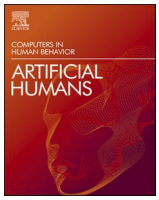
Computers in Human Behavior: Artificial Humans 2 (2024) 100031

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

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



"There Is something Rotten in Denmark": Investigating the Deepfake persona perceptions and their Implications for human-centered AI

Ilkka Kaate [a,*], Joni Salminen [b], João M. Santos [c], Soon-Gyo Jung [d], Hind Almerekhi [d], Bernard J. Jansen [d]

[a] University of Turku, FI-20014, Turun yliopisto, Finland
[b] University of Vaasa, Wolffintie 32, FI-65200, Vaasa, PL 700, Finland
[c] Instituto Universitário de Lisboa (ISCTE-IUL), Avenida das Forças Armadas, 1649-026, Lisboa, Portugal
[d] Qatar Computing Research Institute, P.O. Box: 34110, Education City, Doha, Qatar

ARTICLE INFO

Keywords:
Deepfakes
User perceptions
Human-centered AI
User study

ABSTRACT

Although they often have a negative connotation due to their social risks, deepfakes have the potential to improve HCI, human-centered AI, and user experience (UX). To investigate the impact of deepfakes on persona UX, we conducted an experimental study with 46 users who used a deepfake persona and a human persona to carry out a design task. We collected think-aloud, observant notes, and survey data. The results of our mixed-method analysis indicate that if users observe glitches in the deepfake personas, these glitches have a detrimental effect on the persona UX and task performance; however, not all users identify glitches. Our quantitative analysis of survey data shows that there are differences in how (a) users perceive deepfakes, (b) users detect deepfake glitches, (c) deepfake glitches affect information comprehension, and (d) deepfake glitches affect task completion. Glitches have the most significant impact on authenticity, persona perception, and task perception variables but less impact on behavioral variables. The results imply that organizations implementing deepfake personas need to address perceptual challenges before the full potential of deepfake technology can be realized for persona creation.

1. Introduction

Human-computer interaction (HCI) is increasingly affected by Artificial Intelligence (AI) as information systems are integrating AI components to enhance user experience (UX) (Schmidt, 2021). The remarkable progress of AI technologies, often leveraging machine learning (ML) innovations, has introduced opportunities for empowering users in information systems (Agostinelli, Battaglini, Catarci, Dal Falco, & Marrella, 2019; Barricelli & Fogli, 2021; Catania et al., 2021; Ferrell, Grando, & Zancanaro, 2021), computer-supported collaborative work (Galassi & Vittorini, 2021), and HCI, spawning a new subset of research, human-centered AI (HCAI). One of the promising and ominous technologies for HCAI is deepfake technology. Deepfakes are photo-realistic, computer-generated human representations, typically in the form of videos (Mustak et al., 2023). These deepfakes could enhance user interaction with information systems, although research is still in the early stages of corroborating their potential benefits in terms of empirical evidence. In this research, we investigate deepfake personas (DFPs) that are personas created using deepfake technology. Personas are fictional characters that represent central user groups (An, Kwak, Salminen, Jung, & Jansen, 2018) and are used by humans, for example, in system development (Cooper, 1999), product design (Pruitt & Adlin, 2006), and marketing (Revella, 2015).

While much research on deepfakes has thus far focused on the risks and negative implications of this new technology, including manipulation, misinformation, and fake news (Gamage, Ghasiya, Bonagiri, Whiting, & Sasahara, 2022; Hancock & Bailenson, 2021; Lyu, 2020), it is vital to acknowledge that DFPs also bring about positive opportunities for HCAI (Danry et al., 2022; Mustak et al., 2023). For example, DFPs could enhance the UX in *Metaverse* applications (Tricomi et al., 2023), increase the level of realism in virtual customer service agents, and act as pedagogical agents to inform or educate students about different topics. Therefore, deepfake technology, despite its risks, has potential value in improving users' self-expression and the interaction quality

* Corresponding author.
E-mail addresses: iokaat@utu.fi (I. Kaate), joni.salminen@uwasa.fi (J. Salminen), jmcsm@iscte.pt (J.M. Santos), sjung@hbku.edu.qa (S.-G. Jung), hialmerekhi@hbku.edu.qa (H. Almerekhi), bjansen@hbku.edu.qa (B.J. Jansen).

https://doi.org/10.1016/j.chbah.2023.100031
Received 7 June 2023; Received in revised form 22 November 2023; Accepted 22 November 2023
Available online 1 December 2023

between organizations and their customers.

However, a central antecedent to realizing these potential benefits is that the DFPs are or can be experienced positively by the end-users. If users find DFPs scary, dull, or creepy - as per the uncanny valley effect (Mori et al., 2012) - then the users would likely resist adopting and using the DFPs and instead prefer other interaction techniques. Therefore, how users perceive DFPs is central to integrating them into real information systems. Alas, we know little of this essential human factor: *deepfake user perceptions remain largely unexplored in current HCI research*. Without empirically oriented research informing us of the crucial dimensions of how DFPs are perceived and why, it is difficult to ascertain the pros and cons of implementing them in real information systems towards positive net effects for design.

One obstacle in the way of deepfake technology being widely used in HCI and design tasks is the presence of what we refer to as glitches in DFPs. Glitches are sudden, perhaps temporary irregularities of properties in the deepfake that are perceived as abnormal or unnatural by the users (Appel & Prietzel, 2022; Bode, 2021; Hasan & Salah, 2019). Glitches are found to be a major contribution to automatic (Gupta, Chugh, Dhall, & Subramanian, 2020) and human user deepfake detection (Appel & Prietzel, 2022), that is, glitches usually "give away" deepfakes. Glitches are usually found in several facial features of deepfakes, such as eyes, nose, and mouth (Appel & Prietzel, 2022; Bode, 2021; eSafety, 2022; Gupta et al., 2020) as well as in the voice of deepfakes (Müller et al., 2021). Although deepfake technology has been under development since 1997 (Bregler, Covell, & Slaney, 1997), deepfake technology faces visual and auditive restrictions (Broad, Leymarie, & Grierson, 2020; Weisman & Peña, 2021) resulting in glitches.

In this research, we are interested in determining users' glitch detection abilities, which we call deepfake *glitch perception*. Glitches may signal to the users that "There Is Something Rotten in Denmark" (i.e., something is not right).

To this end, our study aims to address this knowledge gap by exploring central themes in users' deepfake *glitch* perception. Our study is guided by a motivational question, *"How do users perceive deepfakes for a design task?"*. Based on this motivational question, we focus on three research questions (RQs):

- **RQ1:** What type of glitches do users observe in DFPs?
- **RQ2:** How do users' **(a)** perceptions of the persona and **(b)** behavior vary between deepfake and human personas?
- **RQ3:** How does users' glitch perception affect their **(a)** perceptions of the persona and **(b)** behavior?

RQ1 examines what does users' glitch perception consist of. RQ2 examines how does the persona's realness affect user perceptions and behavior. RQ3 examines how does the strength of the glitch perception drive user perceptions and behavior. To address these RQs, we performed a user experiment with 46 users each interacting with two experimental video scenarios, one involving a human persona and one a DFP. Our findings shed light on the nature of the deepfake user perception, particularly the deepfake persona perception (i.e., user perceptions of DFPs), offering avenues for theorization and further empirical work on understanding human-deepfake interaction in greater detail. Our analysis contributes to HCI, particularly to HCAI, a nascent subfield that deals with analyzing how deepfake technologies can be integrated into the design process by leveraging personas.

2. Literature review

2.1. What is known about how users perceive eepfakes?

As the potential implications of deepfakes have become more evident due to progress in AI technology, there has been an increase in research focused on deepfakes in the HCI domain. While most studies focus on deepfake detection (Lyu, 2020), i.e., developing algorithms and models for this task, there has been a gradual rise in the number of studies exploring how deepfakes are perceived by human users and applied in design. Despite being a relatively novel field of research, deepfake perception studies have covered this topic from various angles using a wide range of techniques (Müller et al., 2021). Table 1 presents the central themes in the literature and summarizes said findings presented in the literature so far. Research on deepfake user perception is categorized in Table 1 based on the article's emphasis as *Harmful* (focusing on negative implications of deepfakes, n = 3), *Detecting* (focusing on deepfake detection, n = 12), *Consequences* (focusing on deepfakes' ramifications, n = 6) and *Attributes* (focusing on attributes of deepfakes, n = 5).

Deepfakes are seen in the literature largely through their harmful or negative properties and misuse (Table 1), not so much from the angle of using deepfakes in design tasks or for 'good intentions'. A lot of the findings in the prior literature have negative connotations and/or are dealing with deepfake detection rather than utilizing deepfake technology for good. Deepfake detection has focused on automatic or algorithmic deepfake detection, but human deepfake detection has been less investigated. According to Thaw et al. (2021), users are poor at detecting deepfakes. Pu et al. (2021) found that inconsistent facial features and head movements act as cues for users to detect deepfakes. The ability to detect glitches may vary due to the users' background with deepfakes (Groh et al., 2022), user political views (Appel & Prietzel, 2022), agreement with content (Appel & Prietzel, 2022), and cognitive capabilities (Groh et al., 2022). Generally, users do not perceive reflections, shadows, or other non-credibility cues in deepfake videos well (Groh et al., 2022). On the other hand, detecting abnormalities in personas used in design tasks has been found to deteriorate the design process (Vincent & Blandford, 2014). As users' positive interaction with deepfakes generally increases the value of deepfakes (Seymour et al., 2021), finding the DFP strange, not trustworthy, and not credible would likely lower the persona's usability in design tasks (Seymour et al., 2021).

2.2. Why does deepfake persona perception matter?

Understanding how users perceive deepfakes is vital for advancing HCAI since user perceptions can have deciding consequences for using deepfakes in various applications, including virtual reality environments, virtual assistants, educational applications, and so on (Seymour et al., 2021). However, if deepfakes are to be used in these applications, users must have a positive experience with the deepfakes. Should deepfakes be perceived as untrustworthy or confusing, this may result in a negative UX and diminished acceptance of the technology. Understanding how users perceive deepfakes might thus assist designers in creating more effective and user-friendly systems and applications (Gamage, Ghasiya, et al., 2022; Grodzinsky, Miller, & Wolf, 2011; Kleine, 2022). For example, users may be more distrustful of a video's content and less likely to believe it if they know it is a deepfake.

While deepfakes have received much attention due to their ability to manipulate images, videos, and audio, which has raised concerns about their possible misuse (L. Wang et al., 2022; Westerlund, 2019), when used responsibly, deepfakes have several potential benefits that may give users a positive experience (Cruse, 2006). For example, DFPs can assist users with disabilities by generating artificial sign language, and facial emotions, and recreating the voice of those who cannot speak (Chesney & Citron, 2019). Deepfakes can also enhance players' experience in gaming through in-gaming aids (Westerlund, 2019). Additionally, deepfakes can be utilized for educational purposes by enhancing the learning experience in innovative ways (Cruse, 2006). Deepfakes can improve education and provide a more personalized learning experience by producing educational content (Silbey & Hartzog, 2018). Such applications can make deepfakes less scary and more engaging (Chesney & Citron, 2019).

Deepfake technology can also aid in rehabilitating users with addictions, such as smoking. The World Health Organization has created

Table 1
Research on deepfake user perception and a summary of key findings in the literature on public perception of deepfakes.

The article predominantly sees deepfakes as …	Literature References	Key findings
Harmful	(Cleveland, 2022; Dobber, Metoui, Trilling, Helberger, & de Vreese, 2021; Kugler & Pace, 2021)	• Deepfakes have been perceived by the majority of users as entertaining and impressive but also raising concerns about the potential misuse of deepfake technology (Cleveland, 2022). • Attitude towards public figures can deteriorate after a deepfake video surfaces (Dobber et al., 2021). • Pornographic deepfakes are perceived as more harmful compared to non-pornographic deepfakes. The public is concerned about using deepfakes for privacy violations (Kugler & Pace, 2021).
Detecting	(Barari, Lucas, & Munger, 2021; Groh, Epstein, Firestone, & Picard, 2022; Köbis, Doležalová, & Soraperra, 2021; Lewis et al., 2022; Mink et al., 2022; Müller et al., 2021; Ng, 2022; Pu et al., 2021; Shahid et al., 2022; Sütterlin et al., 2021; Thaw et al., 2021; Vaccari & Chadwick, 2020)	• Almost 80% of users could recognize deepfake videos (Groh et al., 2022). • Users could correctly perceive that the deepfake and deepfake-described videos were less realistic than the real and non-described videos (Ng, 2022). • Users' perception of deepfake audio clips was on almost similar level as an advanced deepfake detection algorithm. Native speakers performed better than non-natives (Müller et al., 2021). • Most users perceived deepfake videos as real. When informed about deepfake videos, users showed no concern (Shahid et al., 2022). • Users tend to believe realistic, artificially made social network profiles, even if they are susceptible to phishing attacks. Users more likely interact with profiles with more connections and information since they tend to be more trustworthy (Mink et al., 2022). • Content warnings did not significantly enhance user perception of deepfakes, implying that other interventions may be required to address the threat of deepfakes (Lewis et al., 2022). • A detection method based on inconsistencies in facial features and head movements between the fake and original videos achieved high accuracy in detecting deepfakes (Pu et al., 2021). • Analysis on public perception regarding deepfake videos, audio, and texts revealed that a higher percentage (44%) of users found fake audios are authentic compared to deepfake videos where 42% of users found the videos and texts as authentic (Barari et al., 2021). • Political leaning of the viewer, level of agreement with the content, and the device used impact deepfake perception. Conservatives and content-agreeing viewers are less likely to differentiate between real and deepfake content (Sütterlin et al., 2021). • The majority of the users perceived deepfake videos as real (Thaw et al., 2021). • Users could not differentiate between real and deepfake videos despite being educated regarding deepfakes and offered financial incentives for recognizing deepfake content (Köbis et al., 2021). • A deepfake video was created to determine users' perceptions. Half (50.8%) of the users identified the video as fake, 33% were uncertain and 16% perceived it as real (Vaccari & Chadwick, 2020).
Consequences	(Ahmed, 2021; Ahmed, Ng, & Wei Ting, 2023; Hughes et al., 2023; Hwang, Ryu, & Jeong, 2021; Ternovski et al., 2022; Wittenberg et al., 2021)	• People thinking deepfakes are true are more likely to share them on social media. People with inferior cognitive abilities were found to be more prone to spread deepfakes (Ahmed et al., 2023). • Potential voters perceived even real videos as fake when they were informed about the existence of deepfakes. It could tarnish their trust in political institutions and strengthen their belief in conspiracy theories (Ternovski et al., 2022). • Users were exposed to real and deepfake audio and videos. Users perceived deepfake videos and audio as real content, which changed their attitudes and intentions (Hughes et al., 2023). • Majority of respondents perceived deepfake political videos as real and found fake videos more believable compared to doctored texts. The videos altered users' political views (Wittenberg et al., 2021). • Users tend to perceive non-political deepfake videos as

(*continued on next page*)

Table 1 (*continued*)

The article predominantly sees deepfakes as …	Literature References	Key findings
		more accurate than they actually are and are more likely to share them on social media compared to genuine videos (Ahmed, 2021). • A deepfake video associating a misleading statement with Mark Zuckerberg attracted a vast audience in comparison to an article mentioning the same statement (Hwang et al., 2021).
Attributes	(Lee et al., 2021; Preu et al., 2022); (S. Wang, 2021; Welker et al., 2020; Korshunov & Marcel, 2020)	• The impact of deepfakes on college students' perceptions of trust, credibility, and identity was examined. Students identified deepfakes as fake, still evaluating them as potentially damaging to individuals' reputations and social identity (Preu et al., 2022). • The framing of YouTube videos influences audience perception of credibility and trustworthiness, with political and entertainment contexts generating more skepticism than in social and educational contexts (Lee et al., 2021). • When deepfake videos have adversarial noise, users are more likely to perceive the videos as authentic. Users are also less inclined to share them on social media (S. Wang, 2021). • Users perceived full face swaps more human-like and less disturbing than partial face swaps (Welker et al., 2020). • Subjective human perception to objective criteria was analyzed to study the quality of deepfake videos, revealing that both subjective and objective measures can be useful for assessing deepfake quality (Korshunov & Marcel, 2020).

"Florence," an AI-based solution that assists users with tobacco addiction, with whom users can engage to boost their confidence in quitting smoking by developing a strategy to track their progress (Organization, 2020). Deepfake technology can also be utilized in arts to critique public figures and celebrities and by activists to convey their message innovatively (Usukhbayar & Homer, 2020). Deepfakes can enhance UX in the HCI context by providing more engaging, personalized, and immersive interfaces. A deepfake interface that uses users' faces and/or voices to create videos, avatars, and other content may provide a personalized experience (Whittaker et al., 2021).

Concerning deepfake user perception and deepfake authenticity, three distinct variables have been recognized: (a) eyes (Mustafa et al., 2022), (b) speech (Jafar, Ababneh, Al-Zoube, & Elhassan, 2020), and (c) emotional authenticity (Tinwell et al., 2011). Additionally, Barari et al. (2021) found audio had a key role in making deepfakes sound more realistic. Mustafa et al. (2022) found that eyes are one of the primary properties that help detect deepfake personas in videos and separate fake videos from real videos. The facial expressions and emotions of deepfakes are valuable properties when making deepfakes realistic (Tinwell et al., 2011). Lack of communicated emotion and unnatural facial expressions make deepfakes look unreal.

2.3. *What are the research gaps in deepfake persona perception?*

Research regarding deepfakes is still in its nascent stage; thus, several open research gaps (RGs) remain unexplored.

RG01: First, the way users perceive deepfakes is a relatively unexplored area yet not completely unexplored. Users' viewing patterns of deepfake videos (Gupta et al., 2020) and perceptions of deepfake audio (Müller et al., 2021) have been studied as well as deepfake user perceptions' effect on perceived empathy and credibility in a design task (Kaate et al., 2023). An important potential application of deepfakes is in the creation of virtual assistants which enables users to interact with AI-powered systems (Canbek & Mutlu, 2016). These systems may generate deepfakes as part of their responses. As users' perceptions regarding deepfakes may impact the usability and effectiveness of various systems, it is pertinent to explore how users perceive deepfakes. For example, research may explore how users respond to deepfakes in conversations, how they perceive deepfakes' credibility, and how these factors impact trust and engagement with various systems.

RG02: Second, there is a lack of empirical research regarding users' perceptions of deepfake technology. Although there have been numerous concerns raised about deepfakes' potential negative consequences, such as their ability to spread misinformation or cause harm to individuals and society (Gamage, Ghasiya, et al., 2022; Hancock & Bailenson, 2021; Lyu, 2020), there has been little research into how users react to deepfakes in real-world settings, such as in design tasks. This research gap matters because understanding users' perceptions and responses to deepfakes can help formulate strategies to counter any negative effects.

RG03: Third, the lack of information or cues regarding deepfakes is another research gap that may not only help enhance UX but also assist users in differentiating real content from deepfake. Some studies, like Groh et al. (2022), employed AI tools to enhance users' experience which helped to a certain extent but failed to provide perfect results. Similarly, Bray, Johnson, and Kleinberg (2022) exposed users to deepfakes before the experiment. Since the users were not provided with specific cues, they were not able to differentiate between real and deepfakes, indicating a need for research targeted at determining visual cues, as well as non-visual factors like behavioral information, and contextual and linguistic cues, which can play a vital role in enhancing deepfake applicability in design.

RG04: Fourth, another research gap is the perception of deepfakes among different demographic groups. It is pertinent to explore whether age, gender, political and/or religious affiliation, education level, and other related factors impact users' perception of deepfakes. Research on deepfakes' effects on voters political views (Ternovski et al., 2022) and political views' effect on deepfake perception has been performed (Sütterlin et al., 2021). Although users' perception regarding deepfakes involving different demographic groups has been explored to some extent (Haut et al., 2022), this line of work is still scarce. Most deepfake perception studies have limited scope and do not depict real-world scenarios. To enhance the accuracy of deepfake detection models and improve user perception, diverse datasets are needed reflecting a variety of demographic backgrounds.

RG05: Finally, research regarding glitches in deepfakes and their impact on the design process is scarce. Abnormalities like unusual face characteristics, inconsistencies in lighting, shadows, and differences in voice pitches, along with other types of glitches, may undermine the impact of deepfakes (Appel & Prietzel, 2022; Li et al., 2018). The effects of perceived deepfake realness on perceived empathy and credibility

and the effects on a design task has also been studied (Kaate et al., 2023). Thus, research to determine how different types of glitches affect users' experience and their ability to detect deepfakes is critical to understand whether certain forms of glitches are more effective in identifying deepfake, or whether the presence of glitches creates a cumulative effect that may affect the UX. There is also a need to investigate ways that deepfake creators can employ to minimize these glitches.

3. Methodology

3.1. Experiment design

To address our RQs and some of the RGs, we conducted a user experiment in January 2023. The experiment followed a mixed design method, which involved dividing the users into two groups that are then assigned to a treatment condition. We tested one male and one female DFP, along with one real male and real female (who were hired actors).

Two DFPs and two human personas expressing the same content as in the DFPs were used in the user study (see Table 2). The two deepfake videos used in the user study were created in a deepfake video creation system called Synthesia[1] (Synthesia, 2022), leveraging personas validated in a study by Carey, White, McMahon, and O'Sullivan (2019) and the both the visual persona and audio of DFP's were artificial. One female (Fiona) and one male (James) persona were chosen for the study for balanced gender representation. The two personas were transformed into a narrative form, a written script, and uploaded to Synthesia to be used in the DFP development. The same script was given to the human actors for recording the acted videos. The lengths of the videos used in the experiment were: human James 144 s, human Fiona 142 s, deepfake James 144 s, and deepfake Fiona 156 s. The experiment was pilot tested by three users who were not included in the analysis of the results.

3.2. Users

A total of 46 users carried out the user study, of whom 16 were female (34.8%), and 30 were male (65.2%). The average age of the users was 37.1 years (SD = 10.4). Users' occupations were many, including research associate, GIS expert, project coordinator, custodian, and software engineer. The average years of experience in users' current profession was 9.8 years (SD = 9.6 years). Users' nationalities were multiple, including Chinese, USA, Qatari, British, Pakistani, Filipino, Tanzanian, and Nepalese. Each study administrator kept notes about noteworthy observations of user behavior concerning deepfakes. In addition, the think-aloud of each session was recorded and transcribed, yielding 92 transcriptions of the users explaining how they perceived the videos to which they were exposed.

3.3. Data collection

Recruiting users for the user study took place via email. In the recruitment email, the invitees were told that we were conducting a user study about the impact of video quality on marketing tasks, not to reveal the real purpose of the study. The study was carried out on the university premises. Two identical workstations were used consisting of two laptops, a mouse, a Sony voice recorder, and a separate 24″ display. Fig. 1 shows the experiment from the user's point of view, while Table 3 shows the guidance given to the users.

The videos containing the DFPs and real humans were uploaded to YouTube and then presented with METRIC, which is a real-time user study and analytics system (Metric, 2023). The videos, including the DFPs and the real humans, are available in the supplementary material.[2]

The user study workstations were conducted by three researchers with previous experience in conducting user studies. Tasks and surveys were conducted in Qualtrics.

To ensure consistency of study administration, a detailed script was prepared for the study administrators to be used in each user study session. The script included detailed instructions on what was to be said to the user study users and what was to be done by the administrator at each stage of the user study. Instructions were read to the users based on their condition groups in Table 4. Study users were invited to the study according to a premade schedule (users could choose their preferred time as they registered for the study), after which they were seated at the workstation. Then, the user was provided with a consent form, and after reading and signing the consent form, the user was read the overall study procedure based on their condition group (*Study introduction and consent* and *Guidance of the study flow* in Fig. 1).

Each user was either alerted or not alerted to pay attention to glitches in the video. Then, the first video shown to a user (*Watching video 1* in Fig. 1) was either a male deepfake (James) video, a video in which a real human male (James) performed, a female deepfake (Fiona) video, or a video in which a real human female (Fiona) performed. Assigning a condition group to a user was decided based on a spreadsheet where all eight sequences were repeated in the same order for new users. If the user's first video was deepfake James, the second video was human Fiona, and vice versa. If the user's first video was deepfake Fiona, the second video was human James, and vice versa.

After viewing the first video, the user was directed to answer the survey where they first complete the design task (*Task 1 completion* in Fig. 1) and answer questions about the video (*Survey 1 completion* in Fig. 1). After completing survey 1, the user was asked for background information (*Background variables* in Fig. 1). After the background variables, the user was given the same instructions as the first time according to the user's condition group and the user watched the second video (*Watching video 2* in Fig. 1) and completed the same design task (*Task 2 completion* in Fig. 1) and answered the same survey for the second video (*Survey 2 completion* in Fig. 1). After completing survey 2, the user was asked for background information (*Background variables* in Fig. 1). After completing the background variable for the second time, the user was thanked for participation, and he/she was asked vocally how familiar he/she had been with deepfakes before this study session on a scale of 1–5, one being not familiar at all and five being extremely familiar. The answer was marked down, and the user was given a gift card as thanks.

3.4. Measures

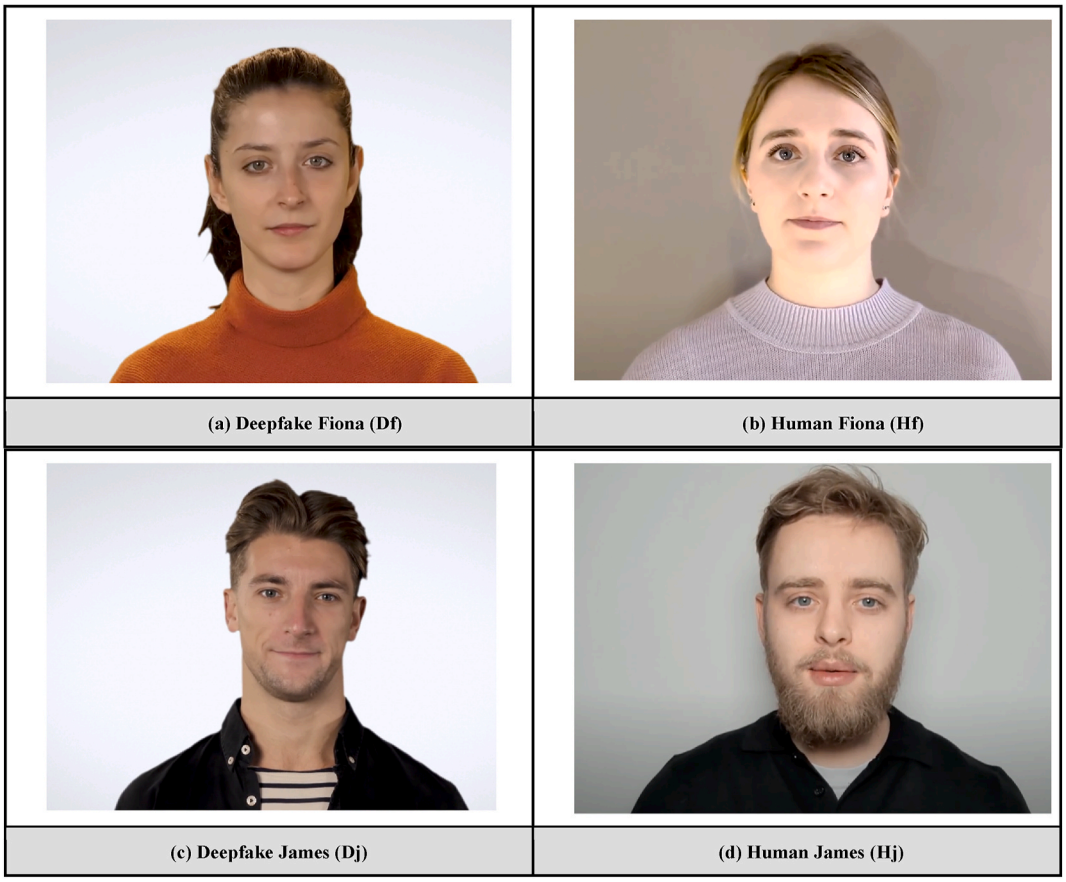
The RQs, dimensions, and variables used in the study are presented in Table 4. *First*, RQ1 addresses the user glitch observation capabilities per facial and body features. *Second*, RQ2a addresses the user perception of the persona between DFPs and human personas by measuring users' perceived trust, humanlikeness, credibility, empathy towards the persona, and willingness to use the persona (persona perceptions) as well as measuring the perceived strangeness, eye authenticity, speech authenticity, and emotional authenticity of the persona (authenticity perceptions), and perceived confidence, glitch severity, and effect of glitches (task perceptions).

Third, RQ2b addresses the variation in user behavior when exposed to DFPs and human personas. Task completion time refers to the time it took the user to complete the design task, persona evaluation time is the time it took the user to complete the survey questions on *trust, humanlikeness, completeness, credibility, empathy, willingness to use, eye authenticity, speech authenticity, emotional authenticity,* and *confidence*. Glitch evaluation time is the time the user spent on answering glitch perception questions in the survey (*How frequently did you observe glitches in the following features of the person? Would you say that the glitches in the video were severe? The glitches affected my task completion of designing the mobile app or game., In what way did the glitches affect your task completion of*

[1] https://app.synthesia.io/.

[2] https://drive.google.com/drive/folders/1Ys4VJOf74kc1By7Rm343zgCaNI9oumdc?usp=sharing.

Table 2
Still images of **(a)** deepfake Fiona, **(b)** human Fiona, **(c)** deepfake James, and **(d)** human James from the videos used in the user study. All videos are available in the supplementary material.



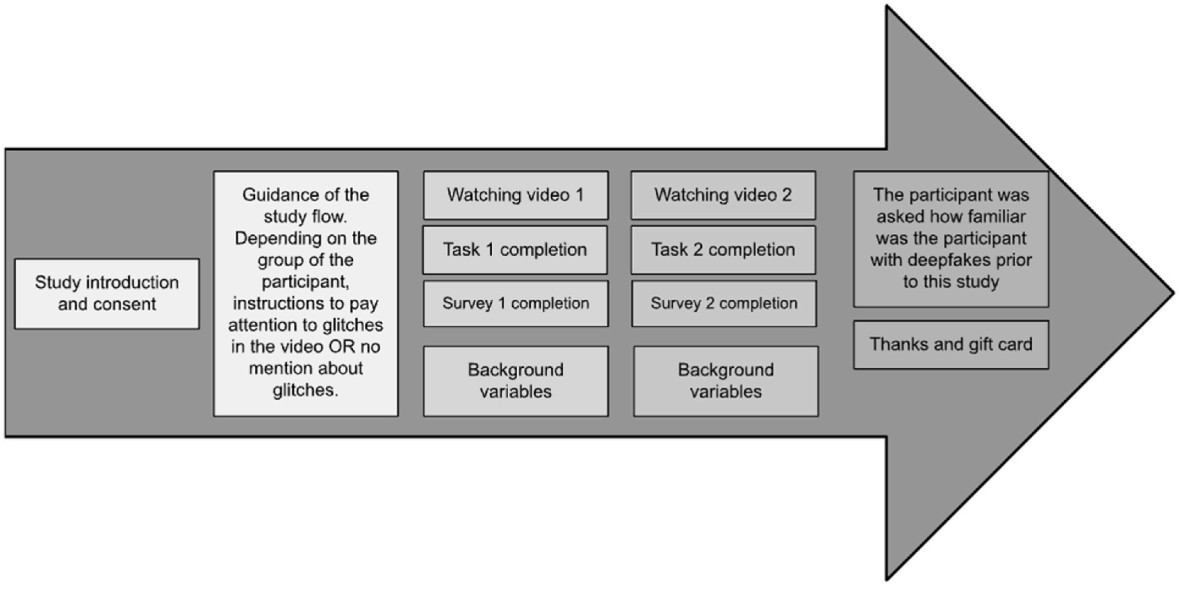


Fig. 1. Study flow procedure from the user's point of view (from left to right).

designing the mobile app or game? (If you observed any), and *Do you think the person in the video was a deepfake?).*

Fourth, RQ3 addresses the effects of the perceived glitches on (a) users' perceptions of the persona and (b) users' behavior.

For RQ1 variables, we used a three-point scale (No glitches at all, Moderate number of glitches, extremely many glitches). For RQ2a variables, we used a seven-point Likert scale (Strongly agree … Strongly disagree). RQ2b and RQ3 variable values were derived as seconds from

Table 3
The instructions were read to the users based on their condition group (alerted or non-alerted).

The users in the alerted group were read these study instructions:	The users in the non-alerted group were read these study instructions:
In this study, you will shortly be presented with two videos. In each of the videos, a person is speaking about themselves. After each video, you will perform a short, written task and answer a few questions related to the person in the video. Inboth videos you see, please pay attention to glitches. A glitch is something unnatural or abnormal about the person in the video (e.g., unnatural eyes). After the first video and task completion, you will be shown another video and complete the same task. Your task is to develop an idea for a mobile app or a game that is compatible with the person's sustainability attitudes. Sustainability attitudes refer to what the person thinks about the environment and his/her role as a consumer.	In this study, you will shortly be presented with two videos. Ineach of the videos, a person is speaking about themselves. After each video, you will perform a short, written task and answer a few questions related to the person in the video. After the first video and task completion, you will be shown another video and complete the same task. Your task is to develop an idea for a mobile app or a game that is compatible with the person's sustainability attitudes. Sustainability attitudes refer to what the person thinks about the environment and his/her role as a consumer.

METRIC and Qualtrics.

3.5. Data preprocessing and analysis

Survey responses were gathered from 46 users in 92 sessions. All 92 survey responses, two from each user, were then validated session by session. All 92 responses were included in the analysis. Two open-ended survey questions were thematically coded and analyzed. These open-ended questions were: *(a) Did you notice anything abnormal/strange in the person in the video? If yes, what was it? (RQ1)* and *(b) In what way did the glitches affect your task completion of designing the mobile app or game? (If you observed any)*. For (a) and (b) separately, themes were formed from the open answers based on three through-readings of the open answers. First, the open answers were analyzed, and each answer was marked whether the user had recognized anything abnormal or not. Then, if the user recognized something abnormal (a) or something had affected his/her task completion (b), those answers were thematically organized after two through-reading of the answers after principles of thematic analysis (Maguire & Delahunt, 2017). For (a), eight themes rose from the open answers as different facial distortions and bodily movements or auditory distortions were mentioned in the open answers. For (b), three themes rose from the open answers as an inability to focus and lack of trust and emotion in the deepfake personas were mentioned by the users in the open answers.

A repeated-measures MANCOVA was conducted due to the between-subjects control (Gender), two continuous controls (Age and Deepfake Familiarity), and the fact that each user was exposed both to a Deepfake and a Human stimulus – which was employed as a within-subjects factor (henceforth, "Type"). Analyses were run on SPSS.

4. Results

4.1. RQ1: what type of glitches do users observe in deepfake personas?

Based on the survey, 44 (95.7%) users observed glitches in the personas in 68 sessions (73.9%) of the 92 sessions 44 (47.8%) sessions of which were DFPs, and 24 (26.1%) sessions were human personas. Some users observed multiple types of glitches and glitches were detected 128 times by the users (Table 5). For all results in sections 4.1-4.3, we also tested the effect of alerting, but it had no significant effect on any of the measures. We also tested the effect of alerting the user about potential

Table 4
Research questions, dimensions, and variables. **PPS** = Persona Perception Scale (Salminen et al., 2020).

RQ	Dimension	Variable
RQ1: What type of glitches do users observe in deepfake personas?	N/A	**Feature glitches:** How frequently did you observe glitches in the following features of the person? A glitch is something unnatural or abnormal about the person in the video (e.g., unnatural eyes). – Mouth/Right eyebrow/Left eyebrow/Right eye/Left eye/ Nose/Jaw/Hair/Neck/ Shoulders/Ears
RQ2a: How do users' perceptions of the persona vary between deepfake and human personas?	Persona perceptions	*Trust:* I trust the information given by the person (Behrend, Toaddy, Thompson, & Sharek, 2012). *Humanlikeness:* The person in the video was humanlike (Macdorman, 2006). *Completeness:* The person in the video provided enough information for me to understand his/her needs (PPS). *Credibility:* I have met people like the person in the video./ The person seemed like a real person./ The video of the person looked authentic (PPS). *Empathy:* I feel like I understood the person./ I felt strong ties to this person./ I can imagine a day in the life of this person (PPS). *Willingness to use:* I found the information given by this person useful for my design task./ I could imagine multiple ways to make use of the person's information in my design task./ The information given by this person improved my ability to make decisions about similar people (PPS).
	Authenticity perceptions	*Strangeness:* The person in the video seemed strange (Macdorman, 2006). *Eye authenticity:* The person in the video was blinking his/her eyes naturally (Mustafa et al., 2022). *Speech authenticity:* The person in the video was speaking naturally (Jafar et al., 2020). *Emotional authenticity:* The person displayed emotion (Tinwell et al., 2011).
	Task perceptions	*Confidence:* I am confident that the person would like the mobile app or game I designed. *Effect of glitches:* The glitches affected my task completion of designing the mobile app or game.
RQ2b: How does users' behavior vary between deepfake and human personas?	N/A	Task completion time (*seconds*) Persona evaluation time (*seconds*) Glitch evaluation time (*seconds*)
RQ3: How does users' glitch perception affect their (a) perceptions of the persona and (b) behavior?	N/A	Task completion time (*seconds*) Persona evaluation time (*seconds*) Glitch evaluation time (*seconds*)

Table 5
Glitch themes, theme definitions, and the number of times glitches were detected. N = Number of times glitches were detected (% of the times glitches were detected).

Theme	Theme definition	N
Unnatural eyes	Persona's eyes seemed abnormal	26 (20.3%)
Unnatural voice	Persona had an abnormal voice	26 (20.3%)
Unnatural body	Persona's whole body seemed abnormal	18 (14.1%)
Unnatural in general	The overall appearance of the persona was abnormal, but the user could define specifically what was strange	16 (12.5%)
Unnatural ears	Persona's ears seemed abnormal	14 (10.9%)
Unnatural mouth	Persona's mouth seemed abnormal	11 (8.6%)
Unnatural hair	Persona's hair seemed abnormal	9 (7.0%)
Emotionless	Persona expressed little or no emotions	8 (6.3%)
		Total = 128

glitches, but this did not significantly affect the dependent variables.

Common themes (Table 5) were 'Unnatural eyes' (n = 26, 20.3%), 'Unnatural voice' (n = 26, 20.3%), 'Unnatural body' (n = 18, 14.1%), 'Unnatural in general (n = 16, 12.5%), 'Unnatural ears' (n = 14, 10.9%), 'Unnatural mouth' (n = 11, 8.6%), 'Unnatural hair' (n = 9, 7.0%), and 'Emotionless' (n = 8, 6.3%). The 'Unnatural eyes' class contained observations about the eye's flickering ("The person was not blinking her eyes normally", P24; "The eyes were taking longer than normal to blink", P34). The 'Unnatural voice' class contained observations about monotonical and robotic-like voice ("Also his tone is quite monotonic, even when he expressed something he dislikes or is excited about.", P19; "The voice was monotone, robotic, P01). The 'Unnatural body' class contained observations about body language and movement ("Body language feels a bit unnatural", P15; "There is also no head or body movement", P19). 'Unnatural in general" class contained observations about "something being not right" or the entirety of the persona being weird ("The human looks unnatural", P34; "I had a strong suspicion that this video was a synthetic rendering", P40).

The 'Unnatural ears' class contained observations about ears looking weird or moving in an unnatural manner ("The ears were not natural, moving too much like disconnected from the face", P41; "I kept noticing his ears moving while he talks", P14). 'Unnatural mouth' class contained observations about the mouth and lips ("The mouth motion was not natural", P10; "The lips were moving weirdly", P27). 'Unnatural hair' class contained observations about the hair movement or looking weird ("The hair was weird", P27; "The movement of the hair, I can tell it's an AI rendering", P43). The 'Emotionless' class contained observations about the emotionlessness in the persona ("He wasn't as expressive or showed any emotions", P45; "They were visibly emotionless while talking about topics and things that should cause some level of emotions to arise.", P20).

There were differences in the detection of glitches by the users in different features of the deepfake personas (Table 6). There were glitches detected in each feature (mouth, nose, etc.) of the deepfake personas. We also tested the effect of alerting the user about potential glitches, but this did not significantly affect the dependent variables. Features, where any glitches were detected (moderate or extreme), were hair (60.9%), ears (56.8%), and right eye (54.3%). The *most severe* glitches were detected most frequently in the ears (36.4% of respondents), both eyes (21.7%), and the mouth (21.7%). The least glitches were detected in the nose (82.6% did not detect any glitches), right eyebrow (73.9%), left eyebrow (67.4%), and shoulders (67.4%).

Table 6
The glitched features and the percentage of glitches detected in the deepfake personas by the users and the severity of the glitches.

Glitches detected in deepfakes	No glitches	Moderate number of glitches	Extremely (many glitches)	Any glitches detected (moderate or extreme)
Mouth	47.8%	30.4%	21.7%	52.2%
Right eyebrow	73.9%	17.4%	8.7%	26.1%
Left eyebrow	67.4%	23.9%	8.7%	32.6%
Right eye	45.7%	32.6%	21.7%	54.3%
Left eye	47.8%	30.4%	21.7%	52.2%
Nose	82.6%	15.2%	2.2%	17.4%
Jaw	56.5%	28.3%	15.2%	43.5%
Hair	39.1%	41.3%	19.6%	60.9%
Neck	63.0%	19.6%	17.4%	37.0%
Shoulders	67.4%	17.4%	15.2%	32.6%
Ears	43.2%	20.5%	36.4%	56.8%
Mean	57.7%	25.2%	17.1%	42.3%

4.2. RQ2: how do users' (a) perceptions of the persona and (b) behavior vary between deepfake and human personas?

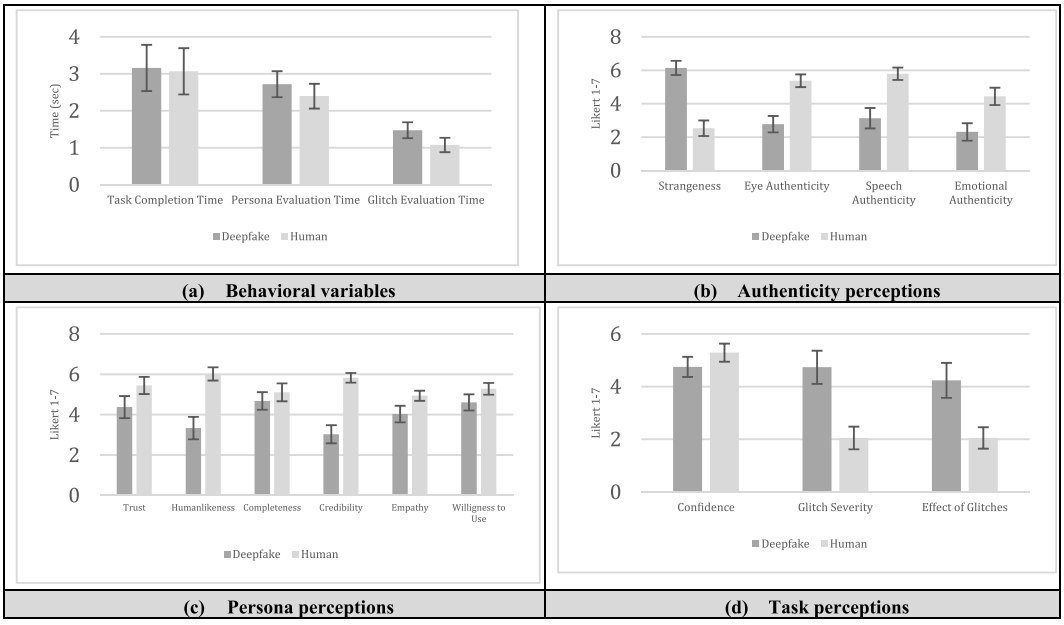
Table 7 presents the results for (a) behavioral variables, (b) authenticity perceptions, (c) persona perceptions, and (d) task perceptions which are then discussed in more detail. Time values in Table 7 are expressed as 1/100th of their actual value for scaling purposes in (a). In Table 7 (b-d) values are in a 1–7 Likert scale. We also tested the effect of alerting the user about potential glitches, but this did not significantly affect the dependent variables.

We begin by looking at the multivariate tests. The multivariate test for Gender indicates no significant effects ($F(16, 27) = 0.687$, $p = .782$), suggesting that none of the studied variables exhibit differences between users of different genders. As for the within-subjects comparison (deepfake vs. human personas), "Type" also does not exhibit a significant effect at a multivariate level ($F(16, 27) = 1.053$, $p = .439$), which indicates that there are no differences between Human and Deepfake stimulus when considering the combined set of dependent variables. However, some specific features might still differ at a univariate level, which we will explore next. Finally, the interaction terms between the controls (gender, age, familiarity with deepfakes) and Type also did not exhibit significant effects. As these variables were only included for control purposes, we will not delve much further into their analysis. Following this, we proceed with the univariate analysis, beginning with the stimulus type comparison.

First, let us note the *variables in which deepfakes did not differ from humans*: Task Completion Time ($F(1, 42) = 0.202$, $p = .655$), Completeness ($F(1, 42) = 1.150$, $p = .290$), Persona Evaluation Time ($F(1, 42) = 0.237$, $p = .629$), Willingness to Use ($F(1, 42) = 1.341$, $p = .253$), Eye Authenticity ($F(1, 42) = 1.731$, $p = .195$), Speech Authenticity ($F(1, 42) = 1.731$, $p = .195$), Confidence ($F(1, 42) = 0.026$, $p = .873$), and Glitch Evaluation Time ($F(1, 42) = 0.016$, $p = .899$). These variables exhibited functionally the same scores for both deepfakes and humans.

However, all other studied variables exhibited significant differences across the two persona types, notably: *Trust* ($F(1, 42) = 5.347$, $p < .05$), *Empathy* ($F(1, 42) = 4.769$, $p < .05$), *Humanlikeness* ($F(1, 42) = 4.736$, $p < .05$), *Strangeness* ($F(1, 42) = 6.986$, $p < .05$), *Credibility* ($F(1, 42) = 6.451$, $p < .05$), *Emotional Authenticity* ($F(1, 42) = 4.592$, $p < .05$), *Glitch Severity* ($F(1, 42) = 4.123$, $p < .05$), and *Effect of Glitches* ($F(1, 422) = 8.991$, $p < .01$). Regarding the nature of these differences, Humans scored higher in terms of *Trust, Empathy, Credibility*, and *Emotional Authenticity*. Deepfakes, on the other hand, exhibited higher scores regarding *Strangeness, Glitch Severity, and Glitch Effect*.

Table 7
Comparison of deepfake and human personas (estimated marginal means; interval bars indicate 95% confidence intervals) on (a) behavioral variables, (b) authenticity perceptions, (c) persona perceptions, and (d) task perceptions.



4.3. RQ3: how does users' glitch perception affect their (a) perceptions of the persona and (b) behavior?

4.3.1. Results for behavioral variables

The data for RQ3 behavioral variables was analyzed through linear regressions, the results of which are summarized in Table 8. We also tested the effect of alerting the user about potential glitches, but this did not significantly affect the dependent variables.

First, the relation between Glitch Perception and Task Completion Time was not significant (B = −13.562, p = .171). Second, the relation between Task Completion Time and User Age was negative and significant (B = −4.263, p < .05) indicating that higher age resulted in faster task completion times. Third, the relation between Glitch Perception and Persona Evaluation time was not significant (B = 0.054, p = .992). Fourth, the relation between Deepfake Familiarity and Persona Evaluation Time was negative and significant (B = −32.200, p < .01), indicating that users' higher familiarity with deepfakes faster persona evaluation time. Fifth, the relation between Glitch Perception and Glitch evaluation time was positive and significant (B = 7.731, p < .05), indicating that perceived glitches increased glitch evaluation time.

4.3.2. Results for authenticity variables

The data for RQ3 authenticity variables were analyzed through linear regressions, the results of which are summarized in Table 9. We also tested the effect of alerting the user about potential glitches, but this did not significantly affect the dependent variables.

First, the relation between Glitch Perception and Persona Strangeness was positive and significant (B = 0.752, p < .001), indicating that perceived glitches increased persona strangeness. Second, the relation between Glitch Perception and Eye Authenticity (B = −0.456, p < .001), Speech Authenticity (B = −0.510, p < .001), and Emotional Authenticity (B = −0.408, p < .001) was negative and significant in all cases, indicating that perceived glitches lowered the authenticity of eyes, speech, and emotions. Thus, all four authenticity variables showed significant relation to glitch perception.

4.3.3. Results for persona perceptions

The data for RQ3 persona perception variables were analyzed through linear regressions, the results of which are summarized in Table 10. We also tested the effect of alerting the user about potential glitches, but this did not significantly affect the dependent variables.

First, a negative and significant relation was found between Glitch Perception and Trust (B = −0.304, p < .001), indicative that perceived glitches lowered trust. *Second*, the relation between Glitch Perception

Table 8
The results of linear regression analysis on the relationship between users' glitch perception and its effect on behavioral variables and the effect of control variables of gender, age, and familiarity with deepfakes on the behavioral variables.

Variable	Task Completion Time	Persona Evaluation Time	Glitch Evaluation Time
User Gender	4.763 (47.687)	10.116 (26.068)	−21.301 (15.773)
User Age	−4.263* (2.072)	−1.634 (1.133)	−0.187 (0.685)
User Deepfake Familiarity	27.542 (20.645)	−32.200** (11.285)	4.094 (6.828)
User Glitch Perception	−13.562 (9.819)	0.054 (5.368)	7.731* (3.248)

Notes: ***p < .001; **p < .01; *p < .05.

Table 9
The results of linear regression analysis on the relationship between users' glitch perception and its effect on authenticity variables and the effect of control variables of gender, age, and familiarity with deepfakes on the authenticity variables.

Variable	Strangeness	Eye Authenticity	Speech Authenticity	Emotional Authenticity
User Gender (male)	0.418 (0.389)	−0.453 (0.406)	−0.392 (0.444)	0.171 (0.432)
User Age	0.005 (0.017)	0.010 (0.018)	0.009 (0.019)	0.033 (0.019)
User Deepfake Familiarity	−0.058 (0.168)	0.118 (0.176)	−0.087 (0.192)	−0.327 (0.187)
User Glitch Perception	0.752*** (0.080)	−0.456*** (0.084)	−0.510*** (0.091)	−0.408*** (0.089)

Notes: ***p < .001; **p < .01; *p < .05.

Table 10
The results of linear regression analysis on the relationship between users' glitch perception and its effect on persona perceptions and the effect of control variables of gender, age, and familiarity with deepfakes on the persona perception variables.

Variable	Trust	Humanlikeness	Completeness	Credibility	Empathy	WTU
User Gender (male)	−0.102 (0.353)	0.201 (0.389)	−0.220 (0.338)	−0.214 (0.303)	−0.336 (0.238)	−0.381 (0.252)
User Age	0.002 (0.015)	0.000 (0.017)	0.045** (0.015)	−0.001 (0.013)	0.003 (0.010)	0.027* (0.011)
User Deepfake Familiarity	0.285 (0.153)	−0.232 (0.168)	0.141 (0.146)	−0.074 (0.131)	0.141 (0.103)	0.028 (0.109)
User Glitch Perception	−0.304*** (0.073)	−0.539*** (0.080)	−0.190** (0.070)	−0.609*** (0.062)	−0.301*** (0.049)	−0.212*** (0.052)

Notes: ***p < .001; **p < .01; *p < .05.

and Humanlikeness was both negative and significant (B = −0.539, p < .001), indicating that perceived glitches lowered humanlikeness. *Third*, the relation between Glitch Perception and Persona Completeness was negative and significant (B = −0.190, p < .001), indicating that perceived glitches lowered persona completeness. *Fourth*, the relation between Completeness and User Age was positive and significant (B = 0.045, p < .01), indicating that older users perceived personas more completely. *Fifth*, a negative and significant relation was found between Glitch Perception and Persona Credibility (B = −0.609, p < .001), showing that perceived glitches lowered persona credibility. *Sixth*, a negative and significant relation between Glitch Perception and Persona Empathy was also found (B = −0.301, p < .001), signifying those perceived glitches lowered persona empathy. *Seventh*, a significant and negative relation between Glitch Perception and Willingness to Use was found (B = −0.212, p < .001), indicating that perceived glitches lowered willingness to use persona. *Eighth*, the relation between Willingness to Use and User Age was positive and significant (B = 0.027, p < .05), indicating that older users were more willing to use personas.

Of the six persona perception variables, all six (100%) showed a significant and negative relation to glitch perception. In addition, the relation between user age and completeness was positive and significant. Also, the relation between user age and willingness to use was positive and significant.

4.3.4. Results for task perception

The data for RQ3 task perception variables were analyzed through linear regressions, the results of which are summarized in Table 11. We also tested the effect of alerting the user about potential glitches, but this did not significantly affect the dependent variables.

First, the relation between Glitch Perception and Confidence was negative and significant (B = −0.138, p < .05) indicating that perceived glitches lowered confidence in task performance. Second, the relation between Glitch Perception and the effect of glitches was positive and significant (B = 0.698, p < .001), demonstrating that perceived glitches raised the effect of glitches on task performance. Therefore, of two task perception variables, both (100%) showed significant relation to glitch perception.

4.3.5. Analysis of open answers

Additional evidence on the user task perception is provided in the open answers. Based on the open answers, more than half (n = 28, 60.9%) of the users reported that the glitches affected their task completion. In this study, the effects of users' glitch detection on task completion were all negative (unwanted or distracting). There is, however, the possibility in other studies that the users' glitch detection could affect task completion in a positive manner. Glitch effects occurred in 33 (35.9%) of the 92 sessions, of which 28 (84.8%) were in deepfake persona sessions and five (15.2%) in human persona sessions. According to a chi-squared test of independence, the users' task performance was more likely to be affected by the glitches in deepfake personas than in human personas, $X^2(1, N = 92) = 25.0$, p < .00001.

Common themes (Table 12) that users recognized for the effects of glitches on task performance were 'Unable to focus' (N = 30, 32.6%), 'Distrust' (N = 2, 2.2%), and 'Lack of emotion' (N = 1, 1.1%). 'Unable to focus' class contains observations about the glitches making it hard to focus on the persona ("The glitches were severe, so I didn't pay much attention to what the man was saying.", P06; "It was hard to follow what she was saying due to the robotic nature of her voice and movement. It was very distracting.", P26). 'Distrust' class contains observations about features making it hard to take the persona seriously ("It was perhaps hard to take the person seriously due to the glitches.", P38; "I would say that the fact that the voice was very unnatural and robotic made it seem as if the person in the video was a robot, which made it somewhat hard to take her seriously.", P46). 'Lack of emotion' class contains observations about emotionless features of the persona ("If the person was more expressive, I feel like I would be able to design an app that relates more to the issue he is most passionate about. In this video the person was very one-note.", P26).

Table 11
The results of linear regression analysis on the relationship between users' glitch perception and its effect on task perception and the effect of control variables of gender, age, and familiarity with deepfakes on the task perception variables.

Variable	Confidence	Effect of Glitches
User Gender (male)	0.301 (0.278)	−0.055 (0.348)
User Age	−0.002 (0.012)	−0.018 (0.015)
User Deepfake Familiarity	−0.112 (0.120)	0.205 (0.151)
User Glitch Perception	−0.138* (0.057)	0.698*** (0.072)

Notes: ***p < .001; **p < .01; *p < .05.

5. Discussion

5.1. Theoretical implications

The progress of AI technologies creates a demand for new and improved interaction. Deepfakes pose opportunities for enhancing UX in many user-facing information systems, making them an example of an HCAI application. However, realizing these opportunities requires that deepfakes are well received by the end-users, as any resistance can mitigate the theoretical or potential practical benefits.

Our study introduces the concept of *deepfake glitch perception* and explores the emergence of this concept concerning DFPs. We identify several impactful themes about deepfake persona perception. Our results show multiple features affecting the deepfake persona perception, including the humanlikeness in, trust with, and empathy in DFPs. Some of these features have been also found to limit the adoption of deepfakes in previous research, such as unnatural eye movement (Li et al., 2018) or

Table 12
Common themes that users recognized for the effects of glitches on task performance. N = Number of times theme was mentioned in the answers (% of the times glitches affected task completion).

Themes	Theme definition	N
Unable to focus	The user felt it was hard to focus on what the persona was saying because of glitches that affected the user's task completion.	30 (23.6%)
Distrust	The user did not trust the persona which affected the user's task completion.	2 (2.2%)
Lack of emotion	The user felt the persona expressed little or no emotion which affected the user's task completion.	1 (1.1%)

a feature that has not been discussed in the deepfake literature, such as gender-related behavior patterns (Ablon, Brown, Khantzian, & Mack, 2015; Wester et al., 2002), but our research extends these findings by providing actual user explanations of how users experience DFPs.

According to our findings, the user perception of DFPs depends on the perceived realness and humanlikeness, which are dependent on the manners, ways of speech, perceived trust, emotional expressions, vocal properties, and lack of perceived connection between the deepfake and the user. Similar themes have been studied in prior research on humanlikeness (Kietzmann, Mills, & Plangger, 2021), trust (Glikson & Woolley, 2020; Jacovi, Marasović, Miller, & Goldberg, 2021), glitches and distortions (Kang, Ji, Lee, Jang, & Hou, 2022), and empathy (i.e., the connection between the user and the deepfake) (Salminen et al., 2021). Previous research has found that deepfakes are most likely recognized by the users (Groh et al., 2022) as they were recognized based on the distortions and unnaturalness in our study. The themes of unnaturalness have been found to lower the deepfake user perception; our results support and expand the results of previous research (Ahmed et al., 2023; Bray et al., 2022; Cleveland, 2022; Groh et al., 2022; Müller et al., 2021; Ng, 2022; Preu et al., 2022; Shahid et al., 2022) by offering quantitative and qualitative insights.

RQ1 dealt with the glitch types that users observe in deepfake personas. In short, the response to RQ1 is that users are generally sensitive to the abnormalities present in deepfake personas, such as unnatural ears, eyes, and voice. As the current deepfake technology still contains some glitches, our findings indicate that these glitches affect users' experience of deepfakes.

RQ2 dealt with how users' (a) perceptions of the persona and (b) behavior vary between deepfake and human personas. In short, the response to RQ2 is that users perceive human personas more realistic than deepfake personas but the users' behavior does not significantly differ between deepfake and human personas. This implies that deepfake personas are used in a similar way to human personas.

RQ3 dealt with how users' glitch perceptions affect users' (a) perception of the persona and (b) behavior. In short, the response to RQ3 is that detecting glitches decreases perceived authenticity, trust, humanlikeness, completeness, credibility, empathy, and willingness to use of personas and increases perceived strangeness of deepfakes. Detecting glitches also increased glitch detection time but had no significant effect on other behavioral measures. Here, it is important to note that as users differ in their ability to detect glitches, these second-order effects are also likely to manifest somewhat differently among the deepfake users.

All RQs RQ1-RQ3 contribute to the RGs RG1-RG5. Namely, our study gave insight to the ways users perceive deepfakes (RG1) and our study offered empirical insights regarding user perceptions of deepfake technology (RG2). Our study contributes to RG3 by offering insight to how users differentiate real content from deepfakes, and we also explored the potential differences in deepfake perceptions between genders (RG4). Finally, we delved into the deepfake glitches' impact on design process (RG5).

5.2. Practical implications

Regarding practical design implications, the UX of information systems can be improved by ensuring deepfakes are integrated ethically and transparently. Users need to know when and how deepfakes are employed and have control over their personal information (Diakopoulos & Johnson, 2021). Given that deepfake technology is utilized responsibly and safely, it can help keep it from becoming a tool for nefarious actors. This includes putting safeguards in place to prevent deepfakes from being used for fraud or other criminal acts, and building powerful detection and verification tools to detect and remove harmful deepfakes (Meskys et al., 2020). Therefore, when creating DFP-based applications, HCI designers and developers must consider user perceptions. They must ensure that the deepfakes they produce are authentic and do not deceive users. Also, designers and developers must ensure that the deepfakes improve the overall UX. By doing this, deepfake applications can positively influence UX, increasing user engagement and trust (Grodzinsky et al., 2011; Pandey et al., 2021) thus conveying design information carried by the deepfakes.

According to our findings, users are sensitive to identifying deepfakes by several features they find unnatural or non-human in the deepfakes. Recognizing deepfakes lowers, on many occasions, users' DFP perception. Thus, to use deepfakes as believable media for users, the quality of deepfakes should be kept at a level where unnatural and non-human features are low in number. Also, users may see unnatural and non-human features in human personas as well, which drives us to believe that it would be beneficial for UX designers to make it clear to the users which persona is a real human being and which persona is a deepfake. This way the disbelief among users towards deepfakes and online videos in general (Ternovski et al., 2022) could be mitigated.

For task perception, we saw in our study that the detection of deepfake glitches influences the users' task perception. Detecting glitches lowered the confidence of the user in the task performance. In practice, using deepfakes that are qualified as believable in the eyes of the users for performing tasks is necessary for successful and quality task completion. The effectiveness of using deepfakes as the media of personas and information distribution is dependent on the quality and credibility of the deepfakes.

In our research, we found that the credibility of the deepfake drops if the user catches a small hint of the deepfake being fake which influences the user's task perception. Detecting glitches in deepfakes made the deepfake seem strange. Also, the authenticity of eyes, speech, and emotions was found to deteriorate when glitches were noticed. Trust, humanlikeness, completeness, credibility, empathy, and willingness to use the persona deteriorated when glitches were noticed by the users. Also, time spent by users on the task, the persona, and glitch detection increased when glitches were noticed by the users. Findings on behavioral, authenticity perception, persona perception, and task perception variables indicate that persona perception was affected by glitch perception. Overall, our findings show that there are differences between DFP and human personas on the user perception of the persona and user behavior. These findings pave the way for UX designers and developers to pay greater attention to the quality of deepfakes.

5.3. Limitations and future work

The limitations of our research give ideas for future research. In the future, if industry and scholars are willing to develop deepfake technology in a user-friendly, more human-centered direction, more emphasis should be put on the issues put forth in our research. The humanlikeness and realism of deepfakes are the first links in the chain that build trust towards deepfakes in deepfake users, and while there are problems in those links now, it is hard to see how deepfakes are going to be the technology that deepfakes have been predicted to be if the problems found in our research are present in deepfakes. Designers utilize DFPs, and deepfake perception is dependent on deepfakes' ability to mimic real human beings. Users did quite well at discovering the abnormalities of deepfakes in our study. Not meeting the expectations of users towards deepfakes lowers the perception of deepfakes.

This research, we did not test the effect of deepfake gender, age, race (demographics), and user demographics and how different demographics might be more sensitive in detecting abnormalities in deepfakes of different demographics. This kind of study setting could give more insight into the demographical differences in deepfake detection abilities which have been studied in automated deepfake recognition (Nadimpalli & Rattani, 2022) on the deepfake demographics in automatic deepfake detection but not from the point of view of real human users and UX. Modifying the deepfake gender and race would make it possible to see whether the same demographics in deepfakes and users would improve the perception of the deepfake. The deepfake being

of the same demographic group as the user has been found important in bringing the deepfake closer to the user (Aljaroodi, Adam, Chiong, & Teubner, 2019; Wagner, 2009; Ågerfalk, 2020). In our study, the deepfake and human videos were of Caucasian people while many of the users were not Caucasian.

One possible limitation is also the guidance provided for the alerted group. The alerted group was guided to pay attention to glitches, e.g., unnatural eyes. Mentioning the unnatural eyes as an example glitch could create a potential bias in the alerted group to look for unnatural eyes specifically. An alternative study design could have the users informed that the persona could be an artificial human, and that the user should be looking for any kind of abnormalities related to the persona, without explicitly addressing any example abnormality.

The relationship between glitch detection and task performance was examined in our research by users' perception of their glitch detection and its effect on users' task performance. Task performance could also be measured by analyzing the task outcomes qualitatively and studying against users' glitch detection for a more objective view (Bray et al., 2022; Gupta et al., 2020). The relationship between deepfake glitch detection and design task performance has not been studied much which opens new possibilities for future empirical research (Gamage, Stomber, Jahanbakhsh, Skeet, & Shahi, 2022).

In our study, we used a mobile app or a game design task for all users regardless of their background. In the future, users could be chosen for the study based on their prior experience in design, or the design task could be more broadly defined such as users could design a variety of real-world applications for the personas. This would reduce the possible effect of the task itself on the task performance.

In our study, the glitches on the video were indicating, for some users, that the persona on the video was not a real human but a character created with AI. Some evidence on different information absorption patterns between human and AI sources have been found in prior research (Vodrahalli et al., 2022). Receiving information from human and AI sources is relevant because the glitches present in our study videos as such may not be affecting task completion negatively, but the reduced task completion could be related to the fact that glitches in the video reveal the person in the video to be artificial. Users could be more prone to act on information given by a human persona rather than a deepfake persona.

Being or not being receptive to new technological innovations could potentially influence the users' attitude also towards deepfake personas and users' deepfake persona perceptions. In future studies, the potential effect of attitude towards technological innovations could be included.

6. Conclusion

Deepfake technology has surfaced in the everyday use of companies and individuals during the past few years. Still, even in the modern age of rapid advancements in information technology and computing power, deepfakes have room to improve before they can be seen as a primary option for facilitating design tasks and substituting real humans in various tasks in society. Users still manage to detect quite easily which personas are deepfakes and which personas are real human beings. Nonetheless, deepfakes are already used in various solutions in society, both good and evil. What we read in the media is usually the downside of deepfakes where someone is trying to make false accusations about people with the help of deepfakes or trying to put words into the mouth of public figures. Downsides are what get the media's attention but deepfakes encase a lot of potential for good. As observed in this study, deepfakes still lack the characteristics for them to be utilized in design tasks and more widely in our societal functions such as customer service or pedagogy. Therefore, more research and technological development are needed to make deepfakes a viable media of communication. There is, still, something rotten in Denmark.

CRediT authorship contribution statement

Ilkka Kaate: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Joni Salminen:** Conceptualization, Data curation, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. **João M. Santos:** Formal analysis. **Soon-Gyo Jung:** Investigation, Software. **Hind Almerekhi:** Methodology. **Bernard J. Jansen:** Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ilkka Kaate reports financial support was provided by Foundation for Economic Education.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbah.2023.100031.

References

Ablon, S. L., Brown, D. P., Khantzian, E. J., & Mack, J. E. (Eds.). (2015). *Human feelings: Explorations in affect development and meaning (first issued in paperback)* (Routledge). Taylor & Francis Group.

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems, 29*(1), 1–8. https://doi.org/10.1080/0960085X.2020.1721947

Agostinelli, S., Battaglini, F., Catarci, T., Dal Falco, F., & Marrella, A. (2019). Generating personalized aarrative experiences in interactive storytelling through automated planning. In *CHITALY'19–Proceedings of the 13th biannual conference of the Italian conference SIGCHI* (pp. 23–25). Padova: Chapter Designing the next Interaction.

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*, Article 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., Salminen, J., Jung, S., & Jansen, B. J. (2018). Imaginary people representing real numbers: Generating personas from online social media data. *ACM Transactions on the Web, 12*(4), 27. https://doi.org/10.1145/3265986

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

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

Barricelli, B. R., & Fogli, D. (2021). Virtual assistants for personalizing iot ecosystems: Challenges and opportunities. In *CHItaly 2021: 14th biannual conference of the Italian SIGCHI chapter, 1–5*.

Behrend, T., Toaddy, S., Thompson, L. F., & Sharek, D. J. (2012). The effects of avatar appearance on interviewer ratings in virtual employment interviews. *Computers in Human Behavior, 28*(6), 2128–2133. https://doi.org/10.1016/j.chb.2012.06.017

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

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. In *Proceedings of the 24th annual conference on computer graphics and interactive techniques - SIGGRAPH, '97* pp. 353–360). https://doi.org/10.1145/258734.258880

Broad, T., Leymarie, F. F., & Grierson, M. (2020). *Amplifying the uncanny*. https://doi.org/10.48550/arXiv.2002.06890. arXiv:2002.06890). arXiv.

Canbek, N. G., & Mutlu, M. E. (2016). On the track of artificial intelligence: Learning with intelligent personal assistants. *Journal of Human Sciences, 13*(1), 592–601.

Carey, M., White, E. J., McMahon, M., & O'Sullivan, L. W. (2019). Using personas to exploit environmental attitudes and behaviour in sustainable product design. *Applied Ergonomics, 78*, 97–109. https://doi.org/10.1016/j.apergo.2019.02.005

Catania, F., Crovari, P., Beccaluva, E., De Luca, G., Colombo, E., Bombaci, N., et al. (2021). Boris: A spoken conversational agent for music production for people with motor disabilities. In *CHItaly 2021: 14th biannual conference of the Italian SIGCHI chapter, 1–5*.

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

Cleveland, K. (2022). *Creepy or cool? An exploration of non-malicious deepfakes through analysis of two case studies*. College Park: M.S., University of Maryland. https://www.proquest.com/docview/2681852015/abstract/D819BB5EC1F54D76PQ/1.

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, 53*, 17–17 http://link.springer.com/10.1007/978-3-322-99786-9_1.

Cruse, E. (2006). Using educational video in the classroom: Theory, research and practice. *Library Video Company, 12*(4), 56–80.

Danry, V., Leong, J., Pataranutaporn, P., Tandon, P., Liu, Y., Shilkrot, R., et al. (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

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society, 23*(7), 2072–2098.

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics, 26*(1), 69–91.

eSafety, C. (2022). *Deepfake trends and challenges—position statement*. eSafety Commissioner. https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes.

Ferrell, D. H. I. H., Grando, G., & Zancanaro, M. (2021). The AI Style Experience: Design and formative evaluation of a novel phygital technology for the retail environment. In *CHItaly 2021: 14th biannual conference of the Italian SIGCHI chapter* (pp. 1–4).

Galassi, A., & Vittorini, P. (2021). Automated feedback to students in data science assignments: Improved implementation and results. In *CHItaly 2021: 14th biannual conference of the Italian SIGCHI chapter, 1–8*.

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*.

Gamage, D., Stomber, J., Jahanbakhsh, F., Skeet, B., & Shahi, G. K. (2022). Designing credibility tools to combat mis/disinformation: A human-centered approach. *CHI Conference on Human Factors in Computing Systems Extended Abstracts, 1–4*. https://doi.org/10.1145/3491101.3503700

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *The Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Grodzinsky, F. S., Miller, K. W., & Wolf, M. J. (2011). Developing artificial agents worthy of trust: "Would you buy a used car from this artificial agent?". *Ethics and Information Technology, 13*(1), 17–27. https://doi.org/10.1007/s10676-010-9255-1

Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences, 119*(1), Article e2110013119. https://doi.org/10.1073/pnas.2110013119

Gupta, P., Chugh, K., Dhall, A., & Subramanian, R. (2020). The eyes know it: FakeET- an eye-tracking database to understand deepfake perception. In *Proceedings of the 2020 international conference on multimodal interaction* (pp. 519–527). https://doi.org/10.1145/3382507.3418857

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

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

Haut, K., Wohn, C., Antony, V., Goldfarb, A., Welsh, M., Sumanthiran, D., et al. (2022). Demographic feature isolation for bias research using deepfakes. In *Proceedings of the 30th ACM international conference on multimedia* (pp. 6890–6897).

Hughes, S., Fried, O., Ferguson, M., Hughes, C., Hughes, R., Yao, X., et al. (2023). *Deepfaked online content is highly effective in manipulating attitudes & intentions [working paper]*.

Hwang, Y., Ryu, J. Y., & Jeong, S.-H. (2021). Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 188–193.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. https://doi.org/10.1145/3442188.3445923, 624–635.

Jafar, M. T., Ababneh, M., Al-Zoube, M., & Elhassan, A. (2020). Forensics and analysis of deepfake videos. In *2020 11th international conference on information and communication systems (ICICS), 053–058*. https://doi.org/10.1109/ICICS49469.2020.239493

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.* , Article 103096. https://doi.org/10.1016/j.ijhcs.2023.103096

Kang, J., Ji, S.-K., Lee, S., Jang, D., & Hou, J.-U. (2022). Detection enhancement for various deepfake types based on residual noise and manipulation traces. *IEEE Access, 10*, 69031–69040. https://doi.org/10.1109/ACCESS.2022.3185121

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

Kleine, F. (2022). *Perception of deepfake technology—the influence of the recipients' affinity for technology on the perception of deepfakes*.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience, 24*(11), Article 103364.

Korshunov, P., & Marcel, S. (2020). *Deepfake detection: Humans vs. machines. arXiv Preprint arXiv:2009.03155*.

Kugler, M. B., & Pace, C. (2021). *Deepfake privacy: Attitudes and regulation. Nw, 116* p. 611). UL Rev..

Lee, Y., Huang, K.-T., Tim), Blom, R., Schriner, R., & Ciccarelli, C. A. (2021). To believe or not to believe: Framing analysis of content and audience response of top 10 deepfake videos on YouTube. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 153–158. https://doi.org/10.1089/cyber.2020.0176

Lewis, A., Vu, P., & Chowdhury, A. (2022). *Do content warnings help people spot a deepfake? Evidence from two experiments*.

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. 1–7*. WIFS).

Lyu, S. (2020). Deepfake detection: Current challenges and next steps. In *2020 IEEE international conference on multimedia & expo workshops (ICMEW)*, 1–6.

Macdorman, K. F. (2006). Subjective ratings of robot video clips for human likeness , familiarity , and eeriness: An exploration of the uncanny valley. In *Proceedings of the ICCS/CogSci 2006 long symposium "toward social mechanisms of android science', 26–29*. http://www.macdorman.com/kfm/writings/pubs/MacDorman2006SubjectiveRatings.pdf.

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *The All Ireland Journal of Teaching and Learning in Higher Education, 3*(3351).

Meskys, E., Kalpokiene, J., Jurcys, P., & Liaudanskas, A. (2020). Regulating deep fakes: Legal and ethical considerations. *Journal of Intellectual Property Law & Practice, 15*(1), 24–31.

Metric. (2023). *About METRIC*. https://metric.qcri.org/about.

Mink, J., Luo, L., Barbosa, N. M., Figueira, O., Wang, Y., & Wang, G. (2022). *{DeepPhish}: Understanding user trust towards artificially generated Profiles in online social networks. 1669–1686*. https://www.usenix.org/conference/usenixsecurity22/presentation/mink.

Mori, M., MacDorman, K., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics and Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

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

Mustafa, A., Usman, M., Gul, N., Kundi, M., Aslam, A., Mir, J., et al. (2022). A comparative analysis for extracting facial features to detect deepfake videos by various machine learning methods. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4202285

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

Nadimpalli, A. V., & Rattani, A. (2022). *GBDF: Gender balanced DeepFake dataset towards fair DeepFake detection*. https://doi.org/10.48550/ARXIV.2207.10246

Ng, Y.-L. (2022). An error management approach to perceived fakeness of deepfakes: The moderating role of perceived deepfake targeted politicians' personality characteristics. *Current Psychology*. https://doi.org/10.1007/s12144-022-03621-x

Organization, W. H. (2020). *Quit tobacco today*. Publisher Full Text.

Pandey, C. K., Mishra, V. K., & Tiwari, N. K. (2021). Deepfakes: When to use it. In *2021 10th international conference on system modeling & advancement in research trends* (pp. 80–84). SMART).

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*.

Pruitt, J., & Adlin, T. (2006). *The persona lifecycle: Keeping people in mind throughout product design* (1 edition). Morgan Kaufmann.

Pu, J., Mangaokar, N., Kelly, L., Bhattacharya, P., Sundaram, K., Javed, M., et al. (2021). Deepfake videos in the wild: Analysis and detection. *Proceedings of the Web Conference, 2021*, 981–992. https://doi.org/10.1145/3442381.3449978

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

Salminen, J., Jung, S.-G., Santos J, M., Mohamed Sayed Kamel, A., & Jansen B, J. (2021). Picturing it!: The effect of image styles on user perceptions of personas. In *Proceedings of the 2021 CHI conference on human factors in computing systems, 1–16*. https://doi.org/10.1145/3411764.3445360

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*, Article 102437. https://doi.org/10.1016/j.ijhcs.2020.102437

Schmidt, A. (2021). The end of serendipity: Will artificial intelligence remove chance and choice in everyday life?. In *CHItaly 2021: 14th biannual conference of the Italian SIGCHI chapter, 1–4*.

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*.

Shahid, F., Kamath, S., Sidotam, A., Jiang, V., Batino, A., & Vashistha, A. (2022). It matches my worldview": Examining perceptions and attitudes around fake videos. *CHI Conference on Human Factors in Computing Systems, 1–15*.

Silbey, J., & Hartzog, W. (2018). The upside of deep fakes. *Maryland Law Review, 78*, 960.

Sütterlin, S., Ask, T. F., Mägerle, S., Glöckler, S., Wolf, L., Schray, J., et al. (2021). *Individual deep fake recognition skills are affected by viewers' political orientation, agreement with content and device used*.

Synthesia. (2022). *60+ languages | different voices & accents | Synthesia*. https://www.synthesia.io/features/languages.

Ternovski, J., Kalla, J., & Aronow, P. (2022). Negative consequences of informing voters about deepfakes: Evidence from two survey experiments. *Journal of Online Trust and Safety, 1*(2). https://doi.org/10.54501/jots.v1i2.28

Thaw, N. N., July, T., Wai, A. N., Goh, D. H.-L., & Chua, A. Y. K. (2021). How are deepfake videos detected? An initial user study. In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), *HCI international 2021—posters, 1419* pp. 631–636). Springer International Publishing. https://doi.org/10.1007/978-3-030-78635-9_80.

Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior, 27*(2), 741–749. https://doi.org/10.1016/j.chb.2010.10.018

Tricomi, P. P., Nenna, F., Pajola, L., Conti, M., & Gamberi, L. (2023). You can't hide behind your headset: User profiling in augmented and virtual reality. *IEEE Access, 11*, 9859–9875.

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), Article 2056305120903408.

Vincent, C. J., & Blandford, A. (2014). The challenges of delivering validated personas for medical equipment design. *Applied Ergonomics, 45*(4), 1097–1105. https://doi.org/10.1016/j.apergo.2014.01.010

Vodrahalli, K., Daneshjou, R., Gerstenberg, T., & Zou, J. (2022). Do humans trust advice more if it comes from AI?: An analysis of human-AI interactions. In *Proceedings of the 2022 AAAI/ACM conference on AI, ethics, and society*. https://doi.org/10.1145/3514094.3534150, 763–777.

Wagner, C. (2009). Action learning with second life – a pilot study. *Journal of Information Systems Education, 20*(2), 249–258.

Wang, S. (2021). *How will users respond to the adversarial noise that prevents the generation of deepfakes?*.

Wang, L., Zhou, L., Yang, W., & Yu, R. (2022). Deepfakes: A new threat to image fabrication in scientific publications? *Patterns, 3*(5), Article 100509.

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

Welker, C., France, D., Henty, A., & Wheatley, T. (2020). Trading faces: Complete AI face doubles avoid the uncanny valley. PsyArXiv https://doi.org/10.31234/osf.io/pykjr.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review, 9*(11).

Wester, S. R., Vogel, D. L., Pressly, P. K., & Heesacker, M. (2002). Sex differences in emotion: A critical review of the literature and implications for counseling psychology. *The Counseling Psychologist, 30*(4), 630–652. https://doi.org/10.1177/00100002030004008

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

Wittenberg, C., Tappin, B. M., Berinsky, A. J., & Rand, D. G. (2021). The (minimal) persuasive advantage of political video over text. *Proceedings of the National Academy of Sciences, 118*(47), Article e2114388118.