

Review

Not peer-reviewed version

---

# Artificial Intelligence as a New Field of Activity for Applied Social Psychology – A Reasoning for Broadening the Scope

---

[Gesa Alena Linnemann](#) \* and [Linda-Elisabeth Reimann](#)

Posted Date: 24 June 2024

doi: 10.20944/preprints202406.1586.v1

Keywords: artificial intelligence; applied social psychology; large language models; eliza effect; generative ai



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

# Artificial Intelligence as a New Field of Activity for Applied Social Psychology—A Reasoning for Broadening the Scope

Gesa Alena Linnemann <sup>1,\*</sup> and Linda-Elisabeth Reimann <sup>2</sup>

<sup>1</sup> Correspondence Catholic University of Applied Sciences

<sup>2</sup> University of Münster; linda.reimann@uni-muenster.de

\* Correspondence: g.linnemann@katho-nrw.com

**Abstract:** Interaction with social agents employing natural language have been discussed especially since Weizenbaum (1966) observed the social effects of his chatbot named 'Eliza', which imitated a client-centered conversation. With Nass and Reeves' media equation theory and computers as social actors paradigm, a well-tested framework has been introduced to the study of human-computer interaction. In the 1990s, there was a renewed focus on natural language analysis, which has become a crucial area of research in machine learning. Recent advancements in data processing and storage capacities have led to the development of AI-generated utterances that closely resemble human utterances. As a result, AI-based systems can now mimic social relationships to a significant extent. In light of these technological developments, the article explores whether quasi-social relationships should be considered within the field of social psychology and examines the potential advantages and challenges of this approach.

**Keywords:** artificial intelligence; applied social psychology; large language models; eliza effect; generative ai

## 1. Introduction

The burgeoning prominence of artificial intelligence (AI) applications in contemporary society warrants increased scrutiny regarding their impact on human behavior and relationships. Especially the introduction of the Transformer model [1] as employed in, for instance, ChatGPT and GPT-4, enabled interactions that are closer to human-human-interaction than any artificial system mimicking human interaction before could provide. This development signified a milestone in generative AI, but also holds implications for the field of psychology and social psychology in particular. In this paper, we understand AI according to Innes and Morrisson [2] who stated that AI imitates human abilities in terms of sensing the environment, learning, thinking, and reacting in response. The word 'imitate' should be emphasized in our conceptualization. We do not assume any semantic understanding. Research started addressing human-AI relationships (e.g., friendships, romantic or therapeutic relationships) [3–5] wherein individuals develop emotional attachments to conversational agents like chatbots, highlighting the potential implications of prolonged engagement with anthropomorphized algorithms. The interest in researching the interaction with social agents is not a recent phenomenon. Even in the 1960s, Weizenbaum [6] examined interaction processes between clients and his chatbot 'Eliza'. In this line of reasoning, frameworks such as media equation theory [7] and computers as social actors paradigm [8] have evolved. This research has reached a new level of relevance due to the increasing importance of understanding how people interact with the new generation of chatbots that are easily available and even customizable. In addition, Gambino et al. [8] noted the possibility of a reserved effect, whereby interactions with social agents may shape ensuing human-human interactions. This is supported by recent work by Guingrich and Graziano, who argue that schema activation may lead persons to ascribe consciousness to AI systems, which in turn affects the way other humans are treated [3]. Given these developments, it becomes crucial to

examine how AI influences our daily lives and relationships, particularly concerning text- and speech-based output irrespective of underlying technical complexities.

Although numerous fields of study, such as human-computer interaction (discipline that examines user experience design when interacting with computers), human-robot interaction (the investigation of collaboration between humans and robots through physical or virtual means), and explainable AI (discipline that focuses on the enhancement of transparency of algorithmic decision-making processes), have emerged in response to rapid advancements in AI and especially machine learning, they predominantly emphasized technological appearances or usability over nuanced exploration of socio-behavioral consequences. Machine learning (ML) refers to a subfield of artificial intelligence that leverages statistical methods and computational algorithms to enable self-improving systems which learn from data patterns for enhanced prediction, decision-making, and problem-solving capabilities [4]. The development of Transformer [1] was an important step to enhance the technology behind Large Language Models (LLM) [5]. LLM represent a form of artificial neural networks that stem from the field of Natural Language Processing, and serve different purposes of language processing, including the generation of natural language, which, in turn, represents a form of generative AI. To date, different forms of generative AI can be combined to multimodal AI, which can process several modalities like image and text.

To fill the – in some regards subtle, in others obvious – research gap regarding social-behavioral consequences, we propose expanding the scope of applied social psychology to incorporate examination of human-AI interaction and elucidate its concomitant effects using key theories anchored in this discipline such as social cognitive theory, decision making, and self-disclosure etc.

Our focus diverges from extant efforts harnessing psychological principles to enhance AI functionality; instead, we underscore the value of delving into social psychology due to its emphasis on intricate facets of social interaction. Furthermore, differentiating ourselves from the use of AI as a research tool within social psychology [6], we concentrate on exploring the reciprocity between AI and human behavior. By doing so, we focus on 'weak' AI – an AI that processes information provided to solve relatively specific tasks – instead of 'strong AI' – which would be an AI with a self-consciousness and own goals and which – to date – does not exist. Whether its existence is possible is the subject of much debate e.g., [7–9]. This approach is in line with Innes and Morrison [10] who stated the importance of further researching weak AI from a psychological perspective. Thus, this review aims to provide compelling evidence supporting the integration of social psychology theory into AI research agendas, ultimately benefiting people in social interaction, users and developers alike.

### 1.1. Social Psychology

Before discussing how social psychology provides a new perspective on AI, it is important to discuss what social psychology entails. Social psychology is a discipline of psychology that focuses on social situations and interactions. In his seminal definition, [11] (p 12) states that social psychology is “the science which studies the behavior of the individual in so far as his behavior stimulates other individuals, or is itself a reaction to their behavior; and which describes the consciousness of the individual in so far as it is a consciousness of social objects and social reactions”. In other words, social psychology is about exploring individuals in social situations and includes thinking and experiencing, attitudes, prosocial and antisocial behavior. The importance of this research approach is evident for various areas of life. However, given that social psychology primarily concerns the interactions between living human individuals, it is unclear how AI fits into this field. To answer this, two additional questions arise:

*RQ1: Is there an “imagined or actual presence of other persons” (following the definition of Allport) when individuals interact with AI?*

*RQ2: Can an AI be perceived as a social actor?*

For addressing this questions, we will briefly overview influential work on this topic from various perspectives in the following sections.

### 1.2. Human-Computer Interaction

In considering the current literature on AI and its interactions with human individuals and vice versa, it is evident that the discipline of human-computer interaction (HCI) represents a central approach to be considered. HCI is a discipline that strives to find answers to the question how interaction with a technology is comfortable or easy for the user [12,13], in particular how aspects of usability can be improved. Moreover, HCI is an interdisciplinary approach that combines “the fields of computer science and psychology and cognitive sciences” [14] (p 4). In a systematic review Nazar et al. further argue that HCI and AI are working closely together because “AI mimics human beings to build intelligent systems, and HCI attempts to understand human beings to adapt the machine to improve safety, efficiency, and user experience” [14] (p 4). A new field emerged from the interaction of HCI and AI which is called explainable artificial intelligence (XAI). The main focus of this field is to solve the black box problem regarding the question of how AI technology comes to specific decisions [14,15]. Such a transparency about decision making is important when humans use AI. However, it does not answer the question of human decision making and the process of (social) interaction with AI. Therefore, it is necessary to take a closer look on another field of research that more specifically addresses the question of how individuals interact with AI.

### 1.3. Human-AI and Human-Chatbot Interaction

Back in the 1960s, Joseph Weizenbaum created a chatbot named 'Eliza' that was intended to emulate interpersonal interaction [16]. With Eliza, Weizenbaum mocked the conversation style of person-centered psychotherapy. However, he observed that people wished to interact with Eliza in a serious way. His secretary, for instance, even asked him to leave the room in order to interact with Eliza without disturbance [16]. Therefore, the effect that people ascribe human qualities to textual output from computers was later coined 'Eliza effect' by Hofstadter [17].

During the 1990s, there was a renewed emphasis on natural language analysis, which has since become a crucial area of research in machine learning [18]. Due to advances in data processing speed and increased storage capacities and the resulting availability of data, AI-generated utterances are now possible that can hardly be distinguished from human utterances [19]. The processing or generation of natural language by algorithms is referred to as natural language processing (NLP). In this context, AI and computational linguistic methods are utilized. The object can be the recognition of moods in texts, the automated generation of language and the recording of the content and context of language to generate summaries [18]. This technology is used in chatbots and spoken dialog systems (e.g., voice assistants Alexa, Microsoft Cortana, Apple Siri, etc.). In recent years, there has been considerable progress in NLP once again. NLP can generate texts that are indistinguishable from human-written (and even human-spoken) texts. Subjectivity has been excluded from this evaluation. However, people differ from NLP models in their understanding of the content, intentions, and emotions conveyed.

The best-known example at the moment is ChatGPT 4o, which has been announced by Open AI in May 2024 and which follows the models 4, 3.5 to the GPT-3 language model published in 2020 . It is a so-called autoregressive language model. It learns from its own data without formalizing rules in advance and is constantly improving [19]. In addition, the model is trained with the help of people who improve it through reinforcement learning and 'train' it away from undesirable results. ChatGPT passed a law review at the University of Minnesota in early 2023 [20] and is currently being discussed at how it changes different life domains such as learning, teaching, and working. In this article, we aim to highlight another aspect of the ongoing discourse by the question whether also personal relationships are influenced when interacting regularly with an AI.

Most experts in various fields agree on one thing which is that AI and NLP in particular will lead to lasting societal changes [21]. Efforts are being made to shape chatbots in an engaging way, to foster longer interactions [22]. When considering this, the two research questions posed above gain further importance. Nevertheless, the question of whether human-like characteristics can be ascribed to machines, computers or other artificial systems is not new. Therefore, we will look at different theories from social psychology that already tried to answer this question.



## 1.4. Theoretical Approaches

### 1.4.1. Media Equation Theory

According to the media equation theory, proposed by Reeves and Nass [23], individuals interact with technology and media, such as computers or voice assistants, in a manner similar to how they would interact with other people. This should be the case even when aware that they are not communicating with a human counterpart. Reeves and Nass conducted several experiments demonstrating that individuals exhibit polite behavior towards computers [23], as well as attribute personality traits, gender, and social roles to them [23]. Additionally, individuals treat computers differently based on whether they are on the same or opposing team. The activation of mental models is cited as one of the underlying mechanisms for the thoughtless application of social scripts.

### 1.4.2. Computers as Social Actors Framework

The CASA framework was developed based on media equation theory [24]. It proposes that computers can be perceived as social actors. The assumptions of this theory were tested using advanced technologies such as chatbots [25] and embodied agents [26]. Gambino et al. [24] also identified characteristics of these newer technical social agents: anthropomorphism, personalization, bandwidth cues, time, and evolving relationships can impact human-social agent interaction. Additionally, increasing experience may lead to differences between human-human and human-social agent interaction. It is important to investigate the nature of these scripts using inductive methods. As stated above, Gambino et al. [24] noted the possibility of a reverse effect, where human-human interactions are influenced by interactions with social agents. Recent research suggests that knowing the interlocutor is an AI may diminish its positive effects and even, compared to human, superior performance in showing emotional support [27]. As with the upcoming of chatbots and embodied agents, the assumptions of CASA and the extensions proposed by Gambino et al. [24] have to be tested with to date Large Language Models (LLM). Heyselaar [28] discovered that CASA seems to be applicable especially to emergent technologies, but vanishes when technology (such as desktop computers) is not novel anymore. These finding suggests that CASA should be tested over a larger period of time [29].

## 1.5. Answering the Research Questions

From a theoretical perspective, there is sufficient evidence to assume that interaction with AI-based speech generation can create quasi-social situations, even though the activated scripts may differ from a comparable human-human interaction due to persons' experience with use. Nevertheless, based on the Media Equation Theory and the CASA framework, the research questions can partially be answered positively. Theoretical assumptions are in line with empirical evidence which suggests that AI can be perceived as a social actor (RQ2). It also seems that individuals imagine or experience the presence of other people when interacting with AI (RQ1), although it is more difficult to answer this question given the state of research. We will argue for various aspects in the field of applied social psychology that should be considered as AI and NLP become more prevalent in daily life. Nevertheless, the incorporation of social psychology into this analysis does not negate the existence of significant distinctions between humans and AI. The growing capacity to emulate human conduct presents a significant challenge to social psychology, compelling it to persistently re-examine and refine its definitions.

## 1.6. Importance of AI for Applied Social Psychology

When trying to understand how individuals interact with AI or NLP technologies, several theories and approaches from social psychology should be reconsidered in this new application area. In the following, we would like to present some exemplary theories and research areas from social psychology and how our understanding of human-AI interaction might benefit from employing them in this context.

### 1.6.1. Social Learning Theory

In the field of social psychology, one influential theory is social learning theory, which posits that individuals learn behaviors through observation and imitation of others in their social environment [30]. This theory explains how individuals adapt their behavior according to behavior of significant others in order to reach a desired outcome [30]. The same may happen when interacting with an AI. As AI systems become increasingly integrated into our daily lives, they have the potential to both model and reinforce certain social behaviors. For instance, an AI-powered personal assistant that prioritizes efficiency over politeness may encourage users to adopt similarly brusque communication styles. Conversely, an AI system that serves as a therapist (see above) models respectful and empathetic behavior can serve as a positive influence on users' interactions with others. Therefore, in line with Gambino et al. [24], we argue that the requirements of the AI do not only influence our communication style and behavior towards the AI itself [31,32] but also towards other people when certain communication styles and behaviors are practiced and internalized (see also [3])

An additional consideration pertains to the opportunity to individualize AI conversational agents based on one's distinct necessities and objectives. The same is generally unattainable in human relationships due to natural limitations in availability and compatibility; no matter how compatible two individuals may be, there will always remain limits to mutual adaptation. In contrast, AI systems offer greater flexibility in terms of responsiveness and availability, adjusting themselves to user demands without tiring or requiring reciprocity; they can even have a positive impact on well-being [33]. Despite these advantages, several challenges arise in the context of human-AI interaction, one example of study is "Replika" [34,35]. Continuous engagement with an entity capable of fulfilling every desire may hinder the development and maintenance of essential social skills required for successful real-life interactions. It may become more complicated to interact with other individuals in real life who naturally do not bring the same level of empathy as the AI does. We argue that it needs more research in this field, for instance longitudinal studies that examine how individual interaction may change when continually using a highly adaptive and responsive AI (in this realm, see also [3]). It may be that such studies reveal that the preservation of social skills is a crucial skill that should be considered from an ethical perspective when designing personalized AIs. Two main issues may fundamentally separate AI companions from humans to date in terms of being able to fulfill humans' social needs: most social AI applications are for now limited to text, voice or an avatar and lack a bodily representation in three-dimensional space (for instance, in form of a robotic representation). Moreover, the lack of semantic understanding [36] limits not only its conversational abilities, but also the perception as a coherent social entity – not to speak of not being a human being.

### 1.6.2. Social Cognitive Theory

Social cognitive theory [37] is another example of application for which challenges arise in the context of human-AI interaction. According to social cognitive theory, individuals learn new behaviors through observation, imitation, and reinforcement, suggesting that AI applications have immense potential to shape both conscious and unconscious thought patterns. A prominent example for this and research area in social psychology are biases. It is important to consider whether biases occur in human-AI interaction in a similar way they do in real life interactions. Biases refer to systematic errors in judgment or decision-making that stem from cognitive processes influenced by factors such as stereotypes, heuristics, or personal experiences [38]. These biases can manifest in various forms and may influence the interaction between humans and NLP technologies. Compelling examples are the confirmation bias or, in this case, automation bias [39,40] and the availability bias [41]. The confirmation bias describes the fact that humans tend to seek out information that confirms their own beliefs while ignoring contradictory information [42]. The availability bias occurs when humans overestimate the likelihood of information based on their ease of recall from memory [41]. Both biases are well researched when it comes to human-human communication. Nevertheless, the same or similar processes may happen when interacting with AI. On one hand, humans may ask questions in a way that increase the likelihood of receiving biased information (e.g., suggestive

question). On the other hand, humans tend to exhibit confirmation bias by selectively interpreting or accepting information generated by the system that aligns with their preconceptions. This could lead to distorted perceptions and reinforce cognitive biases. Moreover, users may be more likely to accept information provided by the system that is readily available or easily retrievable, even if it is not necessarily accurate or representative of the broader context. However, so far, it is unclear to what degree biases from human-human interaction differ to when humans interact with NLP technologies.

Another important point when it comes to biases in human-AI interaction is that AI response mechanisms often reflect underlying biases inherent in the data used to train them [10]. These prejudices stem from various sources, including but not limited to historical discrimination, unequal representation, and flawed algorithmic design [10]. A recent work has been conducted that started addressing the question of whether users of ChatGPT are aware of such biases. They assessed whether trust in ChatGPT, ChatGPT's user perceptions, and perceived stereotyping by ChatGPT was related to users' self-esteem and psychological well-being [43]. Results showed that the participants in this study did not perceive that ChatGPT fosters negative stereotypes and therefore, interacting with the system showed a positive association regarding psychological well-being [43]. This study is a promising start in researching human-AI interaction. Nevertheless, we argue that social psychology should further be utilized as a mean to reveal and rectify bias in AI. Doing so may help to build fairer and more transparent AI applications capable of delivering equitable treatment for all individuals involved.

#### 1.6.3. Decision Making

A familiar research focus to how people interact with AI and LLM is so-called *algorithm awareness*. Research has shown that different aspects are crucial to consider when trying to explain under which circumstances algorithms are disliked or accepted [44]. Important factors are for example the perceived agency of the algorithm or how independent the algorithms perform [45].

A study by Utz et al. [44] explored in an experiment how people make decisions supported with algorithms in scenarios that varied regarding their degree of morality. The study revealed that participants reported the highest levels of algorithm aversion in the scenario with the highest moral-load. However, not only the moral-load of the scenario influenced the preference to include the algorithm in the decision making process, but also individual characteristics such as conventionalism and need for leadership [44]. The study provides fruitful evidence also for the interaction of individuals with LLM. As LLMs give answers to almost every question a human-being can imagine, it is important to gain more knowledge regarding the question which individuals are more likely to believe the answers that LLM provides without questioning it. The research presented by Utz et al. [44] represents a significant advancement in the field, yet it is imperative to extend these findings.

#### 1.6.4. Self-Disclosure

Self-disclosure is another relevant aspect to consider in the field of human-AI interaction and it is of enormous practical importance. First, self-disclosure can be defined as the process in which an individual reveals personal information to other individuals [46]. Self-disclosure in face-to-face interactions is usually performed in a verbal way [47]. Nevertheless, past research has already focused on self-disclosure in non-face-to-face interactions, particularly self-disclosure via messengers or on social media [48] or the use of computer processing [49]. In their meta-analysis, Weisband and Kiesler [49] showed that the use of computer processing can enhance the level of self-disclosure. This finding was subsequently reinforced by three studies conducted by Joinson [50], which also sought to elucidate the mechanisms underlying the phenomenon of self-disclosure being enhanced when interacting with a computer. He found that visual anonymity led to more self-disclosure when (1) private self-awareness is strengthened and (2) public self-awareness is reduced. Therefore, communication via a computer is characterized by anonymity and the reduction of communication channels [50]. These findings are also in line with the hyperpersonal model [51]. Another systematic review by Nguyen et al. [52] compared online (via social media) and offline self-disclosure and explored contradictory result. While some studies reported greater self-disclosure in online contexts,

others reported more self-disclosure in offline contexts [52]. The authors argue that this discrepancy might be due to the different operationalization of self-disclosure in “actual versus self-report of disclosure” [52] (p 104). Another study by Bazarova and Choi [48] compared self-disclosure in private and direct messaging versus public status updates on social media. They found that the main motivation to perform status updates was social validation or self-expression / relief. In addition to that, the level of intimacy was higher in private and direct messaging than in the status updates that can reach a broader audience. These findings can be attributed to the functional theory of self-disclosure [53] that states that an individual can strive for five aspects when performing self-disclosure which are “relationship development and maintenance, self-expression, self-clarification or identify development, social validation, and social control” [54] (p 5).

However, two salient concerns emerge when transferring this theory to human-AI interaction. Firstly, it is necessary to ascertain whether all five aspects of the functional theory of self-disclosure can be observed in human-AI interaction. Secondly, it is important to determine which factors of an AI system lead to high or low levels of self-disclosure. These factors, however, could be used both in ways that benefit people, or that harm people, for example when exploited in social engineering for phishing and fraud. A beneficial application area for human-AI interaction in which high levels of self-disclosure are desirable could be the interaction with a therapeutic application. Interacting with a digital therapist that uses NLP can especially be beneficial as a supplement to therapy sessions with a human therapist or when there are long waiting times to begin therapy or between sessions. Although some digital therapists have shown initial effectiveness [55], empirical evidence regarding their long-term effects on individuals is missing [56]. Furthermore, when speaking with a therapist, whether in person or through digital means, the individual is in a highly vulnerable state. Therefore, the design of the AI should take this into consideration [56]. Therefore, to date, due to its ‘hallucinations’ and uncontrollable utterances, generative AI does not seem applicable, but the usage should be limited to rule-based systems, as employed in ‘Woebot’ [57], for example. Furthermore, it is important to consider whether relationship development and maintenance should occur with a digital ‘therapist’. Regardless of the answer, research should investigate how either outcome can be achieved to avoid unwanted outcomes or side effects. Besides, there is much to suggest that offers such as the AI companion ‘Replika’ [58] are used by laypersons as an unofficial substitute for therapy or at least for comparable purposes [59].

## 2. Discussion

In this paper, we explored the question of whether social psychology, a discipline traditionally focused on human interactions, can provide valuable insights into the emerging field of AI, especially regarding the developments in NLP. Specifically, we examined whether the interactions between humans and AI systems meet the criteria established by early social psychologists, such as Allport [11], who defined social psychology as the scientific study of individual behavior in response to others' behavior or in relation to social objects. To address this issue, we first considered two relevant research questions. Firstly, we asked whether there is an "imagined or actual presence of other persons" during interactions between humans and AI systems (RQ1). Secondly, we investigated whether AI systems can be perceived as social actors (RQ2).

To answer these questions, we drew upon relevant theoretical frameworks such as the Media Equation Theory and the CASA framework which propose that humans tend to treat technology and media as if they were social entities, leading to behavioral patterns similar to those observed in human-human communication [23,24]. However, we further argued that relying on these theories and related studies, there is sufficient ground to consider social psychology as both applicable and relevant for communicative processes involving AI systems, given that people ascribe human-like qualities and perceive a kind of ‘social presence’. Human-AI interaction should be seen as a broad new field which we highlighted by applying human-AI interaction to first theories and research areas of social psychology. Nevertheless, before starting to further discuss the given theories and research areas, we would like to state that our exploration is not extensive. The aspects considered in this work should rather be understood as examples among many possible avenues for future research. This



includes promising approaches from neighboring research fields as well, for instance the analyses of Garfinkel's studies on Eliza and LYRIC by Eisenmann et al. [60].

At first, our study extends social learning theory to the realm of human-AI interaction, demonstrating that individuals can learn and internalize communication styles and behaviors from AI systems. One significant finding in this context is that an AI often provides more satisfactory responses when the used prompt is polite. In particular, when interacting with an AI, it can be advantageous to utilize words such as 'thank you' or 'please' or to offer a gratuity when the response is highly satisfactory. Even utterances like 'Take a deep breath' yield to better results [61].

AI serving as a therapeutic agent can positively influence users' interpersonal interactions by modeling respectful and empathetic behavior [62]. One promising aspect of using AI as a therapist is its ability to model positive social behaviors such as respect and empathy [63]. Unlike human therapists, AI systems can be available around the clock and can be tailored to meet the unique needs of each user. However, the right balance of support and challenge has to be determined to ensure AI supports building new strategies and habits instead of impairing autonomy. Moreover, AI lacks the capacity for personal judgment, which allows it to provide support without passing judgment on the user's experiences or decisions. Despite these differences, users are still able to communicate with AI in a similar way they would with a human therapist. The consistent availability and customizable nature of AI make it a valuable tool for promoting healthy communication patterns, even if the system itself does not possess true understanding or emotional experience. By applying social learning theory to the domain of human-AI interaction, we provide novel insights into the role of AI as a social actor influencing users' behavior. Additionally, our focus on the unique features of AI conversational agents, namely their ability to individualize responses and accommodate diverse user needs, adds depth to existing discourse surrounding the impact of AI on society. Nevertheless, the importance of considering both the benefits and drawbacks associated with increased integration of AI into everyday life should be highlighted.

The intersection of social cognitive theory [37] and human-AI interaction offers unique opportunities to explore the role of biases in shaping user behavior and decision-making. Previous literature has established the existence of various cognitive biases in human-human interaction (e.g., [41]). Another type of bias relevant to human-AI interaction is the automation bias, defined as over-reliance on automated decision making tools [39]. Such an over-reliance becomes more inherent for individuals in daily life with the rise of AI systems and NLP. Automation bias can have significant consequences in domains where accuracy and reliability are critical, such as healthcare or transportation. Understanding and mitigating automation bias will require careful consideration of both technical factors, such as the design of AI algorithms, and human factors, including individual attitudes towards automation and trust in AI systems. Our study encourages further work to examining whether similar biases emerge in human-AI interaction and investigating the potential psychological mechanisms underpinning these phenomena. On the one hand, we argued that humans may exhibit biases in their interactions with AI systems, suggesting that the same cognitive processes at play in human-human interaction may also apply to human-AI interaction. On the other hand, AI may also perpetuate and exacerbate biases through the reflection of underlying prejudices present in training data [64]. Consequently, efforts to reduce biases in human-AI interaction must take into account both human cognitive processes and AI algorithmic design, particularly in relation to ethical considerations regarding marginalized communities. Future research should aim to develop interdisciplinary approaches to understanding the complex interplay between human biases and AI algorithms, taking into account the unique characteristics of each domain such as (social) psychology, computer science, and ethics.

Building upon existing research on algorithm awareness and its determinants [44], we argue that further research is needed into the ways users perceive and engage with LLMs, specifically focusing on the factors influencing acceptance versus rejection of LLM outputs. Additionally, given the increasing prevalence of LLMs in everyday life, research must continue to examine the long-term consequences of repeated exposure to LLM outputs. This entails studying the formation of trust over time and assessing the extent to which users become increasingly dependent on LLMs for decision

support. Finally, addressing issues surrounding transparency and explainability remains paramount, especially since LLMs operate as 'black boxes' whose internal logic eludes comprehension. This becomes even more important when considering the possibilities that so called 'social engineering' provides. Social engineering refers to attempts to exploit human psychology and manipulate individuals through AI agents or systems [65]. AI algorithms are used to create sophisticated conversational bots, deepfakes, or other automated tactics that are aimed at tricking people into disclose sensitive information. These attacks can manifest in various forms such as phishing emails, chatbot scams, or voice impersonation. As AI technology continues to evolve, the capabilities of social engineers are also expanding, making it imperative to develop effective countermeasures to address these evolving threats. Therefore, raising awareness about AI-powered social engineering, educating users on how to identify and resist manipulative techniques, and integrating secure design principles into AI system development are all essential components of maintaining digital trust and safety. Thus again, fostering collaborations among experts from diverse backgrounds, including computer scientists, psychologists, and philosophers, becomes essential for advancing responsible LLM use and minimizing negative impacts on society.

The last exemplary application we have presented in this paper refers to self-disclosure in human-AI interaction. As stated, several research has already focused on the question whether computer-mediated communication or online versus offline settings change the level of self-disclosure that humans provide [52]. Applying this to human-AI interaction raises the question which design aspects of an AI are crucial for enhancing or reducing the human level of self-disclosure. As stated above, high levels of self-disclosure are preferred when AI is used for therapeutic sessions or when getting first medical advice. However, in other contexts, high levels of self-disclosure may be problematic, particularly when there is a lack of transparency regarding the processing and storage of user data and the potential for further use of this data. Past research has already demonstrated that people do not exhibit high levels of self-disclosure simply when being aware that they are interacting with an AI [27]. On the one hand, this finding indicates that people are not uncritically utilizing AI and exhibit at least some degree of skepticism regarding the potential consequences of providing information to an AI system. On the other hand, we posit that this skepticism may diminish over time. Subpopulations, namely adolescents, seem to show very little concerns regarding privacy issues when employing chatbots [66]. The current state of affair finds us at the outset of a prolonged and multifaceted human-AI interaction. For many, the technology is still relatively novel [67]. Consequently, further research is required to ascertain the most effective means of preventing the disclosure of highly sensitive information in circumstances where such disclosure might be detrimental. One potential avenue for investigation is the efficacy of providing a warning message at the outset of a conversation or during the course of a conversation when the algorithm identifies a potential for the disclosure of sensitive information. In conclusion, we have presented a number of examples of applications that are both fruitful and necessary to consider when conducting research into human-AI interaction. However, it is important to note that these examples should not be seen as a comprehensive list of possibilities for research into social psychology and AI.

### *2.1. Implication and Challenges*

Based on our reasoning, several practical implications can be derived and associated challenges should be considered. A widely discussed and fundamental implication is the labeling of whether an AI system is employed, especially in a social or conversational role (as already adapted in the AI Act, although controversial in its implementation [68,69]).

AI has emerged as a promising tool in mental health services, offering accessible support through chatbots. Several studies have already addressed the therapeutic setting (e.g., [59]). Nevertheless, the concrete utilization remains unclear. It needs more research to understand how an ongoing interaction with a therapeutic AI may help therapy progress and how such an interaction may also influence the interaction of a patient with his or her face-to face therapist [70]. However, given the sensitivity of therapeutic contexts, upholding ethical standards of these AIs becomes paramount. Key considerations include privacy, informed consent, transparency, accountability, and

non-maleficence [71,72]. Unfortunately, recent instances reveal how commercial entities have exploited AI chatbots for profit, masquerading them as therapeutic interventions while surreptitiously promoting products such as prescription drugs [73]. This exploitation highlights two critical issues. Firstly, it underscores the urgent need for stringent regulations governing AI applications in mental health care. Secondly, this situation illustrates the importance of fostering collaboration between various stakeholders - including technology companies, healthcare providers, and government agencies - to establish best practices and monitor compliance. For instance, Scorici et al. [74] depict the negative effects of 'human-washing' of AI to create wrong impressions in potential users.

Another practical implication of incorporating AI into various aspects of society is its application in the workplace. As AI becomes increasingly integrated into daily operations, it is essential to consider how this technology affects collaboration among individuals and teams. Specifically, AI systems can significantly alter the ways in which people interact and share information within organizations. In this regard, three key areas are important to consider: first, how AI can facilitate team coordination; second, how AI can enhance knowledge sharing; and third, how AI can promote diversity and inclusion in group settings. One illustrative example is the use of an AI system to monitor intergroup processes and serve as an early-warning system that signals when ongoing processes or interactions are heading in an unbeneficial direction. However, realizing these benefits necessitates careful consideration of the ethical implications associated with deploying AI technologies. Organizations must ensure that AI systems align with core values and principles, prioritize transparency and explainability, and protect user privacy and security.

The last practical implication that is discussed in this work refers to addressing social isolation, especially prevalent among older adults [75]. Loneliness has been identified as a growing public health concern, contributing to physical and mental decline in later life stages [76]. Here, AI can play a bridging role, connecting isolated individuals with support networks and resources tailored to their needs. However, implementing such solutions raises ethical questions regarding autonomy, privacy, and surveillance. Current research indicates that AI-driven interventions may only be effective if users remain unaware they are interacting with a machine [27]. Nonetheless, societal norms surrounding AI could evolve over time, leading to increased acceptance and reliance on artificial companionship. Consequently, further study is required to explore the long-term effects of AI integration in combatting loneliness while balancing ethical concerns.

## *2.2. Future Directions*

In light of the discussion above, several implications for future research directions can be derived. Initially, it was observed that numerous researchers have commenced addressing the question of how AI can be integrated into (social) psychology research. For example, Salah et al. [6] provide a comprehensive summary as to how AI systems such as ChatGPT can be utilized for research in social psychology. Moreover, they also emphasize the fact that following such an approach, it should be in line with existing social psychology theories. However, the ideas regarding how to integrate new methodological approaches that AI provides into the field of social psychology so far have diverged from the approach advocated in this work. While others have concentrated on the design of AI for enhanced interaction with individuals or the utilization of AI for research purposes [6], our focus is on the examination of the interaction between humans and AI. In this context, it is necessary to consider whether the definitions of social psychology require revision, particularly with regard to the question of what research particularly examines when focusing on 'social' interaction in the domain of social psychology. In considering the future research directions regarding the integration of AI into social psychology, it is worth exploring the potential for social psychological theories and approaches to enhance our understanding of AI. Since AI systems rely heavily on human-generated data, using social psychology to analyze their behavior and decision-making processes seems like a logical step. Through this analysis, we might uncover hidden biases or limitations within AI algorithms and shed new light on the intricate relationship between human cognition and machine learning. Moreover, revisiting and expanding the definition of social

psychology to include interactions with AI will keep our theoretical frameworks up-to-date and applicable in today's technology-driven society. In this vein, besides the strong necessity of interdisciplinary work, the collaboration of various psychological disciplines seems promising and fruitful as well (e.g., developmental and cognitive psychology with Theory of Mind [77] and personality psychology with 'traits' of AI and humans respectively, e.g., [78]). In addition, there seem to be differences in how people react to AI due to personal dispositions [79], which should be further explored in order to shape interfaces and to be able to make individualized recommendations. Dealing with AI presents social psychology with the challenge of defining and delimiting its terms more precisely and reviewing its paradigms against this background. Overall, pursuing this line of research holds great promise for advancing both social psychology and AI development [10].

### 3. Conclusion

The fundamental philosophical question of defining humanness has taken on renewed urgency in light of recent advances in AI. As AI becomes increasingly adept at mimicking human thought and behavior, determining the boundary between real people and highly advanced simulations poses a pressing challenge. Social psychology can contribute significantly to this much-needed interdisciplinary dialogue by offering valuable insights into the unique aspects of human behavior and experience. Specifically, social psychologists can investigate the ways in which AI differs from human cognition and social interaction, highlighting areas where simulation falls short of genuine human engagement. Additionally, social psychology can illuminate the ethical implications of creating increasingly lifelike AI entities, helping to guide responsible innovation in this rapidly evolving field. Thus, harnessing the power of social psychology can facilitate a richer understanding of the boundaries between humans and machines while fostering informed discussions about the role of AI in modern society.

### References

1. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. *Attention Is All You Need*, 2017.
2. Elliott, A. *The Routledge Social Science Handbook of AI*; Taylor & Francis Group: Milton, 2021, ISBN 978-0-367-18825-2.
3. Guingrich, R.E.; Graziano, M.S.A. Ascribing consciousness to artificial intelligence: human-AI interaction and its carry-over effects on human-human interaction. *Front. Psychol.* **2024**, *15*, 1322781, doi:10.3389/fpsyg.2024.1322781.
4. Sergios Theodoridis. *Machine Learning*; Elsevier, 2020, ISBN 9780128188033.
5. Bowman, S.R. *Eight Things to Know about Large Language Models*, 2023.
6. Salah, M.; Al Halbusi, H.; Abdelfattah, F. May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research. *Computers in Human Behavior: Artificial Humans* **2023**, *1*, 100006, doi:10.1016/j.chbah.2023.100006.
7. Fjelland, R. Why general artificial intelligence will not be realized. *Humanit Soc Sci Commun* **2020**, *7*, doi:10.1057/s41599-020-0494-4.
8. *Superintelligenz – Szenarien einer kommenden Revolution.*; Bostrom, N., Ed.; Suhrkamp: Berlin, 2020.
9. Russell, S.J.; Norvig, P. *Artificial intelligence: a modern approach*; Pearson, 2016.
10. Innes, J.M.; Morrison Ben W. Artificial intelligence and psychology. In *The Routledge Social Science Handbook of AI*; Elliot, A., Ed.; Routledge, 2021; pp 30–57.
11. Allport, F.H. *Social psychology*; Houghton, 1924.
12. Bansal, H.; Khan, R. A Review Paper on Human Computer Interaction. *IJARCSSE* **2018**, *8*, 53, doi:10.23956/ijarcsse.v8i4.630.
13. Yun, Y.; Ma, D.; Yang, M. Human-computer interaction-based Decision Support System with Applications in Data Mining. *Future Generation Computer Systems* **2021**, *114*, 285–289, doi:10.1016/j.future.2020.07.048.
14. Nazar, M.; Alam, M.M.; Yafi, E.; Su'ud, M.M. A Systematic Review of Human-Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques. *IEEE Access* **2021**, *9*, 153316–153348, doi:10.1109/ACCESS.2021.3127881.
15. Taylor, J.E.T.; Taylor, G.W. Artificial cognition: How experimental psychology can help generate explainable artificial intelligence. *Psychon. Bull. Rev.* **2021**, *28*, 454–475, doi:10.3758/s13423-020-01825-5.
16. Weizenbaum, J. *COMPUTER POWER AND HUMAN REASON : FROM JUDGMENT TO CALCULATION*, 7. PRINT; FREEMAN: SAN FRANCISCO, C1976, ISBN 0-7167-0464-1.



17. Hofstadter, D.R.; Fluid Analogies Research Group. *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*; Basic Books, 1995.
18. Jain, A.; Kulkarni G.; Shah, V. Natural Language Processing. *International Journal of Computer Sciences and Engineering* **2018**, 161–167.
19. Zong, M.; Krishnamachari, B. *a survey on GPT-3*, 2022.
20. Choi, J.H.; Hickman, K.E.; Monahan, A.B.; Schwarcz, D. Chatgpt goes to law school. Available at SSRN **2023**.
21. Stanford University. The AI Index Report. Available online: <https://aiindex.stanford.edu/report/>.
22. Irvine, R.; Boubert, D.; Raina, V.; Liusie, A.; Zhu, Z.; Mudupalli, V.; Korshuk, A.; Liu, Z.; Cremer, F.; Assassi, V.; et al. *Rewarding Chatbots for Real-World Engagement with Millions of Users*, 2023.
23. Reeves, B., & Nass, C. I. The media equation: how people treat computers, television, and new media like real people and places. *Cambridge University Press*, 1996.
24. Gambino, A.; Fox, J.; Ratan, R. Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *HMC* **2020**, 1, 71–86, doi:10.30658/hmc.1.5.
25. Ho, A.; Hancock, J.; Miner, A.S. Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot. *J. Commun.* **2018**, 68, 712–733, doi:10.1093/joc/jqy026.
26. Hoffmann, L.; Krämer, N.C.; Lam-chi, A.; Kopp, S. Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In *Intelligent Virtual Agents*; Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp 159–165, ISBN 978-3-642-04379-6.
27. Yin, Y.; Jia, N.; Waksak, C.J. AI can help people feel heard, but an AI label diminishes this impact. *Proc. Natl. Acad. Sci. U. S. A.* **2024**, 121, e2319112121, doi:10.1073/pnas.2319112121.
28. Heyselaar, E. The CASA theory no longer applies to desktop computers. *Sci. Rep.* **2023**, 13, 19693, doi:10.1038/s41598-023-46527-9.
29. Xu, K.; Chen, X.; Huang, L. Deep mind in social responses to technologies: A new approach to explaining the Computers are Social Actors phenomena. *Computers in Human Behavior* **2022**, 134, 107321, doi:10.1016/j.chb.2022.107321.
30. Bandura, A.; Walters, Richard, H. *Social learning theory*; Prentice Hall, 1977.
31. Meng, J.; Dai, Y. Emotional Support from AI Chatbots: Should a Supportive Partner Self-Disclose or Not? *Journal of Computer-Mediated Communication* **2021**, 26, 207–222, doi:10.1093/jcmc/zmab005.
32. Linnemann, G.A.; Jucks, R. As in the Question, So in the Answer? Language Style of Human and Machine Speakers Affects Interlocutors' Convergence on Wordings. *Journal of Language and Social Psychology* **2016**, 35, 686–697, doi:10.1177/0261927X15625444.
33. Skjuve, M.; Følstad, A.; Fostervold, K.I.; Brandtzaeg, P.B. My Chatbot Companion - a Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies* **2021**, 149, 102601, doi:10.1016/j.ijhcs.2021.102601.
34. Pentina, I.; Hancock, T.; Xie, T. Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior* **2023**, 140, 107600, doi:10.1016/j.chb.2022.107600.
35. Laestadius, L.; Bishop, A.; Gonzalez, M.; Illenčik, D.; Campos-Castillo, C. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society* **2022**, 146144482211420, doi:10.1177/14614448221142007.
36. Titus, L.M. Does ChatGPT have semantic understanding? A problem with the statistics-of-occurrence strategy. *Cognitive Systems Research* **2024**, 83, 101174, doi:10.1016/j.cogsys.2023.101174.
37. Bandura, A. A social cognitive theory of personality. In *Handbook of Personality*; Pervin L., John, O., Eds.; Guilford, 1999; pp 154–196.
38. Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **1974**, 185, 1124–1131, doi:10.1126/science.185.4157.1124.
39. Parasuraman, R.; Riley, V. Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum Factors* **1997**, 39, 230–253, doi:10.1518/001872097778543886.
40. Vered, M.; Livni, T.; Howe, P.D.L.; Miller, T.; Sonenberg, L. The effects of explanations on automation bias. *Artificial Intelligence* **2023**, 322, 103952, doi:10.1016/j.artint.2023.103952.
41. Tversky, A.; Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* **1973**, 5, 207–232, doi:10.1016/0010-0285(73)90033-9.
42. Nickerson, R.S. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* **1998**, 2, 175–220, doi:10.1037/1089-2680.2.2.175.
43. Salah, M.; Alhalbusi, H.; Ismail, M.M.; Abdelfattah, F. Chatting with ChatGPT: decoding the mind of Chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. *Curr Psychol* **2024**, 43, 7843–7858, doi:10.1007/s12144-023-04989-0.
44. Utz, S.; Wolfers, L.; Göritz, A. The Effects of Situational and Individual Factors on Algorithm Acceptance in COVID-19-Related Decision-Making: A Preregistered Online Experiment. *HMC* **2021**, 3, 27–46, doi:10.30658/hmc.3.3.

45. Jussupow, E.; Benbasat, I.; Heinzl, A. Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion **2020**.
46. Jourard, S.M. *Self-disclosure: An experimental analysis of the transparent self*; John Wiley, 1971.
47. Chelune, G.J. Self-disclosure: an elaboration of its basic dimensions. *Psychol. Rep.* **1975**, *36*, 79–85, doi:10.2466/pr0.1975.36.1.79.
48. Bazarova, N.N.; Choi, Y.H. Self-Disclosure in Social Media: Extending the Functional Approach to Disclosure Motivations and Characteristics on Social Network Sites. *J. Commun.* **2014**, *64*, 635–657, doi:10.1111/jcom.12106.
49. Weisband, S.; Kiesler, S. Self disclosure on computer forms. In *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96*. the SIGCHI conference, Vancouver, British Columbia, Canada, 13–18 Apr. 1996; Tauber, M.J., Nardi, B., van der Veer, G.C., Eds.; ACM Press: New York, New York, USA, 1996; pp 3–10, ISBN 0897917774.
50. Joinson, A.N. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *Euro J Social Psych* **2001**, *31*, 177–192, doi:10.1002/ejsp.36.
51. Walther, J.B.; Whitty, M.T. Language, Psychology, and New New Media: The Hyperpersonal Model of Mediated Communication at Twenty-Five Years. *Journal of Language and Social Psychology* **2021**, *40*, 120–135, doi:10.1177/0261927X20967703.
52. Nguyen, M.; Bin, Y.S.; Campbell, A. Comparing online and offline self-disclosure: a systematic review. *Cyberpsychol. Behav. Soc. Netw.* **2012**, *15*, 103–111, doi:10.1089/cyber.2011.0277.
53. Derlega, V.J.; Grzelak, J. Appropriateness of self-disclosure. In *Self-disclosure: Origins, patterns, and implications of openness in interpersonal relationships*; Chelune, G.J., Ed.; Jossey-Bass, 1979; pp 151–176.
54. Trepte, S.; Masur, P.K.; Scharrow, M. Mutual friends' social support and self-disclosure in face-to-face and instant messenger communication. *J. Soc. Psychol.* **2018**, *158*, 430–445, doi:10.1080/00224545.2017.1398707.
55. Durden, E.; Pirner, M.C.; Rapoport, S.J.; Williams, A.; Robinson, A.; Forman-Hoffman, V.L. Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent "Woebot". *Internet Interv.* **2023**, *33*, 100637, doi:10.1016/j.invent.2023.100637.
56. Sedlakova, J.; Trachsel, M. Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent? *Am. J. Bioeth.* **2023**, *23*, 4–13, doi:10.1080/15265161.2022.2048739.
57. Woebot Health. Available online: <https://woebothealth.com/why-generative-ai-is-not-yet-ready-for-mental-healthcare/>.
58. Replika. Available online: <https://replika.com/>.
59. Pham, K.T.; Nabizadeh, A.; Selek, S. Artificial Intelligence and Chatbots in Psychiatry. *Psychiatr. Q.* **2022**, *93*, 249–253, doi:10.1007/s11126-022-09973-8.
60. Eisenmann, C.; Mlynář, J.; Turowetz, J.; Rawls, A.W. "Machine Down": making sense of human-computer interaction—Garfinkel's research on ELIZA and LYRIC from 1967 to 1969 and its contemporary relevance. *AI & Soc* **2023**, doi:10.1007/s00146-023-01793-z.
61. Yang, C.; Wang, X.; Lu, Y.; Liu, H.; Le, Q.V.; Zhou, D.; Chen, X. *Large Language Models as Optimizers*, 2023.
62. Sharma, A.; Lin, I.W.; Miner, A.S.; Atkins, D.C.; Althoff, T. Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat Mach Intell* **2023**, *5*, 46–57, doi:10.1038/s42256-022-00593-2.
63. Brandtzaeg, P.B.; Skjuve, M.; Følstad, A. My AI Friend: How Users of a Social Chatbot Understand Their Human-AI Friendship. *Human Communication Research* **2022**, *48*, 404–429, doi:10.1093/hcr/hqac008.
64. Ntoutsis, E.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejd, W.; Vidal, M.-E.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; et al. Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Min & Knowl* **2020**, *10*, doi:10.1002/widm.1356.
65. Cialdini, R.B. *Influence: Science and practice*, 4th ed.; Pearson education, 2009.
66. Bae Brandtzaeg, P.B.; Skjuve, M.; Kristoffer Dysthe, K.K.; Følstad, A. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21: CHI Conference on Human Factors in Computing Systems, Yokohama Japan, 08 05 2021 13 05 2021; Kitamura, Y., Quigley, A., Isbister, K., Igarashi, T., Bjørn, P., Drucker, S., Eds.; ACM: New York, NY, USA, 2021; pp 1–13, ISBN 9781450380966.
67. Fletcher, R.; Nielsen, R.K. *What does the public in six countries think of generative AI in news?*, 2024.
68. Li, Z. Why the European AI Act transparency obligation is insufficient. *Nat Mach Intell* **2023**, *5*, 559–560, doi:10.1038/s42256-023-00672-y.
69. Gjevvar, B.; Ferguson, N.; Schafer, B. Bridging the Transparency Gap: What Can Explainable AI Learn From the AI Act? **2023**, doi:10.48550/arXiv.2302.10766.
70. Richards, D.; Richardson, T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin. Psychol. Rev.* **2012**, *32*, 329–342, doi:10.1016/j.cpr.2012.02.004.
71. Floridi, L.; Cowls, J.; King, T.C.; Taddeo, M. How to Design AI for Social Good: Seven Essential Factors. *Sci. Eng. Ethics* **2020**, *26*, 1771–1796, doi:10.1007/s11948-020-00213-5.

72. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach. (Dordr)* **2018**, *28*, 689–707, doi:10.1007/s11023-018-9482-5.
73. Wolfangel, E. Der hinterlistige Therapeut. Available online: <https://www.zeit.de/digital/2023-04/chatbot-psychologie-therapie-pharmaindustrie-manipulation> (accessed on 13 June 2024).
74. Scorici, G.; Schultz, M.D.; Seele, P. Anthropomorphization and beyond: conceptualizing humanwashing of AI-enabled machines. *AI & Soc* **2024**, *39*, 789–795, doi:10.1007/s00146-022-01492-1.
75. Pinquart, M.; Sorensen, S. Influences on Loneliness in Older Adults: A Meta-Analysis. *Basic and Applied Social Psychology* **2001**, *23*, 245–266, doi:10.1207/S15324834BASP2304\_2.
76. Coyle, C.E.; Dugan, E. Social isolation, loneliness and health among older adults. *J. Aging Health* **2012**, *24*, 1346–1363, doi:10.1177/0898264312460275.
77. Cuzzolin, F.; Morelli, A.; Cîrstea, B.; Sahakian, B.J. Knowing me, knowing you: theory of mind in AI. *Psychol. Med.* **2020**, *50*, 1057–1061, doi:10.1017/S0033291720000835.
78. Williams, J.; Fiore, S.M.; Jentsch, F. Supporting Artificial Social Intelligence With Theory of Mind. *Front. Artif. Intell.* **2022**, *5*, 750763, doi:10.3389/frai.2022.750763.
79. Gillath, O.; Ai, T.; Branicky, M.S.; Keshmiri, S.; Davison, R.B.; Spaulding, R. Attachment and trust in artificial intelligence. *Computers in Human Behavior* **2021**, *115*, 106607, doi:10.1016/j.chb.2020.106607.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.