Empathic Machines?
Ethical Challenges of Affective Computing from a Sustainable Development Perspective

Cordula Brand, Leonie N. Bossert, Thomas Potthast

1 Introduction

The question whether machines can recognise and simulate emotions is currently researched and discussed intensely in science, industry, and the public.[1] The corresponding technology is called *Affective Computing* (AC). The concept and possible applications of AC gain much attention due to the high potential as well as risks for society such as the potential misuse of highly sensitive data on the one hand or fostering participation within society through sensitive technology on the other (Devillers 2021; Cowie 2015). AC is linked to two main goals in the field of machine learning. Firstly, machines should be enabled to *recognise emotional states* of people to adapt the machine's behaviour to these states, i.e., they should be made 'empathic'. Secondly, and especially in the case of conversational systems like chat bots, avatars, or robots, for example in the care sector, they should be able to *simulate emotions* convincingly to enrich and simplify human-computer interaction (HCI). For implementing both recognition and simulation, the following parameters are mainly used: facial expression, posture, gestures, and speech. Depending on the technical

1 We thank the editors for their valuable feedback and Lukas Weber for assistance in preparing the manuscript. The paper has been developed in the project ("Orientation towards the common good in the digital age - Transformation narratives between planetary boundaries and artificial intelligence" (Gemeinwohlorientierung im Zeitalter der Digitalisierung. Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz), funded by the German Environment Agency (Umweltbundesamt - UBA, FKZ 3718111050).

equipment, recognition may also include the possibility of collecting additional physiological data like skin temperature and skin conductivity (Picard 2000: chapter 2). Various possible applications are tested, developed, and discussed: from the gaming scene to sexbots, from automotive industry to advertisement and possibilities for self-optimization, and within the care- and education sector.

All of them come along with diverse ethical aspects, which need to be considered when discussing AC technologies. We argue in this chapter that the ongoing ethical considerations so far lack an important perspective: orientation towards a justice-based approach – we use the Sustainable Development framework –, which may help in developing and using AC technologies that will benefit all humans, not only privileged groups.

We shall first give an overview on the ethical issues that need to be addressed regarding AC. Some of them are rather general and have already been mentioned regularly (Hagendorff 2020), as they emerge in the overall context of Artificial Intelligence (AI)-technologies and Big Data, like protection against discrimination, equity of access and protection of the privacy of users and cybersecurity. Affective Computing, however, raises additional ethical issues. This technology seems to be able to change our understanding of what it means to be a human being more severely than other applications of AI.

In a second step we will thus summarize these anthropological concerns. Here, aspects are addressed that are rarely mentioned in AI-guidelines as well as in the academic discourse on AI, such as, for example, solidarity, inclusion, and diversity. These aspects are only slowly entering the academic as well as public debate concerning AI. A reason for this can be found in a technology-focussed ethical approach (rather than taking into account the human condition) that at the same is concentrating too much on an individual level. Some of these missing aspects, this is our argument, can be covered by the normative concept of Sustainable Development (World Commission on Environment and Development 1987), which is a justice-based approach.

Therefore, we will show in a third step, how addressing the principles of global inter- and intragenerational justice and the priority for the basic needs of the world's poorest sheds new light on the ethical reflection of AC. Thus, raising ethical issues within the Sustainable Development framework fosters the demand that AI technologies, much more than to date, must follow pathways that serve *all* humans and avoid discrimination and exclusion.

2 Overview on general ethical issues

Ethical issues that apply to all AI technologies (Hagendorff 2020) are also relevant in the context of AC. General points that are central for the specific ethical debate on AC as well are: 1) Protection against discrimination and equity of access and 2) protection of the user-privacy and cybersecurity. Some further points come up particularly when vulnerable people are involved, as in the case of care and education scenarios, with a focus on elderly people in the former and young people in the latter. In these cases, questions concerning 3) autonomy become especially relevant, as many elderly people as well as young children – and other people in certain contexts – are hardly able to take autonomous decisions . In addition, the characteristics of AC technology give rise to specific implications of general aspects, such as the problem of deception, the risk of stigmatization, and a more severe potential for misuse of the data.

2.1 Protection against discrimination and equity of access

Basically, when designing AC systems, it must always be kept in mind that, first, the underlying datasets are oftentimes biased and second, emotions are not value neutral. This point has already been discussed extensively in the literature. The datasets for training AI-systems mainly consist of white people, thus sometimes leading to problematic results, for instance, in search-engines (Makhortykh et al. 2021) or face-recognition-systems (Cavazos et al. 2021) when people of color interact with these systems. Even if developers are aware of that problem, it takes time and a lot of effort to enrich the existing datasets (Endrass et al. 2013) or develop new ones – time and effort that need to be financed and are part of a highly competitive time-critical economic field of innovation. Regarding AC, an additional short-coming needs to be kept in mind. Development and training of the algorithms are mainly based on the scale of universal emotions as suggested by Ekmann (1999), entailing six basic emotions: anger, surprise, disgust, enjoyment, fear, and sadness. However, in complex situations of social engagement, emotional settings are much more diverse than that.

The second aspect of possible discrimination and stigmatization poses a special challenge in the context of AC. The description or processing of emotions usually includes an evaluation of these emotions. Therefore, using the emotional data entails moral deliberating, which is inevitably done by the ma-

chine as well. This becomes highly problematic when people are categorized based on such an artificial procedure, like in the case of semi-intelligent information filters (SIFF systems). These are used to evaluate and interpret a set of data to draw conclusions about persons, marking them to behave suspiciously or be ready to act aggressively (Cowie 2015), conclusions that might have serious consequences.

Both mentioned sources of possible discrimination and stigmatization require intense attention and awareness, on the side of the developers and producers, as well as on the user side. The limitations of AC technologies must be understood and communicated thoroughly, which means that all members of society must be not only informed but educated accordingly.

Intense training and a broad corresponding ethical reflection would also serve to promote the development of applications that might be able to benefit everyone in society in a more suitable way. AC technologies can, for example, help to reduce discrimination by using culturally sensitive systems (Endrass et al. 2013). If developers were able to adapt artificial systems ("agents"[2]) more clearly to cultural settings in which they are to be used, awareness of – and maybe even more respect for – these differences would increase, presumably also on a global North-South scale.

This is an important aspect for AC technologies in *education* as well, where one of the main goals of the developers is to improve the usability of digital technology and the effectiveness of the learning experience. The (rudimentary) emotional sensitivity of the technology should enable people to use it more easily and in a more approachable way, so that users could take better advantage of the benefits the technology offers (Cowie 2015). This includes, for example, the possibility for the artificial AC system to react directly to frustrations on the part of the users, which is especially relevant in the educational context (Troussas 2020: chapter 5). Furthermore,

2 We are preferring the rather neutral term artificial system for practical reasons and to avoid misunderstandings. In the literature on AI, often parlance is about artificial "agents" without discussing the conceptual implications. We do not consider these systems to possess agency according to a philosophical understanding of the term, linking actions to purposefulness and/or moral responsibility. If one day strong AI machines might be developed which are able to make decisions on their own, one would have to discuss further under which circumstances those machines could be (moral) agents. Yet, to date, AI technologies are not since they 'act' upon programmed algorithms. We are aware that this perspective is contested, and other scholars have different opinions on that (cf. Loh 2019: chapter 2.1).

emotionally sensitive technical assistants could help break down barriers to communication and access, enabling a more inclusionary approach towards diverse people. Emotion-sensitive language assistants could benefit people, for example, by enabling them to better cope with everyday life or by reducing communication barriers caused by different national or technical languages (Burchardt/Uszkoreit 2018). Access to various interfaces could be made easier for less technology-savvy people if these systems responded to negative reactions from users and at the same time presented a friendly and sympathetic counterpart (Cowie 2015). In addition, culturally specific adaptations could help to mitigate or overcome cultural hurdles in education scenarios as well. Furthermore, AC technologies are used in different training scenarios, for instance, for patients suffering from autism spectrum disorder (Obe et al. 2020) or people that face difficulties in stressful social settings (Schneeberger et al. 2019). Taken together, AC could contribute to increasing access to educational content within societies and on a global level. By this, AC could strengthen the potential for vulnerable groups to participate.

However, these advantages must be treated with caution, as AC systems could also have the exact opposite effect. This is especially the case when access to the technology is distributed unevenly, demands of transparency are not met or discriminating biases are not revealed. Furthermore, it must be critically questioned whether the enormous development effort can be justified in view of the mentioned ethical challenges, since it is doubtful how big the positive impact of AC systems will be. Therefore, expanded access for less privileged persons must be supported politically as well and goes far beyond the realm of technology development and distribution.

2.2 Protection of user privacy and cybersecurity

In the case of technical systems that record and analyze human emotions, a particular vulnerability is at stake for all users since the processed data is fundamentally intimate, sensitive information. In every-day situations within the public sphere, we show emotions and read the emotions of others frequently. However, in these situations we have some influence on what we want to show and what not, especially in cases we know we are recorded. In the case of AC systems, it is mandatory for the systems to work properly that we do not hide or fake out emotional states. And as is the case for all digitized data, these states are distributable and usable for other than the intended purposes. The demands and difficulties concerning data privacy have

been discussed broadly in several contexts which already led to international standards[3]. In a justice-based approach, here individual privacy links up to social issues of tracking and treating the emotions of certain 'suspec' groups in an even more discriminatory way.

But, because of their intimacy, mass recording, generation, dissemination, and commodification of emotions through AC in the areas of digital imaging platforms and online transaction platforms are particularly critical and require special regulation (Stark/Hoey 2021), Here, it is plausible to follow the precautionary principle[4] (Andorno 2004) to rather slow down the speed of implementation of the AC technologies processing emotions in order to allow for broader technology and ethical assessments. In a practical sense, handling the information generated by these systems with the same caution as it is the case for medical data would allow for a high(er) ethical and legal status of protection.

For the use of digitized medical data, strict regulations already exist and could be adapted.

However, with increased networking and automatic processing, even these regulations face problematic limits. Particularly disputable is the topic of informed consent for (non- disclosure/processing, see Jörg 2018) as the information about all the processes involved is particularly complex and hard to understand even for healthy adult lay people. These problems also occur in the context of AC technologies, as has been already reported in the case of psychological treatment with AC support (Nicholas et al. 2020).

2.3 Autonomy

Factors seen as important for the autonomy of human persons is the ability of self-determination and being as independent in one's decision making as

3 The United Nations digital strategy (2022-25) lists guiding principles to digital transformation: https://digitalstrategy.undp.org/#Guiding-Principles, last access: 02.06.2022.

4 The precautionary principle addresses situations where caution in the light of uncertainty should be given (Jordan/O'Riordan 2004). For example, the EU works with this definition: "The precautionary principle applies where scientific evidence is insufficient, inconclusive or uncertain and preliminary scientific evaluation indicates that there are reasonable grounds for concern that the potentially dangerous effects on the environment, human, animal or plant health may be inconsistent with the high level of protection chosen by the EU". (COM 2006: 90)

possible.[5] We take such forms of procedural independence as a requirement to be considered autonomous. In the case of AC machines, if third parties such as the operators of these machines or virtual systems, have information about or even access to emotional states of persons, this could have an influence on procedural independence (Baumann/Döring 2011), for example, by limiting the options for action. If I knew that a machine recognizes my emotional state, I might act differently than if this information were unknown. It might even be best to not consider some courses of action at all. As long as I am aware which different possibilities of action I have, this is not so much of a problem. However, people who are inexperienced in dealing with AC machines or cannot understand the technical limitations, e.g., children or individuals with mental disabilities or elderly people, are more vulnerable to unintended side-effects of their actions.

As a second point, such vulnerable groups might form unintended and/or unwanted bonds to artificial systems (Wilks 2010) in cases where the machines simulate emotions. To enter such an emotional relationship might limit the scope of self-determination and independent decisions as well.

Moments of misinterpretation can be challenging for autonomous action and decision. Those might occur when virtual systems show options for action that cannot be easily derived from the emotions they simulate at the same time. Or machines interpret emotions they recognize in an inadequate way. In such a case, it is hard for users to decide between different options for action because of not having decent information (Beavers/Slattery 2017). This problem is exacerbated if the machine relies on speech and text files from the Internet for the underpinning of options for action as the genesis and trustworthiness of such information cannot be recognized or verified instantly by the user.

Another form of possible deception can be described as *pars pro toto* acting, which is especially poignant with vulnerable groups. Imagine a teaching bot that gives the impression of caring. This might strengthen the bot in its

5 We are using this rather basic understanding of the autonomy of human beings to illustrate some effects AC might have on our abilities of self-determination. In this context, the enduring and complex debate about different concepts of autonomy within philosophical ethics does not have to bother us for our limited purpose here. Furthermore, here, we are not discussing the question in what sense a machine can become an autonomous agent and what has to be the case to assign morality to robots or other artificial agents (cf. Loh 2019: chapter 2.1).

function. However, it might happen that users infer from one type of caring-behavior (e.g., reacting to the emotion of frustration while struggling with a math-problem) that the device is also capable of other caring contexts (e.g., dealing with the loss of a pet), which it might just not be (Cowie 2015).

Accordingly, it must be ensured for the field of *education* (as well as in the fields of gaming and care) that users have clear options for action. Furthermore, it needs to be transparent for which fields of action the artificial systems were developed and where their limits are, so that users can form realistic and binding expectations.[6] As in the case of all AI technologies, it is essential to assess the various possibilities of intentional or unintentional influence and manipulation, thus preventing corresponding misuse. If this fails, a loss of autonomy of the human actors, which goes hand in hand with possible manipulation, can be expected. In addition, the aspects of understandability and control of the technology, which is strongly emphasized in AI guidelines (Hagendorff 2020), must be taken seriously. The greatest possible transparency of the modes of operation should be given to ensure the most informed use possible. This also and especially applies to the target group of children, young adults, and mentally impaired patients due to their vulnerability. Therefore, the well-established discourse about informed consent, with all its intricate questions concerning vulnerable persons, must urgently be initiated for AC.

3 Anthropological perspectives

Until recently, emotions and emotionality have been regarded as a unique feature of humans or all living sentient beings (Manzeschke/Assadi 2019). At the same time, emotions are private in the sense that they cannot directly be accessed from the 'outside'. Artificial systems that can record as well as simulate human emotions more and more convincingly thus might challenge

6 In this context it is discussed controversially whether it is helpful or not to design robots or avatars as humanlike in their appearance as possible. The Uncanny Valley Thesis (Mori et al. 2012) states that up from a certain point of realistic human appearance people tend to become afraid of artificial agents. However, as the Uncanny Valley is being questioned (Bartneck et al. 2007), it would go too far to elaborate on the topic more deeply in this chapter.

those basic anthropological assumptions.[7] This is relevant from a philosophical perspective, as several classical concepts of personal relations and capacities could be affected, like intersubjectivity, friendship and authenticity.[8] At the same time, it is also highly relevant from a societal perspective. All these concepts entail normative assumptions and moral obligations that not only affect singular people, but also societal ideas which in turn can affect societal transformations. How our understanding of what makes human life human, the *conditio humana* (Arendt 1998; Plessner 2003), can be affected by AC, we will elaborate on in the following.

Intersubjectivity, for instance, refers to the space of shared meanings between subjects that emerges through interpersonal exchange (Husserl 1973). These shared meanings can be understood as foundations of the social world, constituted by its members. So far, this social world has been shaped or created by human beings in dialogue (Buber 2009). If avatars or robots enter an (also) emotionally meaningful dialogue, then it is necessary to discuss whether our standard conception of intersubjectivity must be changed or extended. Furthermore, the question arises which consequences intersubjectively participating machines in our lifeworld would have on our understanding of the concept of a morally responsible person (Brand 2015). Ethically speaking, we must consider solidarity and inclusion as well as diversity-aspects regarding robots and other machines in both directions. This might seem far-fetched but the ongoing debate on demands for assigning a moral status to robots (Decker/Gutmann 2012; Loh 2019) and, accordingly, robot-rights (Gunkel 2021; Gellers 2021) poses some challenging insights. This brings us to the question if it would not be even more urgent to reflect and work on solidarity, inclusion, and diversity of *all* humans in society.[9]

Furthermore, the emotional and parasocial human-machine relationship promoted by AC might complement interpersonal relationships; a contested follow-up question is whether und under which circumstances this could also

7 This is also connected to the discussion of what might happen if machines someday pass the Turing-Test.

8 It might even affect our understanding of the capacity of making moral judgments. Emotions are discussed to take an important part in making moral decisions and having moral convictions. If artificial agents form emotionally based judgments, they might come closer to form moral judgments that are of a similar quality like human ones (Cowie 2015; Baumann/Döring 2011).

9 We are fully aware on the debates on including sentient non-human beings into the moral community, but this is beyond the scope of this paper (cf. Bossert 2022).

lead to replacement of the latter. In this context, the concept of friendship, and along with it also the concept of care, might have to be reconsidered. Avatars or robots as so-called companions are becoming much more realistic both in the care sector and in the education sector as part of the development of AC. A Manzeschke and Assadi (2019: 170) point out that up to now people have been dependent on the functionality of machines of all kinds, insofar some form of dependence already exists. However, if people become emotionally dependent on artificial systems, then these machines might even set other standards in the sense of appropriateness of emotional response. This would open a shift or new dimension of the *conditio humana*. Another changing concept could be intimacy, if machines were able to collect, store, process and transmit human emotions indefinitely. This challenge already arises in the context of private conversations as well as sounds of everyday life that systems like Alexa and Siri permanently record.

Such developments again raise the question of solidarity, inclusion, and diversity. Do we really want to solve societal problems in the care and education sectors by developing simulations of human interaction? Would it be more suitable to work on human based solutions, especially when vulnerable people are concerned? Obviously, the concept of a good life (*eudaimonia*)[10] also must be addressed in this context as a rich social life and stable personal relationships are seen as a part of it, as among others, Turkle (2015) has pointed out with a critical view on the digital age.[11]

Another example of how AC might change some aspects not only of the *conditio humana* but of our worldview is the human relation to a 'real' or 'natural' environment. This might change if AC-enhanced environments become more 'non-artificial' in the sense that the artificial systems seem more and more human to us by simulating emotions. The appeal of AC-enhanced environments could thus be enhanced and ultimately be preferred over non-AC-enhanced environments. For example, students who already have certain difficulties with human social contacts could experience them even more demanding, if they can rely on personalized avatars fulfilling exclusively their

10 The discussion about what constitutes a good life is as old as philosophy and accordingly complex. We do not hold it necessary for the purpose of this chapter to go into details. However, we see Martha Nussbaum's Capability Approach as a promising framework to think about this question (Nussbaum 2000; 2007).

11 We cannot go into the connected issues of objectophilia/object sexuality with regard to AI systems here.

wishes. With this, wishes are fulfilled without the necessity of – sometimes complex and demanding – social interaction. This can be seen as problematic if one holds that social competencies and at least some, even demanding, interactions with fellow humans are valuable and indispensable. The previously mentioned aspects, especially of inclusion and a good life are therefore affected here again.

These bundles of questions need to be further illuminated – not only but also – by in-depth philosophical research, for example, by analyzing the terms and concepts that might change as well as the anthropological implications and implications for moral psychology. What effects such changes might have cannot be foreseen in detail now. Nonetheless, we must reflect upon the ethical implications these developments might have. This should be done for the individual, the organizational, and the societal level (Manzeschke/Assadi 2019). The aspects we want to shed light on here – namely solidarity, inclusion, diversity, and further questions of good life –, are thereby located on the societal level. When addressing them, questions of governance must not only be discussed but also decided. Therefore, we need a normative framework for ethically reflecting AI developments that fundamentally include societal points of view. We suggest doing so by implementing the Sustainable Development perspective into the deliberation.

4 The Sustainable Development Perspective

One might ask why the Sustainable Development (SD) framework is a feasible choice regarding an (ethical) evaluation of AC technologies. Before answering this question, we shall highlight some important aspects of the SD concept.

Numerous academic as well as political documents on SD refer to the so-called Brundtland Report of the World Commission on Environment and Development (WCED 1987). This report brought the two political agendas of development and environmental conservation into a joint focus, two fields that had mostly been treated as contradictory to each other. Both are discussed as belonging together in the concept of SD, while setting the principle of inter- and intragenerational justice as the ethical foundation of SD. It reads:

> Sustainable Development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It contains within it two key concepts: the concept of

> 'needs', in particular the essential needs of the world's poor, to which overriding priority should be given; and the idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs. (WCED 1987: chapter 2.1)

Inter- and intragenerational justice as a principle of equity is complemented by a priority principle regarding the basic needs of the world's poorest – and both justice aspects being explicitly framed by the limits of the non-human environment (ecological systems connected with the socio-technological systems). In the following, we will strengthen our argument – that a normative framework of SD is a suitable approach for AC technologies – by demonstrating how these principles can address the ethical challenges described in section 2. This is important because solidarity and inclusion, diversity, the influence on ideas of 'good life', the need for quality (digital) education, the planetary to local environmental conditions[12], and the question of governance all come to the table when deliberating if, why, and how machines should be trained to analyze and simulate human emotions at all.

4.1 Global inter- and intragenerational justice

Taking the principle of inter- and intragenerational justice seriously also means to strive for just societies today and in the future. We hold taking care of emotional needs of all members of societies as one important aspect for being able to achieve and sustain such just societies that foster a good life. Here, we cannot discuss in detail the – biologically, psychologically, as well as philosophically – complex term 'emotional needs'. Following the capability approach of Nussbaum (2000; 2007), we broadly understand emotional needs as, among others, the need to feel safe and being free from abuse, to be cared for, to be respected and to be self-effective.

For addressing these emotional needs, it is important to analyze from a societal perspective how a lack of emotional needs can be avoided or dimin-

12 As every AI technology, AC requires large amounts of energy. The production of digital end devices needs resources, which often are rare and nonrenewable like rare earths, and the development of these technologies is responsible for an enormous amount of CO_2 emissions (van Wynsberghe 2021). However, in this chapter, we do not focus on these aspects, which by no means should undermine their importance.

ished and what measures are best taken to cope with problems.[13] It might, as a general orientation, be more fruitful to invest in educational structures that are flexible enough to deal on an interpersonal level with emotional needs then to develop an emotion-simulation robot (the usefulness of some specific AC-applications and constellations notwithstanding).

Further, to take emotional needs seriously, the problem of possible user deception is to be considered in the development stage as well as in the implementation of AC technologies. This is necessary due to the danger of abusing an emotional need if the deception is not recognised as such. In order to prevent this from happening, transparency must be maintained. One way of avoiding unintentional possibilities of deception is to allow future users to participate in the development (Cowie 2015: 340).

When investing in technology development, a SD framework calls for focussing on the development of technologies which ease processes to strengthen public welfare. In line with this, it also calls for striving for *solidarity* with as well as the inclusion of *all* humans. Regarding AC technologies this means – as for all AI technologies – developing algorithms which do not revert to racist, sexist, anti-disabled, or other forms of discriminating biases. Instead, one needs to develop culturally sensitive systems (while avoiding discrimination and stigmatization; cf. section 2.1) and systems that mirror diversity (the appearance of the systems play an important role here, but also diversity in interaction should be fostered, as long as the diverse ways of interaction are evaluated as being helpful and valuable).

With this comes the problem that AC technology can reinforce discrimination and stigmatization (e.g., in the case of semi-intelligent information filter (SIFF) systems; cf. section 2.1). However, AC technologies can also serve to mitigate cultural differences, for example, in the area of intercultural training settings or through the use of appropriately sensitive learning companions. For being useful for all people, AC technologies appear to be worthy of support or expansion if they ensure greater access to digital or virtual systems,

13 When discussing the question if and how AC technologies can foster or hinder a good life for as many individuals as possible, one also needs to investigate the question, if (in some cases even unrecognized) simulation of emotions or relations is detrimental to a good life or not (see above). We cannot provide a general answer to this question, especially since that hinges on the setup and context of the specific AI-system, yet we would point out the importance of critical perspectives, which Turkle (2015) and others have elaborated.

if the use of digital systems causes less stress or unpleasant experiences for people, or if cultural barriers can be made visible or be removed.

In sum, AC technologies should be built and used for enabling and empowering people – an aspect which a) needs to become an important litmus test for technologies to be evaluated ethically sound and useful and b) to meet the requirements given by a SD framework, namely, to comply with the principle of inter- and intragenerational justice. In such a way, the technology comes closer to enable good lives for (as many as possible) people.

On the other hand, to truly fulfill intragenerational (in the sense of global justice) means to seriously include the needs of the world's poorest. Without this, global justice cannot be realized. In relation to all technologies – including AC – this means developing them in a way that the technologies meet basic needs and enable basic social participation, rather than being adapted to luxury-oriented needs. In concrete terms, this means, for example, that AC technologies should be developed and used to enable accessible quality education in as many parts of the world as possible, rather than being used, for example, to improve micro-targeting for companies by assessing the emotions of potential customers. According to the Brundtland Report (WCED 1987), the needs of the world's poorest must be prioritized. This prioritization must be mirrored in technology development.

4.2 Education

Nearly 30 years after the publication of the Brundtland Report and following the spirit of the UN-Agenda 21 of the Summit for Environment and Development in Rio de Janeiro 1992, the United Nations' Sustainable Development Goals (SDGs) of 2015 have been agreed upon, which should be implemented in all states to transform societies into more sustainable ones.[14] As pointed out, we hold that education plays an important role for achieving intragenerational justice (cf. section 4.1) and we showed that AC technologies have a high potential to be used for education (cf. section 2.1). The SDGs explicitly address education in SDG No. 4: "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". We want to stress that the full spectrum of education – which mostly is not considered in the

14 For an interesting overview of the linkage of AI and the SDGs cf. Vinuesa et al. (2020) and Sætra (2021). Yet, both are not discussing AC technologies.

AI guidelines (Hagendorff 2020) – is of particular importance in the development of AC technologies. There is a quasi-double educational mandate for developers and users. Developers must be able to know, reflect and communicate the ethical implications of their systems. Users, on the other hand, must be able to inform themselves about the possibilities and limitations of the systems they use and to understand this information. Furthermore, developers must be able to recognize and reflect on ethical aspects of their work, especially regarding vulnerable groups, as it is often the case within the field of education. They need to be empowered to make their motivations and goals transparent to the public to contribute to an informed societal debate. This is especially true when it comes to particularly sensitive applications, such as dealing with emotions and vulnerable groups. All this is important and needs to be considered in relation to digital education programs, which have to be adjusted accordingly to serve establishing just societies.

Accordingly, these ethical and scientific communication skills should already be addressed during the training of further researchers and developers. If this were the case, then – at least in a roundabout way – more ethical reflection could also be incorporated into the development processes of commercial providers than has been the case to date.

4.3 Governance

The overall focus on justice highlights the whole set of principles such as equity of access and distribution, discrimination, and diversity, as well as education as central enablers for developing and using AC devices in a positive way. This eventually leads to governance questions in at least three main aspects.

First, with AC highly sensitive personal data can be collected, copied, and distributed. Therefore, data about emotions should be treated like medical data in general, granting the same high security and transparency standards to avoid misuse and respect the autonomy of the users. In this context, the fundamental question of the interaction between science and politics arises, as well as the question of social and political regulation. One can certainly ask whether, for instance, the further development of SIIF systems is socially or politically desired and whether AC should be permitted for free entrepreneurial use. And one must make sure that especially vulnerable persons are safe to use the technology. However, the decision to treat AC data as medical data is ultimately a governance decision.

Second, the discussion of AC technologies shows how urgently the dual educational mandate that arises in connection with AI technologies must be perceived. There is a need both to implement ethical competencies in the education of researchers and technicians, as well as to further expand digital education for all members of our society. This, again, entails a bundle of diverse governmental decisions, if taken seriously. Within the training of researchers and developers, for example, modules must be embedded that convey ethical competences from the start. When one remembers the effort it often takes to communicate digital knowledge in schools, it is obvious how crucial governmental decisions are in this area.

Third, the need to do justice to the world's poorest means that politics must not be oriented exclusively towards the interests of its own population. It must also consider the effects of local actions on other parts of the world. Measures should be pushed which not only do not have destabilizing consequences for the Global South – through, for instance, resource exploitation or selling of technologies to non-democratic regimes –, but which have the potential to benefit people in the poorest regions of the Global South. This requires highly complex governance activities in general and in the entire field of AI, including the field of AC (for regulatory options to minimize scenarios of AI damaging the SDGs cf. Truby (2020)). As always is the case in technology assessment and ethics, the danger is to look at AI-systems mainly from a technology-driven perspective. Taking into account SD as a guiding ethical groundwork, one would shift towards i) problem-driven and ii) cause-oriented approaches and ask how far AC-systems really could help here (Erdmann et al. 2022). This would also have implications for the way governance of such technologies should be organized and structured in the first place, namely not by separate but integrative regulatory works.

5 Conclusion

The normative framework of Sustainable Development as a basis to investigate and evaluate Affective Computing shows that precisely this perspective can – and should – complement, if not integrate, the common AI-ethical considerations, since it meaningfully broadens and at the same strengthens ethical considerations.

With the strong focus on the principles of inter- and intragenerational justice underlying the SD framework, the recognition and simulation of emo-

tions by machines should be done in a way that empowers all humans and reduces the risk of deception to the lowest possible level.

Moreover, if the prioritization of the basic needs of the world's poorest is taken seriously, this would prevent AC technologies from being used primarily to satisfy market-oriented interests of the Global North, like for example the development of even more realistic computer games or targeted advertising that not only addresses an individual's interests, but also their current emotional state. AC must then be used in a way that promotes fundamental interests such as equal participation in society. This can be implemented for example by using AC for quality digital education programs. However, the use of AC in an ethically acceptable or even desirable way does not only serve people in the Global South. It will also be useful to form more just societies in the Global North if, for example, data about emotions is treated like sensitive medical data in the development and use of this technology. While this is necessary for the general usage of AC, it is specifically important in the field of education. Here, AC has a high potential, but this field also usually affects vulnerable groups. It will also benefit *all* societies if programming takes into account cultural specificities and the avoidance of discrimination and stigmatisation, and if emotion recognition and simulation by machines is developed to provide digital education and training scenarios at a level that would be more difficult without this technology.

References

Andorno, Roberto (2004): "The Precautionary Principle: A New Legal Standard for a Technological Age." In: Journal of International Biotechnology Law 1/1, pp. 11-19.

Arendt, Hannah (1998 [1958]): The Human Condition. 2nd edition, Chicago: University of Chicago.

Bartneck, Christoph/Kanda, Takayuki/Ishiguro, Hiroshi/Hagita, Norihiro (2007): "Is the Uncanny Valley an Uncanny Cliff?" In: RO-MAN 2007 – The 16th IEEE International Symposium on Robot and Human Interactive Communication, pp. 368-373.

Baumann, Holger/Döring, Sabine (2011): "Emotion-Oriented Systems and the Autonomy of Persons." In: Paolo Petta/Catherine Pelachaud/Roddy Cowie (eds.), Emotion-Oriented Systems: The Humaine Handbook, Heidelberg: Springer, pp. 735-752.

Beavers, Anthony F./Slattery, Justin P. (2017): "On the Moral Implications and Restrictions Surrounding Affective Computing." In: Myounghoon Jeon (ed.), Emotions and Affect in Human Factors and Human-Computer Interaction, London: Elsevier Academic, pp. 143-161.

Bossert, Leonie N. (2022): Gemeinsame Zukunft für Mensch und Tier – Tiere in der Nachhaltigen Entwicklung, Baden-Baden: Karl Alber.

Brand, Cordula (2015): "Wie Du mir so ich Dir: Moralische Anerkennung als intersubjektiver Prozess." In: Robert Ranisch/Sebastian Schuol/Marcus Rockoff (eds.), Selbstgestaltung des Menschen durch Biotechniken, Tübingen: Narr Francke Attempto, pp. 21-33.

Buber, Martin (2009): Das dialogische Prinzip, Gütersloh: Gütersloher.

Burchardt, Aljoscha/Uszkoreit, Hans, eds. (2018): IT für soziale Inklusion: Digitalisierung – Künstliche Intelligenz – Zukunft für alle, München and Wien: De Gruyter Oldenbourg.

Cavazos, Jacqueline G./Phillips, P. Jonathon/Castillo, Carlos D./O'Toole, Alice J. (2021): "Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?" In: IEEE Transactions on Biometrics, Behavior, and Identity Science 3/1, pp. 101-111.

COM (2006 [1995]): "Communication from the Commission of 2 February 2000 on the Precautionary Principle (COM (2000) 12.02.2000, P. 1)." In: Philippe Sands/Paolo Galizzi (eds.), Documents in European Community Environmental Law, 2nd edition, Cambridge: Cambridge University, pp. 90-115.

Cowie, Roddy (2015): "Ethical Issues in Affective Computing." In: Rafael A. Calvo/Sidney D'Mello/Jonathan Gratch/Arvid Kappas (eds.), The Oxford Handbook of Affective Computing, Oxford library of psychology, Oxford: Oxford University, pp. 334-348.

Decker, Michael/Gutmann, Mathias, eds. (2012): Robo- and Information-ethics: Some Fundamentals, Wien, Berlin and Münster: Lit.

Devillers, Laurence (2021): "Human-Robot Interactions and Affective Computing: The Ethical Implications." In: Joachim von Braun/Margaret S. Archer/Gregory M. Reichberg/Marcelo Sánchez Sorondo (eds.), Robotics, AI, and Humanity: Science, Ethics, and Policy, Cham: Springer International, pp. 205-211.

Ekmann, Paul (1999): "Basic Emotions." In: Tim Dalgleish/Michael J. Power (eds.), Handbook of Cognition and Emotion, Chichester: Wiley, pp. 45-60.

Endrass, Birgit/André, Elisabeth/Rehm, Matthias/Nakano, Yukiko (2013): "Investigating Culture-Related Aspects of Behavior for Virtual Characters." In: Autonomous Agents and Multi-Agent Systems 27/2, pp. 277-304.

Erdmann, Lorenz/Cuhls, Kerstin/Warnke, Philine/Potthast, Thomas/Bossert, Leonie/Brand, Cordula/Saghri, Stefani (2022): Digitalisierung und Gemeinwohl – Transformationsnarrative zwischen planetaren Grenzen und Künstlicher Intelligenz, Texte 29/2022, Dessau-Roßlau: Umweltbundesamt.

Gellers, Joshua C. (2021): Rights for Robots: Artificial Intelligence, Animal and Environmental Law, Oxon and New York: Routledge.

Gunkel, Harald (2021): "Robot Rights – Thinking the Unthinkable." In: John-Stewart Gordon (ed.), Smart Technologies and Fundamental Rights, Leiden and Boston: Brill Rodopi, pp. 48-72.

Hagendorff, Thilo (2020): "The Ethics of AI Ethics: An Evaluation of Guidelines." In: Minds and Machines 30/1, pp. 99-120.

Husserl, Edmund (1973): Zur Phänomenologie der Intersubjektivität: Texte aus dem Nachlass. Husserliana: Edmund Husserl. Gesammelte Werke Vol. 13, 14, 15, Den Haag: Martinus Nijhoff.

Jörg, Johannes (2018): Digitalisierung in der Medizin: Wie Gesundheits-Apps, Telemedizin, künstliche Intelligenz und Robotik das Gesundheitswesen revolutionieren, Berlin and Heidelberg: Springer.

Jordan, Andy/O'Riordan, Tim (2004): "The precautionary principle: A legal and policy history." In: Marco Martuzzi/Joel A. Tickner (eds.), The Precautionary Principle: protecting public health, the environment and the future of our children, Copenhagen: World Health Organization, pp. 31-48.

Loh, Janina (2019): Roboterethik: Eine Einführung, Berlin: Suhrkamp.

Makhortykh, Mykola/Urman, Aleksandra/Ulloa, Roberto (2021): "Detecting Race and Gender Bias in Visual Representation of AI on Web Search Engines." In: Ludovico Boratto/Stefano Faralli/Mirko Marras/Giovanni Stilo (eds.), Advances in Bias and Fairness in Information Retrieval. Communications in Computer and Information Science Vol. 1418, Cham: Springer International, pp. 36-50.

Manzeschke, Arne/Assadi, Galia (2019): "Emotionen in der Mensch-Maschine Interaktion." In: Kevin Liggieri/Oliver Müller (eds.), Mensch-Maschine-Interaktion: Handbuch zu Geschichte – Kultur – Ethik, Berlin: J.B. Metzler, pp. 165-171.

Mori, Masahiro/MacDorman, Karl/Kageki, Norri (2012): "The Uncanny Valley [From the Field]." In: IEEE Robotics & Automation Magazine 19/2, pp. 98-100.

Nicholas, Jennifer/Onie, Sandersan/Larsen, Mark E. (2020): "Ethics and Privacy in Social Media Research for Mental Health." In: Current psychiatry reports 22/12, 84.

Nussbaum, Martha Craven (2000): Women and Human Development: The Capabilities Approach, Cambridge: Cambridge University.

Nussbaum, Martha Craven (2007): Frontiers of Justice: Disability, Nationality, Species Membership, Cambridge and London: The Belknap Press of Harvard University.

Obe, Olumide/Akinloye, Folasade Oluwayemisi/Boyinbode, Olutayo (2020): "An Affective-Based E-Healthcare System Framework." In: International Journal of Computer Trends and Technology 68/4, pp. 216-222.

Picard, Rosalind W. (2000): Affective Computing, Cambridge and London: The MIT.

Plessner, Helmuth (2003 [1986]): Conditio humana. 4. Auflage. Gesammelte Schriften in zehn Bänden Vol. VIII, Frankfurt am Main: Suhrkamp.

Sætra, Henrik (2021): "AI in Context and the Sustainable Development Goals: Factoring in the Unsustainability of the Sociotechnical System." In: Sustainability 13/4, 1738.

Schneeberger, Tanja/Gebhard, Patrick/Baur, Tobias/André, Elisabeth (2019): "PARLEY: A Transparent Virtual Social Agent Training Interface." In: IUI '19: Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion, New York: Association for Computing Machinery, pp. 35-36.

Stark, Luke/Hoey, Jesse (2021): "The Ethics of Emotion in Artificial Intelligence Systems." In: FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York: Association for Computing Machinery, pp. 782-793.

Troussas, Christos (2020): Advances in Social Networking-Based Learning: Machine Learning-Based User Modelling and Sentiment Analysis, Cham: Springer International.

Truby, Jon (2020): "Governing Artificial Intelligence to Benefit the UN Sustainable Development Goals." In: Sustainable Development 28/4, pp. 946-959.

Turkle, Sherry (2015): Reclaiming Conversation: The Power of Talk in a Digital Age, New York: Penguin.

van Wynsberghe, Aimee (2021): "Sustainable AI: AI for Sustainability and the Sustainability of AI." In: AI Ethics 1/3, pp. 213-218.

Vinuesa, Ricardo/Azizpour, Hossein/Leite, Iolanda/Balaam, Madeline/Dignum, Virginia/Domisch, Sami/Felländer, Anna/Langhans, Simone Daniela/Tegmark, Max/Fuso-Nerini, Francesco (2020): "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals." In: Nature Communications 11/1, 233.

Wilks, Yorick (2010) "Introducing Artificial Companions." In: Yorick Wilks (ed.), Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues Vol. 8, Amsterdam: Benjamins, pp. 11-20.

World Commission on Environment and Development (WCED) (1987): Our Common Future, Oxford: Oxford University.