



Microexpressions in digital humans: perceived affect, sincerity, and trustworthiness

Aliya Tastemirova¹ · Johannes Schneider¹ · Leona Chandra Kruse¹ · Simon Heinzle¹ · Jan vom Brocke¹

Received: 28 September 2021 / Accepted: 10 June 2022 / Published online: 14 July 2022
© The Author(s) 2022

Abstract

Digital agents with human-like characteristics have become ubiquitous in our society and are increasingly relevant in commercial applications. While some of them closely resemble humans in appearance (e.g., digital humans), they still lack many subtle social cues that are important for interacting with humans. Among them are the so-called microexpressions—facial expressions that are short, subtle, and involuntary. We investigate to what extent microexpressions in digital humans influence people's perceptions and decision-making in order to inform the practices of digital human's design. Our two experiments applied four types of microexpressions based on emotion type (happiness and anger) and intensity (normal and extreme). This paper is among the first to design and evaluate microexpressions with different intensity levels in digital humans. In particular, we leverage the possibilities of digitally (re)designing humans and human perception. These possibilities are feasible only in a digital environment, where it is possible to explore various microexpressions beyond real human beings' physical capabilities.

Keywords Microexpression · Digital human · Conversational agent · Human–computer interaction

JEL Classification O33 · O39

Introduction

Have you ever tried to "read someone's mind" by identifying their subtle facial movements? These small movements are called microexpressions. They are brief, involuntary, and very subtle (Ekman & Friesen, 1969). Microexpressions are harder to control than vivid or macro facial expressions,

such as smiling and frowning. They are one of the most human characters, non-verbal cues that convey unspoken truth (Porter & Brinke, 2008). Can we design microexpressions in non-human agents? How do we design them? More importantly, how do microexpressions affect the way people perceive non-human agents? These are the questions that motivate our research.

Recent technological advancement has given rise to a plethora of chatbots, digital assistants, and social robots. They are conversational agents (CAs) — software-based systems that use natural language and are developed to interact with humans (Adam et al., 2021; Feine et al., 2019). CAs function at different levels of sophistication, from a simple, pre-determined input-response dialog flow to adaptive, AI-based pattern recognition and prediction. Organizations have begun harnessing CAs in healthcare, retail, and education. The current pandemic crisis has also accelerated the deployment of CAs—from responding to COVID-related questions (Mehfooz et al., 2021) to aiding remote working (Qiu et al., 2020).

Some CAs are embodied. They have a digital face or even a full body (Loveys et al., 2020). Embodied conversational

Responsible Editor: Hans-Dieter Zimmermann

✉ Johannes Schneider
Johannes.schneider@uni.li

Aliya Tastemirova
aliya.tastemirova@gmail.com

Leona Chandra Kruse
leona.chandra@uni.li

Simon Heinzle
simon@codelane.ch

Jan vom Brocke
jan.vom.brocke@uni.li

¹ University of Liechtenstein, Fürst-Franz Josef-Strasse, 9490 Vaduz, Liechtenstein

agents (ECAs) are capable of communicating with people, as well as with each other, by using intelligible verbal and nonverbal means (Cassel et al., 2000) such as gestures and facial expressions. The most human-like ECAs are referred to as virtual humans or digital humans (Raij et al., 2007). This paper focuses on this instance of ECA. Digital humans were previously primarily used in the video game and film industry. They have recently been introduced for customer support and coaching purposes (Deloitte, 2021). For example, Aimee, a customer assistant designed by UneeQ for Southern Cross Health Society, can explain health insurance to customers from New Zealand through digital face-to-face conversation. Digital humans have a better potential to establish a rapport with people compared to non-embodied CAs (Lucas et al., 2017). But the potential has yet to be unlocked through appropriate design. Specifically, we need to use the end user's perspective in order to understand the attributes they find favorable in a digital human and those they find less favorable. Such an understanding can help digital human designers to design better user interaction and, in turn, more useful digital humans.

The overarching question is, *how to design digital humans that people perceive to be affective, sincere, and trustworthy?*

In a digital world, designers can even experiment with appearance and behavior that humans cannot display. However, care is needed to avoid falling into the uncanny valley. The uncanny valley describes a phenomenon where the affinity towards the digital human grows at first as its human-likeness increases, until a point where the digital human is almost realistic—at which point the level of affinity drops drastically (Mori, 1970). Users report that the digital human evokes eerie and spine-tingling feelings when this happens. Nevertheless, the uncanny valley can be crossed by making the character even more realistic, leading to the highest affinity level. Until recently, people had not considered digital humans to be capable of performing social aspects of a conversation (Clark et al., 2019). We argue that providing natural and rich social cues can help people establish a rapport with digital humans. Social cues can positively influence digital humans' and other CAs' social presence. The richer and more human-like they are, the more social they are perceived (Feine et al., 2019). Naturalness also depends on social cues (Feine et al., 2019). Facial expression is an instance of rich social cues, and it does increase digital human's perceived trustworthiness, as well as its perceived sociability and usefulness (Loveys et al., 2020; Philip et al., 2020).

We investigate the impact of microexpressions on people's perception of a digital human, focusing on the overall perception, perceived affect, perceived sincerity, and perceived trustworthiness, as well as behavioral variables (that is, their decision-making) accompanied by their general

evaluation. We chose two specific scenarios that we deem relevant or closely related to interactions in electronic markets. The first corresponds to a digital human that aims at helping people, e.g., in a buying decision or to handle complaints in an E-commerce setting. The second scenario is related to deciding whether to recruit a digital human and choosing between different digital human candidates. We deem this scenario of particular interest, since there is a strong focus on the digital human itself in a recruitment setting. We believe that in the future digital humans and other CAs might get more and more personalized, effectively letting people choose (or in some sense "hire") their preferred digital human—similar to a physical world, where a person might choose her preferred doctor, sales clerk etc. based on her preferences. With the advent of virtual and hybrid reality, we believe that people might soon be able to not only customize their own avatars but also choose their preferred digital human among representatives of a business when inquiring about a service or asking for customer support. We have produced microexpressions based on two types of emotions, happiness and anger; and two intensity levels, normal and extreme. The two emotions were chosen to be on opposite sides of valence (one positive and one negative). We investigate multiple possible effects of adding these 2×2 combinations of microexpressions to digital humans to address the following questions:

1. How do people perceive digital humans with respect to affect, sincerity, and trustworthiness, as well as overall perception, when microexpressions are added?
2. Is decision-making affected when microexpressions are added, and in which way?
3. Can a designer use the almost infinite design options in a digital world to create more recognizable microexpressions and have a stronger impact on decision-making while still appearing natural?

Our findings have potential design implications. First, positive microexpressions can help encourage a more favorable impression of digital humans. Second, extreme microexpressions may either evoke stronger (desired) impressions or, on the contrary, appear unnatural and consequently make a digital human appear less sincere. Third, it may be more worthwhile to invest in producing appropriate microexpressions, as they are probably equally important (if not more important than) the physical appearance of digital humans. This is interesting for application areas where the choice of specific appearance-related characteristics often receives a lot of attention (e.g., in advertisements). Finally, we evaluate different approaches to designing microexpressions in digital humans. Future designers of digital humans will find our insights useful, so as not to reinvent the wheel. However, these findings represent only the first steps towards understanding the roles

of microexpressions in digital humans. Fellow researchers are invited to continue the endeavor by incorporating more emotion types, intensity levels, and use scenarios.

This paper proceeds as follows. The introduction is followed by a review of related works and research background. Then, we outline the experiment design, including the description of designing microexpressions in digital humans and the experimental setup. Next, we synthesize and discuss the results. Finally, we provide a discussion of results as well as implications for research and practice.

Conceptual background and related works

Digital humans as an instance of embodied conversational agent (ECA)

Conversational agents (CAs) have roots in systems like ELIZA (Weizenbaum, 1966), which simulate natural conversation with a psychotherapist. Although researchers saw the potential benefits of CAs, particularly in healthcare, these agents were only able to accept limited user inputs, such as multiple-choice options (Laranjo et al., 2018). Advances in artificial intelligence have sparked new interest in CAs and embodied conversational agents (ECAs), allowing for more sophisticated conversation and presenting new opportunities (e.g., in assisting patients and the elderly population). But such agents are still rarely utilized in healthcare, where they still rely on safer rule-based approaches, lagging behind CAs applied in service-related areas, where more advanced digital assistants are used to help with booking, inquiring about travel information, and other tasks (Laranjo et al., 2018). Furthermore, fast adoption of VR and AR technology allows designers to introduce more subtle features and gives more context to a conversation to improve social interactions with digital humans and increase the sense of immersion (Sajjadi et al., 2019). The most human-like ECAs are referred to as virtual humans or digital humans (Raji et al., 2007).

The potentials of digital humans have been harnessed in the video game and film industry, as well as in customer support and coaching (Deloitte, 2021). Having a digital human can be advantageous in many situations. Digital humans have a better potential to rapport with people compared to non-embodied CAs (Lucas et al., 2017). For example, learning with an animated agent can be more interesting and help make the material seem less difficult for students when compared to having no digital instructor at all (Lin et al., 2020). In such a case, displaying social cues may be highly important: an ECA that takes on a role of a teacher can influence learning outcomes by appearing warmer and more empathetic (Loveys et al., 2020). Appearance, voice, and non-verbal communication, including facial expressions, can impact students' motivation and learning outcomes

(Horovitz & Mayer, 2021). Students show a higher motivation when the instructor expresses happiness. Fitton et al. (2020) demonstrated that teaching in VR with an ECA tutor leads to the same student performance, engagement, and motivation as teaching in a real classroom with a real teacher. Students reported that they found it strange that the ECA did not show different facial expressions, which might indicate that this is an important non-verbal social cue in this scenario. When ECA poses as a museum guide, creating a friendly personality using non-verbal communication can lead to a positive first impression and overall interaction (Sylaiou et al., 2020).

Anthropomorphism and social cues in ECA

Humans display similar social reactions when interacting with computers and humans, reacting even to basic social cues conveyed by computers. This is further described by the "Computers Are Social Actors" (CASA) paradigm, stating that humans perceive computers as social actors, even when the social cues are rudimentary (Lee & Nass, 2010). According to the CASA paradigm, humans assign a personality and human features even to an artificial voice and treat it as if it were a social actor, even if these traits were not designed. Moreover, this phenomenon remains even if users are fully aware of a computer's artificialness (Garcia & Lopez, 2019). Hence, CAs can invoke such social reactions as well and have their social presence impacted by the social cues, affecting trust, user satisfaction, and the success of a long-term relationship between CA and a human (Feine et al., 2019). Social presence grows when a CA becomes more realistic due to factors including appearance and behavior, such as non-verbal cues (Fitton et al., 2020). The influence of appearance can be so strong that individuals may judge trustworthiness of an ECA based on whether its neutral face is more similar to happiness or anger, where the agent seems to be more trustworthy in the first case and less so in the second. An agent's representation generally affects humans' attitudes and reactions towards a CA, and the impact is positive, especially if humans anthropomorphize the agent (Seymour et al., 2020).

Anthropomorphism is a phenomenon when human-like features are attributed to an object or an animal. Humans generally prefer anthropomorphic assistants over other kinds of agents (Kontogiorgos et al., 2019). In general, making an ECA more human-like visually stimulates the perceptions of anthropomorphism (Seeger et al., 2021). Additionally, there are various social cues that an anthropomorphic agent can convey to aid that perception: verbal, auditory, visual, and even invisible (such as response time) cues (Feine et al., 2019). Verbal cues may involve content-related cues (greetings, for example) or style (such as a formal or informal manner of speaking) while auditory includes pitch, tempo,

etc. Many of them can in fact be designed with a low effort, for example adding a response delay or giving a human name to the ECA (Diederich et al., 2020).

These social cues can contribute to the perception of anthropomorphism and evoke social responses. They could be used to show friendliness and emotional responses. Digital humans are increasingly built to establish emotional bonds, where friendly style and social-emotional cues positively influence the perceived trustworthiness and level of interactivity (Zierau et al., 2020). However, it should be noted that nonverbal cues only lead to positive outcomes when they are implemented along with verbal cues and high human-likeness (Seeger et al., 2021). Non-verbal cues improve user engagement and satisfaction by increasing friendliness and perceived empathy. Non-verbal communication can influence a service worker's perceived competence and friendliness, affecting trust, intention to use the service, and customer loyalty (Gabbott & Hogg, 2000). Appeal can be also influenced by changing proportions of facial features, as it is important to follow realistic placement and facial proportions when designing a human-like character (Zell et al., 2019). The eye size can influence trustworthiness: larger eyes tend to appear more trustworthy compared to narrow eyes.

Designing a robot or a digital human that shows anthropomorphic characteristics can make users believe that they are communicating with a sophisticated agent, which creates a higher expectation for the interaction (Kontogiorgos et al., 2019). Therefore, when a digital human has a highly realistic appearance, its behavior, including non-verbal communication, needs to correspond to the same level of realism (Fitton et al., 2020).

Microexpressions as an instance of social cues

Facial expressions are part of the kinesics group of visual social cues and gestures. The face can produce three kinds of signals that can all be sources of information (Adamo et al., 2019). These are static (long-term features like shape), rapid (muscle movement, such as facial expressions), and slow signals (changes over time like wrinkles). Emotions, conveyed by facial expressions, are a rapid response system that makes it possible to react quickly to events that impact an individual's welfare (Matsumoto & Willingham, 2009). The advanced structure of our facial muscles means humans are able to produce over forty facial actions to display these emotions (Matsumoto & Willingham, 2009). Emotions can also be categorized into two dimensions (Horovitz & Mayer, 2021): (1) Valence that ranges from positive to negative (or pleasant – unpleasant), and (2) Arousal, ranging from active to passive.

The basic facial expressions (including microexpressions) are universal. They are the same for every human:

the culture has no influence here (Ekman, 1999). Even congenitally blind individuals show the same spontaneous facial expressions (Matsumoto & Willingham, 2009). Scholars introduced a taxonomy system for encoding facial movements which was called the Facial Action Coding System (FACS) (Friesen & Ekman, 1978), and was subsequently updated (Ekman et al., 2002). The system is utilized across multiple fields, such as psychology and facial animation. Facial movements according to FACS are represented by Action Units (AU) and temporal segments. Each of the 46 AUs refers to either contraction or relaxation of facial muscles. This way, it is possible to describe each microexpression reliably. The FACS system helped in the development of microexpression databases such as the SAMM dataset (Davison et al., 2018), CASME-II (Yan et al., 2014), and SMIC (Li et al., 2013) that can be used for learning how to recognize or synthesize microexpressions, as well as for automated microexpression recognition and microexpression spotting with Machine Learning (ML) (Liu et al., 2019).

The existence of microexpressions was already conjectured in the nineteenth century, when Darwin stated that some facial movements might be expressed involuntarily, regardless of whether an individual is trying to suppress them (Matsumoto & Hwang, 2011). From the neuroanatomical point of view, facial expressions can be controlled, while involuntary movements originate in distinct neural pathways. When a person is in a high-stake situation, where the expressions need to be controlled, both systems are activated, leading to microexpressions leakage (Matsumoto & Hwang, 2011). Seven universal emotions can be identified in such leakage: anger, sadness, fear, disgust, contempt, happiness, and surprise (Ekman, 2009).

At first the fleeting facial expressions were described as micro momentary expressions after being discovered while scanning video recordings of psychotherapy sessions (Haggard & Isaacs, 1966). The researchers found that running videos at around 1/6th of the normal speed could help the identification of very subtle facial movements. Ekman and Friesen (1969) coined the term microexpressions after examining videos of patients attempting to hide their depression. Even when people try to conceal their microexpressions, they still involuntarily show them (Porter & Brinke, 2008).

When it comes to classifying an expression as macro or micro, duration is the primary distinguishing factor, as microexpressions are complete despite the short duration (Porter & Brinke, 2008). Nevertheless, there is no consensus on the upper and lower limits of different types of facial expressions. Although early studies defined the duration of microexpressions to be between 1/25th and 1/5th of a second (Ekman & Friesen, 1969), more recent publications state that a better boundary of such facial expression is below half a second and onset duration of 260 ms (Yan et al., 2013). Nevertheless, based on how the human brain recognizes

different facial expressions, the upper limit of $1/5^{\text{th}}$ of a second or 200 ms seems correct (Shen et al., 2016). Important terms are the apex, onset, and offset phases. An apex phase is when the expression is at its peak of intensity. Onset phase is the transition between the baseline, which is usually a neutral face, until the apex, while offset phase describes the transition between the apex and the end of the expression. In general, a microexpression involves the muscles around the eyes and the eyebrows, as well as muscles around the nose and mouth. Hence, facial regions play an essential role in recognizing microexpressions, also referred to as necessary morphological patches (NMPs) (Zhao & Xu, 2019).

The roles of facial expressions in human-ECA interaction

Numerous studies have aimed at designing personified ECAs with the help of behavioral models in order to give "human-ness" qualities to digital humans. Such features can be logical, physical expressional, and emotional (Sajjadi et al., 2019). Different research projects demonstrated contradictory claims about the influence of anthropomorphic features on social presence, which is defined as "awareness of co-presence of another sentient being" and can include deep psychological involvement and engagement with the other being (Biocca et al., 2001). It is suggested that users perceive a higher level of social presence when the personality of a CA matches their own (Sajjadi et al., 2019).

On the other hand, research also highlights how important it is for ECAs to be capable of showing a range of emotions, so that people can have a natural interaction with them. To this end, they need to follow the principles of human conversation, which include displaying emotions (Yalçın, 2020). For instance, both verbal and non-verbal communication affect a customer's perception of qualities of a service employee, such as competence and friendliness. Consequently, this perception influences trust, intention to use, and loyalty (Gabbot & Hogg, 2000). Therefore, researchers investigated the influence of facial expressions displayed by ECAs on social interactions, highlighting the importance of animating the upper face in particular (Tinwell et al., 2011). Facial expressions of emotions improve intimacy, and empathic facial action can positively affect perceived trust, intention to use, and usefulness (Loveys et al., 2020). Humans seem to interpret facial signals even if they might be designed to look neutral: the trustworthiness of neutral faces is judged based on whether the expression is closer to happiness or anger (Marzi et al., 2014). Furthermore, digital humans that are able to express various emotions are also helpful for interacting with customers from different cultural backgrounds, making them more useful for companies that have users all over the world. (Miao et al., 2021).

Other research shows that emotions of happiness and anger are better expressed involving facial actions and hand gestures. At the same time, it is necessary to employ full-body movements to convey a more apparent sadness or fear (Zell et al., 2019). Exaggerated movements have been explored before, where realistic characters were found to be more likable, appear more extraverted and intelligent when they showed exaggerated motions, while the cartoon characters were more likable and seemed more intelligent when they displayed dampened movements (Adamo et al., 2019).

However, there are challenges in developing expressive digital humans that can convey rich non-verbal cues. There is a lack of high-quality, large datasets, a lack of standardization in both the creation process and implementation. Researchers also point out a lack of understanding of why social cues need to be implemented at all, as interacting with an expressive digital human can increase users' cognitive load (Wang & Ruiz, 2021).

While there is a plethora of studies regarding facial expressions of digital humans with a focus on macroexpressions (Adamo et al., 2019; Loveys et al., 2020; Tinwell et al., 2011; Zell et al., 2019), research is more limited when it comes to exploring how ECA's microexpressions affect user perception. It has been suggested that users can recognize the presence of a microexpression, being able to categorize happiness in particular (Queiroz et al., 2014). Researchers also explored the technical challenges of creating such non-verbal communication, specifically for assisting warfighters in training to distinguish between potential threats and non-threats (Zielke et al., 2011). However, no studies investigated how microexpressions affect trust, preferences, and the feeling of eeriness in users. In fact, there is even very limited research on how microexpressions shown by individuals influence others. They may cause a stronger emotional response to speech (Stewart et al., 2009) but it has not been investigated whether they can influence perceived sincerity or trustworthiness.

Understanding how expressing emotions influence perception by individuals is highly important for the implementation of digital humans in services. For example, in a scenario of an interaction between an employee and a customer, the positive emotions can influence customer-employee rapport through emotional contagion, i.e., when emotions are transferred from one person to another consciously or subconsciously (Hennig-Thurau et al., 2006). Emotional contagion exists beyond facial expression, e.g., based on textual online comments (Bösch et al., 2018). Emotions can have an especially strong effect if they are seen as authentic and sincere. Higher rapport, in turn, increases customer satisfaction. Hence, corporations have already started developing CAs that can convey various emotions, raising the question of whether AI agents influence user emotions

similarly (Bock et al., 2020). Facial expressions of digital humans seem to influence rapport and trust (Loveys et al., 2020); however, the impact has not been explored regarding microexpressions.

There are no specific digital human development guidelines that could help create digital humans perceived as having certain characteristics (such as being trustworthy), and there are still open areas for research regarding various aspects of the design, including anthropomorphic appearance and form realism (Miao et al., 2021). A highly anthropomorphic digital human is perceived to be more credible, and so evokes a desire to interact. Thus, implementing human-like features is an important part of designing an ECA. As anthropomorphism goes beyond simply appearance, involving capabilities and actions (Murphy et al., 2017), this also concerns realistic facial expressions, including microexpressions.

Learning from the uncanny valley: microexpressions and the perception of digital humans

The uncanny valley is a key concept for the perception of digital humans, expressing a phenomenon where the affinity towards the digital human increases as its human-likeness increases, before dropping when the digital human is almost realistic and raising again when it becomes a perfect imitation (Mori, 1970). The different phases are accompanied by different levels of perceived eeriness of the digital human. Several theories attempt to explain the uncanny valley phenomenon. The dehumanization hypothesis states that the more a digital human is anthropomorphized, the more likely it is that any unnatural feature (appearance or behavior) triggers dehumanization, perceiving this character as "lacking humanness" (Wang et al., 2015). This is related to the idea of infrahumanization which describes how humans categorize others as either in-group members or out-group members; out-group members are, in turn, perceived to be less human-like than in-group members (Wang et al., 2015). In general, these hypotheses suggest that the eeriness is caused by psychological processes which are not always conscious to us.

As the advances in computer graphics allow the design of increasingly realistic-looking digital humans, understanding the uncanny valley is more relevant than ever. To avoid falling into the valley, digital humans should not show any unnatural features if they are to cross the uncanny valley instead of falling into it. For the overall perception, we adopted the measurement from Ho and MacDorma (2017), one which has been used to validate the uncanny valley phenomenon. This measurement assesses the overall perception through four dimensions: perception of humanness, eeriness, spine-tingling feelings, and attractiveness.

Evaluating users' trust in a digital human is important as, for example, it can influence whether they make

a purchase or not (Seymour et al., 2020). Other studies explored the concepts of sincerity and affect in relation to a digital human (Latoschik et al., 2017). Trust can be defined as an individual's willingness to be vulnerable to another individual's actions, regardless of whether the trustor can monitor or control the trustee (Seymour et al., 2019). Trustworthiness can be judged within 33 ms of being exposed to someone (Brambilla et al., 2018). Such a quick evaluation could be explained by the necessity of assessing whether another being is a threat or an opportunity (Brambilla et al., 2018). Trustworthiness can be evaluated based on several components: reliability, capability, and ethicality or sincerity (Woodward et al., 2020). The first two align with machine predictability, defined as behavioral trust, while the third is related to the interpersonal trust, which is more important for creating truly social agents (Woodward et al., 2020).

Truly social ECAs might need to maintain a high level of perceived sincerity and ethicality to appear trustworthy. Humans trust in-group members more than out-group members (Seymour et al., 2020), which echoes the infrahumanization hypothesis. Hence, out-group members are seen as both more eerie and spine-tingling, as well as less trustworthy. Importantly, trust is not constrained by relationships between humans only, as it can also be applied in Information Systems (IS) (Thielsch et al., 2018). When it comes to trust towards digital humans, it is suggested that users judge both the personality and non-anthropomorphic features to evaluate trustworthiness (Silva & Bonetti, 2021). The digital human should have a good task performance, avoid providing misinformation, and mitigate errors (Woodward et al., 2020). Lack of sincerity and unethical behavior have a detrimental effect on trust as well.

Empirical studies

In this section we outline our research design in general followed by specific designs of the two experiments.

Research design

Our research addresses how to design digital humans that people can truly connect with based on their perceived affect, perceived sincerity, and perceived trustworthiness, as well as their overall perception. In so doing, we conducted two experiments. The first experiment focuses on a participant's perception of digital humans when presented with different types of emotion and intensity levels of microexpression as well as verbal information. The second experiment is concerned with how different types and intensity levels of microexpressions in digital humans can influence a participant's decision-making.

We used the same digital humans and microexpressions for both experiments. Each participant takes part in both experiments.

Stimuli

We produced four microexpressions that combine two emotions (happiness and anger) and two different intensity levels (normal and extreme). Happiness is generally expressed in the lips and in the eyes (for a genuine, Duchenne smile), where lip corners are pulled up, raising the cheeks, while eyebrows and eyelids are slightly pulled down. Anger is mainly expressed in the eyes: eye muscles are tightened, eyebrows are lowered and drawn together. Four microexpressions were created at the end and later used in the survey: normal happiness, normal anger, extreme happiness, and extreme anger. Videos of microexpressions of humans were used as a reference, taken from the standard datasets: CASME-II (Yan et al., 2014), SAMM (Davison et al., 2018), and SMIC (Li et al., 2013). We used the Unreal Engine (Version 4.26) and the MetaHumans project, and showed the microexpressions in video format. The normal condition corresponds to average duration and muscle movement, according to Yan et al. (2013) and Ekman and Friesen (1969)—hence, an onset phase of 87.5 ms and a total duration of 200 ms. The extreme one has the same duration but a more profound muscle movement. Figure 1 shows the maximal displacement of facial muscles for each of the four conditions, one emotion for each character. Differences between normal and extreme microexpressions are more noticeable when the videos¹ are played.

Manipulation check

We conducted a manipulation check by interviewing five participants (graduate and postgraduate students). We asked them to watch the videos and identify the microexpressions as well as to provide feedback on the task difficulty, how noticeable the microexpressions are, the video quality, and their opinions on the difference between normal and extreme microexpressions. Based on their feedback, we made minor adjustments to the videos. In the interview sessions, the participants also evaluated all scales and questions in the two experiments.

Participants

We recruited participants for the experiment from Amazon's Mechanical Turk (MTurk). Participants were required to have at least 99% HIT Approval Rate. We collected responses from the final sample of 292 participants consisting of 180 males and 112 females. Thus, data from more than the minimum required number of participants was collected. As shown in Table 1, 230 were from the US, 18 from India, and the rest from various other countries. The biggest age group with 114 participants was 25–34 years old. The highest completed education of 141 participants was a bachelor's degree. 63 participants had achieved a master's degree, 52 had completed high school as their highest education, 5 held a doctorate degree, and the rest had received an associate degree. The two experiments were shown to all participants in a random order to counterbalance them. The data was collected between April 28th 2021 and May 2nd 2021.

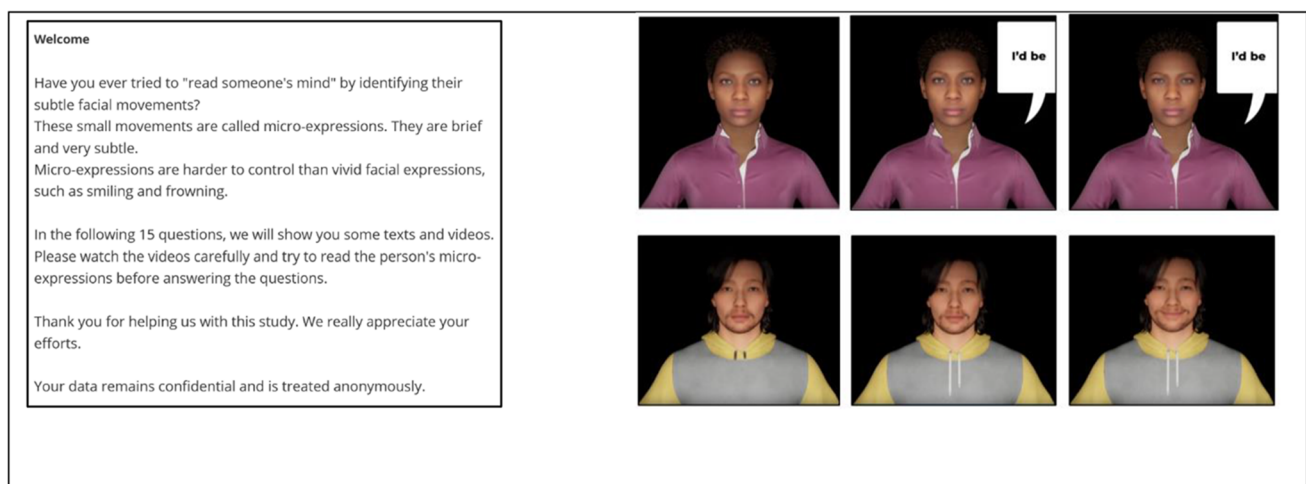


Fig. 1 Left: Welcome screen; Right: Top row: no microexpression, normal anger, and extreme anger; Bottom row: no microexpression, normal happiness, and extreme happiness (Videos are available at

https://drive.google.com/drive/folders/1pLzTKA4eO-hlnjgIT4Lp08_yL4Fbluu_?usp=sharing)

Table 1 Demographic data of the participants

	Frequency	Percent (%)
Countries		
United States of America	230	79
India	18	6
Brazil	15	5
Italy	8	3
United Kingdom	8	3
Other	13	4
Gender		
Female	112	38
Male	180	62
Age groups		
25 – 34	114	39%
35 – 44	67	23%
45 – 54	38	13%
55 – 64	33	11%
18 – 24	27	9%
65 – 74	13	4%
Education		
Bachelor's degree	141	48
Master's degree	63	22
High school	52	18
Associate degree	31	11
Doctorate	5	2

Data analysis

We conducted ANOVA for our between-subject experiment. G*Power (Faul et al., 2007) software facilitates the calculation of the required sample size to achieve statistically significant results. Here, ANOVA was chosen as the statistical test, α was set at 0.05 and power $1 - \beta$ at 0.95, with the medium effect size f of 0.25. There are two types of emotions and two levels of intensity in a 2×2 design. Thus, the number of groups is 4 and $df = 1$. Based on this information, the minimum required sample size was calculated to be 210 participants. For our data analysis, we used Python 3.8, using Statsmodels 0.11, Pandas 1.2, and Numpy 1.20 libraries.

Experiment 1: The impact of microexpressions on the perception of digital human

How do microexpressions of different emotion types and intensity impact the perception of digital humans with respect to affect, sincerity, and trustworthiness?

This experiment was designed to investigate the influence of the emotion type and the intensity of emotions on

participants' perception of a digital human. Participants' perception of a digital human was specified in four dependent variables: perceived affect, perceived sincerity, perceived trustworthiness, and overall perception based on humanness, eeriness, spine-tingle feelings, and attractiveness, adopted from Ho and MacDorman (2017). Experiment 1 has four treatment conditions (emotion type \times intensity: normal happiness, extreme happiness, normal anger, and extreme anger) and two control conditions (the presence of verbal information: microexpression video without text and with text).

Procedure and variables

This experiment was set in a scenario of interacting with a customer support representative represented by a digital human. After the welcome screen (Fig. 1), we presented all of the participants with three stimuli, one text and two videos: (1) a statement that says "I'd be happy to help" as a text; (2) a video with one microexpression showing only the digital human; (3) a video with both the microexpression as in (2) and the same statement as in (1) in a speech bubble together with the digital human (see the top row in Fig. 1). Each stimulus was followed by questions to measure the dependent variable. The microexpression was chosen in a balanced random manner from the four conditions listed above. In the third stimuli, the bubble text was added using Adobe After Effects. We let the microexpression be shown three times to give an opportunity to read the text and look at the digital human. Voice is another strong social cue that provides information through pitch, tone, accent, and intonation. Therefore, showing the text in a bubble allowed focusing on microexpressions and verbal information effects. Based on these stimuli, the experiment can be divided into three phases. Table 2 shows the variables and stimuli that are employed in this experiment. More information on the experiment platform can be found in appendices 2 (Experiment 1) and C (Experiment 2).

Measurement of dependent variables

We measured four dependent variables (DVs). For the first DV, perceived affect, the participants were asked to first evaluate how negative or positive the emotion is of the corresponding stimulus they have just seen (text only or video without text or video with text). Then we asked the participants to rate the overall emotion conveyed in the stimulus on a 10-point Likert scale ranging from very negative to very positive. The second DV, perceived sincerity, we asked the participants to rate the person they see on the video on a 10-point Likert scale, from sincere to ironic.

For the third DV, perceived trustworthiness, was evaluated by asking participants to rate three statements on a 10-point Likert scale: "I think the digital human has good

Table 2 Variables and stimuli of experiment 1

Independent Variables	IV 1: Two types of emotion: anger and happiness IV 2: Intensity of emotion: normal and extreme
Control Variable	The presence of verbal information: with and without text
Dependent Variables	DV1: Perceived affect DV2: Perceived sincerity DV3: Perceived trustworthiness DV4: Overall perception based on humanness, eeriness, spine-tingling feelings, and attractiveness
Order of stimuli	Stimulus 1: Text only Stimulus 2: Microexpression video only Stimulus 3: Microexpression video with text

Table 3 ANOVA Results of experiment 1

Dependent Variables		Effects	SS	F	<i>p</i>
Perceived affect		Emotion	106.66	21.127	0.000
		Intensity	0.17	0.033	0.856
		Verbal	28.18	5.583	0.018
Perceived sincerity		Emotion	0.14	0.024	0.878
		Intensity	24.98	4.263	0.040
Perceived trustworthiness		Emotion	59.18	10.484	0.001
		Intensity	0.83	0.147	0.701
Overall Perception	Humanness	Emotion	0.076	0.012	0.914
		Intensity	9.313	1.436	0.232
	Eeriness	Emotion	16.99	4.791	0.029
		Intensity	13.75	3.876	0.050
	Spine-tingling	Emotion	6.90	1.865	0.173
		Intensity	16.13	4.360	0.038
	Attractiveness	Emotion	0.54	0.183	0.669
		Intensity	0.54	0.183	0.669

intentions”, “I would count on the digital human”, and “I would trust the digital human” (Latoschik et al., 2017). We deem our instrument sufficiently reliable since Cronbach’s alpha for perceived trustworthiness was 0.95. The fourth DV, overall perception, was measured in four dimensions: humanness, eeriness, spine-tingling, and attractiveness, one of which we adopted from Ho and MacDorman (2017), where eerie and spine-tingling combined are referred to as eeriness. Each dimension is an aggregate of multiple adjectives, which were assessed on a 10-point bipolar Likert scale, e.g., inanimate (1) to living (10). We deem our instrument sufficiently reliable since Cronbach’s alpha was 0.93 for humanness, 0.79 for eerie, 0.88 for spine-tingling, and 0.86 for attractiveness. A previous study by Ho and MacDorman (2017) has reported Cronbach’s alpha of 0.87 for humanness, 0.82 for eerie, for 0.81 spine-tingling, and 0.85 for attractiveness. The emotion and affect questions remained the same as for the first stimuli. Table 2 shows a control variable: the presence of verbal information, i.e., whether the microexpression video is accompanied by the text “I’d be happy to help” or not. The rationale is to ensure that the differences in participants’ perceived affect are due to the

difference in microexpression emotion type and intensity levels, not because the text “I’d be happy to help” already suggests a specific affect. For this reason, we consider the control variable only when measuring perceived affect – and not the other three DVs.

Results of experiment 1

We compared mean responses for each of the four conditions, i.e., microexpression and intensity, using two-way ANOVA, and for the treatment with an additional text factor using three-way ANOVA. As shown in Table 3, we investigated the influence of each independent variable on the dependent variables described above. Intensity is more important for happiness than for anger, i.e., extreme happiness is more likely to be noticeable than anger due to larger muscle displacements (see Fig. 1). When examined in detail as shown in Table 4, it can be seen that extreme happiness had a higher affect rating than normal happiness. In contrast, extreme anger got the lowest affect of all microexpressions. It can be suggested that extreme movement possibly amplified the emotion effect, and the affect was rated accordingly,

Table 4 Perceived affect in experiment 1 (Note: We consider the presence of verbal information (video only or video with text) only for participant's perception of affect)

Emotion	Intensity	Verbal Information	Average	SD
Happiness	Extreme	Yes	6.32	2.38
Happiness	Extreme	No	6.31	2.57
Happiness	Normal	Yes	6.15	2.32
Happiness	Normal	No	5.76	2.11
Anger	Extreme	Yes	5.42	2.37
Anger	Extreme	No	4.86	1.91
Anger	Normal	Yes	5.82	2.35
Anger	Normal	No	5.03	1.89

i.e., much higher for extreme happiness and much lower for extreme anger.

Type of emotion and presence of verbal information have a significant effect on how positively a video was rated (affect), while the intensity does not; the details of the differences in responses can be seen in Table 4. This finding confirms prior studies that found happy facial expressions to result in higher interaction quality (Loveys et al., 2020). Intensity influences the perception of sincerity—extreme microexpressions were perceived as more ironic than normal ones, as shown in Table 5. The influence of intensity on perceived sincerity was expected, as it has previously been established that those extreme facial expressions can negatively impact how sincere someone appears (Stephens et al., 2019). This hints that extreme microexpressions are disguised as unnatural and do not convey the original emotion. Thus, intensity is a parameter that must be varied with great care upon designing a digital human. Despite the lack of significant effect of emotion type in this case, it was observed that although normal happiness was rated as the sincerest, the extreme version was considered to be the most ironic. There is potentially a mismatch between the happy emotion, which generally contributes to sincerity (Adamo et al., 2019), and the lack of sincerity of extreme intensity of the facial action, which can create a stronger feeling that the digital human is being ironic.

Type of emotion has a significant impact on perceived trustworthiness (Tables 3 and 6). This is not unexpected since happy facial expressions, in general, have been found to be perceived as significantly more trustworthy than angry ones (Torre et al., 2020). This is aligned with prior studies that showed how happy facial expressions improve self-disclosure (Loveys et al., 2020), which in turn is positively influenced by trustworthiness (Steel, 1991; Wheelless and Grotz, 1977). A possible explanation is that, since microexpressions are not aligned with verbal communication but show suppressed latent thoughts, negative emotions like anger are perceived to represent bad intentions. Referring

Table 5 Perceived sincerity experiment 1

Emotion	Intensity	Average	SD
Happiness	Normal	5.10	2.36
Anger	Normal	5.47	2.47
Anger	Extreme	5.74	2.47
Happiness	Extreme	6.01	2.38

Table 6 Perceived trustworthiness of experiment 1

Emotion	Intensity	Average	SD
Anger	Extreme	5.43	0.09
Anger	Normal	5.74	0.16
Happiness	Extreme	6.20	0.09
Happiness	Normal	6.00	0.17

Table 7 Participant's overall perception of digital humans in experiment 1

Dimension of Overall Perception	Emotion	Intensity	Average
Humanness	Anger	Extreme	5.31
	Anger	Normal	5.59
	Happiness	Extreme	5.26
	Happiness	Normal	5.70
Eeriness	Anger	Extreme	5.20
	Anger	Normal	4.86
	Happiness	Extreme	5.81
	Happiness	Normal	5.24
Spine-tingling	Anger	Extreme	4.78
	Anger	Normal	4.35
	Happiness	Extreme	5.13
	Happiness	Normal	4.61
Attractiveness	Anger	Extreme	6.15
	Anger	Normal	6.30
	Happiness	Extreme	6.34
	Happiness	Normal	6.40

to the in-group/out-group hypothesis described before, the negative emotion possibly made a digital human seem to be an out-group member, leading to lower trustworthiness. The intensity, however, did not seem to influence how trustworthy a digital human was perceived to be. Still, extreme happiness was considered the most trustworthy, and extreme anger the least trustworthy, as shown in Table 6.

Intensity did not have a significant effect on the perception of humanness (Tables 3 and 7). The same applies to emotion type. It is plausible that static facial signals are a stronger source of information than rapid signals in this regard. However, it can also be suggested that the extreme

Table 8 Variables and stimuli of experiment 2

Independent Variables	IV1: Two types of emotion: anger and happiness IV2: Intensity of emotion: normal and extreme
Dependent Variables	DV1: General evaluation DV2: Intention to hire and choice
Order of stimuli	A: Applicant 1: normal happiness, Ap. 2: normal anger B: Ap. 1: normal anger, Ap. 2: normal happiness C: Ap. 1: extreme happiness, Ap. 2: extreme anger D: Ap. 1: extreme anger, Ap. 2: extreme happiness

microexpressions were not extreme enough to change the perception of an otherwise human-looking character.

Both emotion and intensity significantly impacted eeriness of users, who viewed extreme and angry facial expressions as eerier (Tables 3 and 7). Participants favored happiness over anger and normal expressions over extreme, as shown in Table 7. Extreme expressions as used in our experiment are not observable in actual humans. Thus, this is expected. One may also argue that expressing anger is not a typical human emotion in the case of a customer service representative (i.e., someone typically does not express latent anger towards a stranger for no apparent reason). Also, a widely accepted notion related to the uncanny valley phenomenon states that as the digital human becomes more similar to a human, a facial expression that differs from a natural human behavior (such as an extreme microexpression) will make the character look more eerie and spine-tingling (Tinwell et al., 2011). Extreme movements were considered more spine-tingling, which contributes to the feeling of eeriness. This is also aligned with prior studies, as exaggerated facial movements influenced the eerie and spine-tingling feeling when a digital human was designed to look highly realistic (Mäkäräinen et al., 2014).

Neither of the independent variables had an impact on attractiveness. As for humanness, we conjecture that static facial signals are the primary influencing factor, i.e., more relevant than microexpressions.

Participants often confused anger with contempt. It is possible that not all of them were aware of what contempt usually looks like and interpreted a negative attitude towards them rather as contempt instead of anger. However, the verbal information often changed their responses, as expressions on such videos were classified as happiness, although the microexpression was exactly the same. Seeing the significant effect of verbal information on affect, the influence seems to be consistent here as well. In general, a happy microexpression was discerned better, aligned with previous research, where participants recognized microexpressions of happiness very well (Queiroz et al., 2014).

There were no significant differences in responses when comparing gender groups, except affect. Male participants tended to give higher affect scores than female participants ($p=0.0465$). Gender effect is not consistent across

studies, suggesting that it depends on each specific case. The vast majority of participants were residing in the USA. Hence, there was insufficient data to analyze whether geographic locations influence any measured aspects. While the result does not show a significant overall interaction effect, the interaction between emotion and intensity on affect and perceived trustworthiness is quite noticeable ($p=0.085$).

Experiment 2: The impact of microexpressions on decision making

Is decision-making impacted when microexpressions are added, and in which way? The second experiment assesses the impact of microexpressions on decision making, i.e., the intention to hire a digital human showing one of four microexpressions. This experiment also has a 2×2 factorial design to investigate the influence of the emotion type and the intensity of emotions on people's general evaluation of and intention to hire a digital human. The four conditions are as in Experiment 1. There were still the same four microexpressions in total, but here with two digital humans.

Procedure and measurement instruments

In this experiment participants were tasked to recruit a sales assistant. After the welcome screen (Fig. 1), each participant was presented with two applicants that they could choose from. The experiment was conducted in two parts: (1) introduction of the first applicant that showed a microexpression; (2) introduction of the second applicant that displayed a different microexpression. The introduction text was the same for both applicants, with the exception of the name. The two parts included a video of an applicant, followed by two questions on the general evaluation on a 10-point Likert scale ranging from very negative to very positive, and how likely they are to hire this candidate on a 10-point Likert scale (not likely to very likely)—that is, what is their intention to hire. Applicants differed only in physical appearance and gender (Fig. 1). The variables and stimuli employed in this experiment are depicted in Table 8. The participants also took part in Experiment 1, but we counterbalanced the order of experiments. There was no verbal information, but the applicant itself was considered to be an independent variable.

Table 9 ANOVA results of experiment 2

Dependent Variables	Effects	SS	F	p
Intention to hire	Emotion	21.37	4.964	0.026
	Intensity	17.23	4.002	0.046
	Applicant	1.30	0.302	0.583
General evaluation	Emotion	19.28	4.604	0.032
	Intensity	21.53	5.142	0.024
	Applicant	2.92	0.697	0.404

Table 10 Intention to hire in experiment 2

Emotion	Intensity	Diana	Michael	Average
Anger	Extreme	5.72	5.88	5.80
Anger	Normal	6.19	6.20	6.19
Happiness	Extreme	6.09	6.38	6.23
Happiness	Normal	6.56	6.49	6.53
Average		6.14	6.24	6.19

Table 11 General evaluation in experiment 2

Emotion	Intensity	Diana	Michael	Average
Anger	Extreme	5.68	5.93	5.81
Anger	Normal	6.30	6.17	6.24
Happiness	Extreme	6.03	6.40	6.21
Happiness	Normal	6.52	6.59	6.56
Average		6.13	6.27	6.20

As shown in Table 8, each participant was presented with one of the following combinations of microexpressions: (A) Applicant 1: normal happiness, Applicant 2: normal anger; (B) Applicant 1: normal anger, Applicant 2: normal happiness; (C) Applicant 1: extreme happiness, Applicant 2: extreme anger; and (D) Applicant 1: extreme anger, Applicant 2: extreme happiness.

Results of experiment 2

ANOVA tests showed a significant effect of type of emotion and emotion intensity on intention to hire an applicant and how positive their general evaluation was, as seen in Table 9. Here, happiness was preferred over anger, while normal expressions were rated better over extreme ones, as shown in Tables 10 and 11. Finally, the type of applicant did not influence either of these dependent variables, even though digital humans differed strongly in appearance, e.g., they had different clothing and hairstyle, with one digital human dressed more professionally than the other. As the type of applicant did not significantly affect this experiment, it can

be inferred that static facial signals have a lower impact on the decision-making processes than rapid signals.

There were no significant differences in how male and female participants evaluated applicants in this experiment. Furthermore, interaction effects were not significant.

Evaluating different approaches to designing microexpressions in digital humans

Can a designer use the infinite design options in a digital world to create microexpressions that are more recognizable, have a stronger impact on decision-making while still appearing natural? Three different approaches have been tested that are commonly used to synthesize facial images: manual animation by setting parameters in a special tool and two AI-based techniques derived from deep learning, i.e., generative adversarial networks, namely GANimation (Pumarola et al., 2020) and the fs-vid2vid network. For GANimation and fs-vid2vid pre-trained models were used. We first describe the three approaches, followed by a brief discussion on our qualitative evaluation and chosen approach for the experiment.

Approach 1: GANimation

First, an input image and an image with a target facial expression were cropped using the face_recognition library (Geitgey, 2018) to contain only faces. After that, the OpenFace framework (Baltrusaitis et al., 2018) was applied to extract facial Action Units (AU) from these images, which results in activation strength values for each facial AU.

This model was tested on Emma, a digital human developed by our partner company (name anonymized for review). While facial expressions of happiness were produced quite well, others—such as sadness and disgust—were not. This could be explained by the fact that an open-source pre-trained model was used that could differ in the training process from the original GANimation. Also, facial expressions of happiness generally occur more frequently in emotion datasets. Examples of created expressions are shown in Appendix 1, i.e., Figures 2 and 3.

Approach 2: fs-vid2vid

The fs-vid2vid is a motion transfer framework that was developed at NVIDIA and has demonstrated exceptional performance, especially when it comes to speech motion transfer.

The model requires facial landmarks values of the image that contains the target facial expression in order to create

a segmentation mask. This can be extracted using OpenCV and dlib libraries. The model applies the mask on the input image to create the motion transfer.

Fs-vid2vid was tested on the two digital humans developed for the Unreal Engine as part of the MetaHuman Creator plugin. Although this model performs remarkably with speech transfer, overall microexpressions were not captured by the 68 facial landmarks it is tuned to recognize. Furthermore, the identity was not preserved in the result: facial features of the digital human were changed. This is an expected outcome as the authors mention the lower performance when it comes to non-human animation. An illustration is provided in Appendix A, i.e., Figure 4.

Approach 3: Manual animation using special software

There are many 3D modeling software tools available for creating advanced animation. Unreal Engine (UE) is one of the most prominent ones, especially concerning digital humans, ever since they released the MetaHumans project. It includes highly realistic sample digital humans that are fully rigged and are ready to be animated out of the box. There is a lot of control of facial and body muscles, which allows users to design complex and subtle facial and body movements.

For our test animations, UE (Version 4.26) was used. The apex phase of microexpressions was specified using keyframes, where the muscles are activated at the given maximum value, which is low for the normal microexpressions and high for the extreme ones. Then, the neutral baseline was set at the beginning of the onset phase and at the end of the offset phase. The shape interpolation between the baseline and the apex and back to the baseline was calculated by UE. Samples are shown in Fig. 1.

Evaluation and our chosen approach

We concluded a qualitative assessment based on comparing microexpressions generated by all three approaches as illustrated in Fig. 1, Appendix Figure 3. The manual animation using a tool tailored towards the design of digital humans yielded by far the best outcomes of realism of the generated microexpressions, and in maintaining overall facial characteristics. AI solutions tended to either fail to generate the microexpressions appropriately or distort the overall appearance of the digital human. While AI-based methods required relatively little knowledge of microexpressions, they required significant technical expertise.

Discussion

Being mindful of the potential dark sides and avenues for future research

We investigated two emotions expressed in the form of microexpressions using two levels of intensity, alongside text in two contexts, i.e., interacting with a digital human assistant offering help and recruiting a digital human. Clearly, this is only a very limited set of possible configurations, and more work is needed to derive a more complete set of implications for designing microexpressions in digital humans. Our investigation of intensity revealed a potential trade-off along the intensity (design) dimension: More extreme (positive) emotions are perceived as less sincere, but they also result in a more positive rating of the digital human.

It is also interesting to study very subtle microexpressions. Subtle microexpressions might not be consciously notable, but are still perceived subconsciously (Bailey and Henry, 2009). While this might be used with good intentions, it could potentially also be used for manipulations, and we deem a proper understanding of effects critical for the understanding of potential risks when used in commercial settings—in particular, in combination with AI. This is because ECAs are often controlled by AI that is commonly purely performance-driven in an untransparent manner (Maedche et al., 2019; Meske et al., 2020), possibly at the risk of such ethical misconduct as biased decision-making and discrimination and even deceptive explanations (Schneider et al., 2022). The risk of manipulation is even stronger when accounting for the general persuasive power of ECAs, as it was shown that digital humans could make text that is presented to humans seem more credible, making humans more likely to be obedient and less likely to question suggestions and text they heard from the digital human (Sylaoui et al., 2020). Therefore, further work regarding the ethical risks of designing microexpressions in digital humans is necessary.

Our study focused on studying microexpressions, where the digital human exhibited a fixed, i.e., neutral, macroexpression. Studying other (fixed) macroexpressions is interesting; in particular, from a practical (design) perspective, where the goal might be to design digital humans showing emotions that are perceived even more vividly than with text alone. In our setting, the overall (positive) emotion was reduced (averaged) by the presence of a digital human exhibiting a neutral macroexpression, and it is likely increased by a positive macroexpression. Our work primarily contributes to the passive setting, which is largely understudied for digital humans. In other words, we focused on digital humans that either listen or communicate non-verbally, i.e., using

text. However, examining macroexpressions in a more active setting, i.e., where a digital human speaks and facial expressions constantly change, is also of interest.

Even though the type of digital human did not significantly influence the responses in the second experiment, it would be interesting to see whether the same is true in general and whether the appearance would influence perceived trustworthiness, sincerity, and other variables to a greater or lesser extent than microexpressions. Collecting more data would provide more confident insights into whether the type of a digital human does not have a significant impact in comparison with microexpressions.

Microexpressions could be used in digital humans to express compassion in a subtle way. For example, at a funeral showing or mimicking macroexpressions to indicate compassion is often inappropriate for people who are not particularly close to the deceased (Kastendick et al., 2021). Microexpressions might allow a more subtle statement that is likely more appropriate.

Eeriness and trustworthiness

One goal of this study is to assess the impact of adding microexpressions regarding eeriness related to honesty and trustworthiness. Even when a person tries to conceal her true emotion by showing another “fake” emotion or no emotion at all, her microexpression may betray her at least for a short period, involuntarily revealing her true emotion (Ekman & Friesen, 1969). This gave rise to the idea that microexpressions can be used for lie detection (Ekman, 2009). Thus, more generally, microexpressions are a vital sign to detect non-trustworthy behavior, and also impact the overall perception of real humans. In principle, any concealment could be deemed as negative, since it shows a lack of sincerity. However, the type of concealed emotion, situation and even culture might have an impact on perception, i.e., our dependent variables, as well. For example, concealing a positive emotion might be deemed as more or less sincere than a negative one. In some cultures, showing facial expressions is not appropriate in certain situations (Ekman, 2004). For example, mimicking macroexpressions of the grieving can be intended to express a feeling of compassion and rapport, but it is inappropriate for some attendees of a funeral (Kastendick et al., 2021). In such cases, displaying subtle microexpressions instead could help to reap the benefits of showing emotions (trustworthiness, social presence) while staying professional and appropriate.

Joining the discourse on trustworthy digital agents

This paper expands on the research on expressiveness of digital humans. Previous human–computer interaction studies showed that positive facial expressions are generally

perceived as more trustworthy (Torre et al., 2020), while extreme facial expressions lead to feelings of eeriness (Mäkäräinen et al., 2014) and can have a negative influence on the perceived sincerity (Stephens et al., 2019). We found that the same holds true for microexpressions, also confirming that happiness is generally better recognized, aligned with Queiroz et al. (2014). This helps understand how to create digital humans that users perceive well and, thus, can likely connect with, instead of how they view CAs now: in purely transactional terms (Clark et al., 2019).

Positioning our study in a bigger picture, we join the ongoing discourse on designing trustworthy artificial intelligence and digital agents. Prior works have focused on the five principles of beneficence, non-maleficence, autonomy, justice, and explicability (see Thiebes et al., 2021 for a comprehensive literature review). With our works, we hope to further the discourse by also considering the observable and designable characteristics of digital humans (as an instance of both artificial intelligence and digital agents), including the types of emotion and intensity of emotions.

Designing digital humans for the electronic markets

Our study provides insights into designing digital humans that can show subtle facial expressions, helping designers avoid the uncanny valley and increase trustworthiness. This is highly relevant in many application areas ranging from E-commerce to healthcare. The study compared the influence of microexpressions of anger and happiness at normal and extreme intensity levels to explore the differences in perception and potential limits in expressiveness. For instance, in a situation where emotions are not usually expressed, showing a happy microexpression can help make a digital human more trustworthy when compared to one that shows a neutral or a negative facial expression. If it is important to convey a sincere message, subtle microexpressions are preferred. However, if showing that the ECA is very positive then extreme microexpressions of happiness could be displayed instead. Microexpressions of happiness help decrease feelings of eeriness in users. The same is valid for subtle (normal) microexpressions, potentially making it possible to cross the uncanny valley. Furthermore, normal microexpressions of happiness in a digital human may positively influence the decision-making process of people. Nevertheless, for attractiveness and perceived human-likeness, the microexpressions do not play such an important role, and a designer may focus on static and slow facial signals instead. While extreme microexpressions showed an increase in eeriness, this might not necessarily hold for digital humans that appear more cartoon-like, i.e., that also express extreme forms of macroexpressions. Furthermore, negative microexpressions

that might reveal bad intent or at least a contradictory emotion to what other cue suggests, seem likely to trigger a feeling of eeriness in general. In our setup, “I’d be happy to help” suggests a positive attitude, but a microexpression depicting anger suggests that the true feeling is negative. This might inspire fear and appear unnatural since honesty is (arguably) more natural. Therefore, microexpressions show a form of dishonesty that might only be witnessed occasionally in our setup.

Our study mimicked human microexpressions, which were varied across two dimensions, i.e., valency and arousal (Horovitz & Mayer, 2021). However, one might construct additional microexpressions, e.g., motivated by macroexpression. For example, there can be two types of smiles: an enjoyment smile (also known as Duchenne smile), where eye muscles are involved in addition to the lip corner movement; and a non-enjoyment smile (non-Duchenne) that only involves muscles around the lips (Ekman, 2004). Duchenne smiles are seen as more authentic, sincere, trustworthy, and likable when compared to non-Duchenne smiles (Malek et al., 2019). This can also imply that a digital human that shows a Duchenne smile may also seem more genuine.

As a by-product of the experiments, we have systematically compared several design tools, showcasing effectiveness and the high quality of animations developed using the Unreal Engine in comparison with machine learning tools. While novel technologies such as generative adversarial networks (GANs) leveraging key AI technologies such as deep learning have shown impressive results, our findings indicate that existing mature tools that allow a manual adjustment of facial “parameters” are preferable both in terms of quality of outcomes and development time. However, as AI technologies mature further, this might change.

Ultimately, we are interested in identifying ways to contribute to the design of future digital humans for the electronic markets. Digital humans are no longer the object of science fiction prophecies. In fact, they can already be found in the entertainment industries (e.g., the virtual ABBA concert¹), healthcare (e.g., health and wellbeing coach²), and social media marketing (e.g., Miquela Sousa, a digital influencer³) of today. This paper is a first step toward harnessing the full potential of “soft” social cues and high-fidelity representation in enhancing the quality of human-digital human interaction. We believe that the better people can interact with digital humans, the more benefits they can harvest from the interaction.

Conclusions

We conducted two experiments to investigate to what extent microexpressions influence people’s perception of digital humans (perceived affect, sincerity, and trustworthiness, as well as overall perception) and their decision-making (general evaluation and intention to hire). We showed participants two emotion types (happiness and anger) and two intensity levels (normal and extreme microexpressions). We found that emotion type influences perceived affect and trustworthiness, while intensity influences perceived sincerity. We also found that extreme microexpressions and negative microexpressions increase perception of eeriness. The same applies to negative microexpressions. At first, it might appear to contradict the uncanny valley theory and other theories on social presence and anthropomorphism. However, negative microexpressions are natural part of human’s repertoire of facial expressions, so they can be expected to increase the naturalness of digital humans. When it comes to the actual decision-making (intention to hire), microexpressions are probably more telling than the general appearance (e.g., how a digital human is dressed). Our participants did not show any significant tendency towards hiring a male dressed in casual attire or a woman dressed in elegant business attire.

The possibilities of designing digital humans seem infinite. Indeed, even the design-space for microexpressions would appear to be infinite, i.e., humans have 42 facial muscles that can be controlled differently over a specific timespan to yield facial expressions, and, even more, a digital human is not tied to constraints of a real human being. We narrow down the options by investigating microexpressions along two dimensions, i.e., valency and arousal (Horovitz & Mayer, 2021). We designed microexpressions that mimic human microexpressions as closely as possible and microexpressions that show an extreme “intensity” only possible in a digital world where no physical constraints on muscle movements exist. We chose a positive (happiness) and a negative (anger) emotion to cover the valency.

This paper highlights the relevance of microexpressions in practical scenarios and provides first insights to inform the design of ECAs, particularly that of social digital humans capable of showing emotions. It shows that intensity and/or type of emotion impact people’s perception of important characteristics such as affect, sincerity, and trustworthiness, as well as the overall perception. They are also relevant in decision-making, i.e., affecting intention to hire. We discuss our findings with regard to contemporary academic and societal discourse: trustworthy digital agents, the dark sides of high-fidelity digital agents, and designing digital humans for the electronic markets. Our two experiments are only the first steps toward harnessing the full potentials of “soft”

¹ <https://www.forbes.com/sites/bernardmarr/2021/09/06/abbas-virtual-concert-the-metaverse-and-the-future-of-entertainment>

² <https://www.replika.com>

³ <https://www.instagram.com/lilmiquela>

social cues and high-fidelity representation in enhancing the quality of human-digital human interaction. We propose multiple directions for future research, such as investigating more types of emotions along various levels of intensity in different scenarios.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12525-022-00563-x>.

Funding Open access funding provided by University of Liechtenstein.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31, 427–445. <https://doi.org/10.1007/s12525-020-00414-7>
- Adamo, N., Dib, H. N., & Villani, N. J. (2019). *Animated agents' facial emotions: Does the agent design make a difference?* In *international conference on augmented reality, virtual reality and computer graphics* (pp. 10–25). Springer. https://doi.org/10.1007/978-3-030-25965-5_2
- Bailey, P. E., & Henry, J. D. (2009). Subconscious facial expression mimicry is preserved in older adulthood. *Psychology and Aging*, 24(4), 995. <https://doi.org/10.1037/a0015789>
- Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 59–66). IEEE. <https://doi.org/10.1109/FG.2018.00019>
- Biocca, F., Harms, C., & Gregg, J. (2001). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence* (pp. 1–9).
- Bock, D. E., Wolter, J. S., & Ferrell, O. C. (2020). Artificial intelligence: Disrupting what we know about services. *Journal of Services Marketing*. <https://doi.org/10.1108/JSM-01-2019-0047>
- Bösch, K., Müller, O., & Schneider, J. (2018). Emotional contagion through online newspapers. In *European Conference on Information Systems*. https://aisel.aisnet.org/ecis2018_rp/171
- Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, 78, 34–42. <https://doi.org/10.1016/j.jesp.2018.04.011>
- Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. F. (2000). *Embodied conversational agents*. MIT Press. <https://doi.org/10.7551/mitpress/2697.001.0001>
- Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–12). <https://doi.org/10.1145/3290605.3300705>
- Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2018). SAMM: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1), 116–129. <https://doi.org/10.1109/TAFFC.2016.2573832>
- Deloitte. Digital human: elevating the digital human experience. <https://www2.deloitte.com/nl/nl/pages/customer-and-marketing/articles/digital-human.html>. Accessed Apr 2021
- Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing anthropomorphic enterprise conversational agents. *Business & Information Systems Engineering*, 62(3), 193–209. <https://doi.org/10.1007/s12599-020-00639-y>
- Ekman, P. (2009). Lie catching and microexpressions. *The Philosophy of Deception*, 1(2), 5. <https://doi.org/10.1093/acprof:oso/9780195327939.003.0008>
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32(1), 88–106. <https://doi.org/10.1080/00332747.1969.11023575>
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). FACS investigator's guide. *A Human Face*, 96.
- Ekman, P. (1999). Facial expressions. In T. Dalgleish & T. Power (Eds.), *The handbook of cognition and emotion* (pp. 301–320). <https://doi.org/10.1002/0470013494.ch16>
- Ekman, P. (2004). Emotions revealed. *BMJ*, 328(Suppl S5), 405184. <https://doi.org/10.1136/sbmj.0405184>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, 132, 138–161. <https://doi.org/10.1016/j.ijhcs.2019.07.009>
- Fitton, I. S., Finnegan, D. J., & Proulx, M. J. (2020). Immersive virtual environments and embodied agents for e-learning applications. *PeerJ Computer Science*, 6, e315. <https://doi.org/10.7717/peerj-cs.315>
- Friesen, E., & Ekman, P. (1978). Facial action coding system: A technique for the measurement of facial movement. *Palo Alto*, 3(2), 5.
- Gabbott, M., & Hogg, G. (2000). An empirical investigation of the impact of non-verbal communication on service evaluation. *European Journal of Marketing*, 34(3/4), 384–398. <https://doi.org/10.1108/03090560010311911>
- Garcia, M. P., & Lopez, S. S. (2019). Exploring the uncanny valley theory in the constructs of a virtual assistant personality. In *Proceedings of SAI intelligent systems conference* (pp. 1017–1033). Springer. https://doi.org/10.1007/978-3-030-29516-5_76
- Geitgey, A. (2018). Face_recognition [computer software]. *GitHub*. https://github.com/ageitgey/face_recognition
- Haggard, E. A., & Isaacs, K. S. (1966). Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In *In Methods of research in psychotherapy* (pp. 154–165). Springer. https://doi.org/10.1007/978-1-4684-6045-2_14
- Hennig-Thurau, T., Groth, M., Paul, M., & Gremler, D. D. (2006). Are all smiles created equal? How emotional contagion and emotional labor affect service relationships. *Journal of Marketing*, 70(3), 58–73. <https://doi.org/10.1509/jmkg.70.3.058>
- Ho, C. C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. *International Journal of Social Robotics*, 9(1), 129–139. <https://doi.org/10.1007/s12369-016-0380-9>
- Horovitz, T., & Mayer, R. E. (2021). Learning with human and virtual instructors who display happy or bored emotions in video lectures. *Computers in Human Behavior*, 119, 106724. <https://doi.org/10.1016/j.chb.2021.106724>

- Kastendieck, T., Mauersberger, H., Blaison, C., Ghalib, J., & Hess, U. (2021). Laughing at funerals and frowning at weddings: Top-down influences of context-driven social judgments on emotional mimicry. *Acta Psychologica*, 212, 103195. <https://doi.org/10.1016/j.actpsy.2020.103195>
- Kontogiorgos, D., Pereira, A., Andersson, O., Koivisto, M., Gonzalez Rabal, E., Vartiainen, V., & Gustafson, J. (2019). The effects of anthropomorphism and non-verbal social behaviour in virtual assistants. In C. Pelachaud (Ed.), *Proceedings of the 19th ACM international conference on intelligent virtual agents*. Association for Computing Machinery. ACM Digital Library. <https://doi.org/10.1145/3308532.3329466>
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>
- Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. In M. Fjeld (Ed.), *Proceedings of the 23rd ACM symposium on virtual reality software and technology*. ACM. <https://doi.org/10.1145/3139131.3139156>
- Lee, J.-E. R., & Nass, C. I. (2010). Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives* (pp. 1–15). IGI Global. <https://doi.org/10.4018/978-1-61520-901-9.ch001>
- Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikainen, M. (2013). A spontaneous microexpression database: Inducement, collection and baseline. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)* (pp. 1–6). IEEE. <https://doi.org/10.1109/FG.2013.6553717>
- Lin, L., Ginns, P., Wang, T., & Zhang, P. (2020). Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain? *Computers & Education*, 143, 103658. <https://doi.org/10.1016/j.compedu.2019.103658>
- Liu, Y., Du, H., Zheng, L., & Gedeon, T. (2019). A neural microexpression recognizer. In *2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019)* (pp. 1–4). IEEE. <https://doi.org/10.1109/FG.2019.8756583>
- Loveys, K., Sebaratnam, G., Sagar, M., & Broadbent, E. (2020). The effect of design features on relationship quality with embodied conversational agents: A systematic review. *International Journal of Social Robotics*, 12(6), 1293–1312. <https://doi.org/10.1007/s12369-020-00680-7>
- Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI*, 4, Article 51. <https://doi.org/10.3389/frobt.2017.00051>
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. *Business & Information Systems Engineering*, 61(4), 535–544. <https://doi.org/10.1007/s12599-019-00600-8>
- Malek, N., Messinger, D., Gao, A. Y. L., Krumhuber, E., Mattson, W., Joobar, R., Tabbane, K., & Martinez-Trujillo, J. C. (2019). Generalizing Duchenne to sad expressions with binocular rivalry and perception ratings. *Emotion*, 19(2), 234. <https://doi.org/10.1037/emo0000410>
- Mäkäräinen, M., Kätsyri, J., & Takala, T. (2014). Exaggerating facial expressions: A way to intensify emotion or a way to the uncanny valley? *Cognitive Computation*, 6(4), 708–721. <https://doi.org/10.1007/s12559-014-9273-0>
- Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: Evidence from ERPs. *Social Cognitive and Affective Neuroscience*, 9(1), 63–72. <https://doi.org/10.1093/scan/nss102>
- Matsumoto, D., & Hwang, H. S. (2011). Evidence for training the ability to read microexpressions of emotion. *Motivation and Emotion*, 35(2), 181–191. <https://doi.org/10.1007/s11031-011-9212-2>
- Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, 96(1), 1–10. <https://doi.org/10.1037/a0014037>
- Mehfooz, F., Jha, S., Singh, S., Saini, S., & Sharma, N. (2021). Medical chatbot for novel covid-19. In S. Fong, N. Dey, & A. Joshi (Eds.), *Lecture notes in networks and systems: v. 154. ICT analysis and applications: Proceedings of ICT4SD 2020* (Vol. 154, pp. 423–430). Springer. https://doi.org/10.1007/978-981-15-8354-4_42
- Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2020). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management*, 1–11. <https://doi.org/10.1080/10580530.2020.1849465>
- Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2021). An emerging theory of avatar marketing. *Journal of Marketing*, 0022242921996646. <https://doi.org/10.1177/0022242921996646>
- Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7, 33–35. <https://ci.nii.ac.jp/naid/10027463083/>
- Murphy, J., Gretzel, U., & Hofacker, C. (2017, May). Service robots in hospitality and tourism: Investigating anthropomorphism. In *15th APACCHRIE conference* (Vol. 31).
- Philip, P., Dupuy, L., Auriacombe, M., Serre, F., de Sevin, E., Sauteraud, A., & Micoulaud-Franchi, J. A. (2020). Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. *NPJ Digital Medicine*, 3, 2. <https://doi.org/10.1038/s41746-019-0213-y>
- Porter, S., & Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 19(5), 508–514. <https://doi.org/10.1111/j.1467-9280.2008.02116.x>
- Pumarola, A., Agudo, A., Martinez, A. M., Sanfeliu, A., & Moreno-Noguer, F. (2020). Ganimation: One-shot anatomically consistent facial animation. *International Journal of Computer Vision*, 128(3), 698–713. <https://doi.org/10.1007/s11263-019-01210-3>
- Qiu, S., Gadiraju, U., & Bozzon, A. (2020). Remote work aided by conversational agents. In *New future of work symposium 2020* (p. 1). Microsoft Research <https://www.microsoft.com/en-us/research/publication/remote-work-aided-by-conversational-agents/>
- Queiroz, R. B., Musse, S. R., & Badler, N. I. (2014). Investigating macroexpressions and microexpressions in computer graphics animated faces. *Presence: Teleoperators and Virtual Environments*, 23(2), 191–208. https://doi.org/10.1162/PRES_a_00180
- Raij, A. B., Johnsen, K., Dickerson, R. F., Lok, B. C., Cohen, M. S., Duerson, M., Pauly, R. R., Stevens, A. O., Wagner, P., & Lind, D. S. (2007). Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE Transactions on Visualization and Computer Graphics*, 13(3), 443–457. <https://doi.org/10.1109/TVCG.2007.1030>
- Sajjadi, P., Hoffmann, L., Cimiano, P., & Kopp, S. (2019). A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users. *Entertainment Computing*, 32, 100313. <https://doi.org/10.1016/j.entcom.2019.100313>
- Schneider, J., Meske, C., & Vlachos, M. (2022). Deceptive AI Explanations: Creation and detection. International conference on agents and artificial intelligence (ICAART).
- Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2021). Texting with human-like conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems: JAIS*. <https://doi.org/10.17705/1jais.00685>

- Seymour, M., Yuan, L., Dennis, A., & Riemer, K. (2019). Crossing the uncanny valley? Understanding affinity, trustworthiness, and preference for more realistic virtual humans in immersive environments. In *Proceedings of the 52nd Hawaii international conference on system sciences*. <https://doi.org/10.24251/HICSS.2019.213>
- Seymour, M., Yuan, L., Dennis, A., & Riemer, K. (2020). Facing the artificial: Understanding affinity, trustworthiness, and preference for more realistic digital humans. In *Hawaii international conference on system sciences 2020*. ScholarSpace. <https://doi.org/10.24251/HICSS.2020.574>
- Shen, X., Wu, Q., Zhao, K., & Fu, X. (2016). Electrophysiological evidence reveals differences between the recognition of microexpressions and macroexpressions. *Frontiers in Psychology*, 7, 1346. <https://doi.org/10.3389/fpsyg.2016.01346>
- Silva, E. S., & Bonetti, F. (2021). Digital humans in fashion: Will consumers interact? *Journal of Retailing and Consumer Services*, 60, 102430. <https://doi.org/10.1016/j.jretconser.2020.102430>
- Steel, J. L. (1991). Interpersonal correlates of trust and self-disclosure. *Psychological Reports*, 68, (3_suppl), 1319–1320. <https://doi.org/10.2466/pr0.1991.68.3c.1319>
- Stephens, K. K., Waller, M. J., & Sohrab, S. G. (2019). Over-emoting and perceptions of sincerity: Effects of nuanced displays of emotions and chosen words on credibility perceptions during a crisis. *Public Relations Review*, 45(5), 101841. <https://doi.org/10.1016/j.pubrev.2019.101841>
- Stewart, P. A., Waller, B. M., & Schubert, J. N. (2009). Presidential speechmaking style: Emotional response to microexpressions of facial affect. *Motivation and Emotion*, 33(2), 125–135. <https://doi.org/10.1007/s11031-009-9129-1>
- Sylaiou, S., Kasapakis, V., Gavalas, D., & Dzardanova, E. (2020). Avatars as storytellers: Affective narratives in virtual museums. *Personal and Ubiquitous Computing*, 24(6), 829–841. <https://doi.org/10.1007/s00779-019-01358-2>
- Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447–464. <https://doi.org/10.1007/s12525-020-00441-4>
- Thielsch, M. T., Meeßen, S. M., & Hertel, G. (2018). Trust and distrust in information systems at the workplace. *PeerJ*, 6, e5483. <https://doi.org/10.7717/peerj.5483>
- 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>
- Torre, I., Goslin, J., & White, L. (2020). If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior*, 105, 106215. <https://doi.org/10.1016/j.chb.2019.106215>
- Uneeq. Southern Cross. Because health insurance needs a human touch, even when it's done digitally. <https://digitalhumans.com/casestudies/southern-cross>
- Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4), 393–407. <https://doi.org/10.1037/gpr0000056>
- Wang, I., & Ruiz, J. (2021). Examining the use of nonverbal communication in virtual agents. *International Journal of Human–Computer Interaction*, 1–26. <https://doi.org/10.1080/10447318.2021.1898851>
- Weizenbaum, J. (1966). Eliza—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Wheless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research*, 3(3), 250–257. <https://doi.org/10.1111/j.1468-2958.1977.tb00523.x>
- Woodward, N., Nguyen, T., Zhu, L., Fowler, C., Kim, T., Near, S., Thoemmes, S., & Williams, T. (2020). Exploring interaction design considerations for trustworthy language-capable robotic wheelchairs in virtual reality. In *International workshop on virtual, augmented, and mixed reality for human-robot interaction*. <https://doi.org/10.31219/osf.io/72qgt>
- Yalçın, Ö. N. (2020). Empathy framework for embodied conversational agents. *Cognitive Systems Research*, 59, 123–132. <https://doi.org/10.1016/j.cogsys.2019.09.016>
- Yan, W. J., Li, X., Wang, S. J., Zhao, G., Liu, Y. J., Chen, Y. H., & Fu, X. (2014). CASME II: An improved spontaneous microexpression database and the baseline evaluation. *PLoS One*, 9(1), e86041. <https://doi.org/10.1371/journal.pone.0086041>
- Yan, W. J., Wu, Q., Liang, J., Chen, Y. H., & Fu, X. (2013). How fast are the leaked facial expressions: The duration of microexpressions. *Journal of Nonverbal Behavior*, 37(4), 217–230. <https://doi.org/10.1007/s10919-013-0159-8>
- Zell, E., Zibrek, K., & McDonnell, R. (2019). Perception of virtual characters. In *ACM digital library, ACM SIGGRAPH 2019 courses*. Association for Computing Machinery. <https://doi.org/10.1145/3305366.3328101>
- Zhao, Y., & Xu, J. (2019). An improved microexpression recognition method based on necessary morphological patches. *Symmetry*, 11(4), 497. <https://doi.org/10.3390/sym11040497>
- Zielke, M. A., Dufour, F., & Hardee, G. M. (2011). Creating microexpressions and nuanced nonverbal communication in synthetic cultural characters and environments. In *Proceedings of the interservice/industry training, simulation & education conference (IITSEC)*.
- Zierau, N., Engel, C., Söllner, M., & Leimeister, J. M. (2020). *Trust in smart personal assistants: A systematic literature review and development of a research agenda*. WI2020 Zentrale Tracks. https://doi.org/10.30844/wi_2020_a7-zierau

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.