

Can We Trust an AI Agent? Interaction Effects of Its Machine Learning Performance and Digital Character

Taejin Kim
Ajou University
tjnettt@ajou.ac.kr

One-Ki Daniel Lee
University of Massachusetts Boston
Daniel.Lee@umb.edu

Juyoung Kang
Ajou University
jykang@ajou.ac.kr

Abstract

AI-powered digital characters (i.e., AI agents) are expanding their scope of application to various fields. However, research on the key factors influencing AI users' attitude is insufficient. This study investigates the role of machine learning (ML) performance (as the behavioral/ intelligence realism of AI agents) in determining users' trust. The study further investigates the interaction role of different forms of digital character (as the form realism of AI agents) in the relationship between ML performance and trust. The findings achieved from an experimental setting provide a novel understanding of human-AI interaction, expand academic understanding of AI anthropomorphism, and suggest new research directions for AI-powered digital characters. The results will also guide business practitioners in developing various AI services.

Keywords: artificial intelligence, machine learning, digital characters, avatar, digital human, trust

1. Introduction

In the past, AI-powered digital characters, the live entities that resemble humans in terms of their shape, characteristics, and expression (Miao et al., 2022; Silva & Bonetti, 2021), were mainly active in one-way media such as TV and movies. But now, their presence has expanded into various fields such as interactive education and training, counseling, influencer marketing, and customer service, and the scope of their applications is increasing further. For example, in South Korea, major banks, including Shinhan Bank, Kookmin Bank, and NH Bank, are applying AI-powered digital characters to their customer services and transactions such as withdrawal and transfer services¹. And MBN

has produced and organized news within 1-2 minutes with digital humans who learned from real news anchors².

Digital characters are defined as digital entities with an anthropomorphic appearance that can be controlled and interacted with by humans or software (Miao et al., 2022). They are characterized as a life-like being powered by artificial intelligence (AI) and capable of conversation (Silva & Bonetti, 2021). They can be categorized into different types such as avatar and digital human based on their level of human-like realism (Seymour et al. 2021; Silva and Bonetti, 2021). While an avatar is frequently presented as an animated character equipped with human behaviors, a digital human is characterized as a living entity that resembles an actual person. Unlike an avatar, a digital human can generate an illusion that it is a human (Terry, 2018).

Recently, the digital humans have become more applicable to various contexts with various purposes. For example, they are used as an interface for automated chatbots to enhance communication with consumers (e.g., SK Telecom's digital human SUA³). Digital humans are also used for entertainment purposes (e.g., digital human Rozy⁴ as an AI influencer and the RINA⁵ as an AI singer) and for representation of specific professions (e.g., the LUI⁶ as an AI poet). Likewise, digital humans (and also avatars) are actively used already in real-life services and will be more applied to various contexts. For example, NVIDIA launched NVIDIA Omniverse Avatar, an AI assistant creation platform that can be customized for various industries. However, our understanding of human users' feeling or reactions to digital characters are blurred yet.

Extant studies on other types of cognitive agents like service robots have been mostly carried out in the

¹ Regarding the AI Banker, refer to <https://www.etnews.com/20220106000066>

² Regarding the AI Reporter, refer to https://newsis.com/view/?id=NISX20220218_0001764880&cID=10701&pID=10700

³ Regarding the SUA, refer to <https://news.sktelecom.com/181683>

⁴ Regarding the Rozy, refer to <https://www.tatlerasia.com/style/fashion/rozy-virtual-influencer-south-korea>

⁵ Regarding the RINA, refer to <https://biz.chosun.com/it-science/ict/2022/03/29/TA7MOITASZA45HWUMZGV5SARN4/>

⁶ Regarding the LUI, refer to <https://www.aitimes.kr/news/articleView.html?idxno=27343>

field of human-robot interaction (HRI), focusing on their physical features and the uncanny valley effect (Mori et al., 2012). Particularly in the service robot research, the anthropomorphism of robots has been considered a critical factor in shaping consumer attitudes and behaviors. However, due to the technological limitations of service robots, it has been found that human-like robots do not cross the uncanny valley curve and bring discomfort to people.

In the case of digital humans implemented in the virtual space, there is almost no difference in appearance from real humans. Seymour et al. (2022) distorted the appearance of a digital human to form its five different variations and studied changes in consumers' attitudes toward the different forms. Although respondents recognized the difference in the formation of the digital human, it was confirmed that the difference does not affect their affinity, trustworthiness, and intention to purchase, indicating that the uncanny valley effect may not apply to the digital humans (or digital characters in general). Hence, we may need a new approach or perspective in understanding the relationship between human users and AI-powered digital characters.

Miao et al. (2022) proposed two-dimensional taxonomy of digital agents: form realism and behavior realism. Form realism is about the anthropomorphic aspect of digital agents, while behavioral realism is about the actions performed by digital agents. In the case of AI-powered digital characters, the behavioral realism is mainly about their intelligent behaviors. However, most of the prior studies have focused on the physical appearance of digital agents as their form realism (Seymour et al., 2022; Seymour et al., 2021). Thus, the behavioral realism, especially about the intelligent behaviors or capabilities of digital agents, have seldom been discussed, which calls for research on multi-dimensional perspective for further understanding of digital agents.

To fill these research gaps, this study explores the interaction effects of machine learning (ML) performance and digital character type on consumer trust. According to Mayer et al. (1995), trust can be defined as “the willingness of a party to be vulnerable to the actions of another party 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” (p. 712). As AI systems become an important social actor, such a trustworthy relationship has achieved a significant attention relating to the interaction between human users and AI systems in the recent IS literature (e.g., Lee et al., 2021). Hence, it will be important and useful to answer the following question through this study:

RQ: How do the digital character type and the ML performance level interactively affect consumer trust?

To answer this research question, this study employed an experimental research method to investigate whether and how two types of digital character (i.e., caricature avatar and digital human) and three levels of ML performance (low, mid, and high) interact to form consumer trust.

This study will expand academic understanding of human-AI (HAI) interaction and AI-powered digital characters (AI agents) by revealing the interaction effects of two factors on consumer trust. The results of the study will guide developers of AI services.

2. Literature review

2.1 The uncanny valley and digital characters

In the 1970s, Mori's uncanny valley theory developed as a nonlinear response to realism in robotics research. The theory states that the user's affinity increases when a robot resembles a human. However, when a robot's form is almost identical to a human's, a feeling of uncanniness is aroused, and the user's affinity drops dramatically. However, the theory also states that if the robot resembles a healthy human in a significantly realistic manner, it will cross the valley and reach the highest level of affinity.

Recently, this theory has been expanded to digital human avatars used in virtual reality and augmented reality technology—a digital human is an avatar with human-like features and behavior (Miao et al., 2022; Silva & Bonetti, 2021). Digital characters such as avatars are attracting commercial interest through theorizing and evaluating the notion of “crossing the uncanny valley”, and several studies are being conducted (Kätsyri et al., 2015; Wang et al., 2015).

Seymour et al. (2022) found that the realism of existing digital avatars was divided into three levels and his experiment garnered the same non-linear response as the phenomenon of the robot: The non-realistic caricature avatar was located on the left side of the valley while the human-realistic avatar was located on the right side across the valley with the uncanny almost human-realistic avatar shown to be located in the valley, the point where the affinity rapidly falls.

Furthermore, Seymour et al. (2022) applied five different visual distortions centered on the eyes, mouth, and body to the appearance of the human-realistic avatar and subsequently measured the changes in consumer attitudes toward it. Contrary to the researchers' expectations, respondents recognized the differences in the distorted visual quality of the digital human video but did not display any differences in affinity, trustworthiness, and willingness to pay. From this, it can be concluded that a reasonable level of digital-human visual quality and expression fidelity negates any difference in the visual quality in terms of affecting

consumers' behavior and perception, confirming that the realistic digital human does not cause any feelings of uncanniness (Lee et al., 2021).

We can also conclude that it is not possible to find the root cause of changes in consumer attitudes and behavior when experimenting with appearance factors alone. A multidimensional analysis with new factors based on appearance is needed for a deeper understanding.

2.2 Avatars & digital humans

An avatar is defined as a digital entity with an anthropomorphic appearance that can be controlled and interacted with by humans or software (Miao et al., 2022).

In addition, the anthropomorphic appearance of avatars that resemble humans is considered the most important element in the conceptual definition of digital avatars. According to the Computer-as-Social-Actors (CASA) paradigm (Reeves & Nass, 1996), people tend to regard computer technology that exhibits human-like characteristics as a social actor and apply the same social action. In other words, the avatar's anthropomorphic appearance elicits cognitive, emotional, and social responses (Wang et al., 2007) from people. Therefore, digital assistants that lack the anthropomorphic appearance of an avatar are likely to be excluded.

Digital avatars used in business are almost entirely activated and controlled by AI (Miao et al., 2022). Advances in digital technology and AI have enabled the development of complex avatars, and they are now becoming increasingly important, especially in online service experiences such as education, gaming, banking, and shopping (Kim et al., 2016).

According to Silva and Bonetti (2021), a digital human is defined as a living entity that resembles a person more than an avatar in shape, characteristics, and expression and is driven by artificial intelligence (AI) with the ability to communicate. Digital humans can communicate, create emotional connections, and interact with consumers like humans. Above all, digital humans differ from animated character avatars in one key feature: they create the illusion that 'they just live their lives, just like us' (Terry, 2018).

For example, AI-powered digital humans in the fashion industry can better understand consumers' tastes through learning, design clothes accurately to consumers' measurements, and provide consumers with a more productive and efficient shopping experience (McDowell, 2020).

Considering that digital humans are driven not only by human-like forms but also by AI that resembles human learning abilities, Miao et al. (2022) divided a

digital human avatar into two dimensions: avatar's form realism and behavior realism. Form realism refers to an anthropomorphic appearance resembling a person, and behavioral realism refers to intelligent actions performed by digital agents. The intelligence corresponding to the behavioral realism of the digital human avatar can be explained by machine learning, the learning ability of AI.

2.3 Machine learning & recommendation system

A recommendation system is an information filtering system using machine learning or AI algorithms that provides personalized content and services to users by filtering big data with high efficiency (Isinkaye et al., 2015; Resnick & Varian, 1997). However, recommendation systems are limited by data sparseness, but machine learning can mine deeper information between input features, enabling it to provide satisfactory recommendations to users even without complete information. Over the past decade, machine learning algorithms have been integrated into recommendation systems and have effectively handled various recommendation tasks (Watson, 2022; Zauskova et al., 2022).

Today, recommendation systems are integrated into almost all online service websites, providing more revenue and contributing to the development of the recommendation system itself (Chen & Qin, 2021). Many online video platforms, such as YouTube, Netflix, and TikTok employ movie recommendation systems to make personalized content accessible to billions of users.

The performance of the machine learning (ML) algorithm can be checked through ML accuracy. Accuracy, is one of the criteria for evaluating the performance of ML algorithms and is defined by the number of correctly predicted samples among the total number of samples (García et al., 2009). When developing a movie recommendation system using five ML algorithms, Elias et al. (2022) evaluated accuracy to measure ML performance, and several studies including a land cover classification (Jo et al., 2019), diagnostic accuracy in medical imaging (Aggarwal et al., 2021), and a COVID-19 detection model (Vaid et al., 2020) have evaluated accuracy to assess the performance of ML algorithms.

3. Research model and hypotheses

According to Følstad et al. (2018), factors related to human likeness in digital agents like chatbots and digital humans affect users' trust on the agents (Lankton et al., 2015). For the human likeness of AI-powered digital

characters, this study first focuses on the factor of ML performance that resembling the human learning ability as a new factor that causes the difference in attitudes of digital characters.

As reviewed in the literature, ML is essentially a multi-layered neural network that mimics the human brain (LeCun et al., 2015). In particular, the machine learning technology of the CNN algorithm has been proven to be particularly effective in the field of recommendation systems (Watson, 2022; Zauskova et al., 2022). Elias et al. (2022) studied a person's emotional state through various machine learning neural network algorithms, classified it into emotions such as happiness, sadness, anger, neutrality, disgust, and fear, and implemented a system that recommends movies based on emotional state.

Just as movies are an interesting field regardless of age or gender, and people prefer movies when they are in a good or bad mood, recommendation is the only medium that can easily modify a person's emotional state (Ozdemir, 2022). As seen in the theoretical background, trust has been confirmed in many studies as an essential factor in influencing consumers' online purchase decisions in e-commerce (Gefen & Straub, 2004; Van der Heijden et al., 2003; Yoon, 2002). Thus, as a key factor in determining consumers' purchase intention, trust was set as the dependent variable.

It has been suggested that anthropomorphism is essential to technology and that trust can be modified, as we have seen in our review of literature on the topic (Hoff & Bashir, 2015). However, while prior studies have explored interactions with non-humanistic interfaces such as Facebook and MS Access (Lankton et al., 2015), they have not assessed the impact of the level of anthropomorphic attributes that characterize digital agents on trust.

According to Mayer et al. (1995), trust consists of three human-like trust concepts: integrity, benevolence, and competence. Competence is the belief that a trustee has the skills, competencies, and characteristics to exert influence in a particular domain. Integrity is the belief that the trustee adheres to an acceptable set of principles, and Benevolence is the belief that the trustee will demonstrate a willingness to do good even when there is no profit motive (Lankton et al., 2015). Since people tend to anthropomorphize a technology and attribute human motives or human qualities to the technology, researchers have used these human-like beliefs to study trust in technology (Reeves & Nass, 1996). Moreover, researchers found that human-like trust beliefs had a great influence on intention to use (Benbasat & Wang, 2005).

Lankton et al., (2015) has shown that consumers show higher trust in technology that resembles humans rather than technologies that resemble less humans. In

addition, it was found that consumers showed lower trust in technology with low social presence and low social affordance (e.g., Microsoft Excel) and higher trust in technology with high social presence and high social affordance (e.g., Siri). This is because consumers recognize that technologies with high social affordances resemble humans more than technologies with low social affordances.

Research on avatars and digital humans in the digital space so far has remained focused on one-dimensional physical appearance. Therefore, according to the two-dimensional taxonomy of digital agents: form realism and behavior realism (Miao et al., 2022), this study attempted to multidimensionally examine the interaction effect of digital character type (caricature avatar and digital human) and various levels of ML performance (low, mid, and high) on consumer trust. Therefore, we hypothesized as follows:

H1: ML performance levels will have a positive effect on consumers' trust.

H2: ML performance levels and digital character type will have an interactive effect on consumers' trust.

Figure 1 shows the research model based on the above hypotheses' development.

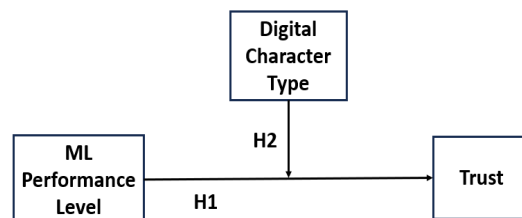


Figure 1. Research model

In addition to the core variables and their relationships, we used two control variables: prior experience and negative emotional reactions to AI. According to a study of the positive effect of prior experience reported in human-robot interaction research (Bartneck et al., 2007) and the mere-exposure effect (Zajonc, 1968), the experience of using avatars or digital humans was used as a control variable. And according to a study that showed the emotional reactions toward robots could be affected by general attitude toward robots and the perceived suitability of robots to a specific context (Savela et al., 2021), we used the negative emotional reactions of AI as a control variable.

4. Methods

4.1 Experimental design

In the experimental situation, a movie recommendation service provided on a membership-based video-on-demand website such as Netflix was utilized.

ML performance was measured by the accuracy of prediction when developing a machine learning model. In particular, we manipulated the ML performance levels by varying the accuracy level of movie recommendations for a specific movie genre that the respondents preferred, i.e., how many accurate movie recommendations among a total of 10 to the respondents, classified as low, mid, and high levels of accuracy.

For manipulation of the accuracy of movie genre recommendation, a preliminary survey was conducted on 200 respondents regarding their favorite movie genre among the following genres. 1) Romance Comedy 2) Action 3) Thriller 4) SF Fantasy 5) Horror 6) Other Genres. As a result of the preliminary survey, the 'SF Fantasy' genre was identified as the most preferred genre by the public (n=55, 27.5%).

In the experimental video produced according to the ML performance level, a total of 10 recommended movie posts are guided, and as shown in Table 1, the prediction accuracy, which is ML performance, was manipulated by the number of times the movie of the most preferred movie genre 'SF Fantasy' was shown. For example, in mid-level ML performance, among a total of 10 movie recommendations, 'SF Fantasy' is provided 6 times, and movies of other genres are recommended for the remaining 4 times.

Table 1. Machine learning performance level

ML Level	Accuracy	Experiment Manipulation Situation
Low	30%	Low-level prediction accuracy of less than 3 per 10 recommendations. *In order: SF Fantasy , Action, SF Fantasy, Horror, Action, Mystery, SF Fantasy , SF Fantasy, Horror, Thriller.
Mid	60%	Mid-level predictive accuracy of less than 6 per 10 recommendations. *In order: SF Fantasy , Action, SF Fantasy , Horror, SF Fantasy , SF Fantasy , SF Fantasy, Comedy, SF Fantasy , Thriller.
High	100%	High-level predictive accuracy, correctly predicting all 10 out of 10

⁷ Regarding the MachineBrainAI, refer to <https://www.machinebrain.io/ko/home>

		recommendations. *In order: SF Fantasy , SF Fantasy , SF Fantasy , SF Fantasy , SF Fantasy , SF Fantasy , SF Fantasy , SF Fantasy , SF Fantasy , SF Fantasy .
--	--	--

To identify the impacts of the various levels of ML performance and the different types of digital characters on users' trust, this study conducted a random analysis with a 3 x 2 factorial design based on the three levels of ML performance (low, mid, and high) and the two types of digital character (caricature avatar and digital human).

In the past, the creation of digital humans involved high production costs and long production times using expensive equipment, studios, and professional manpower, but recently, time and cost have been greatly reduced with the development of CG, AI, and motion capture technologies.

At the DeepBrain AI Studio⁷, digital human images and voices suitable for the movie recommendation service experimental situation were selected, and experimental images were produced by combining the experimental situation background and recommended movie post images.

The caricature avatar video was produced in Midjourney using GAN, a generative model technology. First, a female agent image used in the digital human experiment video was input, with a caricature avatar image suitable for the movie recommendation experiment situation then created from this image. Then, after producing voice files at Typecast, which produces voice content using artificial intelligence voice actor technology, at Studio DID, the caricature avatar video was produced using the created image and audio files.

4.2 Participants and procedure

The experimental participants were recruited through the Invite website, an open survey platform among office staff working in Korea as undergraduates or graduates. The participants for this study were 571 Koreans (287 female) recruited from the ConsumerInsight online panel: 20-29 (24.1%), 30-39 (25.0%), 40-49 (25.0%), and 50-59 (25.7%). As the screening criteria for recruitment, the following criteria were used: 1) Korean residents, 2) university or graduate school graduates, 3) workers, and 4) SF fantasy as the most favorite movie genre. Participants who did not meet the criteria were excluded from the experiment.

As shown in Tables 2 and 3, one of six experimental images combining 2 digital characters (caricature avatar, digital human) and 3 ML performance levels (low, mid, and high) was provided by random assignment (2 x 3 between-subjects factorial design).

First, the digital character (either caricature avatar or digital human) explains the following scenario before the movie recommendation process begins for the video to be watched by the respondents. "Hi, welcome! I've been learning your movie-viewing history. Based on what I have seen so far, I will now start recommending movies in your favorite genre." Then, the digital character recommends 10 movies, including SF Fantasy movies (as the favorite movie genre for respondents) and non-SF Fantasy movies, based on the accuracy-level group that the respondents belong to. The participants were asked to proceed with the questions that followed the video. To confirm the manipulations given to each participant, the respondents were asked to select which type of digital character was provided, caricature avatar or digital human. Then, we used a seven-point scale (1 = "strongly disagree", 7 = "strongly agree") for the trust measurements, which we derived from previous research.

Table 2. Experimental video for caricature avatar

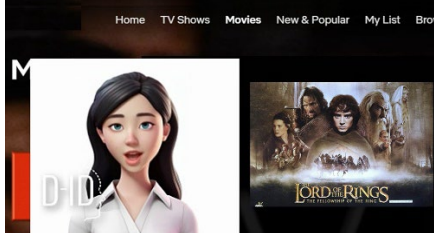

Video	
Caricature Avatar	A caricature avatar presented a total of 10 preferred-genre movie posters, according to the ML performance level.

Table 3. Experimental video for digital human

Video	
Digital Human	A digital human presented a total of 10 preferred-genre movie posters, according to the ML performance level.

5. Results

5.1 Manipulation check

First, of the 282 respondents who received the avatar type, 24 who responded that they were digital humans were failed from the manipulation check, and among 277 respondents who received the digital human type, 12 who responded that they were avatars failed the manipulation check. And 12 respondents who responded that the most preferred genre was not SF Fantasy were excluded as inappropriate respondents. Therefore, a total of 523 samples (267 female) were created: 20-29 (23.1%), 30-39 (26.0%), 40-49 (24.8%), 50-59 (26.0%).

Second, as results of counting the number of times that the genre of SF Fantasy was suggested by the respondent for the manipulation check of the ML performance level, all of them showed a significant level of difference. ($F(2,522) = 277.073, p < 0.000$) as the Table 4.

And the analysis of pairwise comparisons revealed that a high-level ML performance gained higher accuracy than a mid-level ($p = .000$), and mid-level ML performance showed higher than low-level ML ($p = .000$). thus, all were significantly confirmed in the order of ML low < ML mid < ML high.

Table 4. Manipulation check results for ML performance

Level	N	Mean	SD
Low	171	2.977	.9759
Mid	177	5.475	1.373
High	175	7.309	2.444
Total	523	5.272	2.463

5.2 Descriptive statistics

Tables 5, 6 show the descriptive analysis results of each of the ML performances and digital character stimuli on trust, and Table 7 shows the descriptive analysis results of the interactions between the two independent variables.

Table 5. Descriptive analysis of perceived ML performance

Perceived ML Performance	N	Dependent Variable	
		Trust	
		Mean	SD
Low	171	3.823	.945
Mid	177	4.056	.978
High	175	4.125	.981
Total	523	4.003	.975

Table 6. Descriptive analysis of digital character type

Digital Character	N	Dependent Variable	
		Trust	
		Mean	SD
Caricature Avatar	258	3.942	.910
Digital Human	265	4.062	1.033

Total	523	4.003	.975
-------	-----	-------	------

Table 7. Descriptive analysis of digital character and ML performance

Digital Character	ML Performance	N	Dependent Variable	
			Trust	
			Mean	SD
Caricature Avatar	Low	87	3.7816	.921
	Mid	85	4.1529	.851
	High	86	3.8983	.925
Digital Human	Low	84	3.866	.974
	Mid	92	3.967	1.080
	High	89	4.345	.988
Total		523	4.062	1.033

5.3 Two-way ANCOVA

A two-way ANCOVA analysis of trust was performed as shown in Table 8. Pre-AI anxiety and prior experience were entered as control variables. The ANCOVA analysis results for different ML performances were significant ($F_{(2,522)}=4.807, p=.009$). Therefore, H1 was supported.

Since there were three different levels for the ML performance, we further conducted pairwise comparisons among them. The analysis of pairwise comparisons revealed that a high-level ML performance gained higher trust than a low-level ML performance ($p=.010$) but the comparisons with the mid-level were not significant ($p=.778$). A mid-level ML performance showed a higher trust than a low-level one ($p=.062$).

Table 8. Two-way ANCOVA analysis

Source	Dependent Variable			
	Trust			
	Type III Sum	df	F	p
Corrected model	20.484	7	3.164	.003
Intercept	395.629	1	427.754	.000
Pre-Anxiety	1.231	1	1.331	.249
Prior experience	3.552E-7	1	.000	1.000
Digital Character	1.771	1	1.914	.167
ML Performance	8.892	2	4.807	.009*
Digital Character * ML Performance	8.727	2	4.718	.009*
Error	476.323	515	-	-
Total	8878.813	523	-	-

* $p < .05$; ** $p < .01$; *** $p < .001$, df: degree of freedom

However, the results revealed significant interaction effects between the ML performance and the digital character on trust ($F_{(2,515)}=4.718, p=0.009$). Therefore, H2 was supported.

Interestingly, the effects of ML performance levels differed depending on the type of digital character. A low-level ML performance received higher trust in the case of the digital human. A mid-level ML performance received higher trust in the case of the

caricature avatar. However, the caricature avatar with a high-level ML performance received a decrease in trust and the digital human with a high-level ML performance received the highest trust. Therefore, H2 was fully supported. Figure 2 shows the graphical results of our two-way ANCOVA analysis.

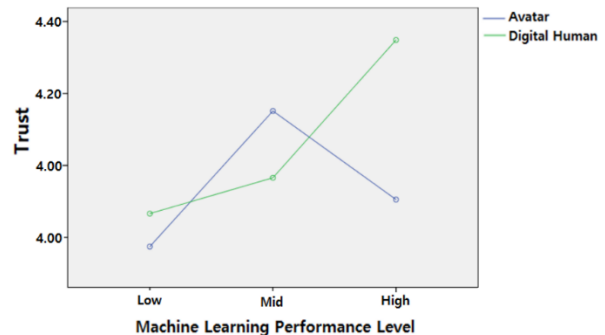


Figure 2. Interaction effects of ML performance and digital character on trust

6. Discussion

Until now, studies on digital characters had been limited to one-dimensional studies focused on physical appearance, and no new factors influencing consumer attitudes and behaviors had been found (Seymour et al., 2022; Seymour et al., 2021). However, this study found a new key factor influencing consumer trust in the digital space, the ML performance factor.

As a result of the study, the ML performance factor showed a significant positive effect on trust. This is an extension of Seymour et al. (2022), providing a novel understanding of AI-powered digital characters and AI agents research and suggesting a new research direction.

In line with Miao et al. (2022) two-dimensional taxonomy classification, this study determined the form realism dimension of AI agents using two types of digital characters, caricature avatars and digital humans, and the behavioral (intelligence) realism dimension of AI agents using their different levels of ML performance, low, mid, and high and confirmed the multidimensional interaction effect of these two dimensions on users' trust.

With a low-level ML performance, caricature avatars received somewhat lower trust than digital humans. This is possibly because people who perceived the caricature avatar as an automated object rather than a person had a high expected trust, but the lack of performance significantly decreased their trust (Alarcon et al., 2021).

However, with a mid-level ML performance, the caricature avatar received higher trust than the digital human, and with a high-level ML performance, the caricature avatar received lower trust, while the digital

human received the highest trust, which is of significant interest.

This was a different result from the expectation of the H2 hypothesis. In the case of the digital human giving a high-level ML performance, although it shows interaction effects, it was expected that trust would decrease significantly due to negative factors such as artificial intelligence anxiety. Thus, the highest trust was expected when the caricature avatar gave a high-level ML performance, but a different result was seen.

This is consistent with a study that predicted that anthropomorphism in technology would affect the trust of digital humans (Lankton et al., 2015), and a study in which more realistic avatars (digital humans) received more trust than less realistic avatars (caricature avatars) (Seymour et al., 2021).

Moreover, even when ML performance increased to a high level, there was no significant negative response. This is consistent with the Ketelaar and Van Balen (2018) study on robots eliciting greater feelings of uncanniness as their human-likeness increases, which showed that digital humans, despite possessing a strong human likeness, minimize negative aspects of technology, including privacy issues.

This may have been related to the fact that in the experimental situation of the movie recommendation system, there was no critical risk related to consumer benefits and costs. And in the Savela et al. (2021) study, the age and gender attributes of young individuals were identified as factors influencing attitudes toward robots, and it is thought that these factors may have had an impact as socio-democratic factors.

In the case of caricature avatars, a shape similar to the uncanny valley curve was confirmed. When ML performance increased from low to mid, trust in caricature avatars increased, but when ML performance increased from mid to high, trust in caricature avatars decreased; thus, a non-linear response was confirmed.

On the other hand, in the case of digital humans, rather, as the ML performance level increased from low to mid and mid to high, all showed a linear response in which users showed the highest trust.

This result shows that a digital human with high anthropomorphism can overcome the uncanny valley phenomenon that occurs in a caricature avatar with low anthropomorphism. This expands the novel understanding of human-AI interaction and academic understanding of AI-powered digital characters (AI agents) and suggests a new research direction.

As Miao et al. (2022) stated, digital agents used in business are almost entirely activated and controlled by an AI engine, and future research should actively conduct ML performance research.

7. Conclusion

7.1 Theoretical implications

Based on Mori's uncanny valley theory, anthropomorphism is a key factor in determining consumer attitudes and behaviors in the field of service robots and digital characters. In particular, visual features have, to date, been the most important factor, and one-dimensional studies focused on physical appearance have been conducted in the digital human domain. However, previous studies have confirmed that the physical appearance of digital humans does not affect affinity, trustworthiness, and bidding behavior, and thus there is a demand for the discovery of new factors that affect consumer attitudes and behaviors in the digital space.

To this end, this study has uncovered the following implications by studying AI agents with a two-dimensional approach that included form realism factors and behavioral realism factors.

First, this study discovered ML performance as a key factor influencing consumers' trust in the digital space. Second, the multidimensional interaction effect of the digital characters (caricature avatars and digital humans) factor and the ML performance factor, on trust was confirmed. Third, While Miao et al. (2022) provided a two-dimensional theoretical framework of form realism and behavior realism for AI Agents, this study classified it into a 2-level digital character and 3-level ML performance level for empirical research through experimental methods.

This will provide a novel understanding of human-AI interactions and AI-powered digital characters (AI agents) and suggest new research directions.

7.2 Practical implications

In this study, the movie recommendation system was used as the experimental environment since the ML performance is the most used recommendation system.

However, AI powered digital characters are not only used in the recommendation system but are used in a wide variety of fields that require two-way communication, such as AI banker, AI news anchor, AI teacher, AI staff, AI doctor, AI nurse, and AI model.

In addition, digital humans are used as an interface for automated chatbots to enhance communication with consumers. SK Telecom's digital human SUA realizes natural voices and facial expressions like real people and is used in customer

centers, voice assistants, and metaverses. A digital human named Rozy is used for a chatbot service called Rozy Chat.

In the field of AI Artists, RINA, a digital human developed by netmarble, is active as an AI singer, and LUI, a digital human developed by MaumAI, is active as an AI poet.

As we have seen, the results of this study can provide implications for creating a combination of optimal ML performance level and digital character type that can lead to consumer trust in various settings where digital characters are utilized.

When developing an AI-powered digital agent, a decrease in ML performance will reduce consumer trust, and a high level of ML performance will increase consumer trust, which in turn will increase consumer purchase intention behavior.

The results of this study on the non-linear interaction effects of digital characters with different levels of anthropomorphism (caricature avatars and digital humans) and various ML performance levels (low, mid, and high) will provide guidance to developers of AI services using machine learning. If the AI service suffers from a low level of ML performance, the use of a digital human will lead to higher consumer trust. Similarly, with a mid-level ML performance, the use of a caricature avatar will garner more consumer trust leading to a higher purchase intention of consumers.

However, in the case of AI services with a high level of ML performance, consumer trust may be lowered with the use of a caricature avatar. Therefore, the use of a digital human can lead to optimal purchase intention behavior by gaining a high level of trust from consumers.

8. Limitations and future research

This study has several limitations in guiding future research directions.

First, this study was conducted in the context of a movie recommendation system where ML algorithm performance can be best utilized. However, this setting may have a limitation in developing practical implications for other relevant contexts that use AI-powered digital characters. Moreover, the current experimental situation can be limited in causing negative customer attitudes as there is no critical risk related to the benefits and costs of consumers. In the future, therefore, it will be useful to study whether ML performance causes positive or negative attitudes of consumers, especially in the experimental situations that increase consumer risks such as financial investment situations.

Second, this study was conducted in a single geographic region. Therefore, it is necessary to conduct this study in another region to take into account potential influences arising from differences in cultural or social background.

Third, to manipulate the three ML performance levels, this study applied the accuracy dimension only, while various alternative dimensions such as precision, recall, and hit rate have also been used for performance evaluation of the ML recommendation system in the literature and real-world settings. Therefore, future research needs to consider such alternative dimensions to detect extra insights regarding the relationship between ML performance and trust.

Fourth, in this study, a movie recommendation service was used as our experimental context, but it was not considered whether the participants in the experiment had experience of watching the movies recommended. In addition, external reviews of the recommended movies could affect the participants' decisions, which was also not considered in this study. In future research, an experimental design that considers the movie viewing experience of survey participants and the potential influence of external reviews will be needed.

Lastly, for our experimental manipulation, only one genre, 'SF Fantasy' was used since it was the most preferred genre by a potential respondent group. However, manipulating the ML performance using actual preferences of each participant at the movie level (involving multiple genres) may help make the experiment more realistic. Moreover, even within a movie genre, the recommended movies can be heterogenous, which can generate a compounding effect. Therefore, in future research, it may be more useful to conduct an experiment using recommendation accuracy based on each respondent's actual preferences at the movie level.

Despite these limitations, this study expands academic understanding of human-AI interaction and AI-powered digital characters as AI agents by proposing a multidimensional approach to form behavioral realism of consumer responses to various AI applications. The study also presents new research directions in the relevant conceptual and empirical developments.

9. Acknowledgement

This work was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0020632, HRD Program for Industrial Innovation).

10. References

- Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S., Karthikesalingam, A., King, D., . . . Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ digital medicine*, 4(1), 65.
- Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2011). The adoption of online shopping assistants: Perceived similarity as an antecedent to evaluative beliefs. *Journal of the association for Information Systems*, 12(5), 2.
- Alarcon, G. M., Gibson, A. M., Jessup, S. A., & Capiola, A. (2021). Exploring the differential effects of trust violations in human-human and human-robot interactions. *Applied ergonomics*, 93, 103350.
- Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & SOCIETY*, 21, 217-230.
- Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social-or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, 90, 315-330.
- Chen, Q., & Qin, J. (2021). Research and implementation of movie recommendation system based on deep learning. 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI),
- Elias, T., Rahman, U. S., & Ahamed, K. A. (2022). Movie Recommendation Based on Mood Detection using Deep Learning Approach. 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT),
- Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. Internet Science: 5th International Conference, INSCI 2018, St. Petersburg, Russia, October 24–26, 2018, Proceedings 5,
- García, S., Fernández, A., Luengo, J., & Herrera, F. (2009). A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. *Soft Computing*, 13, 959-977.
- Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega*, 32(6), 407-424.
- Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4), 712-733.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407-434.
- Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3), 261-273.
- Jo, W., Lim, Y., & Park, K.-H. (2019). Deep learning based land cover classification using convolutional neural network-a case study of korea. *Journal of the Korean geographical society*, 54(1), 1-16.
- Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology*, 6, 390.
- 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.
- Kim, S., Chen, R. P., & Zhang, K. (2016). Anthropomorphized helpers undermine autonomy and enjoyment in computer games. *Journal of Consumer Research*, 43(2), 282-302.
- Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the association for Information Systems*, 16(10), 1.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lee, O.-K. D., Ayyagari, R., Nasirian, F., & Ahmadian, M. (2021). Role of interaction quality and trust in use of AI-based voice-assistant systems. *Journal of Systems and Information Technology*, 23(2), 154-170.
- Liu, J., Choi, W.-H., & Liu, J. (2021). Personalized movie recommendation method based on deep learning. *Mathematical Problems in Engineering*, 2021, 1-12.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, 20(3), 709-734.

- McDowell, M. (2020). A top Silicon Valley futurist on how AI, AR and VR will shape fashion's future. <https://www.voguebusiness.com/technology/a-i-ar-and-vr-shaping-fashions-future-peter-diamandis>.
- Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., & Palmatier, R. W. (2022). An emerging theory of avatar marketing. *Journal of marketing*, 86(1), 67-90.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98-100.
- Mu, Y., & Wu, Y. (2023). Multimodal Movie Recommendation System Using Deep Learning. *Mathematics*, 11(4), 895.
- Ozdemir, M. a. A. B. A. (2022). How to Handle Someone Who Is in a Bad Mood. <https://www.wikihow.com/Handle-Someone-Who-Is-in-a-Bad-Mood>
- Qin, Z., & Zhang, M. (2021). Towards a personalized movie recommendation system: A deep learning approach. 2021 2nd International Conference on Artificial Intelligence and Information Systems,
- Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, 10, 236605.
- Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58.
- Savela, N., Oksanen, A., Pellert, M., & Garcia, D. (2021). Emotional reactions to robot colleagues in a role-playing experiment. *International Journal of Information Management*, 60, 102361.
- Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker Jr, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems*, 114, 94-102.
- Seymour, M., Yuan, L., Dennis, A., & Riemer, K. (2022). Face it, users don't care: Affinity and trustworthiness of imperfect digital humans. Proceedings of the 55th Hawaii International Conference on System Sciences,
- Seymour, M., Yuan, L. I., Dennis, A., & Riemer, K. (2021). Have we crossed the uncanny valley? Understanding affinity, trustworthiness, and preference for realistic digital humans in immersive environments. *Journal of the association for Information Systems*, 22(3), 9.
- Silva, E. S., & Bonetti, F. (2021). Digital humans in fashion: Will consumers interact? *Journal of Retailing and Consumer Services*, 60, 102430.
- Terry, Q. (2018). A primer on digital humans. <https://medium.com/s/story/everything-you-need-to-know-about-digital-humans-aaa4c73b7b04>
- Vaid, S., Kalantar, R., & Bhandari, M. (2020). Deep learning COVID-19 detection bias: accuracy through artificial intelligence. *International Orthopaedics*, 44, 1539-1542.
- Van der Heijden, H., Verhagen, T., & Creemers, M. (2003). Understanding online purchase intentions: contributions from technology and trust perspectives. *European journal of information systems*, 12(1), 41-48.
- Verhagen, T., Van Nes, J., Feldberg, F., & Van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, 19(3), 529-545.
- Wang, L. C., Baker, J., Wagner, J. A., & Wakefield, K. (2007). Can a retail web site be social? *Journal of marketing*, 71(3), 143-157.
- Wang, S., Lilienfeld, S. O., & Roachat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4), 393-407.
- Watson, R. (2022). The virtual economy of the metaverse: Computer vision and deep learning algorithms, customer engagement tools, and behavioral predictive analytics. *Linguistic and Philosophical Investigations*(21), 41-56.
- Yoon, S.-J. (2002). The antecedents and consequences of trust in online-purchase decisions. *Journal of interactive marketing*, 16(2), 47-63.
- Zajonc, R. (1968). Attitudinal effects of mere exposure. *journal of Personality and Social Psychology*, 9. *Monograph Supplement*.
- Zauskova, A., Miklencicova, R., & Popescu, G. H. (2022). Visual imagery and geospatial mapping tools, virtual simulation algorithms, and deep learning-based sensing technologies in the metaverse interactive environment. *Review of Contemporary Philosophy*, 21, 122-137.