



Article

Social Robots Outdo the Not-So-Social Media for Self-Disclosure: Safe Machines Preferred to Unsafe Humans?

Rowling L. Luo, Thea X. Y. Zhang, Derrick H.-C. Chen, Johan F. Hoorn *  and Ivy S. Huang 

School of Design, The Hong Kong Polytechnic University, Hong Kong 999077, China

* Correspondence: johan.f.hoorn@polyu.edu.hk

Abstract: COVID-19 may not be a ‘youth disease’ but it nevertheless impacts the life of young people dramatically, loneliness and a negative mood being an unexpected additional pandemic. Many young people rely on social media for their feeling of connectedness with others. However, social media is suggested to have many negative effects on people’s anxiety. Instead of self-disclosing to others, design may develop alternatives to employ social robots for self-disclosure. In a follow-up on earlier work, we report on a lab experiment of self-disclosing negative emotions to a social media group as compared to writing a conventional diary journal or to talking to an AI-driven social robot after negative mood induction (i.e., viewing shocking earthquake footage). Participants benefitted the most from talking to a robot rather than from writing a journal page or sharing their feelings on social media. Self-disclosure on social media or writing a journal page did not differ significantly. In the design of interventions for mental well-being, human helpers thus far took center stage. Based on our results, we propose design alternatives for an empathic smart home, featuring social robots and chatbots for alleviating stress and anxiety: a social-media interference chatbot, smart watch plus speaker, and a mirror for self-reflection.



Citation: Luo, R.L.; Zhang, T.X.Y.; Chen, D.H.-C.; Hoorn, J.F.; Huang, I.S. Social Robots Outdo the Not-So-Social Media for Self-Disclosure: Safe Machines Preferred to Unsafe Humans?. *Robotics* **2022**, *11*, 92. <https://doi.org/10.3390/robotics11050092>

Academic Editors: Laura Kunold and Linda Onnasch

Received: 27 July 2022

Accepted: 30 August 2022

Published: 7 September 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: social media; social robots; negative mood; self-disclosure; intervention design; anthropomorphic design

1. Introduction

In 165 countries, measures to contain and avoid COVID-19 contamination obstructed more than 1.5 billion young people in their studies and personal life, which is about 90% of the global student population, according to [1]. Although the virus may not have affected children and youngsters as strongly as older adults, preemptive measures such as lockdowns and curfews did exert a major impact on young people’s social lives. In this respect, Fung [2], the President and Executive Committee of the International Association of Child and Adolescent Psychiatry and Allied Professionals (IACAPAP), spoke of “a family and society pandemic.” Those in less advantageous socio-economic circumstances suffered even more (cf. [3]).

The current pandemic causes a negative mood, particularly in young people, aggravated by feelings of loneliness due to social distancing measures. Apart from changing lifestyle habits such as exercise and healthy nutrition [4], a negative mood can be countered by talking to others to relieve stress (cf. [5]). However, people are lonely exactly because they have no one to talk to or at least not at a deep enough level [6]. Social media is a way to self-disclose to others without meeting face-to-face but people on social media are not always too empathetic (e.g., [7,8]). Earlier work has shown beneficial effects on negative mood of self-disclosure to a social robot but in some studies alternative means were preferred and few to none addressed negative mood in adolescents and young adults (see the review by [9,10]).

In an extension of [11], we focus on personal and embodied interactions for disclosing emotions and needs in young adults, and how social robots (and other artificial agents) can

help us, when re-evaluating the act on social media in comparison to social robots equipped with cognitive agents. In our experiment, we will compare current and common tech-based self-disclosure outlets (i.e., social media channels) to upcoming emerging technologies such as social robots that supposedly are more personal and physically present [12].

2. Social Media Aggravates the Problem

In seeking social connection virtually, young people relied (and rely) heavily on social media such as WhatsApp and WeChat. On the one hand, social media potentially could be an effective medium for disclosing negative mood to real people in a benevolent environment. However, research indicates that social media has a host of unfavorable effects on young people's mental and social wellbeing, aggravating the malady. The overuse of social media may lead to overexposure to unreliable information, which breeds an environment of instable interpersonal relationships, biasing youngsters' perceptions and exacerbating their anxiety.

The rapid growth of social media has revolutionized the way people communicate and seek information. The openness and timeliness of social media also allow misinformation to be created and spread quickly. Social media, such as Facebook, YouTube, and Instagram is web-based and app-based sites that let people communicate with each other by sharing messages, images, opinions, and events [13]. Algorithms recommend content to users, shaping the scope of their knowledge and means of knowing [14]. In filtering the Internet and social media, AI determines what people will be exposed to, which raises concerns about 'echo chambers' and 'filter bubbles' [15,16]. The concern is that social media algorithms combined with the tendency to interact with like-minded people create an environment that primarily exposes users to the like-minded. Keeping users in an endless scrolling loop may lead to social comparison behaviors. Viewing other people's carefully curated information, youth may become susceptible to "frequent and extreme upward social comparisons" [17], which may cause negative side effects such as erosion of self-esteem, increase in depression, and reduced life satisfaction.

Group polarization happens when group members end up taking more extreme positions on a given issue after participating in or contacting a discussion [18]. When a group expands and views are exchanged, discussions tend to go into a certain direction, whether supportive or not. The opinions tend to converge and start to be consistent throughout the group, which may lead to group think and a majority of participants sharing the same standpoints. In extreme cases, group polarization occurs [7,8].

Group polarization on social media seems about twice as severe as in real life [19]. Spreading rumors of major emergencies continuously spreads negative emotions. Under the "contagion" of group negative emotions such as dissatisfaction and tension, the event stakeholders are easily affected and may become more anxious and confused psychologically. As one of the major social media in China, Weibo has increasingly become a source of network events. Weibo users show copycat behavior, using other people's choices as standard of judgment. Emotions can be expressed anonymously, causing aggressive responses.

The Internet clearly can facilitate the flow of misinformation (e.g., [19–22]). The massive spread of false information on social media affects public opinion and may threaten social development [23]. In a meta-analysis of 24 studies, [24] points out that social media promotes the spread of misinformation and pose a threat to public health, as in the case of COVID-19.

Social media has a profound impact on the development and maintenance of interpersonal relationships. [25] stated that friendship instability is positively associated with anxiety. Fear of missing out [26] is a phenomenon that emerges particularly among young people and is a fear of social rejection with a sense of self-insecurity [27]. Additionally, [28] present a meta-analysis of 33 studies, indicating that the fear of missing out increases with more intensive use of social networking sites, giving rise to anxiety and depressive symptoms.

However, social media provides new avenues for interpersonal communication and may enable people to connect to those whom they would otherwise not meet in person. Such ‘weak-tie relationships’ show less interaction, are emotionally less intense and intimate, and feel less reciprocal [29]. The meta-analysis of 101 studies conducted by [30] points out that social networking sites mostly provide a platform to strengthen existing relationships and do not turn weak relationships established online into strong ones. Some believe that social media harms well-being because valuable time is consumed that could be spent with existing close relationships. For instance, [31] states that social media fosters a false sense of online “connections” and superficial friendships, depriving families of quality time spent together.

The emotional responses one receives on social media raise a feeling of uncertainty. Emotional messages have an impact on individuals’ judgment and communication [32], which may affect people’s mental health [33]. Human emotions may be temporary but in chatting on social media, there is no possibility to make eye contact, which may increase the inaccuracy of expression [34]. Many misunderstandings may arise that can affect the stability of relationships.

The online social environment resembles and evokes social processes in the real world (including our self-concept) [35]. However, unreal presentations on social media can lead to negative emotions. Self-presentation on social media aims to project and deal with the self to the public, and people try to control their impressions to convey the desired public self-image [36]. People with a less stable self-concept said they experimented with online self-presentation more frequently, presenting an idealized version of themselves and preferring to exhibit themselves online [37]. A systematic review reported a correlation between negative online interaction and both depression and anxiety [38].

Under peer pressure, one feels forced, urged, or dared to do certain things, because peers have pressured, urged, or dared someone to do so [39]. Peer pressure is increasing nowadays, due to regular communication with one another via social media [40]. People are particularly vulnerable to peer pressure because they desire to associate with and compare themselves to other members of their peer group (*ibid.*). The complexity of this comparison and competitiveness is exacerbated by online culture [41]. Further, ref. [42] reports general impacts of social comparison in the context of body dissatisfaction (*cf.* the beautification of photos): The self-presentation of others on social media will increase the relative dissatisfaction of the individual [43].

Authors like [44] find that a growing number of students to be “overusing” and “addicted to” the Internet, combined with other “inappropriate” and “dangerous” activities such as online gambling, pornography abuse, and cyberbullying (also see [45,46]). Social media anonymity has a negative impact on users’ internal censorship, resulting in reduced moral sensitivity [47]. Users’ perceived anonymity may lead to online disinhibition and deindividuation [48], resulting in uncontrolled behavior and ‘flame practices’ in social media interactions [49]. According to several academics, these behavioral characteristics play an important role in cyber violence (*e.g.*, [50]). Researchers found that when users engage in cyber violence, they find it difficult to detect the emotional reactivity of other users directly and simultaneously [51]. These circumstances reduce empathy, which can only be felt when people receive direct feedback that their actions have harmed others [52]. Those who commit or experience cyber-violence become depressed, anxious, and stressed [53].

It seems then that social media is not the best of ways to seek compensation for lack of personal contact. However, writing a diary journal on one’s own also does not seem the solution. How bad is it then that young people want to meet online? After all, it almost goes without saying that people need other people to talk to—even when there are such drawbacks as social comparison and cyber violence.

3. Social Robots as Possible Alternative

Already in 2012, the American Psychological Association reported that virtual therapists, avatars for business consulting, and synthetic personal trainers were found to be as impactful as their human equivalents [54]. Moreover, therapeutic avatars were reported to have extra benefits to young people because of the secure environment (no peer pressure), increasing therapy adherence and more frequent participation in therapeutic activities [54].

More recently, ref. [55] did a systematic review of 97 empirical studies on social robotics and found that participants had a slightly to moderately positive attitude towards robots whereas less than 10% of studies reported a negative attitude. [55] notes that studies with direct robot interaction reported more negative attitudes than studies using indirect contact (e.g., a display screen), maybe because participants had a better assessment of the drawbacks of robotics during direct interaction [55]. Following the pattern of affective and cognitive attitudes, general acceptance of and willingness to use a robot tended towards the positive side. With regard to putting trust in a robot, [55] states that for most studies, trust was related to the robot in relation to other factors, going into either positive or negative directions, canceling each other out, producing an overall neutral net result for trust in robots per se. Most importantly for our objectives, anxiety for social robots overall was fairly neutral [55], probably due to the general non-threatening design of the robots (i.e., Nao machines) under study [55].

That an agent is embodied, physically present, is more important than the disclosure topic [56]. Not surprisingly, participants talked longer and more to humans compared to artificial agents such as social robots and voice assistants. However, they disclosed more information to a humanoid social robot than to a disembodied agent (ibid.), which is important when there is no one to talk to. During a longitudinal experiment (10 sessions over 5 weeks), informal caregivers opened up to a Pepper robot, disclosing more information and talking for a longer period of time [57].

As to appearance and behavior, [58] indicated that perceived human-likeness of social robots positively affected users' preferences, technology acceptance, involvement, and willingness to use the robot. However, this is not an unqualified finding as certain designers wish to avoid the eeriness of a too lifelike robot and certain users prefer a more mechanic embodiment [59]. Additionally, design of conversational agents best includes non-verbal as well as verbal cues to anthropomorphism [60].

Our research question, then, is whether social media is more beneficial for "venting" negative mood than robots and traditional diary writing. From a theoretical perspective, does more human likeness lead to better therapeutic results (i.e., do people need people)? One would expect (H1) social media (i.e., sharing feelings with real people) to be superior to robots (which are but virtual humans), and that robots would outperform journal writing (a non-human medium).

However, evidence accumulates that on the contrary, social media itself gives rise to anxiety (e.g., [43]) and that in fact robots are trustworthy partners to confide in (e.g., [61]). Alternatively then, from a theoretical perspective of functionality or 'affective affordances,' people need trust and a secure environment rather than other people. Therefore, we hypothesize (H2) that social robots outdo journal writing, which outdoes social media, because the latter cannot be relied upon to return supportive feedback upon receiving a disclosure of negative mood.

4. Method

4.1. Participants and Design

Before conducting the experiment, we obtained approval from the institutional Ethical Review Board (filed under HSEARS20200204003). Voluntary participants ($N = 27$; $M_{age} = 22.2$, $SD_{age} = 2.0$, 59.3% female, Chinese nationality) were invited to an experiment of self-disclosure on social media after negative mood induction, not receiving any credits or monetary rewards. Note that the sample size for this condition was already larger

than for the conditions in [11] ($n = 24$ for robot and $n = 21$ for writing), making a total of 72 valid cases for the current study.

Twenty-one participants were master students and six were undergraduate students. Informed consent was obtained formally from all participants. In addition, we used the data sampled in [11] ($N = 45$; $M_{age} = 24.9$, $SD_{age} = 3.29$, 55.6% female, Chinese nationality) to do a comparison with a social robot ($n = 24$; 54.2% female) and a writing condition ($n = 21$; 57.1% female). For compatibility of conditions, we meticulously followed the design, procedure, and measurements in [11] (methods and data available from <https://www.mdpi.com/2218-6581/10/3/98/s1> (accessed on 1 November 2021)).

4.2. Procedure

Participants were taken to a single room and sat in front of a tablet computer with a sheet of paper, explaining the steps of the experimental procedure (Figure 1). The first part of the experiment consisted of negative mood induction (cf. [57]) and the second part was for self-disclosure to a social-media group, after which participants filled out an online questionnaire, using the “Questionnaire star” environment (<https://www.wjx.cn/mobile/index.aspx> (accessed on 13 December 2021)) for administration of surveys and experiments.



Figure 1. Disturbing clips shown on a tablet and self-disclosure thereafter.

During the induction phase, participants were confronted with a 4 min and 57 s long video compilation of three documentaries about a severe earthquake event in Sichuan, China (2008), providing relevant cultural content and authentic experiences. Earlier studies had indicated that viewing negative media, including videos, images and text, indeed evoked negative mood (e.g., [62,63]); with video having the strongest impact [64].

After viewing the shocking footage, participants were invited to join a WeChat group (Figure 2) and share their feelings for 10 min. The WeChat group was not visible before self-disclosure. During this phase, the experimenters acted as six people in the WeChat group, responding to the participant. These were highly skilled social media users, trained for consistency in commenting on WeChat, according to our list of ‘typical responses’ (see Apparatus and Materials). This way, responses by the experimenters closely resembled the ‘typical responses’ on social media, maintaining the empirically established ratio of three positive responses versus two negative responses to one message inputted by the participant.

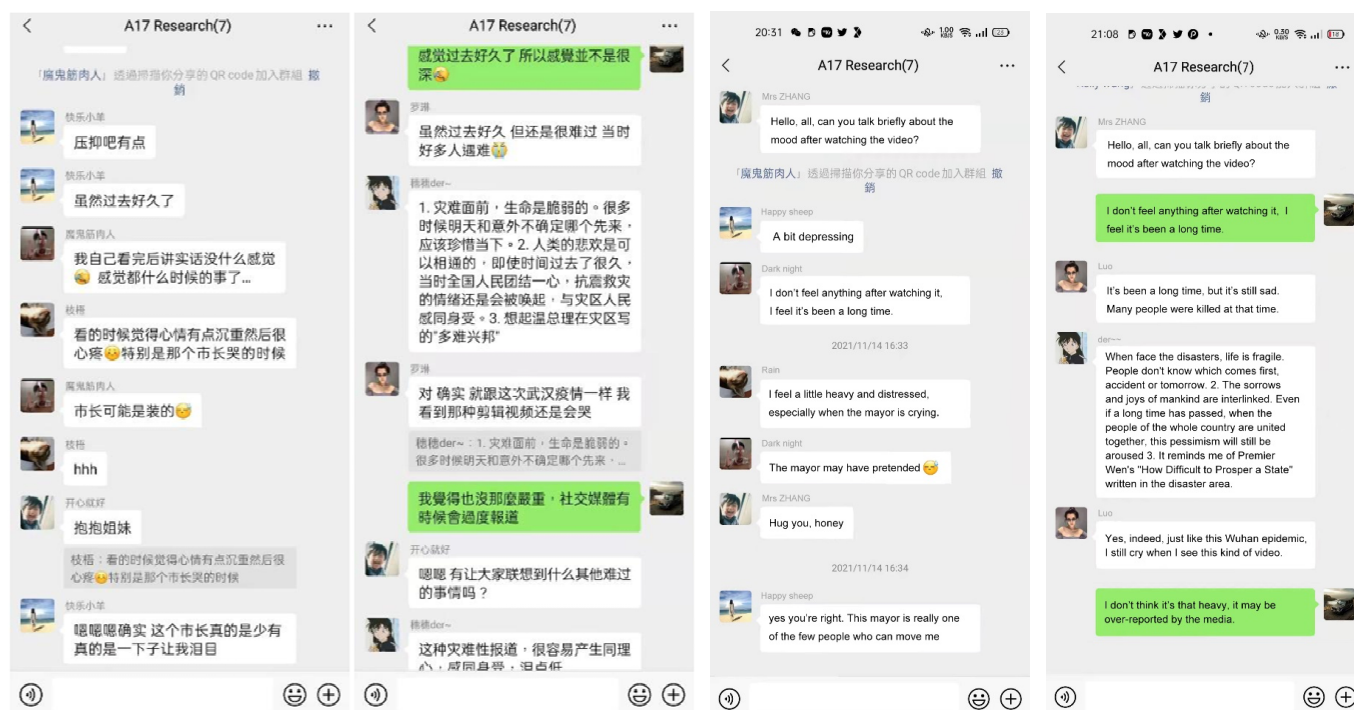


Figure 2. Snippet of a WeChat session (original Chinese and English translation).

After the self-disclosure session on WeChat, participants were asked to fill out a 30-item structured questionnaire [11] (Appendix A) and assess their experiences with the video footage and conversations on social media thereafter. The items on the questionnaire were presented as blocks, and the pseudo-random sequence of items within the blocks was different for each participant. The final part of the questionnaire collected demographic backgrounds. Upon completion of the questionnaire, participants were thanked for their participation and debriefed.

5. Apparatus and Materials

5.1. Video Materials

The negative emotion-induced video was 4 min and 57 s long and consisted of the following 3 video clips from the Sichuan earthquake online documentary:

- Internet video in memory of the Wenchuan Sichuan earthquake 10th anniversary (cut from 00:02–01:19). Available from <https://www.bilibili.com/video/av23087386/> (accessed on 13 June 2019)
- Dazzz2009 (31 December 2008). Internet video record of 512 earthquake in Dujiangyan (cut from 01:20–01:59). Available from https://www.youtube.com/watch?v=vz0ngbl81fmandlist=plf2ppwdjsx1d6r_vuw0vagfzhvir_nro_8andindex=2 (accessed on 13 June 2019)
- Lantian777 (16 May 2008). Internet video 10 min after Wenchuan Sichuan earthquake (in full). Available from <https://www.youtube.com/watch?v=PI5KL7nvU28> (accessed on 14 June 2019)

5.2. Chat-Group Responses

To study the proportion of positive and negative replies on social media, we collected users' thoughts on breaking up a relationship from "Douban-ChoZan," which is a mainstream social media site in China, established in 2005 (<https://www.douban.com/group/topic/83226164/#75043807EPpUK0> (accessed on 2 December 2021)). We used a web crawler for data-mining 10,115 cases with text length of about 200,000 words.

- (1) Data crawling: On Douban, we sampled the texts from group discussions since 2016 around the topic “Let me talk to you about the philosophy behind breaking up and disconnecting.” Information extraction concerned author, time, and contents. We used the request library and tools in the Python programming language to set up a circular structure and record information, which was written to MS Excel documents.
- (2) Data cleaning (word segmentation/de-terminating/tense restoration): We wrote the xls document to the Python IDLE editor and used NLTK/Beautiful soup/NumPy libraries to process the text: 1. Use the word segmentation tool to remove punctuation, paragraphs, etc. 2. Remove function words, such as ‘and,’ ‘or,’ ‘the.’ 3. Restore verb tenses and convert parts of speech.
- (3) Sentiment analysis: We used “The Taiwan University Chinese Semantic Dictionary” (NTUSD) to score the text data after word segmentation, and calculated the total score. Total score = (word score × positive emotion score)—(word score × negative emotion score). Then the positive, neutral, and negative sentences in 10,115 text sentences were counted.
- (4) Statistical results (Figure 3): There were 3633 (36.00%) positive statements, 3562 (34.96%) neutral statements, and 2895 (28.69%) negative statements in total.

Serial num	Release tin	Reply content	Positive vocabulary	Negative words	Degree word	Number of positive sentences	Positive score	Number of negative sentences	Negative score	Total Score
71793		Break up 1 week had redeemed 3 times to say he had not thought good	On time			1	1	0	0	1
71792		The mood is much calmer today	peaceful, normal	concern, worry	Very	10	14	19	-26	-12
71791		He deleted me and I scolded him and told him never to contact me again			Very	0	0	0	0	0
71790		Sorry, I can't help it				0	0	0	0	0
71789		Bingbing, when you broke up with your boyfriend, did you suspect he had		Dobt		0	0	1	-1	-1
71788		If he calls or asks me if I have a boyfriend				0	0	0	0	0
71787		123				0	0	0	0	0
71786		Year 2				0	0	0	0	0
71785		What's the matter, it has been two or three months, I can often see him at				0	0	0	0	0
71784		After breaking up, I always check his Weibo, what should I do? Although I		inferior Embarra	Very	0	1	0	-2	-2
71783		No, his parents have seen me, but we were in KFC that day				0	0	0	0	0
71782		He and I entered into a relationship very quickly. We have very suitable pr	Suitable	lose not good	Very	11	22	7	-9	13
71781		She is really not coming? I'll be here as soon as she left, lol,	HAHA HAPPY		Very	0	0	1	-12	-12
71780		I see				0	0	0	0	0
71779		I don't feel it, she doesn't like me anymore, I didn't care how she felt befo		do not like		0	0	1	-2	-2
71778		At least one semester we have to go to class every day. But why did you i				0	0	0	0	0
71777		Seeing each other every day, it's not easy to forget		not easy		0	0	1	-1	-1
71776		OK				0	0	0	0	0
71775		Yeah, I didn't take care of her feelings for a long time, so selfish.	sense of security	Selfish	Very	1	2	0	0	2
71774		Learned a lot from this post, thank you			Very	0	0	0	0	0
71773		Why is he moaning without illness every day? Why he give me like of my r	promote	Moaning withou		1	1	1	-2	-1
71772		A friend introduced me, he was very enthusiastic to me at first	excellent	self-abasement	Very	1	2	3	-4	-2
71771		I felt so guilty yesterday, so I called him and said sorry	Optimistic	unfair harm	Very	1	10	0	0	10
71770		Many people are anxious for me, but I don't have any tricks to deal with I like many		Deal with		1	0	0	0	1
71769		In fact, his friend before tentatively asked me if I want to get back togeth		unfair	Very	0	0	1	-6	-6
71768		I sent him a message today and asked him to delete all my contact info: Hope				1	1	0	0	1
71767		I think every time he chats with me, he is just like a friend without romanc				0	0	0	0	0

Figure 3. Comments and complaints.

- (5) Typical feedback: From the responses under (4), we compiled a list of hot topics (e.g., wronged, cheated, dissatisfied) (Figure 4) and combined them into ‘typical social-media replies’ to send to our participants. For example, “People bring this on themselves” or “You have to pull yourself together and keep strong” (Appendix B).

Word	Delete	Frequency ↓	Word	Delete	Frequency ↓	Word	Delete	Frequency ↓	Word	Delete	Frequency ↓
Active	<input checked="" type="checkbox"/>	1221	Emotion	<input type="checkbox"/>	806	Tangle	<input type="checkbox"/>	322	Love	<input type="checkbox"/>	216
Mentality	<input checked="" type="checkbox"/>	663	Delight	<input type="checkbox"/>	473	Quarrel	<input type="checkbox"/>	183	Regret	<input type="checkbox"/>	168
Hardworking	<input checked="" type="checkbox"/>	447	Charm	<input type="checkbox"/>	288	Abandon	<input type="checkbox"/>	141	Leave	<input type="checkbox"/>	137
Brilliant	<input checked="" type="checkbox"/>	242	Sad	<input type="checkbox"/>	204	Painful	<input type="checkbox"/>	126	Unbearable	<input type="checkbox"/>	103
Earnest	<input checked="" type="checkbox"/>	194	Sincerity	<input type="checkbox"/>	158	Sad	<input type="checkbox"/>	72	Doubt	<input type="checkbox"/>	66
Family	<input checked="" type="checkbox"/>	158	True love	<input type="checkbox"/>	155	Self-abasement	<input type="checkbox"/>	62	Indifferent	<input type="checkbox"/>	50
Enthusiastic	<input type="checkbox"/>	151	Happy	<input type="checkbox"/>	146	Wronged	<input type="checkbox"/>	49	Boring	<input type="checkbox"/>	42
Angry	<input type="checkbox"/>	141	Fine	<input type="checkbox"/>	127	Violent	<input type="checkbox"/>	41	Upsat	<input type="checkbox"/>	33
Confident	<input type="checkbox"/>	105	Luck	<input type="checkbox"/>	86	Dissatisfied	<input type="checkbox"/>	32	Complain	<input type="checkbox"/>	26
Intelligent	<input type="checkbox"/>	86	Gentle	<input type="checkbox"/>	79	Painful	<input type="checkbox"/>	20	Avoid	<input type="checkbox"/>	19
Security	<input type="checkbox"/>	71				squabble	<input type="checkbox"/>	16	Ridiculous	<input type="checkbox"/>	13
						Silent	<input type="checkbox"/>	12	Pretended	<input type="checkbox"/>	11
						Cheat	<input type="checkbox"/>	9	Exhausted	<input type="checkbox"/>	6
						Speaking rudely	<input type="checkbox"/>	5	Cold	<input type="checkbox"/>	5
						Ruthless	<input type="checkbox"/>	4	Affection	<input type="checkbox"/>	4

Figure 4. Hot topics (e.g., wronged, cheated, dissatisfied).

According to the statistics, the proportion of positive statements was slightly higher than of negative statements. Therefore, when participants self-disclosed, the experimenters replied with three positive responses and two negative responses, in accordance with the contents of the Douban crawler-results. To improve ecological validity, we personalized the typical social-media responses for each participant.

5.3. Measures

For measurement, we employed the structured questionnaire developed by [11], containing four measurement scales: Valence after mood induction (i.e., the earthquake movie) but before treatment (i.e., disclosure to a chat group) and Valence after treatment. Relevance and Novelty served as control variables. The questionnaire ended in inquiring about demographics.

The statements were of Likert type combined with a 6-point rating scale (1 = strongly disagree, 6 = strongly agree). Each measurement scale had four indicative items and four counter-indications. The four indicative items of 'positive Valence before treatment' formed a unipolar conception of Valence, one of the items being "I feel good." The four counter-indications formed a unipolar conception of 'negative Valence before treatment,' for instance, "I feel bad." We also used these items for measurement of Valence after self-disclosure to the social media group, adapting the wording to the situation. Thus, 'Valence after treatment' consisted of four indicative (unipolar positive) and four counter-indicative (unipolar negative) statements as well. The two unipolar scales of Valence combined (with negative Valence recoded) formed the bipolar conception of Valence.

Relevance was measured with two indicative and two counter-indicative items, querying the impact on personal goals and concerns (i.e., one's emotion regulation), in our case, the impact of the typical social-media responses to self-disclosing negative mood. Examples are 'social media is worthwhile' and 'social media is useless'.

The Novelty scale was used as control to see how accustomed participants were to regulating their emotions through social media groups. Novelty was composed of three indicative items (e.g., 'social media is new') and three counter-indicative statements (e.g., 'social media is commonplace'). Raw data can be found in the Technical Report (Supplementary File S1).

We reverse-coded the counter-indicative items on the two Valence scales, Relevance, and Novelty. Because we wanted to compare self-disclosure between social media, robots, and writing, we assessed the reliability of the questionnaire items across these three conditions, thus including the dataset obtained by [11], available from <https://www.mdpi.com/2218-6581/10/3/98/s1> (accessed on 1 November 2021).

Calculated across all three conditions ($N = 72$), measurement scales (all items except Novelty) achieved good to very good reliability in the first run ($0.92 < \text{Cronbach's } \alpha > 0.79$). This was true for the separate subscales of Valence (4 items each) and for their combination (Valence-before and Valence-after, 8 items each), as well as for Relevance (4 items). The control variable of Novelty scored Cronbach's $\alpha = 0.682$ in the first run (all items). Although less than the conventional cut-off of 0.7, we found that the reliability of Novelty could not be improved by eliminating items. Yet, Novelty was a mere control and not of theoretical interest.

We then performed a Principal Components Analysis, using varimax rotation. The component matrix showed that items on the Valence scale and the Relevance scale were arranged nicely, as expected. Novelty showed a certain spread in Relevance but as this was a mere control variable, we left Novelty unchanged. In later analysis, we will check the degree of correlation with theoretical factors. Details of the reliability analysis are tabulated in the Technical Report (Supplementary File S1).

The outliers of Mean Valence, Mean Relevance, and Mean Novelty were participant 9 in Valence (bipolar). Participants 55 and 40 were outliers in Valence-after (bipolar). Participants 5 and 21 were outliers for positive Valence-after. Participants 28, 34, 39, 40, 55,

56 and 64 were outliers for negative Valence-after. Participants 64 and 72 were outliers for Novelty. See the Technical Report (Supplementary File S1).

6. Results

Manipulation Check

We explored whether the shocking video of the earthquake had stirred any emotions in the participants and whether the treatment (Robots, Writing, and Social Media) evoked any change in mood. To check whether emotions (negative or positive) were evoked after mood induction and after treatment, we performed a one-sample *t*-test with 1 as the test value for $N = 72$ and $n = 61$ (outliers removed) (Table 1).

Table 1. One-sample *t*-tests (1 is the test value), checking whether emotions occurred after mood induction and after treatment.

<i>Variables</i>		<i>Mood Induction</i>		
<i>Means</i>	<i>t</i>	<i>p</i>	<i>N</i>	
Positive Valence-before	11.84	<0.001	72	
Negative Valence-before	23.60	<0.001	72	
Positive Valence-before	10.99	<0.001	61	
Negative Valence-before	24.27	<0.001	61	
<i>Variables</i>		<i>Treatment</i>		
<i>Means</i>	<i>t</i>	<i>p</i>	<i>N</i>	
Positive Valence-after	23.42	<0.001	72	
Negative Valence-after	14.91	<0.001	72	
Positive Valence-after	25.28	<0.001	61	
Negative Valence-after	17.22	<0.001	61	

From Table 1, we can conclude that after the earthquake clips (Table 1, *Mood Induction*), more negative than positive mood was induced, as intended, both with $N = 72$ and $n = 61$. For both $N = 72$ and $n = 61$, after Treatment (Table 1, *Treatment*), whether talking to a robot or writing in a journal or chatting with a social group, more positive emotions than negative ones were felt, as expected.

To monitor effects of before and after treatment, we also performed paired-samples *t*-tests in both the $N = 72$ and $n = 61$ datasets (Table 2). Take notice that these tests are manipulation checks; they are not for actual hypothesis testing.

Table 2. Paired-samples *t*-tests for treatment effects on Valence.

<i>Variables</i>		<i>Before-After Treatment</i>		
<i>Means</i>	<i>t</i>	<i>p</i>	<i>N</i>	
Negative Valence before-after	10.88	<0.001	72	
Positive Valence before-after	−9.10	<0.001	72	
Negative Valence before-after	10.89	<0.001	61	
Positive Valence before-after	−10.20	<0.001	61	

From Table 2, we can conclude that participants became less negative after the treatment (i.e., mean negative Valence-before was significantly greater than mean negative Valence-after); furthermore, they became more positive after treatment (i.e., mean positive Valence-before was significantly smaller than mean positive Valence-after). The manipulations were successful: Treatment (whether Robot, Writing, or Social Media) elicited effects into the intended direction.

7. Effects of Media on Valence

7.1. GLM Repeated Measures for Bipolar Valence Before-After

We conducted GLM Repeated measures for bipolar Valence before-after with ($N = 72$) and without outliers ($n = 61$). Table 3 shows the results.

Table 3. GLM Repeated measures for bipolar Valence before-after.

<i>Robots vs. Writing vs. Social Media</i>						
	<i>V</i>	<i>F</i>	<i>df</i> _{1,2}	<i>p</i>	η_p^2	<i>N</i>
Interaction Media \times Valence before-after	0.08	3.01	2,69	0.056	0.08	72
	0.14	4.83	2,58	0.011	0.14	61
Main effect Media (RWS)		2.02	2,69	0.141	0.06	72
		1.96	2,58	0.150	0.06	61
Main effect Valence before-after	0.64	124.90	1,69	0.000	0.64	72
	0.73	152.76	1,58	0.000	0.73	61

Note: Identical results were obtained for unipolar Valence (positive—negative).

The interaction between Media and bipolar Valence (before-after) without outliers was significant and going into the expected direction (more positive after treatment). This interaction effect was supported by a main effect of bipolar Valence but not by the main effect of Media. GLM Repeated measures for unipolar Valence (positive—negative) before and after confirmed these results (Table 3). Paired-samples *t*-tests supported that for each medium, the mood became more positive, the biggest difference being made by Robots and least so by Social Media (Table 4).

Table 4. Paired-samples *t*-tests for bipolar Valence before-after ($n = 61$).

<i>Robots vs. Writing vs. Social Media</i>						
	<i>Difference between Means</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>CI</i>	<i>n</i>
Robot	2.00	−7.87	20	0.000	−2.39 −1.03	21
Writing	1.26	−6.58	16	0.000	−2.31 −0.860	17
Social Media	1.12	−7.41	22	0.000	−2.15 −0.930	23

With $N = 72$ and mean Relevance and mean Novelty as covariates, all interaction and main effects of bipolar Valence and Media vanished but the main effects of Relevance ($F = 1.22, p = 0.244$) and Novelty ($F < 1$) were not significant either. Covariates are dimensions of the participants independent of treatment. Covariates may significantly affect aspects of the analytical model without being significant themselves. However, effect sizes were very low (Relevance $\eta_p^2 = 0.033$; Novelty $\eta_p^2 = 0.003$) and Relevance and Novelty were meant as controls rather than theoretical variables.

GLM Repeated measures for unipolar Valence before-after ($N = 72$) with mean Relevance and mean Novelty as covariates showed significant interactions between negative Valence and Relevance and negative Valence and Novelty. Tests of within-subjects contrasts showed that negative Valence after treatment was lower when the treatment was experienced as more Relevant ($F_{(1,67)} = 5.96, p = 0.017, \eta_p^2 = 0.08$) and as more Novel ($F_{(1,67)} = 5.16, p = 0.026, \eta_p^2 = 0.07$), although effect sizes were small. Relevance and Novelty were positively correlated with each other ($r = 0.47^{**}$).

With $n = 61$ and mean Relevance and mean Novelty as covariates, the interaction effect was still significant ($V = 0.11, F_{(2,56)} = 3.58, p = 0.034, \eta_p^2 = 0.113$). All other effects, including the main effects of Relevance ($F = 2.22, p = 0.142$) and Novelty ($F < 1$) were not significant. GLM Repeated Measures with unipolar Valence (positive—negative) did not change these results.

All in all, it seems that outliers were sensitive to the personal relevance and novelty of the media used, reducing their negative mood. Those are characteristics of this particular subset of participants rather than the media they interacted with or of the larger participant group.

7.2. GLM Univariate (Oneway-ANOVA) for Δ Valence (Bipolar)

To try another perspective, mean difference scores were calculated from the mean values of bipolar Valence before-after and we ran a GLM Univariate analysis (Oneway-

ANOVA) for Medium on bipolar Δ Valence with $N = 72$. The effects were not significant ($F_{(2,69)} = 3.01, p = 0.056, \eta_p^2 = 0.080$). With $n = 61$, the main effect of Media was significant ($F_{(2,58)} = 4.83, p = 0.011, \eta_p^2 = 0.143$). Independent samples t -tests revealed that Robots ($M_{\Delta\text{Val}} = 1.99, SD = 1.16$) made a larger positive difference than Writing ($M_{\Delta\text{Val}} = 1.26, SD = 0.79$) ($t_{(36)} = 2.23, p = 0.016$ (1-tailed), $CI = 0.067$ – 1.41) and even larger than Social Media ($M_{\Delta\text{Val}} = 1.19, SD = 0.77$) ($t_{(42)} = 2.73, p = 0.0045$ (1-tailed), $CI = 0.209$ – 1.40). The difference between Writing and Social Media was not significant ($t_{(38)} = 0.27, p = 0.395$ (1-tailed)).

7.3. GLM Multivariate (Oneway-MANOVA) for Δ Valence (Unipolar)

With $N = 72$, the effect of Media on Δ Valence (positive versus negative) was not significant ($F = 1.95, p = 0.105$). Although this result does not warrant any further exploration, we saw that in the tests of between-subjects effects, Media impacted positive Δ Valence into the desired direction ($F_{(2,69)} = 4.56, p = 0.035, \eta_p^2 = 0.09$) but did not significantly affect negative Δ Valence ($p = 0.177$). Including mean Relevance and mean Novelty into the analysis rendered significant effects for Relevance as covariate ($V = 0.11, F_{(2,66)} = 4.04, p = 0.022, \eta_p^2 = 0.11$) but not so for Novelty. Between-subjects effects showed that mean Relevance correlated positively with positive Δ Valence ($F_{(1,67)} = 5.67, p = 0.020, \eta_p^2 = 0.08$).

Without outliers, $n = 61$, multivariate tests showed significant results of Media ($V = 0.18, F_{(4,116)} = 2.85, p = 0.027, \eta_p^2 = 0.09$). Tests of between-subjects effects showed that Media impacted positive Δ Valence into the desired direction ($F_{(2,58)} = 5.11, p = 0.009, \eta_p^2 = 0.15$) but did not significantly affect negative Δ Valence ($F = 3.01, p = 0.057$). Negativity was not reduced but positivity was increased. Covariate effects of mean Relevance and Novelty were not significant and did not change the pattern of results for $n = 61$.

Independent samples t -tests showed that Robots ($M_{\Delta\text{Valp}} = 1.94, SD = 1.15$) made a larger positive difference than Writing ($M_{\Delta\text{Valp}} = 1.06, SD = 1.07$) ($t_{(35)} = 2.40, p = 0.011$ (1-tailed), $CI = 0.134$ – 1.62) and also larger than Social Media ($M_{\Delta\text{Valp}} = 1.18, SD = 0.85$) ($t_{(41)} = 2.46, p = 0.009$ (1-tailed), $CI = 0.135$ – 1.37). The difference between Writing and Social Media was not significant ($t_{(38)} = -0.42, p = 0.340$ (1-tailed)).

7.4. Exploration: Variance of Valence (VV) as Indicator of Emotional Instability

Out of exploratory curiosity, for $n = 61$, we assessed the variability of the scores within-subjects to the items on the positive Valence and the negative Valence scale before and after treatment. We wanted to evaluate which medium—after negative mood induction—stabilized the variance of affective responses more than other. Therefore, for each participant, we determined the average of squared differences for the scale values of the indicative items (positive Valence) and counter-indicative items (negative Valence). This measure, Variance of Valence, was then submitted to GLM Repeated Measures but did not yield any significant results (Media \times VVpos: $V = 0.12, F_{(3,57)} = 4.04, p = 0.070, \eta_p^2 = 0.12$; Media \times VVneg: $V = 0.05, F_{(3,57)} = 0.95, p = 0.424, \eta_p^2 = 0.05$). For further details, consult the Technical Report (Supplementary File S1).

8. Discussion

H1 stated that negative-mood reduction and/or positive-mood increase would result from the measure of human likeness of the medium, expecting that sharing one's feelings on social media with real people would outperform talking to a social robot and more so writing on paper. H1 was refuted. Although all media increased positive mood, social media did so the least of all with self-disclosure to a social robot outdoing both writing and social media.

H2 took an opposite stance, expecting that human feedback did not have to be the most empathetic so that human-likeness that is trustworthy (i.e., a social robot) would outdo real humans on social media or writing a diary page (which equals talking to oneself). H2 was accepted. In all analyses, the social robot did better on increases in positive mood

than did journal writing or social media; in certain analyses, social media tended to be the least effective of all. People put trust in a robot more so than in humans on social media.

COVID-19 has caused a pandemic of loneliness and depression apart from being a viral disease. This was particularly so for young people, who are in the process of friendship formation and of experimenting with social relationships. The social media they rely on for intimate contact are oftentimes detrimental to young people's mental health. We argued that youngsters overused social media [15,16], receiving too much unreliable information [21,22], breeding an environment of unstable interpersonal relationships [26,28], biasing their cognitions [35], and increasing feelings of anxiety (e.g., [41]). Fear of unstable relationships as well as negative online interaction may lead to depression and anxiety [38]. As our data confirmed, self-disclosure on social media may not exact the best of therapeutic effects.

Although our focus was on youngsters, our findings may be valid for middle-aged and elderly people as well, not just students. It would be worthwhile to check our results with social media relevant to older adults, comparing the effects of diary - robot - living people as we did today. Another aspect worth investigating is in how far our results are limited by their Asian context, knowing that Western users have more ethical objections against robots and AI than Asians have.

On that note, the type of robotics and AI we investigated is advancing and entering the audience at large (cf. *Replika, My AI Friend*, Luka Inc.: <https://replika.ai/>; *beingAI*, <https://beingai.com/> (accessed on 31 August 2022)). That causes ethical, legal, and social issues that designers and developers should be aware of before launching their applications [65]. Safety would be a concern as social robots may be employed without taking care of a human's physical and psychological integrity. Affective bonding in general seems to be beneficial but there may be situations in which vulnerable humans (cf. dementia, autism) develop an attachment to the artificial being that is preferred over actual human contact, which some may deem as 'inappropriate.' With vulnerable patients, the core value of dignity is at stake (e.g., [66]), with the objectification of human friendship, deception by artificiality, a faked identity (of the robot), and trust [65]. If indeed, people trust a robot better than humans on social media; that could stimulate evasive behaviors, escapism even, not dealing with issues humans have among each other. Such avoidance tendencies could actually increase social isolation. Having the user 'all to yourself' could also tempt industries to increase the autonomy of the robot, thus seducing customers into purchasing behaviors they otherwise would not have (cf. *My Friend Caila*, ref. [67]).

If done responsibly, social robots are a design alternative for mental-health interventions. Robots do not invite social comparison (cf. [17]) so that youth does not have to pay attention to the evaluation of others and to the carefully Photoshopped ideal lives presented to them. Social robots do not exert peer pressure (cf. [40]) and do not induce anxiety. If designed with integrity, robots do not give judgmental or opinionated feedback, do not use profane language, and therefore, do not provoke (online) violence (cf. [53]), inducing negative emotions.

In earlier research, we saw that humans empathize with and project feelings onto robots that are maltreated (e.g., [68]). In future work, we wish to study why people sometimes are more empathic with a virtual character (e.g., a robot disclosing it is suffering) than with a real person that says to be suffering (see the Introduction to this study)? Perhaps that with virtual creatures, one has access to all information that is available (WYSIWYG). With humans, one can never be sure if more is concerning than what meets the eye. The suffering of a virtual being has no real-life consequences: The bystander does not need to help and one may reason 'If I don't want to help I don't need to feel guilty about it.' Like animals, a virtual creature is like another 'species', ignorant, innocent, therefore, other behavioral scripts are in place: The virtual character has no hidden agendas as humans do; it doesn't want to get something out of its misery like real humans so to the observers, it may feel like: 'My kindness will not be abused. Therefore, I can freely empathize and feel good about it without suffering the consequences of being kind'.

With respect to design, admittedly, social media, writing, and robots each led to a better mood than before. However, social robots had the most positive impact. With these lessons in mind, we are in the process of designing a number of intervention alternatives, featuring social robots and virtual agents.

People who express negative emotions on social media may encounter negative or indifferent feedback. We attempt a feature that can be embedded in any social media platform (Figure 5): When the Natural Language Processing (NLP) recognizes negative vocabulary and sentiment, it will automatically pop up a social-robot dialog box, inviting users to switch the chat group and talk with the chatbot (cf. *Replika, My AI Friend*, Luka Inc.: <https://replika.ai/> (accessed on 31 August 2022)). In addition, users may want to wear a smart watch that monitors heart-rate variability to indicate possible stress. If confirmed by the user, the watch switches on a smart speaker in the form of a social robot, asking what is wrong. If the smart speaker is built into a mirror, users may want to ‘talk to themselves’ while preparing for the new day or making ready for bed. This ‘mirror for self-reflection’ (Figure 6) may have integrated cameras to detect facial expressions that together with voice analysis and NLP may indicate if mood is improving as the user discloses daily events to their mirror image. This set-up for an empathic smart-home may be applied to care facilities and hospitals alike (cf. [12]).

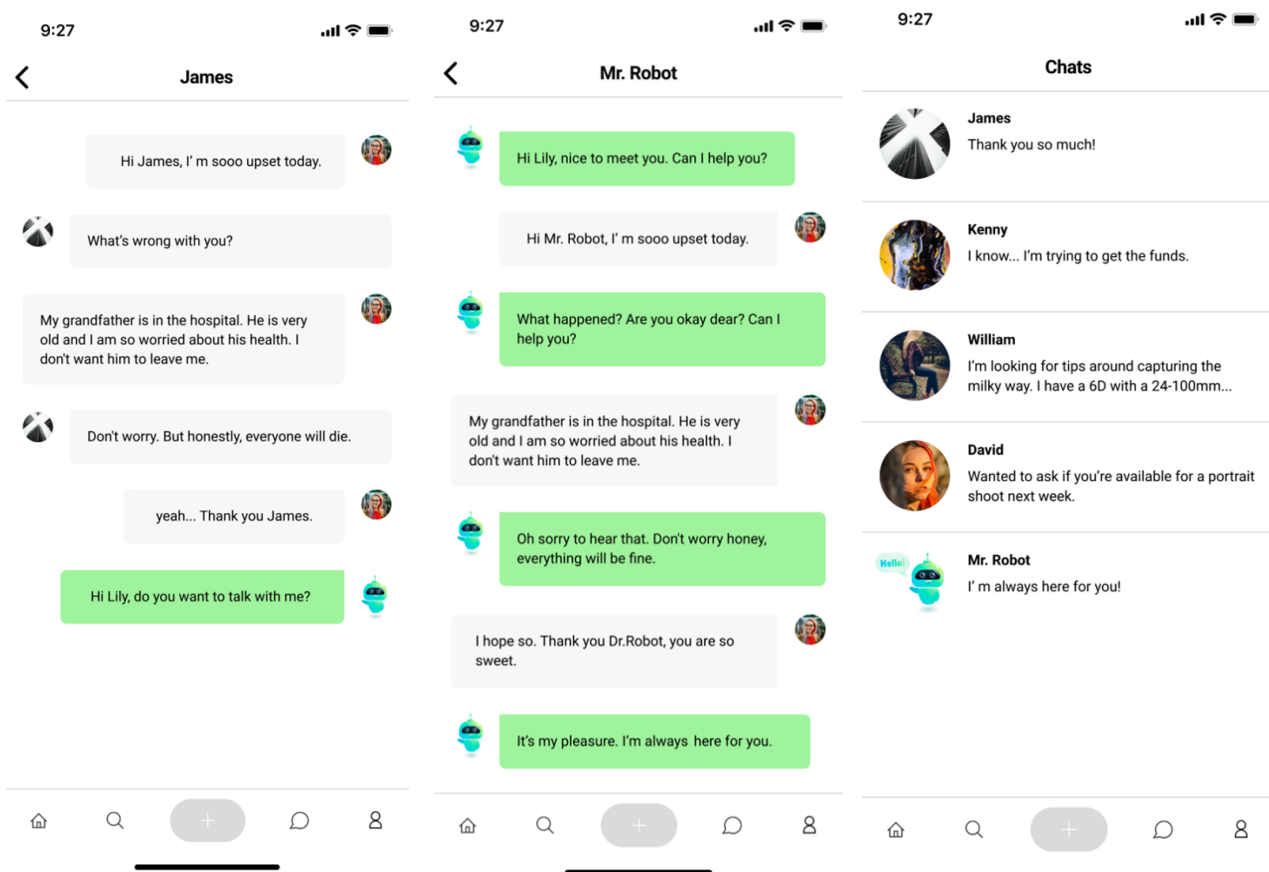


Figure 5. Prototype social-media feature to interfere with unsupportive feedback.

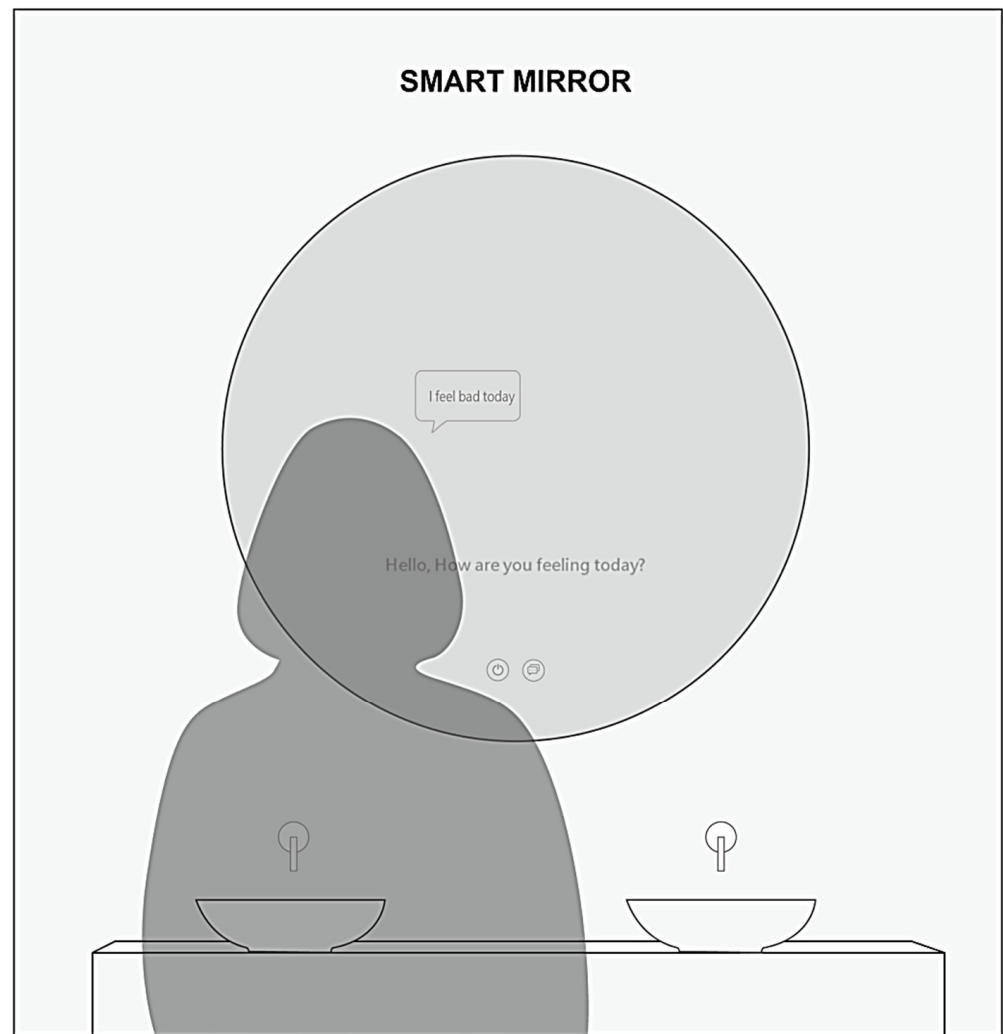


Figure 6. A mirror for self-reflection.

9. Conclusions

The current study was a follow-up of [11], comparing a robot as confidante with writing a diary page. Both these conditions did not include other humans as the robot's voice agent was driven by AI and paper is not interactive. We showed that including other humans into the chat does not necessarily lead to desired outcomes as humans may be unpredictable in responding negatively to disclosing bad mood. If designed responsibly, robots reliably answer in a supportive way.

We started this study from the assumption that more human likeness of an agency would invite self-disclosure by people in a negative mood but there seems to be an optimum. Social robots seem to hit the sweet spot, whereas real humans have too many negative qualities to be fully trusted. We could verify this by using items from [69] (pp. 213–217) survey on self-disclosure. Designers, then, do not need to worry about extremely high 'human fidelity;' a hint into the proper direction may suffice. Even unintended emotional support may instigate users to confide in a software system:

About 10 years ago I set up a fairly popular website for people to play the ancient oriental game of Go, Baduk or Weiqi 围棋 against what was at the time a fairly strong AI program (good old fashioned AI). The website had a rather simple chat feature, with two comments: 'good call' and iirc 'try harder' based on a simple extrapolation of the game position. Searching the logs one day, to get a handle on usage, I was rather disturbed to find that one player had

developed a long conversation, with lengthy self-disclosure, always promoted by one of these expressions, usually the latter. (S)he appeared to consider there was a living person responding. (Jonathan Chetwynd, personal communication, 25 September 2021)

Seen from the above quotation, we want to conclude this paper with a reflection of humanism and humanistic ideas of care: In 16th century Europe, human beings became the measure of all things but with the rise of social robots, functionality—not human beings—will be.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/robotics11050092/s1>, Technical Report S1: Self-disclosure to Social Media.

Author Contributions: Conceptualization, R.L.L. and J.F.H.; methodology, J.F.H.; validation, R.L.L., T.X.Y.Z. and D.H.-C.C.; formal analysis, J.F.H. and I.S.H.; investigation, R.L.L., T.X.Y.Z. and D.H.-C.C.; resources, R.L.L., T.X.Y.Z. and D.H.-C.C.; data curation, J.F.H.; writing—original draft preparation, R.L.L.; writing—review and editing, J.F.H.; visualization, R.L.L., T.X.Y.Z. and D.H.-C.C.; supervision, J.F.H.; project administration, J.F.H.; funding acquisition, J.F.H. All authors have read and agreed to the published version of the manuscript.

Funding: The contributions by Johan F. Hoorn and Ivy S. Huang were supported by the project Negative-mood reduction among HK youth with robot PAL (Personal Avatar for Life) of the Artificial Intelligence in Design Laboratory under the InnoHK Research Clusters, Hong Kong Special Administrative Region Government (grant number: RP2P3).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Subjects Ethics Sub-committee of the university filed under HSEARS20200204003.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are enclosed in Supplementary File S1 and can be made available by the corresponding author.

Acknowledgments: We thank the anonymous reviewers for commenting on an earlier draft of this paper. Kenji Yimin Wang is kindly acknowledged for his help with the data mining. Jia-Yuan Chen is thanked for proposing a new direction for future research.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Structured questionnaires for self-disclosure to social media in Chinese and English.

Appendix A.1 Social Media Questionnaire in Chinese

先生/女士你好:

感謝您參與我們的實驗。這裡我們希望花費你短短幾分鐘回答幾條問題。

你有權隨時終止填寫問卷而不需作任何解釋。你可電郵至

AUTHORS email 與我們的首席調查員 Thea 討論這個研究項目。

當你點擊以下按鈕,即表示同意你是 18 歲以上人士,並自願參與此項目。你了解你有權隨時及以任何原因終止參與這項研究。由參與者提供的數據將會作匿名處理,分析後的結果會記載在此研究的論文中。

這項研究是由香港理工大學監督。

感謝你的參與。

Social Media 團隊

- o 我同意參與這項研究
- o 我不同意參與這項研究

I. 在看了这段影片后,请如实告诉我们您的感受:

Vb1i 我感覺良好

完全不同意 不同意 有點不同意 有點同意 同意 完全同意
1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6

Vb2i 我覺得舒服

完全不同意 不同意 有點不同意 有點同意 同意 完全同意
1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6

Vb3i 我有產生正面積極的情緒

完全不同意 不同意 有點不同意 有點同意 同意 完全同意
1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6

Vb4i 我感到樂觀

Vb5c 我感覺不好

Vb6c 我感到不適

Vb7c 我有產生負面的情緒

Vb8c 我感到悲觀

II. 透過社交媒體聊天後,您感覺如何?

Vb1i 我感覺良好

Vb2i 我覺得舒服

Vb3i 我有產生正面積極的情緒

Vb4i 我感到樂觀

Vb5c 我感覺不好

Vb6c 我感到不適

Vb7c 我有產生負面的情緒

Vb8c 我感到悲觀

III. 我認為通過社交媒體聊天對我的情緒調控

Re1i 有用

Re2i 有效

Re3c 無效

Re4c 沒用

IV. 我認為通過社交媒體聊天這種方式

No1i 是新穎的

No2i 是原創的

No3i 是意想不到的

No4c 是在我的預想之內

No5c 是普通的

No6c 是老土的

V. 其它信息

De1 性別

女

男

其它

De2 年齡

De3 學歷 (最高學歷或現時正修讀)

小學或以下

中學

大專 / 副學士 / 文憑

大學本科

碩士

博士或以上

De4 種族
 亞洲
 非洲
 歐洲
 北美洲
 南美洲
 澳洲/大洋洲
 南極洲

Appendix A.2 Social Media Questionnaire in English

Dear Sir/Madam,

Thank you for your time for our experiment. We would like to ask you to answer a few questions. Answering these questions will only take a few minutes.

You have the right to withdraw at any point during the study, for any reason, and without any prejudice. If you would like to contact the Principal Investigator in the study to discuss this research, please e-mail AUTHOR email.

By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason. The data provided by the participants of the study will be processed and published anonymously in the results sections of the paper.

This study is supervised by The Hong Kong Polytechnic University.

Thank you for your participation.

With kind regards,

Team Social media

I agree to participate in this study

I do not agree to participate in this study

I. After seeing the film samples

Vb1i I feel good

Totally	Disagree	Agree a	Totally	
Disagree	Disagree a little	little	Agree	agree
1 ———	2 ———	3 ———	4 ———	5 ——— 6

Vb2i I am well

Totally	Disagree	Agree a	Totally	
Disagree	Disagree a little	little	Agree	agree
1 ———	2 ———	3 ———	4 ———	5 ——— 6

Vb3i I have positive feelings

Totally	Disagree	Agree a	Totally	
Disagree	Disagree a little	little	Agree	agree
1 ———	2 ———	3 ———	4 ———	5 ——— 6

Vb4i I am optimistic

Vb5c I feel bad

Vb6c I am unwell

Vb7c I have negative feelings

Vb8c I am pessimistic

II. After talking on social media

Vb1i I feel good

Vb2i I am well

Vb3i I have positive feelings

Vb4i I am optimistic

Vb5c I feel bad

Vb6c I am unwell

Vb7c I have negative feelings

Vb8c I am pessimistic

III. To regulate my emotions, talking on social media is ...

Re1i useful

Re2i worthwhile

Re3c worthless

Re4c useless

IV. Talking on social media is ...

No1i novel

No2i original

No3i unexpected

No4c predictable

No5c commonplace

No6c old-fashioned

V. Other information

De1 Gender

Female

Male

Other

De2 Age

De3 What is your highest completed education or current education level?

Primary school or below

Secondary school

Post-secondary school / Associate Degree / Diploma

University undergraduate

Master degree

Doctoral degree or above

De4 Ethnicity

Asia

Africa

Europe

North America

South America

Australia/Oceania

Antarctica

If you have any further questions or remarks about this questionnaire, please let us know.

You can write your feedback below.

Kind regards,

Social media team

AUTHOR email

Appendix B

Typical feedback on social media

Positive feedback:

1. Let me hug you. Don't be sad.
2. How do you feel now?
3. Are you ok?
4. You can talk to me if you are upset.
5. It would be very sad for me to see such content.
6. It was really sad.
7. You have to pull yourself together and keep strong.
8. Yeah, it makes me sad to see them in pain in the video.
9. Human beings are small in the face of disaster.
10. We should cherish life, life is unpredictable.
11. We never know which will come first, the accident or tomorrow.

12. We still have to believe in ourselves.
13. Don't worry so much. Everything will be fine.
14. I understand you. I have a similar experience.
15. Love you, hug you!

Negative feedback:

1. Well, it's okay, why are you so sad about it?
2. You are a crybaby.
3. That's a bit of a stretch.
4. It's been so long, why make you so sad?
5. It serves them right.
6. Social media exaggerates it.
7. People bring this on themselves.
8. Humans are inexorable.
9. It serves you right.
10. In fact, I doubt that you are really sad?
11. Think before you act.
12. It's all your fault.
13. I am so tired from your reply.
14. What you say is so boring.

References

1. United Nations Educational, Scientific and Cultural Organization (UNESCO). *COVID-19 Educational Disruption and Response*; UNESCO: Paris, France, 2020; Available online: <https://en.unesco.org/news/covid-19-educational-disruption-and-response> (accessed on 10 December 2021).
2. Fung, D. *IACAPAP Update: President's Message on COVID-19*; Executive Committee of the International Association of Child and Adolescent Psychiatry and Allied Professionals (IACAPAP): Geneva, Switzerland, 12 April 2020; Available online: <https://iacapap.org/iacapap-update-covid-19/> (accessed on 10 December 2021).
3. Xafis, V. 'What is inconvenient for you is life-saving for me': How health inequities are playing out during the COVID-19 pandemic. *Asian Bioeth. Rev.* **2020**, *12*, 223–234. [[CrossRef](#)]
4. Owens, M.; Watkins, E.; Bot, M.; Brouwer, I.A.; Roca, M.; Kohls, E.; Visser, M. Nutrition and depression: Summary of findings from the EU-funded MoodFOOD depression prevention randomised controlled trial and a critical review of the literature. *Nutr. Bull.* **2020**, *45*, 403–414. [[CrossRef](#)]
5. Laban, G.; Morrison, V.; Kappas, A.; Cross, E.S. Informal caregivers disclose increasingly more to a social robot over time. *CHI Conf. Hum. Factors Comput. Syst. Ext. Abstr.* **2022**, 329, 1–7. [[CrossRef](#)]
6. De Jong Gierveld, J.; Van Tilburg, T.G. Social isolation and loneliness. In *Encyclopedia of Mental Health*, 2nd ed.; Friedman, H.S., Ed.; Academic Press: Oxford, UK, 2016; pp. 175–178.
7. Wang, Q.; Yang, X.; Xi, W. Effects of group arguments on rumor belief and transmission in online communities: An information cascade and group polarization perspective. *Inf. Manag.* **2018**, *55*, 441–449. [[CrossRef](#)]
8. Lee, E.J. Deindividuation effects on group polarization in Computer-Mediated Communication: The role of group identification, public-self-awareness, and perceived argument quality. *J. Commun.* **2007**, *57*, 385–403. [[CrossRef](#)]
9. Robinson, N.L.; Cottier, T.V.; Kavanagh, D.J. Psychosocial health interventions by social robots: Systematic review of randomized controlled trials. *J. Med. Internet Res.* **2019**, *21*, e13203. [[CrossRef](#)] [[PubMed](#)]
10. Scoglio, A.A.; Reilly, E.D.; Gorman, J.A.; Drebing, C.E. Use of social robots in mental health and well-being research: Systematic review. *J. Med. Internet Res.* **2019**, *21*, e13322. [[CrossRef](#)] [[PubMed](#)]
11. Duan, E.Y.; Yoon, J.M.; Liang, E.Z.; Hoorn, J.F. Self-disclosure to a robot: Only for those who suffer the most. *Robotics* **2021**, *10*, 98. [[CrossRef](#)]
12. Laban, G.; Ben-Zion, Z.; Cross, E.S. Social robots for supporting Post-Traumatic Stress Disorder diagnosis and treatment. *Front. Psychiatry* **2022**, *12*, 752874. [[CrossRef](#)] [[PubMed](#)]
13. Wang, R.; Wang, X.; Hao, F.; Zhang, L.; Liu, S.; Wang, L.; Lin, Y. Social identity-aware opportunistic routing in mobile social networks. *Trans. Emerg. Telecommun. Technol.* **2018**, *29*, e3297. [[CrossRef](#)]
14. Gillespie, T.; Boczkowski, P.J.; Foot, K.A. *Media Technologies: Essays on Communication, Materiality, and Society*; MIT: Cambridge, MA, USA, 2014.
15. Pariser, E. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*; Penguin: London, UK, 2011.
16. Gillani, N.; Yuan, A.; Saveski, M.; Vosoughi, S.; Roy, D. Me, my echo chamber, and I: Introspection on social media polarization. In Proceedings of the 2018 World Wide Web Conference (WWW'18), Lyon, France, 23–27 April 2018; Champin, P.-A., Gandon, F., Eds.; pp. 823–831. [[CrossRef](#)]

17. Festinger, L. A theory of social comparison processes. *Hum. Relat.* **1954**, *7*, 117–140. [[CrossRef](#)]
18. Isenberg, D.J. Group polarization. *J. Personal. Soc. Psychol.* **1986**, *50*, 1141–1151. [[CrossRef](#)]
19. Iandoli, L.; Primario, S.; Zollo, G. The impact of group polarization on the quality of online debate in social media: A systematic literature review. *Technol. Forecast. Soc. Chang.* **2021**, *170*, 120924. [[CrossRef](#)]
20. Bordia, P.; Rosnow, R.L. Rumor rest stops on the information highway transmission patterns in a computer-mediated rumor chain. *Hum. Commun. Res.* **1998**, *25*, 163–179. [[CrossRef](#)]
21. Kata, A. A postmodern Pandora’s Box: Anti-vaccination misinformation on the Internet. *Vaccine* **2009**, *28*, 1709–1716. [[CrossRef](#)] [[PubMed](#)]
22. Lewandowsky, S.; Ecker, U.K.H.; Seifert, C.M.; Schwarz, N.; Cook, J. Misinformation and its correction: Continued influence and successful debiasing. *Psychol. Sci. Public Interest* **2012**, *13*, 106–131. [[CrossRef](#)] [[PubMed](#)]
23. Guo, B.; Ding, Y.; Yao, L.; Liang, Y.; Yu, Z. The future of false information detection on social media. *ACM Comput. Surv.* **2020**, *53*, 1–36. [[CrossRef](#)]
24. Walter, N.; Brooks, J.J.; Saucier, C.J.; Suresh, S. Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Commun.* **2021**, *36*, 1776–1784. [[CrossRef](#)] [[PubMed](#)]
25. Croft, C.D.; Zimmer-Gembeck, M.J. Friendship conflict, conflict responses, and instability. *J. Early Adolesc.* **2014**, *34*, 1094–1119. [[CrossRef](#)]
26. Casale, S.; Fioravanti, G. Factor structure and psychometric properties of the Italian version of the fear of missing out scale in emerging adults and adolescents. *Addict. Behav.* **2020**, *102*, 106179. [[CrossRef](#)]
27. Stead, H.; Bibby, P.A. Personality, fear of missing out and problematic Internet use and their relationship to subjective well-being. *Comput. Hum. Behav.* **2017**, *76*, 534–540. [[CrossRef](#)]
28. Fioravanti, G.; Casale, S.; Benucci, S.B.; Probstamo, A.; Falone, A.; Ricca, V.; Rotella, F. Fear of missing out and social networking sites use and abuse: A meta-analysis. *Comput. Hum. Behav.* **2021**, *122*, 106839. [[CrossRef](#)]
29. Rademacher, M.A.; Wang, K.Y. Strong-tie social connections. In *Encyclopedia of Social Media and Politics*; Harvey, K., Ed.; Sage: Los Angeles, CA, USA, 2014; pp. 1213–1216. Available online: <https://us.sagepub.com/en-us/nam/encyclopedia-of-social-media-and-politics/book239101> (accessed on 13 December 2021).
30. Balkundi, P.; Bentley, J.; Kilduff, M.J. Culture, labor markets and attitudes: A meta-analytic test of tie-strength theory. *Acad. Manag. Proc.* **2012**, *2012*, 14728. [[CrossRef](#)]
31. Amedie, J. The Impact of Social Media on Society. *Pop Cult. Intersect.* **2015**, *2*, 1–20. Available online: http://scholarcommons.scu.edu/engl_176/2 (accessed on 23 November 2021).
32. Brown, J.D.; Cai, H. Thinking and feeling in the People’s Republic of China: Testing the generality of the “laws of emotion”. *Int. J. Psychol.* **2010**, *45*, 111–121. [[CrossRef](#)]
33. Lerner, J.S.; Li, Y.; Valdesolo, P.; Kassam, K.S. Emotion and decision making. *Annu. Rev. Psychol.* **2015**, *66*, 799–823. [[CrossRef](#)]
34. Lapidot-Lefler, N.; Barak, A. Effects of anonymity, invisibility, and lack of eye contact on toxic online disinhibition. *Comput. Hum. Behav.* **2012**, *28*, 434–443. [[CrossRef](#)]
35. Firth, J.; Torous, J.; Stubbs, B.; Firth, J.A.; Steiner, G.Z.; Smith, L.; Alvarez-Jimenez, M.; Gleeson, J.; Vancampfort, D.; Armitage, C.J.; et al. The “online brain”: How the Internet may be changing our cognition. *World Psychiatry* **2019**, *18*, 119–129. [[CrossRef](#)]
36. Schlenker, B.R. Self-presentation. In *Handbook of Self and Identity*; Leary, M.R., Tangney, J.P., Eds.; Guilford: New York, NY, USA, 2012; pp. 542–570.
37. Fullwood, J.B.M.; Chen-Wilson, C.-H. Self-concept clarity and online self-presentation in adolescents. *Cyberpsychology Behav. Soc. Netw.* **2016**, *19*, 716–720. [[CrossRef](#)]
38. Seabrook, E.M.; Kern, M.L.; Rickard, N.S. Social networking sites, depression, and anxiety: A systematic review. *JMIR Ment. Health* **2016**, *3*, e50. [[CrossRef](#)] [[PubMed](#)]
39. Brown, B.B.; Clasen, D.R.; Eicher, S.A. Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents. *Dev. Psychol.* **1986**, *22*, 521–530. [[CrossRef](#)]
40. Marino, C.; Gini, G.; Angelini, F.; Vieno, A.; Spada, M.M. Social norms and e-motions in problematic social media use among adolescents. *Addict. Behav. Rep.* **2020**, *11*, 100250. [[CrossRef](#)]
41. Bloemen, N.; De Coninck, D. Social media and fear of missing out in adolescents: The role of family characteristics. *Soc. Media + Soc.* **2020**, *6*, 1–11. [[CrossRef](#)]
42. Myers, T.; Crowther, J. Social comparison as a predictor of body dissatisfaction: A meta-analytic review. *J. Abnorm. Psychol.* **2009**, *118*, 683–698. [[CrossRef](#)] [[PubMed](#)]
43. Fan, X.; Deng, N.; Dong, X.; Lin, Y.; Wang, J. Do others’ self-presentation on social media influence individual’s subjective well-being? A moderated mediation model. *Telemat. Inform.* **2019**, *41*, 86–102. [[CrossRef](#)]
44. Hong, F.Y.; Chiu, S.-L. Factors influencing Facebook usage and Facebook addictive tendency in university students: The role of online psychological privacy and Facebook usage motivation. *Stress Health* **2016**, *32*, 117–127. [[CrossRef](#)]
45. Lusk, B. Digital natives and social media behaviors: An overview. *Prev. Res.* **2010**, *17*, 3–6.
46. Bhat, C.S.; Ragan, M.A.; Selvaraj, P.R.; Shultz, B.J. Online bullying among high-school students in India. *Int. J. Adv. Couns.* **2017**, *39*, 112–124. [[CrossRef](#)]
47. Schlosser, A. Self-disclosure versus self-presentation on social media. *Curr. Opin. Psychol.* **2020**, *31*, 1–6. [[CrossRef](#)]

48. Suler, J. The online disinhibition effect. *CyberPsychol. Behav.* **2004**, *7*, 321–326. [[CrossRef](#)] [[PubMed](#)]
49. Kiesler, S.; Sproull, L. Group decision making and communication technology. *Organ. Behav. Hum. Decis. Processes* **1992**, *52*, 96–123. [[CrossRef](#)]
50. Mishna, F.; Saini, M.; Solomon, S. Ongoing and online: Children and youth's perceptions of cyber bullying. *Child. Youth Serv. Rev.* **2009**, *31*, 1222–1228. [[CrossRef](#)]
51. Kowalski, R.M.; Giumetti, G.W.; Schroeder, A.N.; Lattanner, M.R. Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychol. Bull.* **2014**, *140*, 1073–1137. [[CrossRef](#)]
52. Pettalia, J.L.; Levin, E.; Dickinson, J. Cyberbullying: Eliciting harm without consequence. *Comput. Hum. Behav.* **2013**, *29*, 2758–2765. [[CrossRef](#)]
53. Šincek, D. The revised version of the committing and experiencing cyber-violence scale and its relation to psychosocial functioning and online behavioral problems. *Societies* **2021**, *11*, 107. [[CrossRef](#)]
54. DeAngelis, T. A Second Life for Practice? *Monit. Psychol.* **2012**, *43*, 48. Available online: <https://www.apa.org/monitor/2012/03/avatars> (accessed on 10 December 2021).
55. Naneva, S.; Sarda Gou, M.; Webb, T.L.; Prescott, T.J. A systematic review of attitudes, Anxiety, acceptance, and trust towards social robots. *Int. J. Soc. Robot.* **2020**, *12*, 1179–1201. [[CrossRef](#)]
56. Laban, G.; George, J.-N.; Morrison, V.; Cross, E.S. Tell me more! Assessing interactions with social robots from speech. *Paladyn J. Behav. Robot.* **2021**, *12*, 136–159. [[CrossRef](#)]
57. Laban, G.; Kappas, A.; Morrison, V.; Cross, E.S. Protocol for a mediated long-term experiment with a social robot. *PsyArXiv* **2021**. [[CrossRef](#)]
58. Stroessner, S.J.; Benitez, J. The social perception of humanoid and non-humanoid robots. *Int. J. Soc. Robot.* **2019**, *11*, 305–315. [[CrossRef](#)]
59. Blut, M.; Wang, C.; Wunderlich, N.V.; Brock, C. Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *J. Acad. Mark. Sci.* **2021**, *49*, 632–658. [[CrossRef](#)]
60. Seeger, A.-M.; Pfeiffer, J.; Heinzl, A. Texting with humanlike conversational agents: Designing for anthropomorphism. *J. Assoc. Inf. Syst.* **2021**, *22*, 8. [[CrossRef](#)]
61. Pu, L.; Moyle, W.; Jones, C.; Todorovic, M. The effectiveness of social robots for older adults: A systematic review and meta-analysis of randomized controlled studies. *Gerontologist* **2019**, *59*, e37–e51. [[CrossRef](#)] [[PubMed](#)]
62. Bolls, P.; Lang, A.; Potter, R. The effects of message valence and listener arousal on attention, memory, and facial muscular responses to radio advertisements. *Commun. Res.* **2001**, *28*, 627–651. [[CrossRef](#)]
63. Lang, A.; Shin, M.; Lee, S. Sensation seeking, motivation, and substance use: A dual system approach. *Media Psychol.* **2005**, *7*, 1–29. [[CrossRef](#)]
64. Siedlecka, E.; Denson, T. Experimental methods for inducing basic emotions: A qualitative review. *Emot. Rev.* **2019**, *11*, 87–97. [[CrossRef](#)]
65. Sweeney, P. Trusting social robots. *AI Ethics* **2022**, 1–8. [[CrossRef](#)]
66. Felber, N.A.; Pageau, F.; McLean, A.; Wangmo, T. The concept of social dignity as a yardstick to delimit ethical use of robotic assistance in the care of older persons. *Med. Health Care Philos.* **2022**, *25*, 99–110. [[CrossRef](#)]
67. Roberts, J.J. Privacy Groups Claim These Popular Dolls Spy on Kids. *Fortune*, 9 December 2016. Available online: <http://fortune.com/2016/12/08/my-friend-cayla-doll/>; (accessed on 16 January 2022).
68. Suzuki, Y.; Galli, L.; Ikeda, A.; Itakura, S.; Kitazaki, M. Measuring empathy for human and robot hand pain using electroencephalography. *Sci. Rep.* **2015**, *5*, 15924. [[CrossRef](#)] [[PubMed](#)]
69. Jourard, S.M. *The Transparent Self*; Van Nostrand Reinhold: New York, NY, USA, 1971; Available online: <https://sidneyjourard.com/TRANSPARENTSELFFINAL.PDF> (accessed on 27 July 2022).