

Chapter

Digital Partnerships: Understanding Delegation and Interaction with Virtual Agents

Ningyuan Sun and Jean Botev

Abstract

With recent advances in artificial intelligence and the metaverse, virtual agents have become increasingly autonomous and accessible. Due to their growing technological capabilities, interaction with virtual agents gradually evolves from a traditional user-tool relationship to one resembling interpersonal delegation, where users entrust virtual agents to perform specific tasks independently on their behalf. Delegating to virtual agents is beneficial in numerous ways, especially regarding convenience and efficiency. Still, it poses problems and challenges that may drastically harm users in critical situations. This chapter explores the trust and delegation relationships between users and virtual agents, introducing a trust-based conceptual model to abstract and differentiate users' delegation decisions based on three major dimensions covering the impact of rationality, affection, and technology. Practical guidance for virtual agent designs and potential applications of the model for metaverse development are also presented, followed by an outlook and an overview of future research opportunities.

Keywords: delegation, interaction, trust, virtual agents, metaverse

1. Introduction

Over the last few decades, significant progress has been made in the technologies that underpin virtual environments. For example, computing devices have become exponentially more powerful and capable of highly realistic real-time image rendering. Newer telecommunications technologies like 5G can transmit data at a much greater scale and speed, enabling shared experiences with more users and lower latency. Immersive media devices are also evolving rapidly, allowing users to experience virtual environments more naturally and immersively than is possible with traditional display technologies. With these technological advances, building a metaverse is becoming increasingly realistic and has become the subject of intense debate [1].

In contrast to virtual environments that isolate users from physical reality, the current metaverse vision comprises a collection of 3D digital worlds closely connected to

the real world [2]. It aims to merge the virtual and the real in a blended space so that interpersonal activities (e.g., meetings, teaching, socializing) can be facilitated and complemented by digital elements.

While interactions between users are a significant focus, other interactions with digital entities (e.g., with non-player characters in video games) are also essential to the metaverse experience. These entities are typically controlled by scripts and are often referred to as *software agents* due to their autonomy. Traditionally, interaction with a software agent occurs via a panel-like or 3D user interface following the WIMP¹ paradigm, which represents and operates an information system with abstract widgets such as icons and buttons. Nevertheless, there are increasing efforts to integrate human communication channels—including gestures, facial expressions, natural language, and more—to make the interaction feel natural and personable [3, 4]. In research, these human-like agents in digital environments are more specifically called *virtual agents*.

The vision of a metaverse as an embodied virtual environment where users are represented with avatars is one of the main driving forces behind the burgeoning research and development of virtual agents. Using avatars can enhance physical and social presence, promoting an immersive and sociable environment [5, 6]. Consequently, software agents are also increasingly represented by avatars and, using artificial intelligence (AI) technologies, can already perform complex tasks with minimal human supervision. For example, large language models (LLMs) allow users to describe a task without specifying how it can be accomplished. This type of interaction is similar to interpersonal delegation, where a *principal* (i.e., the person delegating) authorizes an agent to act independently on the principal's behalf. As a result, the need for fine-grained control is decreasing, while natural human communication is becoming more critical.

In the metaverse, delegation-like interactions with virtual agents that embody a virtual character, communicate using natural language, and perform tasks set by the user are likely to predominate. Delegating tasks to virtual agents in such a manner benefits users in several ways, particularly in terms of efficiency and convenience. Users can easily transfer their tasks to virtual agents without having to tender strategies and initiatives. On the other hand, delegation inherently involves risks as users must transfer part or all of their authority to virtual agents, making themselves vulnerable to the agents' actions. These risks can be particularly pronounced in a metaverse scenario, where virtual agents may influence users' delegation decisions with their advanced communication abilities [7, 8]. Therefore, it is essential to investigate this delegation relationship and identify its underlying factors, especially for critical tasks with potentially far-reaching, real-world implications, to facilitate user-agent interaction and resolve potential issues.

This chapter explores the trust and delegation relationships between users and virtual agents. Following a discussion of delegation and agency relationships, a trust-based conceptual model is introduced to explain users' trust in and delegation to virtual agents. Taking a macroscopic perspective, the model abstracts and differentiates three major dimensions covering the impact of rationality, affection, and technology. Practical guidance for virtual agent designs is further derived from the model, and potential applications of the model for metaverse development are presented with an outlook and overview of related research opportunities.

¹ WIMP stands for “windows, icons, menus, and pointer”.

2. Delegation

Delegation is one of human societies' most important and common interpersonal relationships. The operation of most organizations requires effective management, i.e., delegation from managers to subordinates, and successful collaboration, i.e., delegation among specialized members or departments within an organization. With its significance and prevalence, delegation has been extensively studied in many disciplines, particularly economics, politics, and management [9]. Recent years have also seen the emergence of a new research line on delegation to intelligent artifacts, such as robots and software agents [10].

2.1 Interpersonal delegation

Delegation is a dyadic relationship between a principal and an agent, where the principal lets the agent carry out specific tasks on the principal's behalf while remaining responsible or accountable for the task outcomes [11, 12]. Delegation can provide principals with several benefits, for example, allowing them to access and utilize agents' expertise and knowledge [13] or increase efficiency by shifting some of their workloads to agents [14]. Delegation can also be used strategically as a commitment device to express principals' determination in a negotiation [15]. Conversely, principals must bear the risk of transferring part or all of their authority to agents during delegation. Such a risk can be even higher if there is a lack of effective measures to monitor and regulate agents' actions [16].

Among the research on delegation, *agency theory* has been one of the most cited and fundamental contributions [16–18]. It originates from early studies on risk-sharing problems within an organization whose members hold different attitudes toward risk [19]. Later, these studies evolved into a more inclusive theory, i.e., agency theory, mainly addressing the agency relationship between two rational entities. The theory posits that delegation is susceptible to the so-called *agency problems*, which often arise in two situations: (a) when it is difficult for principals to verify what agents are actually doing and (b) when principals and agents have conflicting goals or different attitudes toward risks [16]. These problems are detrimental to delegation, as exemplified by the so-called *moral hazard*, a situation where agents do not act as agreed or even furtively impair principals' interests [17]. To mitigate the influences of these problems, agency theory proposes to formulate some pre-defined protocols—i.e., a *contract*—to regulate the interaction between the principal and the agent during delegation [16, 18]. There are mainly two types of contracts. With *behavior-oriented* contracts, agents receive a fixed amount of money or equivalent upon finishing the tasks delegated to them, regardless of the task outcomes. *Outcome-oriented* contracts give agents minimal remuneration for finishing the tasks, but they can take a share of the task outcomes (e.g., bonus, commission, stocks). The two types of contracts fit different situations depending on several factors, including agent observability [20, 21], task programmability [22, 23], the measurability of task outcomes [24], and more. For example, in programmable tasks, principals are more likely to employ behavior-oriented contracts because agents' performance in such tasks is relatively easy to observe and assess [16].

Much of the literature on delegation is related to agency theory, viewing the principal-agent relationship mainly as an economic issue [25–27]. Nevertheless, many researchers assume a psychological stance and focus on principals' delegation decision-making process. For example, eight factors were found underlying a manager's

decisions on delegating a task to subordinates, including the manager's workload, the task's importance, and subordinates' age, gender, trustworthiness, performance, job capability, and job tenure [11]. This set of factors was later validated in broader contexts and also expanded to include subordinates' experience in managing, the goal congruence between managers and subordinates, and whether there is a favorable exchange relationship [14, 28]. In some cases, decisions on delegation are dominated by subjective causes. For example, people may delegate a task to others to avoid feeling responsible or blamed for negative task outcomes [29, 30]. A manager may refuse to delegate to appear "busy" [31]. Overall, these psychology-oriented studies suggest that delegation is a challenging and prone-to-failure task for principals. Managers of all levels, from leaders of small groups to CEOs of successful corporations, are not free from making sub-optimal or wrong decisions on delegation [31]. People generally prefer making decisions themselves over delegating them to others, even if retaining control would lose potential benefits and incur additional costs [32].

2.2 Delegation to software agents

Although the term "delegation" has been predominantly used and studied in interpersonal contexts, recently it is increasingly assumed by computer scientists to describe the relationship between human users and software agents.

The interaction between human users and software agents is mainly characterized by the latter being designed to support and streamline human tasks. As discussed in **Section 1**, this relationship evolves with increasingly intuitive interfaces and advanced computational capabilities. Software agents leverage AI, natural language processing, and machine learning to interpret and respond to user inputs naturally and efficiently. They aim to improve productivity and user experience by automating routine tasks, providing personalized recommendations, and facilitating decision-making processes. As these technologies advance, the boundary between human and machine interaction and collaboration becomes increasingly seamless, fostering an environment where users can focus on more complex endeavors instead of devising lower-level strategies and initiatives.

Most of the research on software agents today originates from the well-established literature on human-automation interactions, which mainly takes the perspectives of use [33, 34] and reliance [35, 36]. As software agents become increasingly autonomous and capable, a different perspective based on the notion of delegation recently received more attention, as reflected in the quote below.

"[...] the delegation lens will yield more relevant and nuanced insights regarding human agent and agentic IS² artifact relationships, and this lens will be increasingly needed as the agentic capabilities of IS artifacts increase." [10]

The pioneering works on delegation to software agents can be dated back to the 1990s, during which there were initial concerns about the "dramatic change" of software user interface from tool-oriented to delegation-oriented ones [12]. As a result of this rapid change, problems common to interpersonal delegation might similarly occur during interaction with software agents, which imposed new challenges on user interface designs. To overcome these drawbacks, five major dimensions must be

² IS = information systems.

considered when designing delegation-oriented user interfaces, including trust, communication, performance control, user demographics, and cost-benefit analysis [12].

There was also an early theoretical work that formalized delegatory relationships within a multi-agent system [37]. Despite its focus on delegation between software agents, the theory still lends us a unique perspective to explaining users' delegation decisions. Differing from other approaches, the theory defines delegation as a state of the principal, where "an agent A needs or likes an action of another agent B and includes it in its own plan" [37]. This definition moderates the aspect of responsibility and instead views delegation more of the principal's expectation of the agent. With this definition, delegation can be further divided into three subtypes (cf. **Table 1**) according to the following two criteria: whether there is an agreement between a principal and an agent and whether the principal actively induces certain behaviors in the agent.

Following these early studies, research on delegation to software agents was limited only until recent years, during which this topic received more attention due to the development of AI. A frequently discussed topic is whether people prefer human agents or software agents. The evidence is mixed; some studies found that software agents are preferred [38–40], whereas others showed the opposite [41]. Many factors can influence this preference. For example, people may prefer letting software agents carry out tasks involving sensitive data (e.g., credit card information) due to their user-centered design [42, 43] and limited intentional capacity [43, 44], whereas human agents may exploit the sensitive data for their own interests. On the other hand, tasks involving moral decisions (e.g., life-and-death decisions in law, military, or medicine) are less likely delegated to software agents given their lack of empathy [45]. The term *delegability* was recently conceptualized to describe people's general preference for delegating a task to AI [46]. In a survey investigating the delegability of 100 different tasks, it was found that the ones with the highest and lowest levels of delegability were "moving & packing merchandise in a warehouse for shipping to customers" and "picking out and buying a birthday present for an acquaintance", respectively [46].

Another often-mentioned topic is the factors governing users' decisions on delegation to software agents. Research shows that some factors have similar influences on delegation to human and software agents. To give a few examples, both interpersonal and human-software delegation were found to be positively correlated with perceived controllability [47], perceived attachment [48], and agents' trustworthiness [46, 47]. Among them, perceived controllability seems to be particularly crucial; people are more likely to delegate a decision to an algorithm when they are allowed to modify the decision made by the algorithm, even if the modification is severely restricted [49]. On the other hand, certain factors may exert different impacts on delegation depending on whether the agent is a human or a software application. For instance,

Delegation type	Agreement exists?	Behavior inducing?
Weak delegation	No	No
Mild delegation	No	Yes
Strict delegation	Yes	Yes or No

Table 1.
The delegation classification in [37].

high-level task accountability can encourage delegation to software agents but inhibit delegation to human agents [47].

Several studies took an economic perspective and focused on the user-agent dyad [40, 50–52]. An interesting implication from these studies is that software agents may be more proficient at delegation than humans are. In certain tasks, a hybrid team of a human user and a software agent can perform best if the software agent assumes the leading role and delegates tasks to the human user [40]. In contrast, the team may perform less when the human user is the leader or when any party has full control [40].

3. A trust-based model of delegation to virtual agents

Trust is a critical aspect underlying most interpersonal relationships, whether between individuals or within an organization. Owing to its significance and omnipresence, trust has been the subject of extensive study in many disciplines, including economics, psychology, and sociology. While the majority of these studies focus on interpersonal trust, a large body of research on the trust relationship between users and intelligent artifacts such as robots, automation, and software agents has emerged in recent decades. Numerous factors were identified and demonstrated to impact users' trust in these artifacts, constituting a vast and complicated parameter space.

Delegation to virtual agents is potentially governed by a similar parameter space, which, unlike that of trust, remains largely unexplored due to limited research. Nevertheless, given the connection between trust and delegation [11, 28, 46, 47, 53], factors underlying trust in virtual agents may have similar influences on delegation to virtual agents, either directly or through the mediation of trust. To approach and explore these factors, we introduce a trust-based conceptual model of delegation to virtual agents. The model considers various factors and explains how they collectively shape users' trust in and delegation to virtual agents.

3.1 Definitions of trust

The definition of trust has been a contentious topic and varies across disciplines. Some researchers approach it from a sociopsychological perspective, considering trust as a major social relation among individuals [54]. Others take a more psychological stance and view it mainly as the product of affective states [55, 56]. There are also theoretical models quantifying trust, primarily in economics and computer science [57].

“The definition of trust [...] is 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.” [58]

This definition regards trust as an individual's willingness to be vulnerable to others' actions. This basic idea was later broadly accepted in trust research [59, 60]. By the beginning of the millennium, the publication from which this definition originates had already been cited “far more frequently than others on the topic of trust” [61]. Today, it remains the most influential one among different definitions of trust.

“Trust: confidence that [one] will find what is desired rather than what is feared.” [62]

This definition argues that trust is rooted in the cost-benefit analysis of future events. A rational individual always desires and thus pursues benefits, but fears and avoids costs. Trust arises when an individual “is confronted with an ambiguous path, a path that can lead to an event perceived to be beneficial or to an event perceived to be harmful” [63]. When trusting, an individual believes that taking this path will be beneficial.

“Interpersonal trust is defined here as an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon.” [64]

From a sociological perspective, trust can be seen as an expectancy that others will do what they state or promise. Social interactions with others (like parents, teachers, and peers) provide an individual with rich feedback to validate whether their expectancy is accurate. With accumulating feedback, an individual develops a *generalized expectancy* about the extent to which other people can be relied upon. Some individuals may have been surrounded by honest persons and believe that people are trustworthy in general. Others may have experienced much betrayal, which makes them less trusting. This generalized expectancy changes slowly in the long term and becomes a personal trait [64].

“Trust can be defined as the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability. [...] an agent can be automation or another person that actively interacts with the environment on behalf of the person.” [61]

The notion of *trust in automation* has been gaining relevance in trust research as automated systems are extensively used today. There are several definitions of trust in automation. For example, in the one quoted above, the trustee is considered an agent that “can be automation or another person”. The use of the term “agent” suggests that the trustee’s agency, rather than their identity, contributes to the trustor’s perception of uncertainty and vulnerability.

3.2 Trust in Virtual Agents

As discussed above, trust has been defined in diverse ways, encompassing concepts such as willingness, attitude, confidence, and expectancy. The different definitions share a common view that trust is a mental construct, which can be modeled as an information-processing procedure [65]. As **Figure 1** illustrates, the processor functions like a black box, perceiving information from the external world as input and generating mental constructs as output. Based on this model, trust in a virtual agent is essentially the product of users’ perception and processing of information related to their interaction with the agent.

The processor determines how perceived information is processed into trust. Each processor represents an individual user and may respond to perceived information in a unique way. Consequently, the same information may result in different levels of trust depending on the processor, which makes it challenging to predict trust in virtual agents at an individual level. Nevertheless, there are still some patterns in the

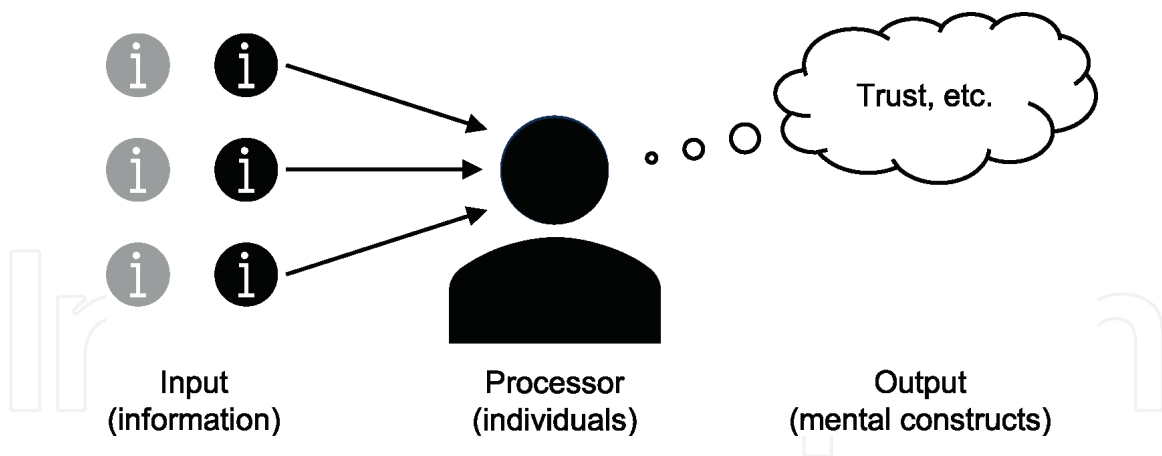


Figure 1. Trust in virtual agents as information processing. The gray and black information icons represent information overlooked and perceived by the processor, respectively.

trusting attitudes among demographically different groups. For instance, females tend to put a higher level of trust in virtual agents than males [66]. Older people consider a virtual agent more trustworthy than young people when the agent uses non-verbal behavior to communicate emotions [67].

On the other hand, the same processor may produce different levels of trust if perceived information differs. The literature has documented various pieces of information that, when perceived by users, can impact their trust in virtual agents. These can be generally classified into two categories—*analytical information* and *affective information*—based on a commonly used dual division of trust into cognitive and affective aspects [68, 69]. Analytical information (e.g., the probability of whether an agent can achieve a task) forms the basis for users to make rational assumptions about the trustworthiness of a virtual agent. These assumptions give users “good reasons” to trust or distrust another [68] and are widely regarded as a fundamental component of trust [58, 70, 71]. Affective information influences trust in virtual agents mainly through psychological and social channels related to emotions, feelings, or stereotypes, etiquette, etc. For example, the Uncanny Valley effect can impair a virtual agent’s trustworthiness by inducing a sense of eeriness [72].

The two categories (i.e., analytical and affective information) are not mutually exclusive. Certain information can influence trust in both rational and affective ways. For instance, task urgency is not only an important factor to be considered in rational thinking but also a stressor that may bias users’ trust [73].

The list below provides several examples of analytical and affective information. Many of them are drawn from research on trust in virtual agents, and the remaining ones come from studies on trust in automated systems or software agents, to which virtual agents also belong. To facilitate the discussion, within the list, we use the term “agentic system(s)” or simply “system(s)” to generally refer to an artificial entity with some degree of autonomy for carrying out users’ tasks. Such an entity can be an automated system, a software agent, or a virtual agent.

3.2.1 Analytical information examples

- *Reliability.* The reliability of an agentic system is an influential factor in its trustworthiness [34, 61, 71]. Users tend to trust reliable systems, whereas signals indicating low reliability, such as errors or task failures, are generally detrimental

to their trustworthiness [74, 75]. Notably, users are particularly attentive to mistakes and errors due to the stereotype that machines can perform tasks perfectly [76].

- *Predictability*. Predictability describes the extent to which an agentic system's behavior, whether reliable or not, is predictable to its users. Predictability is considered a fundamental factor of both interpersonal trust [58] and trust in software agents [77].
- *Benevolence*. In certain critical scenarios, agentic systems may appear more trustworthy than human agents owing to their user-centered design. For example, people are more likely to reveal their credit card information to a software agent than to a human agent [43]. Similarly, in massive multiplayer online games, players often prefer to trade valuable virtual items through a non-player character escrow rather than a third player [78].
- *Transparency*. An agentic system is considered more trustworthy if its algorithms are transparent [79, 80]. Transparency is particularly relevant nowadays as algorithms have grown increasingly complex.

3.2.2 Affective information examples

- *Anthropomorphism*. Agentic systems are more likely to be rated as trustworthy when they exhibit anthropomorphic features [81, 82]. For example, virtual agents with human-like visual representations are perceived as more trustworthy than those embodied in non-human characters [7, 83]. The same applies to the auditory channel: users tend to put a higher level of trust in agents dubbed with a natural human voice than a low-quality synthesized voice [84].
- *Similarity*. Individuals tend to hold favorable opinions of those who share similarities with them [85] and, consequently, are more likely to trust those similar to others [86, 87]. This similarity effect exists not only interpersonally but also between users and agentic systems. A user is more likely to trust a virtual agent if its face resembles the user's face [88] or if it mimics the user's body movements [89].
- *Politeness*. When communicating with users, agentic systems that conform to social etiquette can appear more trustworthy than systems disregarding it. Users are more likely to trust an automated system that communicates in a relatively non-interruptive manner, for example, postponing the notification of non-critical messages when users are focusing on important tasks [90].
- *Attractiveness*. Humans intrinsically tend to label physically attractive individuals with more favorable characteristics, such as being more trustworthy [91], compared to average individuals. Similarly, virtual agents with human-like visual representation are also considered more trustworthy if their face looks attractive [92].

From a macroscopic perspective, there are mainly three dimensions governing virtual agents' trustworthiness in this information-processing model, including an

analytical dimension, an *affective dimension*, and a *technological dimension*. The analytical dimension impacts trust in virtual agents by being directly involved in rational thinking, while the affective dimension biases user trust through psychological and social channels. During interactions with virtual agents, users act as an information processor that continuously perceives and evaluates information from the interaction. The perceived information, in turn, shapes users' trust in virtual agents via the analytical and affective dimensions, as **Figure 2** illustrates. Some information exerts its influences mainly through a single dimension, whereas others can have a significant impact through both dimensions.

The technological dimension accounts for the indirect influence of technologies underlying virtual agents. The perception of a virtual agent may vary depending on the media device used (e.g., desktops vs. head-mounted displays), which in turn may lead to different levels of trust according to the model. Furthermore, technologies underlying virtual agents are highly relevant to data security. Trust in a virtual agent may decrease if the agent runs on an insecure infrastructure threatening personal data and privacy. This can be crucial when experiencing the Metaverse on immersive media devices whose embedded sensors are usually more invasive than those in laptops or smartphones [93].

3.3 From trust to delegation

Trust and delegation are intrinsically related concepts. Theoretically, delegation entails two essential components of trust defined in [58], including *uncertainty* (agents' actions and the ensuing outcomes are neither entirely predictable nor completely unknown) and *vulnerability* (users are accountable or responsible for

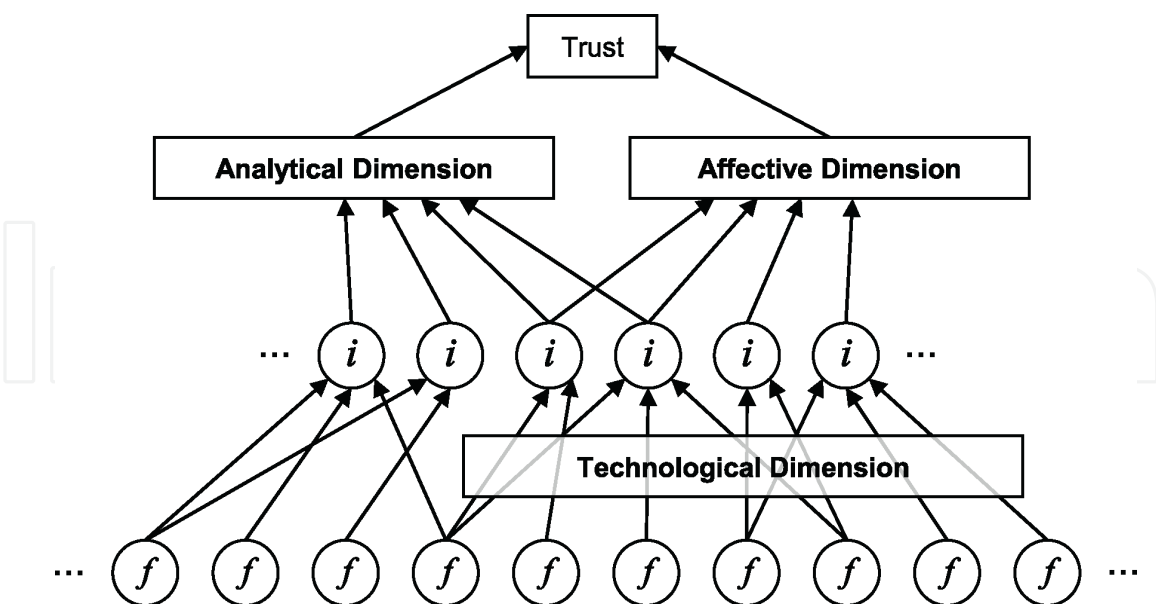


Figure 2.

A conceptual model of trust in virtual agents. The circled *f* in the figure bottom represents facts about interaction with virtual agents. These facts are perceived by users and become information (illustrated as circled *i*). A piece of information may originate from several facts; for example, an agent's capability of financial investment may result from an appraisal of its trading histories in different markets. The bulk of information (represented as the four circled *i* on the right) is perceived through media devices and thus subject to the technological dimension. The remaining information (represented as the two circled *i* on the left) is perceived via other channels, such as conversations with other people about the agent's reputation. Perceived information, in turn, influences trust in virtual agents through the analytical and affective dimensions.

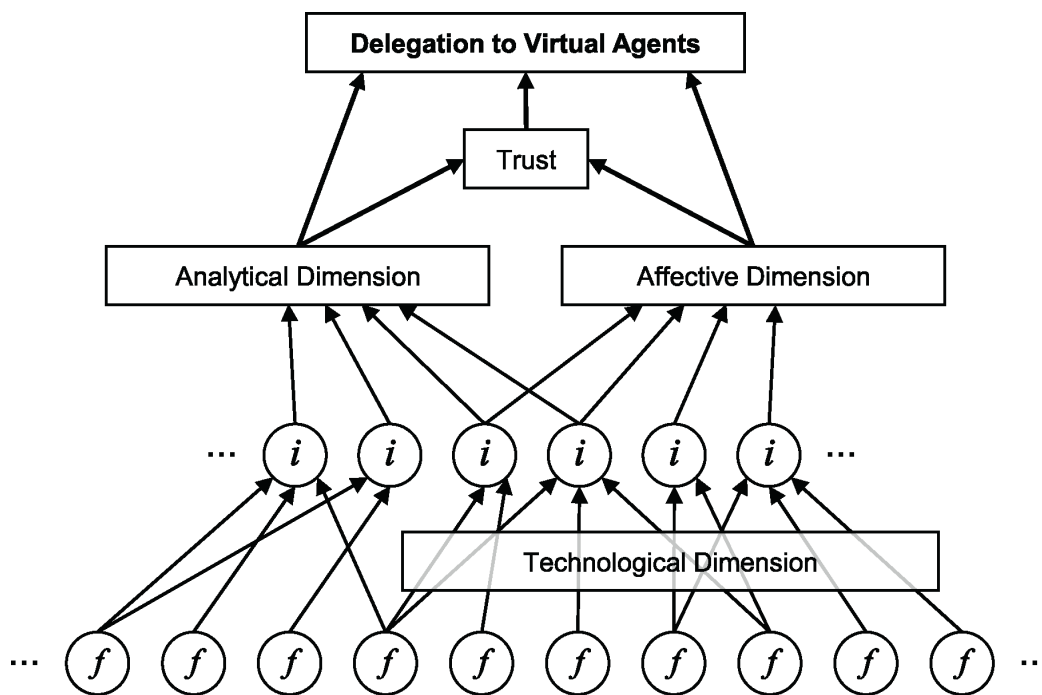


Figure 3.
 The conceptual model extended from trust to delegation.

task outcomes). There is also empirical evidence substantiating the connection. For example, trust was found to be correlated with the level of *decentralization*, i.e., the extent to which top managers of an organization empower subordinates to make decisions as opposed to micro-managing. Research shows that the more trust there is within an organization, the more decentralized the organization is, and the more delegation therein [94]. As noted in [95], “a high-trust society can organize its workplace on a more flexible and group-oriented basis, with more responsibility delegated to lower levels of the organization. Low-trust societies, by contrast, must fence in and isolate their workers with a series of bureaucratic rules”. Similar results can also be found in psychological studies, where trust in subordinates constitutes a vital factor behind managers’ decisions on delegation [11, 28]. In the context of human-computer interaction, trust also plays an important role in delegation to artificial agents [46]. For instance, trust was found to be correlated with students’ willingness to let a software agent arrange travels for their job interviews [47].

With the evidence mentioned above, the conceptual model can be further extended as **Figure 3** illustrates, where the three dimensions also influence delegation to virtual agents directly or through the mediation of trust. The extended model provides a theoretical foundation to systematically view and explore factors potentially governing delegation to virtual agents.

4. Discussion

The conceptual model offers a theoretical foundation explaining how different factors collectively shape users’ decisions regarding delegation to virtual agents. Based on this, further practical insights can be derived to facilitate the design of virtual agents for delegation and explore future metaverse-related research directions.

4.1 Design guidance

The model provides a systematic approach to designing virtual agents for delegation. As previously discussed, the three dimensions decompose a delegation decision into relatively individual aspects of rationality, affection, and technology. Each dimension includes a set of factors that can be modulated through agent design. For example, to increase the impact of the affective dimension, developers may consider increasing the agents' anthropomorphism, which can be achieved by changing their visual representation to more human-like characters. When designing virtual agents for delegation, developers can focus on the most influential dimensions and adjust the factors associated with these dimensions.

The importance of each dimension varies depending on the nature of the delegated tasks. For instance, the analytical dimension is imperative in performance-critical tasks. Our previous studies investigated the delegation of financial investment and showed that analytical information plays a decisive role [8, 53]. Affective aspects like rapport and human likeness are somewhat influential in this context [7, 96], whereas the technological dimension only has a limited impact [97].

These results suggest that, in performance-critical scenarios, virtual agents are still primarily seen as tools rather than social actors. Delegation to virtual agents therefore remains predominantly a matter of cost-benefit analysis, differing from interpersonal delegation where the social connection between principals and agents also weighs [31]. Thus, developers may only need to focus on performance-related aspects when designing virtual agents for critical tasks. A limited number of affective cues (e.g., facial expression, gesture) can make the interaction more natural and personable. As exemplified by LLMs such as ChatGPT, their apparent near-human performance makes many users willing—sometimes unquestioningly—to delegate critical tasks (e.g., thesis writing, exam essays, medical consultation) to them despite their textual interfaces.

On a different note, it is good practice to consider trust when designing virtual agents for delegation. As illustrated in **Figure 3**, trust constitutes a major factor of delegation, and studies show that they are positively correlated [11, 28, 46, 47, 53]. A lack of trust in virtual agents may likely lead to users' reluctance to delegate. Thus, developers should avoid trust-impairing designs and consider incorporating an adequate number of trust-building elements in virtual agents. There is a substantial body of research on trust in intelligent artifacts [80], which provides more specific guidelines that developers may refer to for agent designs.

4.2 Challenges and open issues

Users are more susceptible to manipulation in an agency relationship with virtual agents than in the traditional user-tool relationship with software programs. Developers may use exploitative features or algorithms in virtual agents to entice users to behave in certain ways for the sake of profit. For example, a virtual agent for financial investments may embody a trustworthy human character and use rapport-building body language to gain users' trust and collect capital. This problem is exacerbated by the fact that when users delegate, they give up some or all of their authority, making them vulnerable to virtual agents and the developers behind them. In addition, virtual agents are increasingly controlled by LLMs whose decisions are less predictable and controllable than those of non-AI programs. It remains debatable whether these models have a sufficient level of empathy and integrity when

communicating with users to remain neutral and not influence users to delegate unethically.

Legislation plays a central role in mitigating the above-mentioned ethical concerns. As users increasingly interact with virtual agents in the metaverse and delegate critical tasks with real-world consequences, appropriate regulations should be implemented to penalize manipulation. On the other hand, the regulations could create a legal infrastructure to promote responsibility and risk sharing between virtual agent users and developers. This not only encourages developers to improve the quality of their services but also prevents unethical behavior by aligning the interests of both parties.

Navigating the metaverse and managing engagement with virtual agents is also a major cross-cultural challenge, as different cultural norms and values can clash in interactions. Specific forms of representation, different social etiquette, language barriers, and different levels of digital literacy all impact trust and delegation and can hinder inclusion if they are not taken into account.

4.3 Future research

With virtual agents becoming more capable and autonomous, interactions with them increasingly resemble interpersonal delegation. One of today's most debated topics is identifying influential factors in users' delegation decisions. Future studies can continue this line of research by exploring new factors or re-evaluating known factors in different contexts. The results will provide further insights into users' decision-making process, thereby expanding the theoretical foundation to address the above-mentioned challenges (cf. **Section 4.2**) and facilitating user-agent interactions in a metaverse context. For example, it might be interesting to investigate the delegation of relatively performance-uncritical tasks. Influential factors in critical tasks, such as performance and capability, may degrade to insignificant ones, whereas the impact of the affective and technological dimensions may become more relevant.

From a sociological perspective, the metaverse as a concept is gradually evolving from an inter-user platform to a hybrid digital society of humans and AI, where delegation occurs arbitrarily among a massive number of users and virtual agents. The underlying dynamics at this macroscopic level are complicated and largely unexplored, but merit further investigation for this hybrid society's overall efficiency and welfare. Thus, researchers may consider using multi-agent and multi-task scenarios for future studies to confirm the findings obtained from minimal settings (e.g., one user delegates a single task to a virtual agent). The results will lend us more sociological and economic perspectives on the future of virtual agents and the metaverse, in addition to the psychological stance that many studies today have assumed.

5. Conclusion

As technology advances rapidly, the digital sphere, including the metaverse, will become increasingly populated with AI-powered, intelligent virtual agents. With their enhanced autonomy and capability, these agents are more and more assuming the role of delegate in addition to their traditional role as mere tools. This agency relationship has already found wide acceptance in recent years due to the emergence and adoption of generative AI applications and LLMs like ChatGPT. However, the relevant research can hardly keep pace with technological developments and is still limited today.

Thus, this chapter introduced a trust-based conceptual model to explain how different factors collectively influence and shape users' decisions on delegation to virtual agents. The model distinguishes three dimensions, each related to a set of factors that uniquely impact trust and delegation decisions. Practical implications for virtual agent designs and potential applications of the model for metaverse development were discussed, together with future research opportunities.

Acknowledgements


This research was funded by the Luxembourg National Research Fund (FNR) under Grant 12635165.

Author details

Ningyuan Sun and Jean Botev*
University of Luxembourg, Esch-sur-Alzette, Luxembourg

*Address all correspondence to: jean.botev@uni.lu

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Ning H, Wang H, Lin Y, Wang W, Dhelim S, Farha F, et al. A survey on the metaverse: The state-of-the-art, technologies, applications, and challenges. *IEEE Internet of Things Journal*. 2023;**10**(16):14671-14688
- [2] Weinberger M. What is metaverse? – A definition based on qualitative meta-synthesis. *Future Internet*. 2022;**14**(11):310
- [3] Lugrin B, Pelachaud C, Traum D, editors. *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*. 1st ed. Vol. 1. New York, NY, United States: Association for Computing Machinery; 2021
- [4] Lugrin B, Pelachaud C, Traum D, editors. *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*. 1st ed. Vol. 2. New York, NY, United States: Association for Computing Machinery; 2022
- [5] McDonnell R, Mutlu B. Appearance. In: *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics: Methods, Behavior, Cognition*. 1st ed. Vol. 1. New York, NY, United States: Association for Computing Machinery; 2021. pp. 105-146
- [6] Kim J. Advertising in the metaverse: Research agenda. *Journal of Interactive Advertising*. 2021;**21**(3):141-144
- [7] Sun N, Botev J. Virtual agent representation for critical transactions. In: *Proceedings of the 13th International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE)*. New York, NY, United States: Association for Computing Machinery; 2021. pp. 25-29
- [8] Sun N, Botev J. Why do we delegate to intelligent virtual agents? Influencing factors on delegation decisions. In: *Proceedings of the Ninth International Conference on Human-Agent Interaction (HAI)*. New York, NY, United States: Association for Computing Machinery; 2021. pp. 386-390
- [9] Lupia A. Delegation of power: Agency theory. In: *International Encyclopedia of the Social and Behavioral Sciences*. San Mateo, California, United States: Pergamon; 2001. pp. 3375-3377
- [10] Baird A, Maruping LM. The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*. 2021;**45**(1):315-341
- [11] Leana CR. Power relinquishment versus power sharing: Theoretical clarification and empirical comparison of delegation and participation. *Journal of Applied Psychology*. 1987;**72**(2):228-233
- [12] Milewski AE, Lewis SH. Delegating to software agents. *International Journal of Human-Computer Studies*. 1997;**46**(4):485-500
- [13] Jensen MC. Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review*. 1986;**76**(2):323-329
- [14] Yukl G, Fu PP. Determinants of delegation and consultation by managers.

Journal of Organizational Behavior. 1999;**20**(2):219-232

[15] Sengul M, Gimeno J, Dial J. Strategic delegation: A review, theoretical integration, and research agenda. *Journal of Management*. 2012;**38**(1):375-414

[16] Eisenhardt KM. Agency theory: An assessment and review. *Academy of Management Review*. 1989;**14**(1):57-74

[17] Shapiro SP. Agency theory. *Annual Review of Sociology*. 2005;**31**:263-284

[18] Jensen MC, Meckling WH. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*. 1976;**3**(4):305-360

[19] Wilson R. The theory of syndicates. *Econometrica*. 1968;**36**(1):119-132

[20] Fama EF. Agency problems and the theory of the firm. *Journal of Political Economy*. 1980;**88**(2):288-307

[21] Fama EF, Jensen MC. Separation of ownership and control. *The Journal of Law and Economics*. 1983;**26**(2):301-325

[22] Eisenhardt KM. Control: Organizational and economic approaches. *Management Science*. 1985;**31**(2):134-149

[23] Eisenhardt KM. Agency- and institutional-theory explanations: The case of retail sales compensation. *The Academy of Management Journal*. 1988;**31**(3):488-511

[24] Anderson E. The salesperson as outside agent or employee: A transaction cost analysis. *Marketing Science*. 1985;**4**(3):234-254

[25] Ross SA. The economic theory of agency: The Principal's problem. *The American Economic Review*. 1973;**63**(2):134-139

[26] Grossman SJ, Hart OD. An analysis of the principal-agent problem. In: Dionne G, Harrington SE, editors. *Foundations of Insurance Economics: Readings in Economics and Finance*. Netherlands: Springer; 1992. pp. 302-340

[27] Alonso R, Matouschek N. Relational delegation. *The RAND Journal of Economics*. 2007;**38**(4):1070-1089

[28] Aggarwal P, Mazumdar T. Decision delegation: A conceptualization and empirical investigation. *Psychology & Marketing*. 2008;**25**(1):71-93

[29] Steffel M, Williams EF, Perrmann-Graham J. Passing the buck: Delegating choices to others to avoid responsibility and blame. *Organizational Behavior and Human Decision Processes*. 2016;**135**:32-44

[30] Steffel M, Williams EF. Delegating decisions: Recruiting others to make choices we might regret. *Journal of Consumer Research*. 2018;**44**(5):1015-1032

[31] Jenks JM, Kelly JM. *Don't Do, Delegate!* London, United Kingdom: Franklin Watts; 1985

[32] Bobadilla-Suarez S, Sunstein CR, Sharot T. The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *Journal of Risk and Uncertainty*. 2017;**54**(3):187-202

[33] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. *MIS Quarterly*. 2003;**27**(3):425-478

[34] Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*. 1997;**39**(2):230-253

- [35] Dixon SR, Wickens CD. Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors*. 2006;**48**(3):474-486
- [36] Riley V. Operator reliance on automation: Theory and data. In: Parasuraman R, Mouloua M, editors. *Automation and Human Performance: Theory and Applications*. Boca Raton, Florida, United States: Taylor & Francis, CPC Press; 1996. pp. 19-35
- [37] Castelfranchi C, Falcone R. Towards a theory of delegation for agent-based systems. *Robotics and Autonomous Systems*. 1998;**24**(3-4):141-157
- [38] Candrian C, Scherer A. Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*. 2022;**134**:107308
- [39] Logg JM, Minson JA, Moore DA. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*. 2019;**151**:90-103
- [40] Fügner A, Grahl J, Gupta A, Ketter W. Cognitive challenges in human-artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research*. 2022;**33**(2):678-696
- [41] Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*. 2015;**144**(1):114-126
- [42] Fogg BJ. A behavior model for persuasive design. In: *Proceedings of the Fourth International Conference on Persuasive Technology (PERSUASIVE)*. New York, NY, United States: Association for Computing Machinery; 2009. pp. 1-7
- [43] Sundar SS, Kim J. Machine heuristic: When we trust computers more than humans with our personal information. In: *Proceedings of the 37th CHI Conference on Human Factors in Computing Systems*. New York, NY, United States: Association for Computing Machinery; 2019
- [44] Harbers M, Peeters MMM, Neerincx MA. Perceived autonomy of robots: Effects of appearance and context. In: *Proceedings of the 2015 International Conference on Robot Ethics (ICRE)*. Cham, Switzerland: Springer, Cham; 2017. pp. 19-33
- [45] Bigman YE, Gray K. People are averse to machines making moral decisions. *Cognition*. 2018;**181**:21-34
- [46] Lubars B, Tan C. Ask not what AI can do, but what AI should do: Towards a framework of task delegability. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*. Red Hook, NY, United States: Curran Associates Inc.; 2019. pp. 57-67
- [47] Stout N, Dennis AR, Wells TM. The Buck stops there: The impact of perceived accountability and control on the intention to delegate to software agents. *AIS Transactions on Human-Computer Interaction*. 2014;**6**(1):1-15
- [48] Leyer M, Aysolmaz B, Iren D. Acceptance of AI for delegating emotional intelligence: Results from an experiment. In: *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*. Honolulu, Hawaii, United States: ScholarSpace; 2021. pp. 6307-6316
- [49] Dietvorst BJ, Simmons JP, Massey C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly)

- modify them. *Management Science*. 2018;**64**(3):1155-1170
- [50] Fernández, Domingos E, Terrucha I, Suchon R, Grujić J, Burguillo JC, Santos FC, et al. Delegation to artificial agents fosters prosocial behaviors in the collective risk dilemma. *Scientific Reports*. 2022;**12**(1):8492. Available from: <https://www.nature.com/articles/s41598-022-11518-9>
- [51] Fügener A, Grahl J, Gupta A, Ketter W, Taudien A. Exploring user heterogeneity in human delegation behavior towards AI. In: *Proceedings of the 42nd International Conference on Information Systems (ICIS)*. Atlanta, Georgia, United States: Association for Information Systems; 2021
- [52] Hukal P, Berente N, Germonprez M, Schechter A. Bots coordinating work in open source software projects. *Computer*. 2019;**52**(9):52-60
- [53] Sun N, Botev J, Khaluf Y, Simoens P. Theory of mind and delegation to robotic virtual agents. In: *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. New York, NY, United States: IEEE; 2022. pp. 454-460
- [54] Luhmann N. *Trust and Power*. Hoboken, New Jersey, United States: John Wiley & Sons; 1979
- [55] Dunn JR, Schweitzer ME. Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*. 2005;**88**(5):736-748
- [56] Myers CD, Tingley D. The influence of emotion on trust. *Political Analysis*. 2016;**24**(4):492-500
- [57] Marsh SP. *Formalising Trust as a Computational Concept [dissertation]*. University of Stirling; 1994
- [58] Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. *The Academy of Management Review*. 1995;**20**(3):709-734
- [59] Friedman B, Khan PH Jr, Howe DC. Trust Online. *Communications of the ACM*. 2000;**43**(12):34-40
- [60] Cook KS, Yamagishi T, Cheshire C, Cooper R, Matsuda M, Mashima R. Trust building via risk taking: A cross-societal experiment. *Social Psychology Quarterly*. 2005;**68**(2):121-142
- [61] Lee JD, See KA. Trust in automation: Designing for appropriate reliance. *Human Factors*. 2004;**46**(1):50-80
- [62] Deutsch M. *The Resolution of Conflict: Constructive and Destructive Processes*. New Haven, Connecticut, United States: Yale University Press; 1973
- [63] Deutsch M. Cooperation and trust: Some theoretical notes. In: Jones MR, editor. *Nebraska Symposium on Motivation*. Lincoln, Nebraska, United States: University of Nebraska Press; 1962. pp. 275-320
- [64] Rotter JB. A new scale for the measurement of interpersonal trust. *Journal of Personality*. 1967;**35**(4):651-665
- [65] LaViola JJ Jr, Kruijff E, McMahan RP, Bowman DA, Poupyrev I. *3D User Interfaces: Theory and Practice*. 2nd ed. Boston, United States: Addison-Wesley Professional; 2017
- [66] Khalid HM, Shiung LW, Sheng VB, Helander MG. Trust of virtual agent in multi actor interactions. *Journal of Robotics, Networking and Artificial Life*. 2018;**4**(4):295-298
- [67] Hosseinpanah A, Krämer NC, Straßmann C. Empathy for everyone? The effect of age when evaluating a

- virtual agent. In: Proceedings of the Sixth International Conference on Human-Agent Interaction (HAI). New York, NY, United States: Association for Computing Machinery; 2018. pp. 184-190
- [68] Morrow JL Jr, Hansen MH, Pearson AW. The cognitive and affective antecedents of general trust within cooperative organizations. *Journal of Managerial Issues*. 2004;**16**:48-64
- [69] Punyatoya P. Effects of cognitive and affective trust on online customer behavior. *Marketing Intelligence & Planning*. 2019;**37**(1):80-96
- [70] Rempel JK, Holmes JG, Zanna MP. Trust in close relationships. *Journal of Personality and Social Psychology*. 1985;**49**(1):95-112
- [71] Lee J, Moray N. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*. 1992;**35**(10):1243-1270
- [72] Song SW, Shin M. Uncanny valley effects on Chatbot trust, purchase intention, and adoption intention in the context of E-commerce: The moderating role of avatar familiarity. *International Journal of Human-Computer Interaction*. 2022;**40**:441-456
- [73] Potts SR, McCuddy WT, Jayan D, Porcelli AJ. To trust, or not to trust? Individual differences in physiological reactivity predict trust under acute stress. *Psychoneuroendocrinology*. 2019;**100**:75-84
- [74] de Visser EJ, Parasuraman R. Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*. 2011;**5**(2):209-231
- [75] Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*. 2012;**6**(1):57-87
- [76] Dzindolet MT, Pierce LG, Beck HP, Dawe LA. The perceived utility of human and automated aids in a visual detection task. *Human Factors*. 2002;**44**(1):79-94
- [77] Daronnat S, Azzopardi L, Halvey M, Dubiel M. Impact of agent reliability and predictability on trust in real time human-agent collaboration. In: Proceedings of the Eighth International Conference on Human-Agent Interaction (HAI). New York, NY, United States: Association for Computing Machinery; 2020. pp. 131-139
- [78] Lehdonvirta V, Castronova E. *Virtual Economies: Design and Analysis*. Cambridge, Massachusetts, United States: MIT Press; 2014
- [79] Glikson E, Woolley AW. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*. 2020;**14**(2):627-660
- [80] Hoff KA, Bashir M. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*. 2015;**57**(3):407-434
- [81] Waytz A, Heafner J, Epley N. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*. 2014;**52**:113-117
- [82] Natarajan M, Gombolay M. Effects of anthropomorphism and accountability on trust in human robot interaction. In: Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI). New York, NY,

- United States: Association for Computing Machinery; 2020. pp. 33-42
- [83] Matsui T, Koike A. Who is to blame? The appearance of virtual agents and the attribution of perceived responsibility. *Sensors*. 2021;**21**(8):2646. Available from: <https://www.mdpi.com/1424-8220/21/8/2646>
- [84] Chiou EK, Schroeder NL, Craig SD. How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education*. 2020;**146**:103756
- [85] Montoya RM, Horton RS, Kirchner J. Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*. 2008;**25**(6):889-922
- [86] DeBruine LM. Facial resemblance enhances trust. *Proceedings of the Royal Society B: Biological Sciences*. 2002;**269**(1498):1307-1312
- [87] Verosky SC, Todorov A. Differential neural responses to faces physically similar to the self as a function of their valence. *NeuroImage*. 2010;**49**(2):1690-1698
- [88] Verberne FMF, Ham JRC, Midden CJH. Familiar faces: Trust in facially similar agents. In: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Richland, South Carolina, United States: International Foundation for Autonomous Agents and Multiagent Systems; 2014
- [89] Launay J, Dean RT, Bailes F. Synchronization can influence trust following virtual interaction. *Experimental Psychology*. 2013;**60**(1):53-63
- [90] Parasuraman R, Miller CA. Trust and etiquette in high-criticality automated systems. *Communications of the ACM*. 2004;**47**(4):51-55
- [91] Patzer GL. Source credibility as a function of communicator physical attractiveness. *Journal of Business Research*. 1983;**11**(2):229-241
- [92] Yuksel BF, Collisson P, Czerwinski M. Brains or beauty: How to engender trust in user-agent interactions. *ACM Transactions on Internet Technology*. 2017;**17**(1):1-20
- [93] Adams D, Bah A, Barwulor C, Musaby N, Pitkin K, Redmiles EM. Ethics emerging: The story of privacy and security perceptions in virtual reality. In: *Proceedings of the 14th Symposium on Usable Privacy and Security (SOUPS)*. Berkeley, California, United States: USENIX Association; 2018. pp. 427-442
- [94] Gur N, Bjørnskov C. Trust and delegation: Theory and evidence. *Journal of Comparative Economics*. 2017;**45**(3):644-657
- [95] Fukuyama F. *Trust: The Social Virtues and the Creation of Prosperity*. New York, NY, United States: Simon and Schuster; 1996
- [96] Sun N, Botev J, Simoens P. The effect of rapport on delegation to virtual agents. In: *Proceedings of the 23rd International Conference on Intelligent Virtual Agents (IVA)*. New York, NY, United States: Association for Computing Machinery; 2023
- [97] Sun N, Botev J. Technological immersion and delegation to virtual agents. *Multimodal Technologies and Interaction*. 2023;**7**(11):106