given Unanswerable Questions and Risk of Hallucination. *Proceedings of the 30th International Conference on Intelligent User Interfaces*, 1624–1638. https://doi.org/10.1145/3708359.3712160

Kaate, I., Salminen, J., Santos, J., Jung, S.-G., Olkkonen, R., & Jansen, B. (2023). The realness of fakes: Primary evidence of the effect of deepfake personas on user perceptions in a design task. *International Journal of Human-Computer Studies*, *178*, 103096. https://doi.org/10.1016/j.ijhcs.2023.103096

Kaate, I., Salminen, J., Santos, J. M., Jung, S.-G., Almerekhi, H., & Jansen, B. J. (2024). "There Is something Rotten in Denmark": Investigating the Deepfake persona perceptions and their Implications for human-centered AI. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100031. https://doi.org/10.1016/j.chbah.2023.100031

Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4396–4405. https://doi.org/10.1109/CVPR.2019.00453

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00390

Ketamo, H., Kiili, K., & Alajääski, J. (2010). Reverse Market Segmentation with Personas. *Proceedings of the 6th International Conference on Web Information Systems and Technology*, 63–68. https://doi.org/10.5220/0002781300630068

Kietzmann, J., Mills, A. J., & Plangger, K. (2021). Deepfakes: Perspectives on the future "reality" of advertising and branding. *International Journal of Advertising*, *40*(3), 473–485. https://doi.org/10.1080/02650487.2020.1834211

Kim, R. Y. (2024). Anthropomorphism and Human-Robot Interaction. *Communications of the ACM*, *67*(2), 80–85. https://doi.org/10.1145/3624716

Kohli, A. K., & Jaworski, B. J. (1990). Market Orientation: The Construct, Research Propositions, and Managerial Implications. *Journal of Marketing*, *54*(2), 1–18. https://doi.org/10.1177/002224299005400201

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Korsgaard, D., Bjørner, T., Sørensen, P. K., & Burelli, P. (2020). Creating user stereotypes for persona development from qualitative data through semi-automatic subspace clustering. *User Modeling and User-Adapted Interaction*, *30*(1), 81–125. https://doi.org/10.1007/s11257-019-09252-5

Korshunov, P., & Marcel, S. (2018). *DeepFakes: A New Threat to Face Recognition? Assessment and Detection* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1812.08685

Kozar, K., & Miaskiewicz, T. (2009). Designing the introductory is course using student personas: Lessons learned from product design. *AMCIS 2009 Proceedings*, 454.

Kunur, M., Langdon, P. M., Bradley, M. D., Bichard, J.-A., Glazer, E., Doran, F., Clarkson, P. J., & Loeillet, J. J. (2016). Reducing Exclusion in Future Cars Using Personas with Visual Narratives and Design Anthropology. In P. Langdon, J. Lazar, A. Heylighen, & H. Dong (Eds.), *Designing Around People* (pp. 269–277). Springer International Publishing. https://doi.org/10.1007/978-3-319-29498-8_28

Latour, B. (2007). *Reassembling the social: An introduction to Actor-Network-Theory* (1. publ. in pbk). Oxford Univ. Press.

Leavy, S. (2018). Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 14–16. https://doi.org/10.1145/3195570.3195580

Li, M., Liu, B., Hu, Y., & Wang, Y. (2021). Exposing Deepfake Videos by Tracking Eye Movements. *2020 25th International Conference on Pattern Recognition (ICPR)*, 5184–5189. https://doi.org/10.1109/ICPR48806.2021.9413139

Li, Y., Chang, M.-C., & Lyu, S. (2018). In ictu oculi: Exposing ai created fake videos by detecting eye blinking. *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–7.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Nachdr.). Sage.

Long, F. (2009). Real or imaginary: The effectiveness of using personas in product design. *Proceedings of the Irish Ergonomics Society Annual Conference*, *14*.

López-Gil, J.-M., Gil, R., & García, R. (2022). Do Deepfakes Adequately Display Emotions? A Study on Deepfake Facial Emotion Expression. *Computational Intelligence and Neuroscience*, *2022*, 1–12. https://doi.org/10.1155/2022/1332122

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, *146*, 190–205. https://doi.org/10.1016/j.cognition.2015.09.019

Mane, S., & Shah, G. (2019). Facial Recognition, Expression Recognition, and Gender Identification. In V. E. Balas, N. Sharma, & A. Chakrabarti (Eds.), *Data Management, Analytics and Innovation* (pp. 275–290). Springer Singapore.

Marsden, N., & Pröbster, M. (2019). Personas and Identity: Looking at Multiple Identities to Inform the Construction of Personas. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300565

Martingano, A. J., Hererra, F., & Konrath, S. (2021). Virtual reality improves emotional but not cognitive empathy: A meta-analysis. *Technology, Mind, and Behavior*, *2*(1). https://doi.org/10.1037/tmb0000034

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Massanari, A. L. (2010). Designing for imaginary friends: Information architecture, personas and the politics of user-centered design. *New Media & Society*, *12*(3), 401–416. https://doi.org/10.1177/1461444809346722

Matthews, T., Judge, T., & Whittaker, S. (2012). How do designers and user experience professionals actually perceive and use personas? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1219–1228. https://doi.org/10.1145/2207676.2208573

McDuff, D., & Czerwinski, M. (2018). Designing emotionally sentient agents. *Communications of the ACM*, *61*(12), 74–83. https://doi.org/10.1145/3186591

McGinn, J. (Jen), & Kotamraju, N. (2008). Data-driven persona development. *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, 1521. https://doi.org/10.1145/1357054.1357292

Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2022). An Emerging Theory of Avatar Marketing. *Journal of Marketing*, *86*(1), 67–90. https://doi.org/10.1177/0022242921996646

Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, *32*(5), 417–430.

Mihailova, M. (2021). To Dally with Dalí: Deepfake (Inter)faces in the Art Museum. *Convergence: The International Journal of Research into New Media Technologies*, *27*(4), 882–898. https://doi.org/10.1177/13548565211029401

Miller, K., & Taddeo, M. (Eds.). (2017). *The ethics of information technologies*. Routledge.

Mir, U., Kar, A. K., & Gupta, M. P. (2022). AI-enabled digital identity – inputs for stakeholders and policymakers. *Journal of Science and Technology Policy Management*, *13*(3), 514–541. https://doi.org/10.1108/JSTPM-09-2020-0134

Misztal-Radecka, J., & Indurkhya, B. (2020). Persona Prototypes for Improving the Qualitative Evaluation of Recommendation Systems. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 206–212. https://doi.org/10.1145/3386392.3399297

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, *3*(2), 2053951716679679.

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research*, *1*(1), 48–76. https://doi.org/10.1177/2345678906292462

Mori, M. (1970). The Uncanny Valley. *Energy*, *7*(4), 33–35.

Mulder, S., & Yaar, Z. (2006). *The User is Always Right: A Practical Guide to Creating and Using Personas for the Web*. New Riders.

Müller, N. M., Pizzi, K., & Williams, J. (2021). *Human Perception of Audio Deepfakes*. https://doi.org/10.48550/ARXIV.2107.09667

Murphy, G., Ching, D., Twomey, J., & Linehan, C. (2023). Face/Off: Changing the face of movies with deepfakes. *PLOS ONE*, *18*(7), e0287503. https://doi.org/10.1371/journal.pone.0287503

Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research*, *154*, 113368. https://doi.org/10.1016/j.jbusres.2022.113368

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., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78. https://doi.org/10.1145/191666.191703

Nielsen, L. (2004). *Engaging personas and narrative scenarios* [PhD Thesis]. Samfundslitteratur.

Nielsen, L. (2019). *Personas—User Focused Design* (2nd ed. 2019 edition). Springer.

Nielsen, L., Hansen, K. S., Stage, J., & Billestrup, J. (2015). A Template for Design Personas: Analysis of 47 Persona Descriptions from Danish Industries and Organizations. *International Journal of Sociotechnology and Knowledge Development*, *7*(1), 45–61. https://doi.org/10.4018/ijskd.2015010104

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Park, D., & Kang, J. (2022). Constructing Data-Driven Personas through an Analysis of Mobile Application Store Data. *Applied Sciences*, *12*(6), Article 6. https://doi.org/10.3390/app12062869

Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, *3*(12), 1013–1022. https://doi.org/10.1038/s42256-021-00417-9

Ping, R. A. (2004). On assuring valid measures for theoretical models using survey data. *Journal of Business Research*, *57*(2), 125–141. https://doi.org/10.1016/S0148-2963(01)00297-1

Preu, E., Jackson, M., & Choudhury, N. (2022). Perception vs. Reality: Understanding and Evaluating the Impact of Synthetic Image Deepfakes over College Students. *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 0547–0553. https://doi.org/10.1109/UEMCON54665.2022.9965697

Pruitt, J., & Adlin, T. (2010). *The Persona Lifecycle* (1st edition). Morgan Kaufmann.

Pruitt, J. S., & Adlin, T. (2006). Persona Conception and Gestation. In *The Persona Lifecycle* (pp. 162–271). https://linkinghub.elsevier.com/retrieve/pii/B9780125662512500058

Rafael, G., Kusuma, H., & Tasripan. (2020). The Utilization of Cloud Computing for Facial Expression Recognition using Amazon Web Services. *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, 366–370. https://doi.org/10.1109/CENIM51130.2020.9297974

Raji, I. D., & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 429–435. https://doi.org/10.1145/3306618.3314244

Ramos, H., Fonseca, M., & Ponciano, L. (2021). *Modeling and Evaluating Personas with Software Explainability Requirements*. https://doi.org/10.48550/ARXIV.2108.04640

Renier, L. A., Shubham, K., Vijay, R. S., Mishra, S. S., Kleinlogel, E. P., Jayagopi, D. B., & Schmid Mast, M. (2024). A deepfake-based study on facial expressiveness and social outcomes. *Scientific Reports*, *14*(1), 3642. https://doi.org/10.1038/s41598-024-53475-5

Revella, A. (2015). *Buyer personas: How to gain insight into your customer's expectations, align your marketing strategies, and win more business*. Wiley.

Rönkkö, K., Hellman, M., Kilander, B., & Dittrich, Y. (2004). Personas is not applicable: Local remedies interpreted in a wider context. *Proceedings of the Eighth Conference on Participatory Design Artful Integration: Interweaving Media, Materials and Practices - PDC 04*, *1*, 112. https://doi.org/10.1145/1011870.1011884

Salminen, J., Guan, K., Jung, S.-G., & Jansen, B. J. (2021). A Survey of 15 Years of Data-Driven Persona Development. *International Journal of Human–Computer Interaction*, *0*(0), 1–24. https://doi.org/10.1080/10447318.2021.1908670

Salminen, J., Guan, K., Nielsen, L., Jung, S., Chowdhury, S. A., & Jansen, B. J. (2020). A Template for Data-Driven Personas: Analyzing 31 Quantitatively Oriented Persona Profiles. In S. Yamamoto & H. Mori (Eds.), *Human Interface and the Management of Information. Designing Information. HCII 2020.* (Vol. 12184, pp. 125–144). Springer.

Salminen, J., Jansen, B. J., An, J., Kwak, H., & Jung, S. (2018). Are personas done? Evaluating their usefulness in the age of digital analytics. *Persona Studies*, *4*(2), 47–65. https://doi.org/10.21153/psj2018vol4no2art737

Salminen, J., Jung, S.-G., Chowdhury, S., Sengün, S., & Jansen, B. J. (2020). Personas and Analytics: A Comparative User Study of Efficiency and Effectiveness for a User Identification Task. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376770

Salminen, J., Jung, S.-G., M. Santos, J., Mohamed Sayed Kamel, A., & J. Jansen, B. (2021). Picturing It!: The Effect of Image Styles on User Perceptions of Personas. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445360

Salminen, J., Jung, S.-G., Nielsen, L., & Jansen, B. (2022). Creating More Personas Improves Representation of Demographically Diverse Populations: Implications Towards Interactive Persona Systems. *Nordic Human-Computer Interaction Conference*, 1–11. https://doi.org/10.1145/3546155.3546654

Salminen, J., Liu, Y.-H., Şengün, S., Santos, J. M., Jung, S., & Jansen, B. J. (2020). The effect of numerical and textual information on visual engagement and perceptions of AI-driven persona interfaces. *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 357–368. https://doi.org/10.1145/3377325.3377492

Salminen, J., Mustak, M., Corporan, J., Jung, S., & Jansen, B. J. (2022). Detecting Pain Points from User-Generated Social Media Posts Using Machine Learning. *Journal of Interactive Marketing*, 10949968221095556. https://doi.org/10.1177/10949968221095556

Salminen, J., Nielsen, L., Jung, S.-G., An, J., Kwak, H., & Jansen, B. J. (2018). "Is More Better?": Impact of Multiple Photos on Perception of Persona Profiles. *Proceedings of the 2018 CHI*

Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3173574.3173891

Salminen, J., Santos, J. M., Kwak, H., An, J., Jung, S., & Jansen, B. J. (2020). Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas. *International Journal of Human-Computer Studies*, *141*, 102437. https://doi.org/10.1016/j.ijhcs.2020.102437

Seymour, M., Riemer, K., Yuan, L., & Dennis, A. (2021). *Beyond deep fakes: Conceptual framework, applications, and research agenda for neural rendering of realistic digital faces*.

Shah, T. R., Purohit, S., Das, M., & Arulsivakumar, T. (2025). Do I look real? Impact of digital human avatar influencer realism on consumer engagement and attachment. *Journal of Consumer Marketing*. https://doi.org/10.1108/JCM-05-2024-6853

Shahbazi, N., Lin, Y., Asudeh, A., & Jagadish, H. V. (2023). Representation Bias in Data: A Survey on Identification and Resolution Techniques. *ACM Computing Surveys*, *55*(13s), 1–39. https://doi.org/10.1145/3588433

Sharma, I., Jain, K., Behl, A., Baabdullah, A., Giannakis, M., & Dwivedi, Y. (2023). Examining the motivations of sharing political deepfake videos: The role of political brand hate and moral consciousness. *Internet Research*, *33*(5), 1727–1749. https://doi.org/10.1108/INTR-07-2022-0563

Sharma, P., Kumar, M., & Sharma, H. K. (2024). GAN-CNN Ensemble: A Robust Deepfake Detection Model of Social Media Images Using Minimized Catastrophic Forgetting and Generative Replay Technique. *Procedia Computer Science*, *235*, 948–960. https://doi.org/10.1016/j.procs.2024.04.090

Shin, S. Y., & Lee, J. (2022). The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes. *Digital Journalism*, *10*(3), 412–432. https://doi.org/10.1080/21670811.2022.2026797

Smith, E. R., Šabanović, S., & Fraune, M. R. (2021). Human–robot interaction through the lens of social psychological theories of intergroup behavior. *Technology, Mind, and Behavior*, *1*(2). https://doi.org/10.1037/tmb0000002

Stojnić, A. (2015). Digital anthropomorphism: Performers avatars and chat-bots. *Performance Research*, *20*(2), 70–77. https://doi.org/10.1080/13528165.2015.1026733

Stumpf, S., Taka, E., Nakao, Y., Luo, L., Sonoda, R., & Yokota, T. (2024). The Need for User-centred Assessment of AI Fairness and Correctness. *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, 523–527. https://doi.org/10.1145/3631700.3664912

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions, 2nd ed.* (pp. xii, 314). Cambridge University Press.

Synthesia. (2022). *60+ Languages | Different Voices & Accents | Synthesia*. https://www.synthesia.io/features/languages

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

Tajfel, H., & Turner, J. C. (1986). The Social Identity Theory of Intergroup Behavior. In *Psychology of Intergroup Relation*.

Turner, P., & Turner, S. (2011). Is stereotyping inevitable when designing with personas? *Design Studies*, *32*(1), 30–44.

Usukhbayar, B., & Homer, S. (2020). *Deepfake Videos: The Future of Entertainment*.

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media+ Society*, *6*(1), 2056305120903408.

Vassilakopoulou, P., & Hustad, E. (2021). Bridging Digital Divides: A Literature Review and Research Agenda for Information Systems Research. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-020-10096-3

Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things* (Online-Ausg). University of Chicago Press.

Wagner, T. L., & Blewer, A. (2019). “The Word Real Is No Longer Real”: Deepfakes, Gender, and the Challenges of AI-Altered Video. *Open Information Science*, *3*(1), 32–46. https://doi.org/10.1515/opis-2019-0003

Wang, L., Zhou, L., Yang, W., & Yu, R. (2022). Deepfakes: A new threat to image fabrication in scientific publications? *Patterns*, *3*(5), 100509. https://doi.org/10.1016/j.patter.2022.100509

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Weisman, W. D., & Peña, J. F. (2021). Face the Uncanny: The Effects of Doppelganger Talking Head Avatars on Affect-Based Trust Toward Artificial Intelligence Technology are Mediated by Uncanny Valley Perceptions. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 182–187. https://doi.org/10.1089/cyber.2020.0175

White, P. J., & Devitt, F. (2021). Creating personas from design ethnography and grounded theory. *J. Usability Studies*, *16*(3), 156–178.

Whittaker, L., Letheren, K., & Mulcahy, R. (2021). The Rise of Deepfakes: A Conceptual Framework and Research Agenda for Marketing. *Australasian Marketing Journal*, *29*(3), 204–214. https://doi.org/10.1177/1839334921999479

Zhang, X., Brown, H.-F., & Shankar, A. (2016). Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16*, 5350–5359. https://doi.org/10.1145/2858036.2858523

Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). *Bias in Generative AI* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2403.02726

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0428-0 (PRINT)
ISBN 978-952-02-0429-7 (PDF)
ISSN 2343-3159 (Print)
ISSN 2343-3167 (Online)