2020

Trust in Digital Humans

Annette M. Mills
University of Canterbury, annette.mills@cantebury.ac.nz

Lori F. Liu
University of Canterbury, lorifliu@gmai.com

Trust in Digital Humans

Research-in-progress

Annette M. Mills
UC School of Business
University of Canterbury
Christchurch, New Zealand
Email: annette.mills@canterbury.ac.nz

Lori Liu
University of Canterbury
Christchurch, New Zealand
Email: lorifliu@gmail.com

Abstract

With technology advances, the interaction between organisations and consumers is evolving gradually from 'human-to-human' to 'human-to-machine', due, in part, to improvements in Artificial Intelligence (AI). One such technology, the AI-enabled digital human is unique in its combining of technology and humanness and is being adopted by firms to support customer services and other business processes. However, a number of questions arise with this new way of interacting, among which is whether people will trust a digital human in the same way that they trust people. To address this question, this study draws on technology trust theory, and examines the roles of social presence, anthropomorphism, and privacy to understand trust and people's readiness to engage with digital humans. The results aim to benefit organisations wanting to implement AI-enabled digital-humans in the workplace.

Keywords digital humans, trust in technology, anthropomorphism, privacy, social presence

1 Introduction

The continual improvement of Information Technology (IT), especially in Artificial Intelligence (AI), has led to significant advancements in the interface and interaction between humans and digital machines (Adam et al., 2020; Meuter et al., 2000). The digital human in particular is emerging as one of the most sophisticated of AI technologies, and is poised to significantly change and indeed, disrupt how humans and machines interact. Digital human technology aims to create completely autonomous, animated individuals that are able to interact in ways that are personalised, dynamic and similar to real people (https://www.soulmachines.com/). Such technologies afford multiple opportunities for organisations to radically transform communications with customers and employees by replacing the more disembodied forms of live chat interfaces and conversational agents (chatbots) with highly embodied forms such as the digital human.

While there has been significant research on human interactions with intelligent interfaces such as AI-enabled chatbots (Kessler & Martin, 2017), the emergence of an embodied agent in the form of the digital human is a significant game-changer with its potential for more seamless forms of verbal and nonverbal behaviours that people display, and enhanced social and emotional presence not currently afforded in unembodied forms of conversational agents such as chatbots and other less human-like representations, including embodied avatars. Beyond motivators (such as ease, speed and convenience of getting help and information (Brandtzaeg & Følstad, 2017; Wang & Benbasat, 2005)), a number of considerations emerge in these contexts, among which are impacts on trust and privacy concerns.

Prior research on human-technology interactions and interpersonal exchanges suggest that trust, privacy concern and social presence have significant roles in determining how people respond to digital agents (Araujo, 2018; Ng et al., 2020; Van Pinxteren et al., 2020). In the case of trust, research typically distinguishes between trust in technology and trust in humans (or institutions). For trust in technology, research has suggested that the extent to which people attribute a level of humanness to a technology may impact how trusting beliefs influence outcomes such as use intentions (Lankton et al., 2015; McKnight et al., 2002; Wang & Benbasat, 2005). But, as technology and human-likeness continue to merge (as with digital humans) to become increasingly indistinguishable from real people (in the digital world), the humanness may have an increasing impact on outcomes.

To assess these impacts, in this paper we focus on key elements of the technology-human exchange as these relate to forming trust in digital humans (i.e. privacy, anthropomorphism, and social presence) and explore their role in framing peoples' readiness to engage with digital humans.

2 Literature Review

Artificial intelligence (AI) refers to a programmed behaviour that simulates human cognitive functions by machines or systems such as problem-solving and decision making (Russell & Norvig, 2009). To mimic well human performance, the behaviour of good AI-enabled machines should be indistinguishable from that of a real person. An AI-enabled agent more specifically, is a system or a machine which has the competence to act properly in an uncertain circumstance, where proper action is that which increases the chance of success, that is, the accomplishment of behavioural sub-goals which support the ultimate goal of the system (Albus, 1991, p. 474). With a performance component and the built-in knowledge which enables its performance, an AI agent can become autonomous and act based on its own experience of past interactions with an environment.

One specific type of AI agent is embodied in the conversational system of the digital human (Ciechanowski, et al., 2018), which in organisations, provides a vision-based way for interacting with and supporting customers and various business processes. More specifically, the term 'digital human' refers to a new format of digital agents, which has a human face, voice, and expression. The digital human interface allows machines to communicate with people face to face and express personality and character along with responding to human emotions in a human-like-way (Teicher, 2018). As the technology advances, conversing with the digital human is expected to provide consumers with increasingly warm and engaging interactions, differing significantly when compared with unembodied interactions using a voice-only or typed chatbot interfaces. With the help of technologies such as IBM Watson, industries are benefiting from a reimagining of how humans and computers interact, and how technology can be brought "to life by creating lifelike, emotionally responsive artificial humans, with personality and characters that allow machines to talk to humans literally face-to-face" (https://www.ibm.com/case-studies/soul-machines-hybrid-cloud-ai-chatbot).

Digital humans are poised to become a part of daily life, as people get used to interacting with chatbots or other forms of digital employees that provide customer services. Examples include ANZ New Zealand bank's 'Jamie', which was initially programmed to answer the 30 most frequent questions and respond to frequently searched for online topics (e.g. how to open a bank account); 'Jamie' is able to answer about 60% of customer queries. 'Mia' works for Madera Residential, a real estate company in Texas; her role is to answer customer queries on renting an apartment; their vision for Mia is one where she is walking and can take clients on a tour of an apartment. In education, 'Will' the digital teacher is being trialled to deliver energy education in schools.

The ability of digital humans to work 24/7 and, in customer services, to deal with many frequently asked questions and provide relevant advice, is of significant benefit. For customers, this means significant improvements in the online experience, with personalised recommendations, and an opportunity to interact with organisations online in a way that conveys warmth, friendliness, empathy, and trustworthiness. For businesses, the productivity of digital humans means quicker answers with less effort and freeing up staff to handle more complex issues and provide personalised and consistent care at scale." (Source: https://www.ibm.com/blogs/ibm-anz/soul-machines/).

The use of AI-enabled conversational agents is not new. Prior research has examined people's interactions and attitudes towards AI devices and assistants such as Siri, Google Assistant, and other forms of chatbots that attempt to interact with users by mimicking conversations with human (Russell & Norvig, 2009). These show that factors such as social presence and anthropomorphism impact trust in chatbots (Yen & Chiang, 2020). However, while there has been significant work on conversational agents, few examine the more virtually interactive and physically embodied forms, which when compared with most conversational agents (whether static, disembodied, or embodied) will provide additional cues including verbal, paralinguistic and nonverbal cues such as 'thinking out loud' phrases, voice pitch, speech rate, facial gesture and posture which can impact interactions and engagement with people (Diederich et al., 2019; Van Pinxteren et al., 2020).

At the same time, the rise of the digital human raises many questions, such as: What kinds of conversations are suitable for the digital employee interaction? What control do individuals have over the information that is captured, stored and used in these systems? Ultimately, the key question that this study seeks to address is whether people will trust a digital human in the same way that they trust people? Understanding trust drivers (such as purpose, social presence and privacy) in the digital human - human interaction is important as trust shapes interactions; it facilitates behaviour and enables people to cope with risk. To address this question, we draw on trust theory (in particular tech0logy trust), and antecedents of trust to determine whether and to what extent people would trust digital humans.

2.1 Technology Trust

Trust is defined as *the willingness of a party (trustor) to be vulnerable to the actions of another party (trustee) based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party* (Mayer et al 1995). In this study, the trustor is the individual (or customer), and the trustee is the digital human. Trust is significant as it not only mediates the relationship between a person and others (including organisations), but it also plays an important role in the technology domain, with studies showing that people tend to use technologies that they trust (Lankton et al., 2015; Muir, 1987; Wang & Benbasat, 2005).

For a new technology like digital humans to be successfully adopted by organisations and used by people, the role of trust is critical. However, while prior research indicates that trust in technology may play a significant role in the digital human–person interaction, and determining people's willingness to use it (Lankton et al., 2015), the impacts on trust in an embodied-AI has received less attention. Further, given the unique properties of digital humans compared with chatbots and less human-like technologies, it cannot be assumed that the prior understanding of trust in the human-machine relationships will apply. Furthermore, it is anticipated that trust in technology is likely to differ according to the context, situational factors, and the role that the technology plays. For example, it is expected that users' views would differ for questions like: 'Would you accept advice from a digital human concerning the company's products and services' (e.g. on which bank account to choose), versus 'Would you trust a digital human to advice you on your health'? Situationally, if someone were concerned about their health they may be less willing to interact and share their personal information with a digital human, than if they were making a more general enquiry about 'good eating habits'. Understanding the trust boundaries and factors that impact these are important for understanding

people's willingness to engage with digital humans and, in making decisions about contexts in which these are deployed (e.g. as one moves from entry/low-level interactions to higher-level interactions), and how they are designed (Muir, 1987).

This leads to the following research question: *What factors influence trust in digital humans*? To address this question, this study draws on trust theory (Lankton et al., 2015; McKnight et al., 2002; Wang & Benbasat, 2005), which addresses trust in the context of the human-computer interaction.

3 Research Model

3.1 Technology Trust

Lankton et al. (2015) posit that peoples' perceptions of different technologies vary in terms of the perceived humanness of that technology. They proposed two sets of trusting beliefs as influencing trust in a technology – human-like trusting beliefs and system-like trusting beliefs. System-like trust is impersonal and focuses on the technology itself describing how a technology engenders trust in terms of functionality, reliability, and helpfulness of a technology (Langton et al. 2015). Human-like trust on the other hand is regarded as an interpersonal trust; such trust may be ascribed to a technology because people are inclined to ascribe human characteristics or human motivations to the trust object (Langton et al., 2015; Mayer et al., 1995). Such trust comprises competence, benevolence, and integrity. *Competence* refers to the extent to which the trustee has the competency and skills to accomplish what the trustors expect. *Benevolence* refers to the degree to which the trustee acts in the best interest of the trustor besides the egocentric profit motive. *Integrity* refers to the degree to which the trustee will adhere to an acknowledged set of principles and keep its promises. These three trusting beliefs are conceptually congruent with functionality, reliability, and helpfulness (Langton et al. 2015). Since digital humans, are brought to life by technology, human-like trust is considered the more salient form of technology trust (Lankton et al., 2015; McKnight et al., 2002; Wang & Benbasat, 2005), and is expected to influence engagement with a digital human. Hence it is expected that:

H1: Technology trust is positively related to willingness to engage with digital humans.

3.2 Anthropomorphism

Anthropomorphism refers to the general tendency of people to attribute human characteristics, motivations, intentions and emotions to nonhuman objects (Epley et al., 2007). Prior research shows that people tend to anthropomorphise various objects and machines such as computers and robots. This becomes even more prominent when they interact conversationally with a system, in particular one that has human-like properties such as personality and other human embodiments (e.g. emotions, facial expressions) (Epley et al., 2007). By anthropomorphising nonhuman objects, the entity's behaviour can be anticipated which increases the likelihood of a positive interaction (Waytz et al., 2010). According to the 'uncanny valley theory' as the anthropomorphic properties of an object (i.e. the movements and appearances of humanoid robots or visual/audio simulations such as digital humans and chatbots) become less distinguishable from human beings, it is expected people would show more positive emotional responses and see them as more acceptable than their counterparts up to a certain point (Richert, 2018). Beyond a point however, the response of viewers drops sharply to intense repulsion if the humanoid tries but fails to mimic a real human. For digital humans, the increasing likeness to real people is expected to minimise the 'drop off' in perceptions that may lead to negative feelings towards a technology, including privacy concerns (Ketelaar, & Van Balen, 2018).

Prior research further suggests that anthropomorphism is important in and can modify trust in a technology (Hoff & Bahir, 2015; Langton et al., 2015). However, these studies were quite narrow and did not look at technologies that have the level of anthropomorphic properties that characterise embodied digital humans. For example, Langton et al. (2015), examined people's interactions with Facebook and MS Access; while some may ascribe varying levels of 'humanness' to these interfaces, this is likely to be far less than when people interact with a digital human. We have similar expectations of interactions with disembodied agents (Wang & Benbasat, 2005). Indeed, studies also show that factors related to the human-likeness of an agent, including self-presentation and professional appearance also impact trust (Følstad et al., 2018). Coupled with the known 'drop-off' that can occur in the acceptability of visual simulations of humans (Richert, 2018), it is important to examine the extent to which anthropomorphic properties enhance trust in technology and also privacy concerns (Ketelaar, & Van Balen, 2018. It is expected that anthropomorphism will impact trust in a digital human (Langton et al., 2015), hence:

H2: Anthropomorphism is positively related to trust in digital humans.

3.3 Social Presence

Drawing on social response theory (Nass & Moon, 2000), human-computer interactions, are in general, social, such that people tend to perceive computers as social actors, even when they know they do not have feelings or intentions. Given this tendency, people tend to orient towards social interaction norms with computers similar to humans, triggering perceptions of social presence, that in turn impact trust (Yen & Chiang, 2020) and likelihood to engage with a technology (Adam et al., 2020). Furthermore, prior studies show that social presence is determined by anthropomorphic properties (Adam et al., 2020); hence it is expected that social presence will partially mediate the impact of anthropomorphism on trust and engagement with a technology (Adam et al. 2020; Jiang et al., 2019; Yen & Chiang, 2020). Hence it is expected that:

H3: Social presence of digital humans is positively related to trust in digital humans.

H4: Anthropomorphism is positively related to social presence of digital humans.

3.4 Privacy Concern

Privacy concerns arise in the context of digital human-human interactions from the collection of user data, raising questions such as what data is collected, how long it is stored for, and who has access to it (Smith et al., 1986). In the context of the digital human, data collection extends beyond the words that are spoken (i.e. speech) to include that which can be inferred from the conversation and interaction as whole including paralinguistic cues (e.g. tone), and nonverbal cues (e.g. facial expressions, proxemics, body language, gestures). While there are clear benefits of having access to such voluminous amounts of user data for development and for organisations to be able to infer key aspects of the person to digital human interaction and understanding, further ethical challenges may arise such as whether individuals' opinions may be swayed by these interactions (Følstad & Brandtzæg, 2017). Furthermore, while anthropomorphism can improve the human-technology interaction and people, their exposure to this experience may elevate privacy concerns (Xie et al., 2020) given the knowledge that the digital human they are interacting with is listening, watching and able to collect, transfer and use their information in ways that are unexpected or not desirable. Hence:

H5: Privacy concern is inversely related to trust in digital humans.

H6: Anthropomorphism is positively related to privacy concerns

With the above consideration, the following research model is proposed:

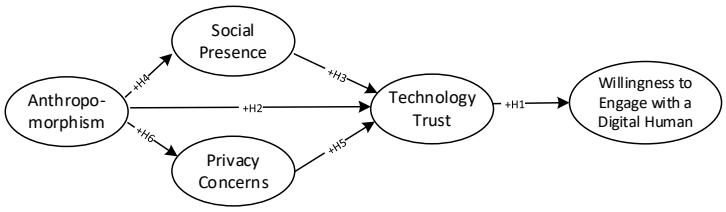

Figure 1: Research Model

4 Discussion and Conclusion

The aim of this research is to investigate trust in digital humans, and as a consequence, people's willingness to engage with a digital human. Prior research suggests that people are more likely to use technologies that they trust (Lankton et al., 2015; Muir, 1987). However, in the context of digital humans this is expected to be highly context and situationally-specific. To understand the extent of trust that people are willing to ascribe to interactions with digital humans, it is important to determine the roles of various trust dimensions (e.g. perceived benevolence, integrity and competence of the digital human) in building engagement with organizations through their digital agents.

From a practical standpoint, this study expects to contribute to current understanding of how and to what extent does the humanness of a technology engender trust and assures people that their needs will be met when interacting with a digital human. Indeed, in order not to disappoint consumers in their interactions it will be important that people know beforehand what the digital human assistant

does, and what they do not do (McTear et al., 2016). As such, transparency and appropriate calibration of the digital human with user expectations is important (Muir, 1987) to assure their trustworthiness.

From a theoretical perspective this research expects to make the following contributions. First the study aims to extend the technology-trust model in to the domain of emerging technology-human agency as supported by the AI-enabled human agent (Lankton et al., 2015). Theoretically, the findings are expected to provide insights into the human-technology interaction and a starting point for further examining the phenomenon. So although digital humans use is currently rare, as their deployment increases, ideally it is expected that this study will provide valuable insights into the role and agency of digital humans in the human-technology interface, the opportunities and the challenges.

For this paper an initial model of trust in digital humans is proposed. We further posit that anthropomorphism together with trusting beliefs will influence outcomes, in this case peoples' willingness to engage with a digital human agent. To assess the model, an experiment will be conducted in which contexts (and hence agents, products and services) are varied, and a survey used to capture people's perceptions in relation to trust dimensions, its antecedents (i.e. privacy concern anthropomorphism, and social presence), and willingness to engage with the product or service. The structural models will then be analysed and the results compared. As the research progresses, additional factors and context will be explored, so as to understand the boundaries of trust as these apply to digital human agents. For example, it is expected that the relative importance of various trust dimensions and other factors such as privacy concerns, will be situational and be influenced by other considerations including the complexity of the interaction and the sensitivity of the context. Thus there is significant opportunity to contextualize and test the technology-trust framework in the merging space of humans and technology automation.

5 References

Adam, M., Wessel, M., & Benlian, A. (2020). AI-based Chatbots in Customer Service and their Effects on User Compliance. *Electronic Markets*, 1-19.

Brandtzaeg, P. B., & Følstad, A. (2017, November). Why People use Chatbots. In *International Conference on Internet Science* (pp. 377-392). Springer, Cham.

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the Shades of the Uncanny Valley: An Experimental Study of Human–Chatbot Interaction. *Future Generation Computer Systems*, *92*, 539-548.

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2019). On Conversational Agents in Information Systems Research: Analyzing the Past to Guide Future Work. *Proceedings: 14th International Conference on Wirtschaftsinformatik*, February 24-27, Siegen, Germany

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-factor theory of Anthropomorphism. *Psychological Review*, 114(4), 864-885.

Følstad, A., & Brandtzæg, P. (2017). Chatbots and the New World of HCI. *Interactions*, *24*(4), 38-42.

Følstad A., Nordheim C.B., Bjørkli C.A. (2018) What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In: Bodrunova S. (eds) *Internet Science. INSCI 2018. Lecture Notes in Computer Science,* 11193. Springer, Cham. https://doi.org/10.1007/978-3-030-01437-7_16

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors that Influence Trust. *Human Factors*, *57*(3), 407-434.

Ketelaar, P. E., & Van Balen, M. (2018). The Smartphone as Your Follower: The Role of Smartphone Literacy in the Relation between Privacy Concerns, Attitude and Behaviour towards Phone-Embedded Tracking. *Computers in Human Behavior*, 78, 174-182.

Lankton, N. K., McKnight, D.H., Tripp, J., Marshall. (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. *Journal of the Association of Information Systems*, *16*(10), 880-918.

Mayer, R. C., Davis, J. H., & Schoorman, F.D. (1995). An Integration Model of Organizational Trust. *Academy of Management Review*, 20(3):709–734.

McKnight, D.H., Choudhury, V. & Kacmar, C. (2002a) "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology" *Information Systems Research*, (13)3, 334-359

McTear M., Callejas Z., Griol D. (2016) Creating a Conversational Interface Using Chatbot Technology. *In: The Conversational Interface: Talking to Smart Devices*. Switzerland: Springer.

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-service Technologies: Understanding Customer Satisfaction with Technology-based Service Encounters. *Journal of Marketing*, 64(3), 50-64.

Muir, B. (1987). Trust between Humans and Machines, and the Design of Decision Aids. *International Journal of Man-Machine Studies*, 27, 527-539

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81-103.

Ng, M., Coopamootoo, K. P., Toreini, E., Aitken, M., Elliot, K., & van Moorsel, A. (2020). Simulating the Effects of Social Presence on Trust, Privacy Concerns & Usage Intentions in Automated Bots for Finance. *arXiv preprint arXiv:2006.15449*.

Richert, A., Müller, S., Schröder, S., & Jeschke, S. (2018). Anthropomorphism in Social Robotics: Empirical Results on Human–Robot Interaction in Hybrid Production Workplaces. *AI & Society*, 33(3), 413-424.

Russell, S., & Norvig, P. (2009). Artificial Intelligence: A Modern Approach.

Teicher J. (2018). *Could your next Employee be an AI-powered digital human?* Retrieved July 9, 2018, from https://www.ibm.com/blogs/ibm-anz/soul-machines/

Van Pinxteren, M. M., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like Communication in Conversational Agents: A Literature Review and Research Agenda. *Journal of Service Management*, 31(2), 203-225.

Wang, W. & Benbasat, I. (2005). Trust in and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems*, 6(3), 72-101.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees Human? The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219-232.

Xie, Y., Chen, K., & Guo, X. (2020). Online Anthropomorphism and Consumers' Privacy Concern: Moderating Roles of Need for Interaction and Social Exclusion. *Journal of Retailing and Consumer Services*, 55, 102119, 16pp.

Yen, C., & Chiang, M. C. (2020). Trust me, if you can: A Study on the Factors that Influence Consumers' Purchase Intention Triggered by Chatbots based on Brain Image Evidence and Self-Reported Assessments. *Behaviour & Information Technology*, 1-18. DOI:10.1080/0144929X.2020.1743362

Copyright