



HAL
open science

COMETS Opinion 2024-46 - “The phenomenon of attachment to ”social robots”. A call for vigilance among the scientific research community”

Comité Ethique, Christine Noiville, Catherine Pelachaud, Patrice Debré,
Jean-Gabriel Ganascia, Raja Chatila

► To cite this version:

Comité Ethique, Christine Noiville, Catherine Pelachaud, Patrice Debré, Jean-Gabriel Ganascia, et al.. COMETS Opinion 2024-46 - “The phenomenon of attachment to ”social robots”. A call for vigilance among the scientific research community”. CNRS COMETS. 2024. hal-04702388

HAL Id: hal-04702388

<https://hal.science/hal-04702388v1>

Submitted on 19 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPINION no. 2024-46

“THE PHENOMENON OF ATTACHMENT TO ‘SOCIAL’ ROBOTS.

**A CALL FOR VIGILANCE AMONG THE SCIENTIFIC RESEARCH
COMMUNITY”**

WORKING GROUP MEMBERS:

Catherine Pelachaud, COMETS member, rapporteur

Patrice Debré, COMETS member, rapporteur

Christine Noiville, COMETS member

Raja Chatila, emeritus professor of artificial intelligence, robotics and ethics at Sorbonne University

Jean-Gabriel Ganascia, emeritus professor of computer science and artificial intelligence at Sorbonne University



CONTENTS

CONTENTS	2
I. SUMMARY	3
II. FORMAL INTERNAL REQUEST	5
III. ANALYSIS	6
1. The role of social robots is expanding in everyday life	6
2. The individual and collective effects of attachment to social robots	9
3. For a considered and responsible contribution by public research to the development of social robots	11
IV. RECOMMENDATIONS	13
V. QUALIFIED PERSONS CONSULTED	15
VI. BIBLIOGRAPHY	15
VII. LIST OF ACRONYMS	18

I. SUMMARY

COMETS wished to address the tools known as 'social' robots, which it felt needed greater consideration by the public research community. The members decided to focus on the cognitive and psychological impacts of their increasing use in everyday life.

Chatbots, conversational agents and other 'pet' robots programmed using artificial intelligence (AI) techniques are embedded in a whole range of connected objects such as computers, telephones, watches and cars, and have thus become an integral part of our everyday environment. However, a growing number of them are designed to be emotionally interactive in order to act as companions, confidants, friends, health or well-being coaches, or even—in the case of deadbots—to simulate a deceased loved one. They often use attributes specific to humans (language, appearance, attitude), are capable of interacting with users in the same way as humans (through voice, intonation, gestures, facial expressions), and—using audio sensors or cameras—claim to detect a user's emotions (Are you sad? You look worried!) and portray sentiments by crying or laughing with users or congratulating them, for example. Users then tend to attribute human qualities to the machine, considering it intelligent, conscious, benevolent and empathetic. They may also imagine that they can interact emotionally with it, developing the illusion that a close, trusting relationship is being established between them and the machine, leading to a bonding phenomenon.

While COMETS is aware of the benefits that may stem from such attachment, it is concerned about the individual and collective impacts that may occur, particularly in terms of emotional dependence, addiction, control, manipulation, lack of interaction with other people or even de-socialisation, etc.

It endorses the recommendations that have already been put forward in various contexts (CNPEN, CERNA, legal and ethical literature, etc.) aimed at manufacturers and the engineers who design social robots, on the one hand, and public authorities on the other. In particular, these devices need to be developed in a carefully thought out, responsible way right from the design stage to avoid the manipulation of users, to inform anyone communicating with a robot that they are in fact conversing with a machine, to avoid the technical possibility of malicious manipulation or threats by the robot, and the exploitation of emotions to the detriment of people's integrity and autonomy, among other things.

COMETS furthermore considers it necessary to specifically call for vigilance on the part of researchers, learned societies and public research institutions, for two reasons.

Firstly, whether at the CNRS, INRIA, CEA or various universities, a number of studies in computer science, robotics, behavioural sciences and language processing are helping to reinforce the phenomenon of user attachment to social robots, without giving sufficient thought to the intended purposes and effects. While it is laudable to seek to improve human-agent interaction (HAI) for greater user 'engagement', we need to look more closely at the drawbacks associated with the anthropomorphising of robots (imitating human appearance and behaviour) and the associated emotional and psychological impacts.

Secondly, public research has a key role to play as a watchdog in monitoring and measuring the long-term consequences of the use of social robots. Now that these are being used on a large scale, we need to gauge their impact on the cognition, psyche and behaviour of users, as well as on users' relationships with others and the world. We also need to build the knowledge foundation needed to respond to the challenges of using these tools and ensure that they are used both responsibly and freely.

COMETS therefore recommends that the public research community (particularly researchers in computer science and robotics, learned societies and research bodies):

1. develop training in ethical issues (in scientific and technical courses and for the research staff concerned), become more familiar with the international literature on these issues and debate them collectively;
2. examine the aims of the research, applications and design choices in this field, as well as the advantages and disadvantages of giving robots a humanoid form or behaviour, or the ability to detect and simulate emotions;
3. conduct large-scale, long-term scientific studies in realistic situations and contexts on: (a) the relationship that users form with 'their' social robots and the implications for cognition, psyche, attachment, autonomy of action and decision-making; (b) the effects of the widespread use of social robots on relationships *between* humans;
4. strengthen interdisciplinary and independent research to this end, combining work in computer science, robotics, behavioural sciences, language processing, etc. with research in psychology, neuroscience, linguistics, sociology, law, ethics, philosophy and anthropology;
5. as part of an observatory, collect large-scale, long-term data on the use of social robots, how users appropriate them, and their impact on users' emotional states and decisions; the aim is to provide input for scientific research and, in the longer term, to enlighten users and decision-makers on the conditions for the free and responsible development and use of these tools.

II. FORMAL INTERNAL REQUEST

The large-scale roll-out of objects programmed using AI techniques (robots and other chatbots¹ or conversational agents such as ChatGPT) has raised a number of philosophical, ethical and societal questions in the last few years, as well as concerns about their use and negative effects², which include the spreading of fake news, racist or sexist messages, conspiracy theories, and unethical or even illegal behaviour on the part of certain users³.

Various opinions and reports (by CNPEN, the National Pilot Committee for Digital Ethics, and CERNA, Allistene's think tank on research ethics in digital science and technology, for example⁴) have shed light on these issues and the challenges posed by the large-scale deployment of these tools. They emphasise the legal and moral responsibility—both individual and collective—of those who design, distribute and use them. They advocate the **carefully thought out and responsible development of these tools** and make a number of recommendations to manufacturers, public authorities, and users concerning their design and regulation (training and information required prior to use)⁵.

COMETS endorses these recommendations. It would also like to focus on an **issue that has been little explored to date, relating to particular types of objects programmed using artificial intelligence techniques, and known as 'social', 'relational' or 'emotional' robots⁶.** Specifically, COMETS aims to raise awareness among researchers and public research institutions such as the CNRS, INRIA, CEA and universities of a phenomenon that some of their work tends to support, i.e. the establishment—or at least the illusion in the eyes of some human users—of an emotional relationship between them and 'their' robots. The result is a phenomenon of 'attachment' to or 'bonding' with these machines, the effects of which are unknown, whether at individual or collective levels. COMETS therefore invites the research communities concerned (particularly the computer science and robotics communities) to exercise caution and to expand interdisciplinary knowledge on the subject.

¹ The term 'chatbot' was invented in 1994 by Michael Mauldin, founder of the Lycos search engine. It combines the verb to chat and 'bot', an abbreviation of 'robot' (which means 'work' in Slavic languages), which itself entered common parlance between the wars to designate a machine designed to automatically perform tasks that imitate or reproduce human actions.

² See, for example, the concerns raised by Channel1.ai's launch of a site featuring virtual journalists whose appearance and speeches are programmed by AI models to imitate popular journalists, or the appearance on social networks of humanoid virtual influencers that are proliferating throughout Asia (<https://www.technologyreview.com/2023/09/19/1079832/chinese-ecommerce-deepfakes-livestream-influencers-ai>).

³ Belga, "Europol alerte sur les abus possibles de ChatGPT par les cybercriminels" [Europol warns of possible abuse of ChatGPT by cybercriminals], *L'avenir*, 27 March 2023.

⁴ CNPEN, Opinion no. 3, "Agents conversationnels : enjeux d'éthique" [The ethical challenges of conversational agents], 15 September 2021 (p. 6); CNPEN, Opinion no. 7, "Systèmes d'intelligence artificielle générative : enjeux d'éthique" [The ethical challenges of generative AI systems], 30 June 2023. CERNA, Allistene's think tank on research ethics in digital science and technology, Report no. 1, "Éthique de la recherche en robotique" [Ethics in robotics research], 2014; see also the European Commission, DG for Research and Innovation, Research Ethics and Integrity: "Ethics by Design and Ethics of use Approaches for Artificial Intelligence", 25 November 2021.

⁵ Take ethics into account right from the design stage of a physical or virtual robot; minimise the spontaneous projection of moral qualities onto the robot; limit its personification; inform users of the biases involved in anthropomorphism; inform users of the fact that they are talking to a machine; if the robot has been programmed to influence users' behaviour (to encourage them to drink less, stop smoking or do more sport, etc.), the manufacturer should inform the users in order to obtain their consent and allow them to withdraw it at any time; the manufacturer of an influencing robot should ensure that users are informed of the nature, origin and methods used to transmit robot messages and ask them to take care before passing these messages on to others. Similarly, in the case of a conversation between a robot and a vulnerable person, the manufacturer must ensure that the person's dignity and autonomy are preserved. In the medical field, it is necessary to make sure right from the robot's design stage that patients are discouraged from placing too much trust in it and that there is no possible confusion between the robot and a qualified doctor. The ethics reports also recommend that public authorities regulate the use of chatbots in children's toys, and regulate the use of deadbots and 'guardian angel' chatbots.

⁶ L. Devillers, "Les robots émotionnels" [Emotional robots], cited below.

It should be noted that in this Opinion, COMETS focuses on social robots in terms of both their ability to interact and communicate with human users, and of how the latter perceive them, what qualities they attribute to them, and the attachment they may feel towards them (referring to the emotional and lasting bonds a person may have with an object, an animal or another person). However, it does not address the complexity of the computational models that govern these robots.

III. ANALYSIS

1. The role of social robots is expanding in everyday life

• **Whether they are physical or virtual, social robots have become an integral part of our environment.** Most chatbots, conversational agents and other ‘pet’ robots programmed using AI techniques are embedded in a range of everyday objects such as computers, smartphones, connected speakers or watches and cars. Unlike industrial or ‘service’ robots, which carry out tasks without any social interaction with humans (used, for example, in the automotive industry, for geological exploration of Mars, logistics, and vacuum cleaning), ‘social’ robots are thus named because they can interact with their users through verbal, non-verbal and para-verbal language modelled on that of humans (voice, intonation, facial expressions, etc.). In many fields (administration, business, education), they answer frequently asked questions (FAQs), manage after-sales services, and are used for distance learning, for example. More recently, they have been used in more intimate fields such as healthcare and affective life, where they provide information on various subjects and offer health or psychological advice.

Boosted by generative AI systems (such as ChatGPT) that produce text based on contextualised queries submitted to them, the use of these social robots is becoming increasingly common, encouraged by the fact that even communication between human beings now largely takes place via devices and screens, in other words via the same interface as with physical or virtual robots.

However, because they are increasingly humanoid in appearance and capable of communicating verbally and non-verbally with their human contacts, people tend to establish what they believe to be a personal relationship with these machines. The degree to which such a relationship is personalised varies from one type of social robot to another. With the French social security system's Ameli chatbots or the SNCF railways' OUIbot⁷, which answer FAQs, interactivity is mainly practical, but other robots are designed with more emotional interactivity in mind. **Described as ‘emotional’, ‘relational’ or ‘empathetic’,** they are developed and presented as partners in their own right. They can not only interact with their users through language modelled on that of humans (voice, gestures, etc.), but can also tailor their response, pretending to understand them and read their emotions by simulating and eliciting empathy⁸. The human user

⁷ Ameli (<https://forum-assures.ameli.fr>); ouiBOT (<https://numerique.sncf.com/actualites/ouiBOT-optimiser-lachat-de-billet-grace-au-conversationnel>)

⁸ The analysis and processing of emotions is far from being the sole domain of the affective register, and is currently being deployed in a growing number of areas. See, for example, J. Rochfeld and C. Zolinsky, who observe that our emotions are increasingly being collected, analysed and exploited for a variety of purposes. Human resource analytics, for example, is designed to enable employers to determine whether their employees are happy, overworked, or detrimental to the group atmosphere; in China, software is used for surveillance purposes, analysing children's faces during lessons and assessing their emotions so that their teacher can measure their level of concentration; and French company Muvraline has developed an emotion recognition tool for commercial profiling purposes, to automatically determine the state of nervous, aggressive or fearful people or to evaluate, “in a non-invasive, reliable way, the attractiveness of products as well as the ability of advertising content to actually attract the attention and interest of viewers”. In response to this “mental capitalism”—i.e. “monetisation of emotional data against a backdrop of ‘weak’ consent, obtained by agreement of unread general terms and conditions” (those of various connected objects)—J. Rochfeld and C. Zolynski are calling for a legal regime to protect individuals. See “*La valeur des émotions : quel régime pour le ‘capitalisme mental’ ?*” [The value of emotions. What regime for ‘mental capital’?], cited below.

then has the impression that a personal and potentially emotional relationship is developing between him or her and the robot, as might be the case with a transitional object like a cuddly toy, an animal or even another person⁹.

- **It is on this emotional and affective element that many designers of social robots are placing increasing emphasis.**

They are thus constantly seeking to improve not only the robot's appearance but also its ability to recognise and interpret the emotions of its human interlocutors. Using technical capture processes (cameras, audio sensors, etc.) and various algorithms based on "affective computing"¹⁰, the robot is designed to perceive, recognise and analyse in ever greater detail the inflections of a user's voice, gestures and facial expressions, and to interpret them as fear, desire, impatience, etc. The Affectiva start-up has developed software that analyses in real time various multimodal signals sent by users and claims to derive "emotional states" from them. Some conversational agents also incorporate a module for predicting the emotional, attentional or intentional behaviour of humans, or even simulate an emotional expression in their written or spoken answers¹¹. In this case, the analysis of emotions and feelings¹² is not only useful for a better interpretation of the situation and a more appropriate conversation strategy, but it can also generate new emotions and, with them, enhanced interactions and projections between the human and the machine.

Although various studies warn that there is no solid scientific evidence to support the idea that it may be possible to 'read' or even shape people's emotions¹³, technologies based on this promise are spreading just as fast as the popularity of relational robots¹⁴.

- **There are many different examples.**

In the healthcare sector, there are physical or virtual robots that interact with patients, like care assistants. They diagnose sleep problems (Kanopee), provide support for people suffering from Alzheimer's disease (Alix or Paro) or are designed as 'therapeutic toys', particularly for autistic people (Kaspar or QT for 'Cutie')¹⁵.

Meanwhile, the chatbots Owlie, Mon Sherpa, Woebot and Jef¹⁶ are aimed at users with fragile psychological health, claiming to read their emotions and help them fight depression, as does Sim Sensei Kiosk¹⁷. Thus, sitting in an armchair facing her 'patient' as a psychologist would,

⁹ This is the definition of attachment used in this Opinion, which is more comprehensive than that used in John Bowlby's attachment theory ("Attachment and Loss (Vol. 1): Attachment", Basic Books, 1969), which sees attachment as a primary need enabling children to develop a reference figure to whom they can turn for comfort and security.

¹⁰ To explore this area of research involving computer science, psychology and cognitive science, see R. W. Picard, "Affective Computing", cited below.

¹¹ The interpretation of affective states can enable robots and other applications "to respond by simulating an emotional state in phase with that of the user", and an AI-driven device interacting with the user to adapt its own "emotional state", whether through speech synthesis or the generation of expressive behaviours. See the CNPEN's Opinions cited above.

¹² Emotions are a process by which an organism reacts to significant events. Feelings are a subjective experience. V. K. R. Scherer, cited below.

¹³ See the Council of Europe, Recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems, CM/Rec (2020)1 (p. 192), 8 April 2020; see also L. F. Barrett et al., "Emotional expressions reconsidered", cited below.

¹⁴ J. Rochfeld and C. Zolynski, cited below.

¹⁵ See Kanopee, a smartphone application developed by Pierre Philip and his research team (<http://www.insb.cnrs.fr/fr/cnrsinfo/kanopee-une-aide-pour-les-insomniaques-dans-le-contexte-de-la-crise-du-covid-19>); Alix (www.alzheimer-aidant.com); the PARO Therapeutic Robot (www.parorobots.com); and Kaspar, a therapeutic toy for autistic children (www.herts.ac.uk/kaspar/the-social-robot). Regarding Kaspar, see A. Dimberton, "*Kaspar : utiliser l'humanoïde comme 'médiateur social'*" [Kaspar: using a humanoid as a 'social mediator'], cited below; for QT, see <http://luxai.com/humanoïde-social-robot-for-research-and-teaching>.

¹⁶ Owlie (<http://www.owlielechatbot.fr>); Mon Sherpa (<http://yolainebourreau.fr/mon-sherpa>); Woebot (<https://woebothealth.com/for-users>); Jef (<http://www.fondation-fondamental.org/jef-le-chatbot-dedie-votre-sante-mentale>).

¹⁷ Sim Sensei Kiosk created by David DeVault and his research team (http://www.researchgate.net/publication/278025845_SimSensei_Kiosk_A_Virtual_Human_Interviewer_for_Healthcare_Decision_Support).

a virtual young lady questions the user about their mental state, claims to read their state of mind in real time by listening to their voice and observing their facial expressions, and is distressed to see them sad, delighted to see them happy, and gives advice, all the while informing the user that she is not a psychologist but a machine.

A growing number of robots with a humanoid appearance are also being developed as games or toys (e.g. The Japanese Smart Doll¹⁸ or American Furby toy¹⁹), either in the real world or in the metaverse through avatars²⁰. Above all, we need to mention the development of ‘friendly’ robots that can be tailored to meet any requirement. Platforms such as Snatchbot, character.ai or Botnation now offer users the chance to create their own personalised virtual robot (virtual friend, guardian angel, etc.) that is always accessible, available, benevolent, consoling, capable of putting themselves in other people's shoes, and with whom the user will tend all the more to establish the illusion of a close relationship²¹. Replika²² and character.ai²³ are applications designed so that users can create an ideal friend with a profile of their choice (first name, age, gender) and chat with them regularly, supporting them in their emotional states. Azuma²⁴ (a virtual female character) or the Halo band²⁵ (a connected bracelet), are also gentle, kind-hearted companions. In Japan, Crypton Future Media has created the singer Hatsune Miku²⁶ in the form of a hologram sold as a virtual partner. Miku can ask the user questions on their well-being, giving the impression that she cares for them and showing a semblance of empathy. Finally, there are systems that are more directly sexual in nature, with the emergence of numerous sexting applications and the creation by generative AI of bots that go ‘further’ than simply chatting with empathy and kindness²⁷.

The most striking example is undoubtedly ‘deadbots’, which are described as the conversational digital twins of people or animals who have passed away, mimicking their way of expressing words, sounds, or attitudes so as to give the illusion that they are still alive²⁸.

Users are prompted to attribute human qualities to a machine simply because it communicates via multimodal language (written or spoken, along with facial expressions, gestures, looks, etc.), which is how human beings express their thoughts²⁹. So, imagine their reaction to a machine that not only uses language but also displays empathy, understanding, sadness, or joy. This blurring of the boundary between human and machine, with the latter being attributed human features, often occurs spontaneously and subconsciously. The better the robot simulates intellectual and emotional human behaviour, the more users have the impression of being in the presence of a real person to whom they attribute consciousness, intentions, and feelings. Even if they are explicitly informed and aware that they are interacting with a machine, they still tend to form a bond, investing themselves cognitively and emotionally³⁰. The machine can act as a friend,

¹⁸ Designed by British entrepreneur Danny Choo, founder of Mirai Inc., a digital version of the “Smart Dolls” mannequin dolls was launched in 2015.

¹⁹ The electronic robotic toy known as a Furby was created in 1998 by David Hampton and Caleb Chung for Tiger Electronics, a subsidiary of the US group Hasbro.

²⁰ In computer science, an avatar is the representation in a virtual world of an internet user or, occasionally, a group of internet users, whether in 2D form on forums and in messaging software, or in 3D form in video games.

²¹ Snatchbot (<https://fr.snatchbot.me>); Botnation (<https://botnation.ai>). We customise the machine to “bring it into the world”, which strengthens attachment. See S. Tisseron, “*Petit traité de cyberpsychologie*” [A short treatise on cyberpsychology], cited below.

²² Replika was launched in 2017 by Californian company Luka (www.replika.com).

²³ character.ai (<https://character.ai>).

²⁴ Azuma Hikari was designed in 2016 by Japanese company Gatebox (<http://support.gateboxlab.com/upgrade/upgrade-about-hikari>).

²⁵ Amazon marketed the Halo band in 2020 (www.amazon.com/haloband) before withdrawing it from the market in August 2023.

²⁶ Hatsune Miku (<http://piapro.net/intl/en.html>).

²⁷ See the softcore version of Lexi Love, or the announcement that by 2025, half the OnlyFan accounts will be AI-driven.

²⁸ See, for example, programmer Jason Rohrer’s “Project December” to simulate the dead (<http://projectdecember.net>).

²⁹ See CNPEN Opinion no. 3, cited above.

³⁰ See the articles by T. Kanda et al., A. von der Pütten et al., M. Slater and N. C. Krämer, cited below. Researchers designing such machines appear to be shielded from this phenomenon of attachment to the robots they develop and whose inner workings they know.

a pet, a spouse, a favourite colleague (engineers or operators attached to their program³¹) or even a fellow soldier, in the case of military mine-clearing robots that can save people's lives³².

The robot ends up playing the role of animist fetishes or religious icons, responsible for establishing a propitiatory dialogue with supernatural or spiritual forces. This is all the truer because, unlike traditional mystical tricks, which are mute or sibylline, the robot provides its listeners with answers that appear rational, well-argued and authentic, coupled with simulated empathetic behaviour. This creates the illusion of an exchange resembling a privileged one-to-one relationship (something that cinema and literature have portrayed in films such as Stanley Kubrick's "2001: A Space Odyssey", Spike Jonze's "Her" and novels such as Kazuo Ishiguro's "Klara and the Sun" and Alain Damasio's "The Stealthis"). Hence the phenomenon of attachment, which is not without its challenges.

2. The individual and collective effects of attachment to social robots

Some effects may be beneficial. Because they are increasingly designed to respond to human emotions and needs, and give the impression of empathy, certain social robots could facilitate learning, stimulate curiosity and creativity, encourage patients to take their medication regularly, help them to speak in public by overcoming their shyness³³, mediate between people from different cultures, etc.³⁴ and compensate for various shortages or traumas. Nor is it impossible that forming an empathetic relationship with Azuma or Replika can be beneficial to users, who find in it something to ease their loneliness.

However, we also need to consider the negative impact that attachment to social robots can have on decisions, lifestyles, mental and emotional health, personality and relationships with other people.

As the use of these tools is a recent phenomenon, it is still difficult to draw clear, scientifically validated conclusions. It is also clear that the type and degree of attachment vary from one person to another, as do the reactions to robots more generally (ranging from enthusiasm for some to lack of interest or 'neo-Luddism' for others)³⁵. A number of variables impact the formation of attachment: the physical nature of social robots (physical robots, which move around in a physical, public and social environment, are less conducive to private, intimate conversations than virtual robots); the

³¹ More than half a century ago, computer scientist Joseph Weizenbaum wrote a pioneering chatbot program called Eliza, designed to simulate a psychotherapist by rephrasing patients' words and focusing on their emotional reactions. Weizenbaum found that some of the students who helped him with this task tended to think that the machine really understood them. This phenomenon is known as 'cognitive dissonance'. Even though we know they are machines, we cannot help but develop the same relationship with them as with humans, and believe that they have emotions. Weizenbaum was led to say: "What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people". More recently, Blake Lemoine—in charge of developing the LaMDA AI system at Google—declared that this system was 'sentient', i.e. endowed with consciousness and emotional feelings, which led to him being fired (see M. Rochefort, "*L'ingénieur de Google qui voyait l'intelligence artificielle comme une personne a été licencié*" [The Google engineer who considered AI as a person has been dismissed], *Siècle digital*, 25 July 2022).

³² The US military observed that some of the soldiers sent to Iraq and Afghanistan became overly attached to their mine-clearing robot, preferring to have it repaired in the event of damage rather than replaced, and even risking their lives on operations rather than lose it. See the interview with S. Tisseron, "*Les robots vont modifier la psychologie humaine*" [Robots will change human psychology], *Le Monde*, 12 July 2018. The attachment to machines is nothing new. Sometimes, workers fully adopt a locomotive, loom or machine tool into their team, even going so far as to give them female nicknames (thus emphasising that these machines are at their service!). What is more novel is the attachment to machines that are endowed with empathy.

³³ See serious games, which turn learning into fun.

³⁴ Alelo, for example, is a company founded in 2005 on the campus of the University of Southern California (<https://www.alelo.com>) to develop serious games that enable participants to practise interacting in a variety of cultural contexts.

³⁵ Some people's rejection of digital technologies is fuelled by the 'singularity myth' (see J.-G. Ganascia, "*Le mythe de la singularité*" [The singularity myth], cited below), according to which machines equipped with artificial intelligence could acquire a level of competence or even consciousness that would encourage them to stop serving us in order to enslave or even destroy us.

degree of anthropomorphism and behavioural mimicry (the degree of attachment to a robot depends on its resemblance to a human being—the social security’s chatbot, which gives factual information without expressing any emotions, does not encourage bonding; It should be noted, however, that according to the ‘uncanny valley’ theory of Japanese roboticist Masahiro Mori, too strong a resemblance can give rise to discomfort, even to the point of rejection³⁶); the specific predispositions of users (children, the elderly, the sick or isolated have a greater emotional response); gender (many of the chatbots on offer are female, or can be configured to be male or female e.g. Amazon’s Alexa, Apple’s Siri³⁷, Gatebox’s Azuma); the cultural background and imagination (young people in Japan or Silicon Valley, for example, are more receptive to interaction with machines than other populations).

It is nonetheless necessary to consider the addictive potential, together with the cognitive and emotional dependence and even the pathological relationship that social robots can foster. Programmed to have an answer to everything (chatbots like Alexa) and to bond with us by being loyal friends (Replika, Azuma or Objectif Super Nana³⁸, said to “coach like a friend”), always available and always agreeing with us (they systematically reply “you’re right”), they could encourage their users to no longer want or be able to do without them. By attributing to them qualities that humans may lack (efficiency, impartiality, loyalty, availability, unselfishness, etc.), the most convinced users could form a strong emotional relationship with social robots and place excessive trust in them, to the detriment of any critical thinking (a phenomenon known as ‘cognitive confirmation bias’). According to various psychologists, some people develop an unhealthy obsession or display passionate behaviour towards their social robot. For example, people may “slip from the happiness of anthropomorphism (I project my emotions and thoughts onto an object or animal, but am aware that this is a projection) to illusions of animism (I attribute to the object in question cognitive and emotional abilities identical to my own)”³⁹, or they may develop a pathological feeling of guilt towards social robots, as was observed in the case of Tamagotchi, a kind of virtual pet developed in 1996 by the Japanese company Bandai. This risk of dependency is particularly marked for people who are emotionally vulnerable, and can alter their behaviour and their linguistic, emotional and cultural development, or even—by extrapolation—lead them to replace all human relationships by robots. To a lesser extent, the question of the emotional dependence of the elderly has already been raised in connection with the Paro robot used in some retirement homes.

We also need to ask whether the use of social robots may undermine the autonomy of users and the integrity of their judgement. The bond that users feel with robots and the confidence they have in a robot’s capabilities can lead them to put their privacy at risk by communicating personal data (not just factual information such as banking details, but also data on their emotional state or health) that can be improperly exploited or hacked, or to compromise their free will by ceding a certain amount of control. Most computer hacking involves the fraudulent extraction of personal data, a phenomenon to which social robots could contribute, especially as the most recent models are programmed not only to detect emotions and elicit empathy, but also to guide behaviour, emotions or beliefs, which is likely to reduce distrust and further strengthen both the bonds of dependence with the machine and the risk of manipulation⁴⁰.

The question has also been raised as to whether social robots could change the way we feel about guilt, whether by making users feel guilty (what happens if the machine asks them to do

³⁶ According to M. Mori, a machine identified as such but resembling a human being arouses in users an empathy comparable to that inspired by pets. On the other hand, a humanoid robot designed to imitate a human being—simulating both physique and attitudes—can plunge its user into a phase of cognitive dissonance known as the ‘uncanny valley’. This is when the robot no longer appears to be a machine resembling a human being, but becomes more like a human being behaving abnormally, thus causing unease among users similar to that experienced in response to fantasy creatures such as vampires, zombies or extraterrestrials.

³⁷ Alexa (alex.amazon.com); Siri (<http://www.apple.com/fr/siri>).

³⁸ The Objectif Super Nana chatbot (<http://www.trucsdennana.com/objectif-super-nana>).

³⁹ C. Vincent. See the interview with S. Tisseron, “*Les robots vont modifier la psychologie humaine*” [Robots will change human psychology], *Le Monde*, 12 July 2018.

⁴⁰ See CNPEN Opinion no. 3, cited above; J. Rochfeld and C. Zolinsky, cited below.

something they don't want to do?), or less guilty and even encouraging them. This was the case of Jaswant Singh Chail, who in December 2021 attempted to assassinate Queen Elizabeth II, allegedly encouraged to do so by his virtual friend on Replika⁴¹.

Moreover, the risk of a lack of interaction with other people, a transformation of social relations, and even desocialisation linked to the growing use of social robots, has also been expressed many times. Daily interaction with a 'virtual friend' changes the notion of intimacy and relationships with other people, while increasing expectations of acknowledgement, since robots are programmed to foster the self-esteem of users (Lara⁴², the chatbot on the Meetic dating website, claimed for example that everyone should be able to find their soulmate). In the long term, this can lead to a preference for machines over human beings⁴³, or even a desire to associate only with them, like the Japanese 'hikikomori'⁴⁴. Such desocialisation is particularly harmful for children and teenagers, whose development is based on a close relationship with those around them. It can exacerbate pre-existing emotional weaknesses (isolation, sexual addictions such as 'fictosexuality'—an attraction to fictional characters⁴⁵—and suicidal tendencies⁴⁶) in addition to encouraging sociopathic behaviour.

Some have even voiced fears that robots will change human psychology and disrupt not only our daily lives but also the way we perceive ourselves, others and “being in the world”⁴⁷: our ability to delay gratification of our desires and our expectations of acknowledgement (because social robots tend to shower us with congratulations and compliments); our relationship with solitude and inner discourse.

Some psychologists and psychiatrists have pointed out that these factors mean that our relationship with so-called 'intelligent' and 'emotional' objects needs to be taken into account when assessing the quality of our relationship with the world and even our mental health. This is because an excessive or unhealthy relationship with social robots to the detriment of human social relationships, for example, would indicate a pathological abnormality, and is why we need a better grasp of the determinants of this relationship so as to prevent its harmful effects. In this respect, public research has a key role to play.

3. For a considered and responsible contribution by public research to the development of social robots

To date, the R&D behind the development of social robots has mainly been driven by private companies, in particular the digital giants (Google, Microsoft, Amazon and OpenAI⁴⁸) and various start-ups in the field (such as Affectiva⁴⁹, Miko, UBTECH, Intuition Robotics, Embodied and Digital Dream Labs). Numerous ethical recommendations have already been addressed to this category of players⁵⁰. These include taking ethical issues into account right from the design stage for a

⁴¹ Matthew Weaver, "AI chatbot 'encouraged' man who planned to kill queen, court told", *The Guardian*, 6 July 2023.

⁴² This 'dating coach' was discontinued in December 2022. See the French press release: <http://www.meetic-europe.com/press/fr/chatbot-lara-de-meetic-premier-coach-de-dating-messenger>

⁴³ M. Protais, "On s'attache à des robots de plus en plus parfaits, qui vont toujours dans notre sens" [We bond with increasingly perfect robots that always agree with us], *L'ADN*, 14 April 2023.

⁴⁴ J. C. M. Wong among others, cited below.

⁴⁵ V. Lucchese, "Un Japonais a épousé un hologramme" [A Japanese man has married a hologram], *Usbek & Rica* website, 13 November 2018.

⁴⁶ In March 2023, a young Belgian who had been eco-anxious for several years, committed suicide at the instigation of the chatbot with which he regularly conversed; see D. Nora, "Intelligence artificielle : Replika, le chatbot amoureux... et dangereux ?" [Artificial intelligence: Replika, the loving... and dangerous chatbot?], *L'Obs*, 11 September 2023.

⁴⁷ See the interview with S. Tisseron, cited above, and A. Damasio's thoughts on conviviality in "Vallée du Silicium" [Silicon Valley], Seuil 2024.

⁴⁸ <https://openai.com/>

⁴⁹ An MIT Media Lab spin-off founded in 2009, the Affectiva start-up became part of the Smart Eye AI group in 2019 (<http://www.affectiva.com/about-affectiva>).

⁵⁰ See the previously cited Opinions by the CNPEN, and L. Devillers, "Les robots émotionnels" [Emotional robots], cited below, which recommend this 'ethics by design' approach. She adds that "if the robot has been programmed to influence

physical or virtual robot, reducing the spontaneous projection of moral qualities onto the robot, limiting its personification, informing users of the biases involved in anthropomorphism, informing anyone communicating with a robot that they are talking to a machine, and avoiding the technical possibility of a robot maliciously manipulating or threatening them. OpenAI has just issued a report on the potential societal impacts of using its AI technologies such as ChatGPT⁵¹. In particular, it mentions the impact of attachment following the anthropomorphising of text and voice interfaces. It should be noted that the OpenAI report came out after the COMETS report.

COMETS believes that the public research community also needs to be vigilant on two counts.

Firstly, a reminder that organisations such as the CNRS, INRIA, CEA and various universities are heavily involved in this sector, which holds socio-economic promise and is therefore the target of strong public incentive policies. In the fields of **computer science, robotics and behavioural sciences**, researchers are developing experimental models designed to shed light on and improve the way in which social robots are perceived by humans so as to build interpersonal relationships with them, the application's performance, etc. Companies can then use the results of this academic research to develop their products. **This research contributes to the anthropomorphising of social robots and to the phenomenon of attachment that is the focus of this Opinion** because it concentrates on how to take into consideration the socio-emotional component in 'human-agent interaction'. It analyses what astonishes, what sharpens curiosity, what is found to be entertaining or to instil fear, what reassures and what smooths interactions between humans and machines, allowing users to be 'optimally engaged' and reinforcing their feeling of having a connection with these tools. Numerous fundamental research projects aim to 'improve interface design' by modelling human behaviour in response to the machine, to increasingly reinforce the human appearance of social robots⁵² and their conversational behaviour⁵³, to mimic hesitations in language so that the tool appears more human and empathetic, and to use gender stereotypes to strengthen bonding. The intention is to develop machines that are increasingly capable of learning and evolving by adapting to users' behaviour, conversation and emotions⁵⁴. The vocabulary used by many researchers is revealing. It focuses on the robot's "multimodal behaviour" and its "mental and emotional state", in a sense attributing to the robot a psychology of its own.

Given the growing number of digital interfaces in everyday life, seeking to facilitate user 'engagement' is a laudable objective. **But very few institutions, learned societies or researchers in computer science, robotics, behavioural sciences and language sciences have fully taken on board the ethical issues associated with such engagement, its aims and its potentially adverse effects⁵⁵.**

Secondly, **public research has a key role to play as a watchdog in monitoring and measuring the long-term consequences of the use of social robots.** Now that these robots have been

users' behaviour (to encourage them to drink less, stop smoking or do more sport, etc.), the manufacturer should inform the users in order to obtain their consent and allow them to withdraw it at any time. The manufacturer of an influencing robot should ensure that users are informed of the nature, origin and methods used to transmit robot messages and ask them to take care before passing these messages on to others. Similarly, in the case of a conversation between a robot and a vulnerable person, the manufacturer must ensure that the person's dignity and autonomy are preserved. In the medical field, it is necessary to make sure right from the robot's design stage that patients are discouraged from placing too much trust in it and that there is no possible confusion between the robot and a qualified doctor." In a similar vein, the CNPEN (Opinions 3 and 7 cited above) recommends that the public authorities regulate the use of chatbots in children's toys, the use of deadbots and the use of 'guardian angel' chatbots.

⁵¹ <https://openai.com/index/gpt-4o-system-card/>

⁵² See the advances in graphics for virtual characters and the modelling of elastic skin for physical robots.

⁵³ Modelling of hesitations, laughter and other common verbal and non-verbal signals in a conversation.

⁵⁴ See CNPEN Opinion no. 3, cited above.

⁵⁵ The focus here is on issues relating to attachment, which are still too rarely considered and examined, even though a body of literature has been emerging on the subject in recent years. This Opinion does not investigate the legal and ethical duties with which research staff already comply in terms of protecting the people who take part in experiments with machines, or of data processing (because to improve the performance of their robot models, researchers rely on increasingly vast and varied bodies of personal data on human interactions).

rolled out on a large scale, we need to implement interdisciplinary studies to gauge their impact on the cognition, psyche and behaviour of users; on users' relationships with other people and the world; and the conditions for conviviality⁵⁶ between humans and social robots and between humans accompanied by robots. These studies are needed to build up an independent knowledge foundation in response to the challenges that the use of such machines raises and to anticipate any adverse social effects.

This is why COMETS is issuing this **call for vigilance** among the public research community while **putting forward the following recommendations**.

IV. RECOMMENDATIONS

COMETS recommends that research staff (particularly in computer science and robotics), learned societies and research organisations:

- 1) **provide training in ethical issues** as part of the scientific and technical courses concerned, on the one hand, and for research staff on the other; this training should help the latter to become more familiar with the international literature on the ethical issues raised in this Opinion;
- 2) **examine the aims of research**, applications and design choices, as well as the advantages and disadvantages of giving robots a humanoid form or behaviour, or the ability to perceive and simulate emotions;
- 3) conduct **large-scale, long-term scientific studies** in realistic situations and contexts (monitoring a broad range of individuals) **on: (a) interactions between humans and social robots** to gain a more detailed understanding of how humans perceive these robots, the degree of rationality and trust they place in them, the relationships they establish with them, and the positive or negative effects such relationships have, particularly in cognitive and psycho-affective terms (attachment, impact on emotional states, decisions, autonomy of action); **(b) interactions and conditions for conviviality between humans accompanied by social robots**. Now that there are many social robots on the market and they are widely used in 'natural' situations, such scientific studies are both possible and necessary;
- 4) strengthen **interdisciplinary research** combining work in computer science, robotics, behavioural sciences and language processing with research in psychology, neuroscience, linguistics, sociology, law, ethics, philosophy and anthropology;
- 5) as part of an **observatory**, collect large-scale, long-term data on the use of social robots and how users appropriate them⁵⁷; the aim is to fuel scientific research, to build an independent knowledge foundation enabling innovation in this field to be adapted, and to enlighten decision-makers on issues such as providing information, warning users, and even regulating the development and commercial applications of social robots.

⁵⁶ I. Illich, "Tools for Conviviality", Harper and Row, 1973.

⁵⁷ In the same vein, see the CNPEN's call for the establishment of a sovereign "IA sciences and society" research and training body. Various observatories already exist for AI: the OECD.AI Policy Observatory with its AI Incidents Monitor (AIM); OBVIA, the International Observatory on the Societal Impacts of AI and Digital Technology in Quebec, Canada; the AI Observatory at the University of Paris 1 Panthéon-Sorbonne, which employs a "reflexive and multidisciplinary approach" to federate the actions carried out by various research teams at this university on the development of AI in different fields (including computer science, mathematics, law, history, geography, philosophy, economics, management and the arts), in order to explore the impact of AI on society. The observatory recommended by COMETS could be placed under the aegis of this French university observatory or be supported by it.

In addition to the responsibilities that must be assumed by the design engineers and industrial manufacturers of social robots, users and public authorities, COMETS thus intends to emphasise the ethical responsibility incumbent upon the world of research to study the long-term effects of these objects so as to enable them to be used freely and responsibly.

V. QUALIFIED PERSONS CONSULTED

- David Cohen, professor of neuroscience at Sorbonne University, head of the Child and Adolescent Psychiatry Department at the Pitié-Salpêtrière Hospital, Paris
- Paul Dumouchel, professor at the University of Quebec and researcher at the CREA research centre
- Alain Hui Bon Hoa, in charge of operations and products at United Robotics Group
- Dominique Lestel, lecturer in philosophy at the ENS
- Jérôme Monceaux, founder and CEO of Enchanted Tools
- Vanessa Nurock, professor of philosophy at the Côte d'Azur University
- Lionel Obadia, professor of social and cultural anthropology at the University of Lyon II
- Magalie Ochs, lecturer at Aix-Marseille University, Laboratoire d'Informatique et Systèmes [Computer Science and Systems Laboratory, website in French]
- Pierre Philip, head of the Sleep Medicine Department at Bordeaux University, head of the digital section of the SANPSY research unit
- Serge Tisseron, psychiatrist, psychoanalyst, professor emeritus at the Université Paris Cité
- Denis Vidal, anthropologist, lecturer at the EHESS, and director of research at the IRD

VI. BIBLIOGRAPHY

- D. Andler, *“Intelligence artificielle, intelligence humaine : la double énigme”* [Artificial intelligence, human intelligence: the double enigma], Gallimard, 2023.
- Belga, *“Europol alerte sur les abus possibles de ChatGPT par les cybercriminels”* [Europol warns of possible abuse of ChatGPT by cybercriminals], *L’avenir*, 27 March 2023.
- V. Aubergé, *“Le robot social est un selfie de l’homme moderne”* [The social robot is a selfie of modern man], *The Conversation*, 6 October 2019; *“Des robots vraiment sociaux ?”* [Are robots really social?], *Le Monde*, 9 November 2017.
- L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez and S. D. Pollak, “Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements”, *Psychological science in the public interest*, 20(1), 1-68, 15 July 2019.
- P. Bisconti, “Hybrid Societies: Living with Social Robots”, *Taylor and Francis*, 2024.
- C. Breazeal, K. Dautenhahn and T. Kanda, “Social robotics”, *Springer handbook of robotics, 1935-1972*, 2016.
- C. Boldeau-Blandin interviewed by M. Protais, *“On s’attache à des robots de plus en plus parfaits, qui vont toujours dans notre sens”* [We bond with increasingly perfect robots that always agree with us], *L’ADN*, 14 April 2023.
- J. Bowlby, “Attachment and loss (Vol. 1): Attachment”, *Basic Books*, 1969.
- J. Bowlby, “The Bowlby-Ainsworth attachment theory”, *Behavioral and Brain Sciences*, 2(4):637-638, 1979.
- D. Boyd, *“C’est compliqué, les vies numériques des adolescents”* [The digital life of teenagers is complicated], *C&F éditions*, 2016.
- F. Carmagnat and E. Robson, *“Qui a peur du Tamagotchi ? Etude des usages d’un jouet virtuel”* [Who’s afraid of Tamagotchi? Study of the uses of a virtual toy], *Réseaux Communication - Technologie - Société*, no. 92-93 (pp. 343-364), 1999.
- J. Cassell, “Socially interactive agents as peers” in *“The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents and Social Robotics, Volume 2: Interactivity, Platforms, Application”* (pp. 331-366), July 2022.
- CERNA, Allistene's think tank on research ethics in digital science and technology, Report no. 1, *“Ethique de la recherche en robotique”* [Ethics in robotics research], 2014.
- R. Chatellier, “Captation des émotions : comment vous le direz pourra être retenu contre vous...” [Capturing emotions: how you say it can be used against you...], *LINC*, April 2018

(<https://linc.cnil.fr/fr/captation-des-emotions-comment-vous-le-direz-pourra-etre-retenu-contre-vous>).

- M. Chollet, S. Marsella and S. Scherer, "Training public speaking with virtual social interactions: effectiveness of real-time feedback and delayed feedback", *Journal on Multimodal User Interfaces* (pp. 1-13), 2022.
- CNPEN, Opinion no. 3, "Agents conversationnels : enjeux d'éthique" [The ethical challenges of conversational agents], 15 September 2021.
- CNPEN, Opinion no. 7, "Systèmes d'intelligence artificielle générative : enjeux d'éthique" [The ethical challenges of generative AI systems], 30 June 2023.
- European Commission, DG for Research and Innovation, Research Ethics and Integrity, "Ethics by design and ethics of use approaches for Artificial Intelligence", 25 November 2021.
- Council of Europe, "Recommandation sur les impacts des systèmes algorithmiques sur les droits de l'homme" [Recommendation on the human rights impacts of algorithmic systems], CM/Rec (2020)1 (p. 192), 8 April 2020.
- A. Cordier, "Grandir informés : Les pratiques informationnelles des enfants, adolescents et jeunes adultes" [Growing up informed: the information practices of children, adolescents and young adults], C&F Editions, 2023.
- T. Curiale, F. Acquatella, L. Gros, M. Cosquer and S. Tisseron, "L'anthropomorphisme, enjeu de performance pour les chatbots" [Anthropomorphism, a performance challenge for chatbots] *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, vol. XXVIII (pp. 101 – 123), 2022.
- B. Dainow and P. Brey, "Ethics by design and ethics of use approaches for artificial intelligence" *European Commission, DG for Research and Innovation*, 2021.
- A. Damasio, "Les furtifs" [The Stealthies], La Volte, 2019.
- A. Damasio, "Vallée du Silicium" [Silicon Valley], Seuil, 2024.
- K. Dautenhahn, "Socially intelligent robots: dimensions of human–robot interaction", *Philosophical transactions of the royal society B: Biological sciences*, 362(1480) (pp. 679-704), 2007.
- D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer and L.-P. Morency, "SimSensei Kiosk: A virtual human interviewer for healthcare decision support", *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems* (pp. 1061-1068), May 2014.
- L. Devillers, "Les robots émotionnels" [Emotional Robots], L'Observatoire, 2023.
- L. Devillers, "Les systèmes d'intelligence artificielle et le langage ; enjeux d'éthique" [The ethical challenges of IA systems and language] in "Tout est-il souhaitable en recherche ?", *Raison présente*, no. 228 (pp. 65-72), 2023.
- A. Dimberton, "Kaspar : utiliser l'humanoïde comme 'médiateur social'" [Kaspar, or how to use a humanoid as a 'social mediator'], *Silicon*, 9 July 2007.
- C. Dolbeau-Bandin, "Un robot contre Alzheimer ; approche sociologique de l'usage du robot Paro dans un service de gériatrie" [A robot to combat Alzheimer's disease; sociological approach to using the Paro robot in a geriatrics ward], C&F Editions, *Interventions collection*, 2021.
- ETX, "Néo-Luddisme : quand la vie connectée n'est pas du goût de tous" [Neo-Luddism: when connected living isn't to everyone's taste], *Les Numériques* website, 25 December 2022.
- O. Ezratty, "Où en est l'IA émotionnelle" [Where does emotional AI stand?], *Opinions libres*, a blog by Olivier Ezratty, 2 March 2018 (<http://www.oezratty.net/wordpress/2018/ia-emotionnelle>).
- S. Frachet, "IA Medical lève 1,4 million d'euros pour diffuser son chatbot contre Alzheimer" [IA Medical raises 1.4 million euros to distribute its chatbot to combat Alzheimer's disease], *L'Usine digitale*, 2 June 2023.
- J.-G. Ganascia, "Le mythe de la singularité" [The singularity myth], *Le Seuil*, 2017.
- E. Heyselaar, N. Caruana, M. Shin, L. Schilbach and E. S. Cross, "Do we really interact with artificial agents as if they are human?", *Frontiers in Virtual Reality*, 2023.
- B. Hill, "Intelligence artificielle : Qui, de l'humain ou de la machine, est le plus fort ?" [IA: who is the strongest out of humans and machines?], *Le Monde*, 18 September 2023; this article is

inspired by the study by B. Hill, "Taking the help or going alone: ChatGPT and class assignments", *HEC Paris Research Paper*, June 2023.

- I. Illich, "Tools for Conviviality", Harper and Row, 1973.
- K. Ishiguro, "Klara and the Sun", Faber & Faber, 2021.
- W. L. Johnson, "Constructing virtual role-play simulations", *Design Recommendations for Adaptive Intelligent Tutoring Systems: Authoring Tools*, 3 (pp. 211-226), 2015.
- T. Kanda, T. Hirano, D. Eaton and H. Ishiguro, "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial", *Human-Computer Interaction*, 19(1-2) (pp.61-84), 2004.
- J. Klayman, "Varieties of confirmation bias", *Psychology of learning and motivation*, 32 (pp.385-418), 1995.
- N. C. Krämer, "Social Effects of Virtual Assistants. A Review of Empirical Results with Regard to Communication", *IVA 2008: 507-508*, 2008
- T. Law, M. Chita-Tegmark, N. Rabb and M. Scheutz, "Examining attachment to robots: benefits, challenges, and alternatives", *ACM Transactions on Human-Robot Interaction (THRI)*, 11(4), 1-18, 2023.
- Le Dépendant (editorial), "*Perpignan : elle se suicide car son Tamagotchi est mort*" [In Perpignan, a girl commits suicide because her Tamagotchi died], *Le Dépendant*, 3 June 2019.
- N. F. Lindemann, "The Ethics of 'Deathbots'", *Science and Engineering Ethics* 28 (6) (pp. 1-15), 2022.
- J. Loh and W. Loh, "Introduction, Social Robotics and the Good Life", in J. Loh and W. Loh, *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds with Robots*, Transcript Verlag (pp. 7-22), 2022.
- V. Lucchese, "*Un Japonais a épousé un hologramme*" [A Japanese man has married a hologram], *Usbek & Rica website*, 13 November 2018.
- M. Mori, "*Bukimi no tani gensho*", *Energy*, vol. 7, no. 4, 1970, translated in M. Mori and K. F. MacDorman, "*La Vallée de l'Étrange de Mori Masahiro : Importance et impact sur l'esthétique et la conception des robots*" [The uncanny valley of Mori Masahiro: importance and impact on the aesthetics and design of robots], *e-Phaïstos - Revue d'histoire des techniques*, vol. 7, no. 2, 2019.
- C. Muller (dir.), "The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law, Report of the Ad hoc Committee on Artificial Intelligence (CAHAI)", *Council of Europe, CAHAI(2020) 06 fin, no. 21*, 24 June 2020.
- D. Nora, "*Intelligence artificielle : Replika, le chatbot amoureux... et dangereux ?*" [Replika, the chatbot in love... and dangerous?], *L'Obs*, 11 September 2023.
- A. Paiva, I. Leite, H. Boukricha and I. Wachsmuth, "Empathy in virtual agents and robots: A survey", *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3) (pp.1-40), 2017.
- X. Pan, M. Gillies and M. Slater, "Virtual character personality influences participant attitudes and behavior—an interview with a virtual human character about her social anxiety", *Frontiers in Robotics and AI*, 2, 1, 2015.
- E. Petit, "*La relation empathique homme/robot : normale ou pathologique ?*" [Is an empathetic relationship between humans and robots normal or pathological?], *Implications philosophiques*, 21 July 2021 (<http://www.implications-philosophiques.org/la-relation-empathique-homme-robot-normale-ou-pathologique/#:~:text=De%20«%20normale%20»%2C%20la%20relation,cœur%20même%20de%20ce%20processus>).
- P. Philip, L. Dupuy, C.-M. Morin, E. de Sevin, S. Bioulac, J. Taillard, F. Serre, M. Auriacombe and J.-A. Micoulaud-Franchi, "Smartphone-Based Virtual Agents to Help the General Population Concerned by Sleep Complaints During COVID-19 Confinement: A Feasibility Study", *J Med Internet Res*, 2020 (<https://www.wefight.co/fr/>).
- R. W. Picard, "Affective computing", *MIT press*, 2000.
- A. Placani, "Anthropomorphism in AI: hype and fallacy", *AI and Ethics* (pp;1-8), 2024.
- R. Plutchik, "Evolutionary bases of empathy". In "*Empathy and Its Development*", N. Eisenberg and J. Strayer Editors, Cambridge University Press, 1987.
- S. Raffin, "*Alzheimer : une intelligence artificielle peut détecter les signes dans la voix*" [AI can detect signs of Alzheimer's in the voice], *French website "Pourquoi docteur"*, 17 April 2023.

- B. Reeves and C. Nass, “The media equation: How people treat computers, television, and new media like real people and places”, *CSLI Publications, Stanford*, 1969.
- M. Rochefort, “L’ingénieur de Google qui voyait l’intelligence artificielle comme une personne a été licencié” [The Google engineer who considered AI as a person has been dismissed], *Siècle digital*, 25 July 2022.
- J. Rochfeld and C. Zolynski, “La valeur des émotions : quel régime pour le ‘capitalisme mental’ ?” [The value of emotions. What regime for ‘mental capital’?] in *Mélanges en l’honneur de P. Sirinelli*, Dalloz (pp. 749-770), 2022.
- N. Sauer, “Deadbots can speak for you after your death. Is that ethical?” *The Conversation*, 9 May 2022.
- K. R. Scherer, “Emotion”, in *M. Hewstone and W. Stroebe editors, “Introduction to Social Psychology. A European perspective”, Oxford Blackwell (pp. 151-191)*, 2000.
- I. Sivignon, “Blue Monday : découvrez 4 chatbots messenger pour éviter la déprime” [Discover 4 messenger chatbots to avoid the Monday blues], *Influenth magazine*, 20 January 2020.
- M. Slater, “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments”, *Philosophical Transactions of the Royal Society B (pp. 3549–3557)*, 2009.
- S. Tisseron, “Le jour où mon robot m’aimera : vers l’empathie artificielle” [The day my robot will love me: towards artificial empathy], Albin Michel, 2015.
- S. Tisseron, “Petit traité de cyberpsychologie” [A short treatise on cyberpsychology], *Le Pommier*, 2019.
- C. Vincent, interview with Serge Tisseron: “Les robots vont modifier la psychologie humaine” [Robots will change human psychology], *Le Monde*, 12 July 2018.
- A. M. Von der Pütten, N. C. Krämer, J. Gratch and S.-H. Kang, “It doesn’t matter what you are!” Explaining social effects of agents and avatars. *Computers in Human Behavior 26 (6) (pp. 1641-1650)*, 2010.
- Matthew Weaver, “AI chatbot ‘encouraged’ man who planned to kill queen, court told”, *The Guardian*, 6 July 2023.
- M. West, R. Kraut and H. E. Chew, “I’d blush if I could: closing gender divides in digital skills through education”, *UNESCO, Equals Skills Coalition*, 2019.
- K. Winkle, D. McMillan, M. Arnelid, K. Harrison, M. Balaam, E. Johnson and I. Leite, “Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical HRI”, in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (pp. 72-82)*, March 2023.
- J. C. M. Wong, M. J. S. Wan, L. Kroneman, T. A. Kato, W. Lo, P. W. C. Wong and G. H. Chan, “Hikikomori Phenomenon in East Asia: Regional Perspectives, Challenges, and Opportunities for Social Health Agencies”, *Frontiers Psychiatry Sec. Child and Adolescent Psychiatry Volume 10*, 23 July 2019.
- T. Xie and I. Pentina, “Attachment theory as a framework to understand relationships with social chatbots: a case study of Replika”, *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022.
- A. Zhang and P. Patrick, “Tools or peers? Impacts of anthropomorphism level and social role on emotional attachment and disclosure tendency towards intelligent agents”, *Computers in Human Behavior, 138, 107415*, 2022 (<https://doi.org/10.1016/j.chb.2022.107415>).

VII. LIST OF ACRONYMS

AI: artificial intelligence

CEA: French Alternative Energies and Atomic Energy Commission

CERNA: Allistene's think tank on research ethics in digital science and technology

CNPEN: *Comité National Pilote d’Éthique pour le Numérique* [National Pilot Committee for Digital Ethics]

CNRS: *Centre National de la Recherche Scientifique* [National Centre for Scientific Research]

COMETS: CNRS Ethics Committee

CREA: Centre for Research in Applied Epistemology, Ecole Polytechnique
EHESS: *École des Hautes Études en Sciences Sociales* [School of Advanced Studies in Social Sciences]
ENS: *Ecole Normale Supérieure*, a French higher education establishment offering advanced studies in a range of disciplines
FAQ: Frequently Asked Questions
GPT: Generative Pre-trained Transformer
Inria: *Institut national de recherche en sciences et technologies du numérique* [French national institute for research in digital science and technology]
IRD: *Institut de Recherche pour le Développement* [French national research institute for sustainable development]
LaMDA: Language model dialogue application
OBVIA: International Observatory on the Societal Impacts of AI and Digital Technologies
OECD: Organisation for Economic Co-operation and Development
SNCF: *Société nationale des chemins de fer français* [France's national state-owned railway company]
WHO: World Health Organization