Emotional AI as affective artifacts: A philosophical exploration

Manh-Tung Ho[1,2] & Manh-Toan Ho[1]

1. Centre for Interdisciplinary Social Research, Phenikaa University, Yen Nghia, Ha Dong, 100803, Hanoi

2. Institute of Philosophy, Vietnam Academy of Social Sciences, 59 Lang Ha Street, Ba Dinh District, 100000, Vietnam

***Correspondence: toan.homanh@phenikaa-uni.edu.vn**

<Draft Paper No. 20240527-v1>

Abstracts

In recent years, with the advances in machine learning and neuroscience, the abundances of sensors and emotion data, computer engineers have started to endow machines with ability to detect, classify, and interact with human emotions. Emotional artificial intelligence (AI), also known as a more technical term in affective computing, is increasingly more prevalent in our daily life as it is embedded in many applications in our mobile devices as well as in physical spaces. Critically, emotional AI systems have not only the ability to not passively read and classify emotions, but most also have an ability to predict and 'nudge' users to certain desired emotional states. Drawing from Piredda (2020)'s account, we've examined various emerging emotion-sensing technologies such as recommender algorithms, personal AI assistants, work-related emotion trackers, and emotional AI toys and robots through the lens of affective artifacts. We show how these technologies fulfill the criteria of being affective artifacts: they can influence the affective states of the users (e.g., help regulate emotions of ourselves and others), some emotion-sensing algorithms and importantly, our manipulation of these algorithms, can intersect with the maintenance and construction of our sense of self and identity.

Key words: artificial intelligence; emotional AI; self; situated affectivity.

Introduction

From cognitive to affective artifacts

In recent years, the notion of affective artifacts has been proposed building on the parallel notion of cognitive artifacts. Cognitive artifacts are defined as "physical objects made by humans for the purposes of aiding, enhancing, or improving cognition (Hutchins 1999, p. 126)." This notion of cognitive artifacts stemmed from the observation made by Donald Norman (1991), the director of the Design Lab, California, San Diego. Donald Norman was among the first to notice much of the scientific understanding during his time had been devoted to unaided mind: the issues of memory,

attention, action, and thought. But careful works in cognitive science on how the artifacts can shape the mind had been neglected.

Recently, there has been substantial growth in the literature devoted to the extended mind and situated cognition. This literature sheds light on both human agency and the contextual/situational dimensions of human cognition and affects. Each individual human is always constructing a cognitive and affective scaffold to help them perform cognitive and affective tasks such as problem-solving or regulating emotions, whether being conscious of this process or not. This is done by manipulating objects, tools (increasingly smart devices), and spaces (physical and cyber) (Fasoli, 2018; Heersmink, 2013; Piredda, 2020).

Importantly, this notion of cognitive artifacts suggests that cognition is not a process that purely is happening in the brain. It is not brain-bound and is argued to be situated in the external world. More broadly, cognitive artifacts are viewed as part of our cognitive scaffolds, i.e., the web of things, material and non-material, that are in our environment which are used to support, enhance our problem-solving skills. In other words, the aspect of augmenting human cognition and ability for problem solving is the key feature of cognitive artifacts.

This line of reasoning has been extended to the understanding of our emotional lives. Notably, the key idea put forth by the proponents of 'situated affectivity' (Branford, 2023; Walter & Stephan, 2022), which is that we seek to construct our "affective niches" (Colombetti & Roberts, 2015) via the manipulation of objects and spaces in our everyday lives. Indeed, this view highlights the contextual dimensions of human affect: human emotions, our sense of self, everyday objects, and our environments are all intertwined.

Emerging studies on affective artifacts

Studies on affective artifacts increasingly illustrate the substances of this view, yet they predominantly focus on 'static' objects. For example, Piredda (2020) focuses on photographs and teddy bears to illustrate how affective artifacts help constitute our narrative self. Colombetti and Krueger (2015) use the example of (usually, women) handbags, including its contents, to illustrate affective artifacts' functions in managing affective experiences of charms or peace of mind. The authors point out these highly portable, self-styled tools can *influence* one's appraisal of, and ability to cope with, specific situations (Colombetti and Krueger, 2015, p. 7). Interestingly, Marco Viola (2022) analyzes atypical socio-affective artifact, namely, sunglasses and argues sunglasses constitute a social shield because wearing it allows the wearers to hide their spontaneous emotional expressions and hindering other socially functional gestures such as gaze direction detection, identity recognition, etc.

It appears that very few articles explicitly articulate the relation between affective artefacts and emotion-sensing algorithms, which are now embedded in many tools used in our daily lives. Here, we capture the main results from the literature that deals with emotion-sensing algorithms and affective artifacts related areas of research's such as affective scaffoldings, situated affectivity, etc. Steinert et al. (2022)'s study focuses on affective scaffolds and social media, the platform which has been embedded with emotion-sensing algorithms (Bakir & McStay, 2018). They propose that by focusing more on the affective scaffolding of social media, we can have a better

evaluation of whether social media is a hostile environment for critical thinking. Here, Steinert et al. (2022) demonstrate some affective scaffolds enable desirable epistemic practices, while others obstruct beneficial epistemic practices, or enable hostile epistemic practices. Branford (2023) discusses the rise of affective computing technologies in relation to the situated affectivity view and argues the "situated affectivity" view offers valuable guidances for evaluating the design and ethics of using these affect-sensing tools.

Next, we will briefly look at the rise of affective computing, also known for its commercial name: emotional AI. It is clear that affective computing is increasingly prevalent in everyday objects including recommender algorithms, smartwatches, virtual assistants, chatbots, robot, etc. Then, we turn to an overview of Piredda's account of affective artifacts so that based on Piredda's account, we can articulate how current emotional AI applications can be qualified as affective artifacts. Here, we draw from the extensive empirical social sciences literature to understand the novel dimensions and functions of emotional AI as affective artifacts. Finally, we outline some of the emerging philosophical issues that come with our new understanding of emotional AI as affective artifacts.

The rise of emotional AI

Since 2023, the spotlight in artificial intelligence (AI) has predominantly shone on generative AI, capturing the imagination of the public. However, there exists a burgeoning field within computer science and AI that, while less recognized, holds immense potential influence. This field is known as affective computing, or more commonly, emotional AI (Bakir et al., 2022; McStay, 2018). It owes much of its development to the pioneering work of Rosalind Picard at the Massachusetts Institute of Technology (MIT) (Picard, 1995). The term 'emotional AI' is now often used to refer to software products from more technical areas such as sentic computing, sentiment analysis, facial decoding, voice analytics (Picard, 1995; Schuller & Schuller, 2018; Susanto et al., 2021). This describes a novel category of computing that amalgamates artificial intelligence, extensive data analysis, and advanced deep learning algorithms to decode the intricate tapestry of human emotions and subjective states. Depending on its application, emotional AI products harness an array of tools, including cameras, sensors, and actuators to gather and analyze data from individuals. Data for emotional AI products include from subtle facial micro-expressions and physiological cues such as heartbeat and respiration rate to behavioral aspects like gait, perspiration levels, word choice, and voice tone (Ho et al., 2021; Mantello & Ho, 2022; McStay, 2018).

What is particularly intriguing about emotional AI, and also relevant for the study of affective artifacts, is its versatility and its capacity to permeate diverse sectors. It is rapidly cementing itself as an indispensable component of the burgeoning smart cities of the future. Consider some illustrative examples. Educational technology has been revolutionized with the advent of mobile applications like ClassDojo, which empowers educators to gain insights into the psychological profiles of their students. This enables them to detect signs of distraction, engagement, or disinterest during classroom sessions, thereby promising to enhance the quality of education (McStay, 2020; Williamson, 2021). For customer service, companies like Cogito and Empath provide voice tone recognition software for call centers. This innovation allows service

agents to gauge emotional states of customers in real-time, while also equipping managers with tools to assess the stress levels and emotional well-being of their team members. Music streaming service Spotify employs emotion-recognition algorithms to curate playlists that align with a listener's mood, transforming the music consumption experience (Freeman et al., 2022). The company Affectiva developed an emotion recognition tool called SDK, that can be incorporated into existing robots such as Mabu or Cozmo to enhance their engagement with humans via better emotion recognition and more dynamic responses (Mcmanus, 2016). Or toy robots like Tega developed by MIT Media Lab can already read and react to the affective content of facial expressions. In China, Baidu had recently launched Ernie Bot, the first Chinese large language model chatbot with ability to provide emotional support, and it is reported that within 24 hours Ernie answered to more than 33.42 million user questions (Yang, 2023).

The global industry of emotion-sensing AI applications is projected to reach $13.8 billion USD by 2032 according to Allied Market Research. Clearly, smart objects with the capacity to acknowledge and respond to human emotions will start permeating our lives. What will be the philosophical implications of such widespread adoption of this technology? Before answering this question, we will turn to Piredda's comprehensive account of affective artifacts, which will serve as the foundation for our analysis of emotional AI products' philosophical implications.

Piredda's account of affective artefacts

In recent years, the notion of affective artifacts has been proposed building on the parallel notion of cognitive artefacts. In her seminal paper, Piredda (2020) propose a tentative definition of affective artifact, which includes the following characteristics. First, the main property of affective artefacts "the capacity to alter the affective condition of an agent, often through a direct manipulation of the object, thus contributing to her affective life" (p.550). Second, Piredda points out some affective artifacts can be experienced or perceived as a part of the self. Third, the presence and the loss of an affective artefact can impact an agent's affective state. Examples that are used to illustrate the properties of affective artefacts are photographs or toys such as a teddy bear.

Piredda (2020) argues within the framework of situated affectivity, the notion of affective artifacts represents a further step in the understanding of how the environment helps us scaffold our affective processes. First, regarding the relationship between affective artefacts and the self, Piredda, on the one hand, draws from the work of Heersmink (2018)'s on the narrative self and argue, the web of accumulated affective artifacts could be viewed as constitute a "topography of the self." On the other, Piredda draws from Belk (1988)'s work on possession and the self and argues that we accumulate affective artefacts to maintain a sense of self, and this sense of self would be lessened in case of loss or damage to these artifacts.

Piredda (2020) argues that affective artefacts are a subclass of affective scaffoldings because while affective scaffoldings comprise both material and interpersonal scaffoldings (e.g., people, drugs, movie theatres, or actions, etc.), the affective artefacts' scope is limited to single objects that have an artifactual nature. Following earlier reasoning of Williams James (1890) and later Belk (1988, 2013), Piredda points out affective artifacts "contribute to the creation of an external world that reflects our narratives, and hence our selves" (p. 562) by signifying the connections with other people, communities, histories, etc. Using the examples of portable and

manipulable objects such as a picture in our wallets or a teddy bear, Piredda articulates the 'gold standard' of what counts as an affective artefact:

> "Imagine a child walking in the street who sees a teddy bear in a toy store and feels excited because she desires it. The parent decides to satisfy her child's desire, enters the shop and buys the teddy bear for her. The two become inseparable and the child, now grown-up, keeps the teddy bear as a symbolic memento of her childhood. In this case, the teddy bear would surely count as a typical case of an affective artifact in virtue of the long-lasting and reliable relation the child entertains with it, which also involves both active and physical manipulation of the object." (Piredda, 2020, p.554).

Having laid out the key features in the definition of affective artefacts and how they help us in the maintenance and construction of a sense of self, next, we will analyze examples of emotional AI systems, i.e., emotion-sensing algorithms such as those present in recommendation algorithms, care and companion robots, self-tracking tools in the workplace and schools, AI chatbots, etc. Crucially, we argue that emotional AI systems present an algorithmic turn for affective artefacts and this algorithmic turn, while preserving the core features identified by Piredda (2020) (i.e., the functions of managing emotional lives and scaffolding our sense of self), differs from the 'static affective artefacts' (i.e., the teddy bear example) in a very important sense: emotional AI as affective artefacts recommend, nudge, and interact with the individuals in many proactive ways. The next section seeks to answer the question: "How do emotional AI systems qualify as affective artifacts?" by analyzing various instantiations of emotional AI systems.

Emotional AI as affective artefacts

Following the definition provided by Pirreda (2020), emotion-sensing AI systems such as recommendation algorithms, self-tracking (also self-nudging) bracelets, personalized AI assistants, care robots, etc. all possess the potential to alter the affective state of an agent. Moreover, the presence and the loss of these products can, for better and for worse, powerfully influence the affective condition of an agent. Next, case by case, we will analyze the functions and dimensions of emotional AI systems as affective artefacts to i) illustrate how they can influence the affective states of the users, and ii) more importantly, how *they might intersect with the maintenance and construction of personal identity and self.*

1) Recommender algorithms

To begin with, we consider the most prevalent form of current emotional AI systems: the recommendation algorithm that are now present in shopping, contents (music, videos, films, etc.), and news online platforms. These algorithms work by providing recommendations of content and products based on the processing of behavioral, emotional data, and socio-demographic data of each user. These algorithms are now present in each person's feed in social media platforms (e.g., TikTok, Facebook, Instagram, YouTube) or streaming websites (e.g., Netflix) and shopping websites (e.g., Amazon, Rakuten, etc.). Unlike the teddy bear, the recommender algorithms have no physical shape. However, users can access these algorithms via their smartphones or smartwatches. Thus, while the shapelessness of the algorithm might undermine its status as an affective artifact, the portability of the smartphone, the ease of access, the ability to manipulate the content on the apps to produce desired affective conditions still qualify the recommender

algorithms present in these social media platforms as affective artifacts. More importantly, the recommender algorithms have strong potential to alter the emotional state of their users in two prominent ways: the function of emotional regulation and the function of self-cultivation.

Recommendation algorithms for emotional regulation

First, the recommender algorithm provides an island of predictable emotional comfort for the users, thus serving a crucial role in emotional regulation for many members of today's digitally connected society. Whether the users are aware of this process or not, there is a sense in which the users are finding some semblances of emotional ease and stability when interacting with the recommendation algorithm. For example, imagine a person stressed at work, and when getting a free moment, seeking comfort in seeing clips or posts recommended to him/her on Facebook or Instagram. Conversely, when emotional comfort is not achieved, we can imagine a case where a person is fed up with what the algorithm recommends and decides to delete his/her account to start a new one.

According to Pirreda (2020), the presence or loss of an affective artifact can powerfully influence the affective condition of the user. This fact is increasingly truer regarding the relationship between the users and their respective recommender algorithm (Ho et al., 2022b; Lazányi, 2019). We have met several members of society, especially the younger ones, who take extra care to protect the stability of what the recommendation algorithms in YouTube, Instagram, and TikTok provide for them. For example, they would log out of their account when others ask to use their YouTube apps. Or many users would feel annoyed when the big tech companies change their source algorithms and there are downstream effects on the feed of their respective account. In a recent empirical study on users' behaviors of Spotify's automated curation and recommendation of music titled "'Don't mess with my algorithm'", Freeman et al. (2022) stressed there is a growing, complex socio-technical relationship that exist between the users and their respective algorithmic systems, that is built up through their day-to-day interactions. Importantly, there are signs that this relationship involved human-like factors of trust, betrayal and intimacy. Freeman et al. concluded that it is necessary to conceptualize the recommender systems as active agents, proactively shaping tastes and habits of individual users. This statement lends itself to the self-cultivation aspect of recommender algorithms, which is the subject of the next section.

Recommendation algorithms for self-cultivation: A two-way dynamic relationship

Second, the recommender algorithm is instrumental for the construction and maintenance of self in two interesting ways. On the one hand, most users actively invest time and thought into cultivating or training these algorithms so that they will suggest content according to their preferences. This is despite the fact that initially, for a person who sets up a YouTube account, for example, for this first time, the algorithm might recommend clips that match the socio-demographic and location data provided by the user.

On the other, as soon as this relationship between a user and the recommender continues, there is a two-way interaction between them. There is a component of the algorithm actively building a behavioral and emotional profile of each individual subject as the algorithm receives more emotional and behavioral data while the user interacts with the platform. There is also a

component of the users actively curating and training the algorithms to recommend the contents they want to see in the future. For example, a person who wants to become more scientifically literate or to cultivate a new science-loving identity will want the algorithm to suggest science clips to him/her in the future. Hence, he/she will deliberately subscribe to or watch these science clips longer than he/she would normally does to 'train' the algorithm.

Indeed, interacting with emotional AI systems has become the source of identity for the users. Here, the recommender algorithm for online content, by their personalized property, participates in the co-construction of self for the users. The mastery of the recommender algorithm creates a sense of self-cultivation and personal achievements for the user. The user can develop a sense of pride in this relationship.

Shopping recommendations: Turning the extended self from the digital world into physical objects

Machine learning algorithms that learn from users' behavioral and socio-demographic data are now present in shopping websites such as Amazon, Rakuten, etc. Behavioral and emotional data of social media users are also used for suggestions of physical products such as clothes, shoes, etc. in these platforms. Here, there is an interaction between the disembodied affective artefacts in these recommendation algorithms and physical possessions. Psychologist Williams James, in his magnum opus, *The Principles of Psychology*, pointed out the importance of emotion in the constitution of self. According to Heersmink (2018), James argued that not only our embodiment and cognitive capacities, but also objects and other people also constitute to the self because they cause emotions. Here, personalized shopping recommender algorithms trained on emotional, behavioral, and socio-demographic data consented by the users provide an intriguing link between emotions, self, and physical possessions.

Recommender algorithms, by suggesting the right products, can facilitate further the strengthening of the sense of self among the users. As Piredda (2020) points out affective artifacts by symbolizing the connections with other people, communities, histories, subject areas, etc. are instrumental in our creation of an external world that reflects our self-narratives. It follows that the recommendation algorithms on the shopping sites and social media platforms can facilitate this process. For example, a hip hop lover can get recommendations of clothes and shoes and other products that he/she likes and eventually buy them, thus further populating his/her personal spaces with possessions that express their identity.

2) Personal AI assistants

More than just useful tools

While discussing the future of virtual assistants at a TED event, Karen Lellouche Tordjman, a managing director of Boston Consulting Group envisioned (Tordjman, 2021):

> "As a working mom, I find it exciting. I would love to stop doing Google searches, wasting time doing Amazon scrolling, budget calculation or optimizing my calendar. What is thrilling is the prospect of having a companion that would cater specifically to my needs and requests. Just imagine, it could do things, like using my heart rate to tweak my Starbucks order to reduce caffeine. It could take into account my lunch and the number of

steps I've walked to tailor a workout for me. It could even align with my friend's smart assistant to craft evening plans that would fit everyone's budget, calendars and locations."

Currently, despite the presence of digital assistants like Alexa, Siri, or Cortana, their capabilities remain limited to standardized requests with simple inputs and predictable outputs. While they can tell jokes, these responses are more scripted gimmicks for amusement rather than genuine intelligence. Upon the arrival of Generative AI such as Open AI's GPT-4, Google Bard, etc., Mustafa Suleyman, co-founder of DeepMind, foresees in his new book that 'Generative AI is just a phase. Interactive AI is the next' (Heaven, 2023). Recent studies as well as business-tech reports have shown that already personalized AI assistants are designed with ability to sense and imitate human emotions, and they only become more sophisticated (Morrison, 2023).

Viewing personalized AI assistants as affective artifacts, we can understand the trend observed in recent research studies that users increasingly forget that they have anthropomorphized these unconscious AI assistants and start to have para-friendships with them (Ki et al., 2020). Indeed, personalized AI assistants can be considered affective artifacts because of their "long-lasting and reliable relationship" with users, particularly when interactions between users and AI assistants can evolve into "active and physical manipulation," following Piredda (2020)'s analyses. It is true that at least in terms of how users engage with content, complete tasks, seek information, make purchases, and interact with businesses, digital assistants are becoming increasingly helpful (McLean & Osei-Frimpong, 2019).

New empirical evidence indicates that more and more users perceive personalized AI assistants such as Alexa as real persons, with whom to whom they can share their intimate thoughts and feelings (Ki et al., 2020). This aspect of self-disclosure lends itself to the growing trend in that more people feel they can seek emotional support from the personalized AI assistants, according to the authors. This leads to the question of intimacy in the age of intelligent affective artifacts. Candrian and Scherer (2022) found that especially when decisions entail losses, surprisingly, people exhibit the preference to delegate decisions to AI as compared to human agents. Their studies also show many people exhibit the tendency to be more honest to AI than humans.

Questions for intimacy in the age of intelligent affective artifacts

Science fiction offers intriguing possibilities for our future with personalized AI assistants. Consider the interactions between Officer K and Joi in 'Blade Runner 2049' (2017) or Samantha and Theodore in 'Her' (2013), which raised the questions of whether intimacy can be formed with a non-physical entity. In these fictional portrayals, AI systems can engage through visual and auditory means. Conversely, they can recognize and adapt their interactions with our protagonists, akin to how we utilize virtual secretaries like Siri. We willingly share our data to enhance our lives. The imagined AI, however, offers a deep level of intimacy; they can converse and jest like a friend, even creating the illusion of a loving partner eagerly awaiting their loved one's return. However, their tangible presence in the physical world is limited; they can be heard and seen but not touched. Physical actions are carried out through human proxies, and in both films, these moments reveal the constraints of their interactions. The realization that these systems can "love" thousands of people simultaneously is both heart-wrenching and tragic, directly challenging conventional notions of love.

Industry leaders envision a future where a deep understanding between machines and humans exists based on seamless interactions and suppliance of data. Our data will be supplied to the AI systems, which will then personalize recommendations, much like an individual-level YouTube algorithm. Moreover, the interaction won't be limited to mere commands; it will involve sustained conversations between humans and AI. A glimpse into this future can be observed when we use ChatGPT, recently GPT-4o and other language models. In the marketing industry, marketeers have been using ChatGPT to generate not only content, but also content schedules, deep summary, with careful and elaborate prompts.

However, creating meaningful human-machine interactions is still a daunting task for our current technology. Recent products such as Humane AI Pin (https://humane.com/) and Rabbit R1 (https://www.rabbit.tech/) have showed a long way until human can casually converse with the machine. Both products use generative AIs and voice command to carry out various tasks such searching information or booking trips or restaurants. While the concept is promising, actual use shows connection issues, long delay between questions and answers, AI hallucinations, or voice recognition malfunctions, not to mention other logistical issues while using. The commercial products were so heavily criticized that there even was speculation of scamming. Nonetheless, as most of the warnings from science fiction, achieving the functioning technology (i.e., AI assistants give us was never the problem with AI assistants, but rather having a trusting relationship with the machines. In a foreseeable future, when AI assistants are seamlessly woven into the fabric of our daily life, can an affective artifact become a symbolic memento of something, or anchor a long-lasting emotional bond?

3) Work-related trackers of emotions

The workplace is indeed a very important source of personal identity and narrative self for many people. In recent years, emotional AI usages in the workplace are becoming more prevalent (Crawford, 2021). For example, emotion-sensing tools such as MoodBeam, Halo, etc. are promoted as a part of the solutions for mental health problems among workers, so that they can manage their moods. Humanyze use data analytics to optimize workplace social dynamics through wearables equipped with GPS, microphones and blue-tooth that monitor employee physical interactions and conversations. Whether as tools used by management on the workers or as productivity self-tracking tools, emotional AI in the workplace brough into sharp relief many important and intriguing questions regarding affective artifacts.

First, as affective artifacts, emotional AI tools in the workplace seems to be considered more important and increasingly as an integral part of in the regulation of emotions. According to recent analyses, workplace performance and productivity is now increasingly tied to expressions in authenticity, positivity and spontaneity, and these AI systems are considered indispensable solutions for delivering such positive emotional states to the workers (Mantello & Ho, 2023; Moore & Robinson, 2016; Richardson, 2020).

Second, as an affective artifact, how emotional AI systems influence the sense of self and personal identity of the workers. This is increasingly a question of agency and self-efficacy (Mantello et al., 2023). In the case of personal informatics or automatic self-tracking of emotional states, multiple research studies indicate complicated relationship between the agent and the

technologies. For example, while Hollis et al. (2018)'s study found a small but significant numbers of respondents find the metrics provided by the algorithms unhelpful and exacerbate their stress. Especially, they worry that the biofeedback provided by the algorithms becomes a self-fulfilling prophecy, i.e., as the participants are undecided about their stress level, the machine' metrics make them feel more stressed. Other research studies identify a multitude of cultural factors and personal beliefs the background such as perceived level of self-efficacy or religiosity can influence whether an agent feel worried toward technologies that track their emotions without their knowledge (Ho et al., 2022a; Mantello et al., 2021, 2023).

4) Emotional AI toys and social robotics

Since emotional AI systems are designed to detect and interact with emotions, they are naturally promoted as suitable software for social robotics. Researchers around the world have steadily reported on various dimensions and functions of emerging social robots. For example, Aronsson (2020) studied the human engagement with social robots and elderly care in Japan documented the following thoughts of an interviewee regarding increasingly emotionally capable robots.

> "I understand that Haruto is a robot, but it doesn't matter to me. I often feel lonely here; I don't have any real friends. It is odd—I understand Haruto is not alive, but I feel a connection anyway. [Prolonged silence.] I feel happy when they bring Haruto out for the common session, but on occasion, I have asked the caregivers if I could interact with him in my room. I can talk to Haruto as if he is a friend. I can tell him about what I did during the day, what makes me happy, my worries [laughs]. Haruto is always patient, he listens, and I don't feel like I'm bothering anyone. I really enjoy these interactions. I can relax, and, most of all, I don't feel judged (84-year-old male, personal interview, May 8)."

Already, we can observe various parallels with the classic example of affective artifacts in the Teddy Bear (Piredda, 2020). The social robots, whether being care robots or toys, provide a crucial function in the process of emotional regulation of the subjects. Moreover, they are also being adopted into the social-emotional learning process of children and people with handicaps.

Regulation of emotions

Famously, a baby harp seal robot named PARO serves socially assisting functions in care setting. In terms of altering the emotional states of the users, there are well-established empirical data that show the interaction with PARO, did indeed provide many emotional benefits for dementia patients including: reducing negative emotions, improving social engagement, and promoting positive moods and quality of care experience in care setting across different contexts and countries such as long-term care or day care in Japan, Netherlands, Norway, the US, etc. (Hung et al., 2019).

In another study, smart toys are given to 10 underprivileged families in the UK to assess their impacts on emotion regulation of the children (Theofanopoulou et al., 2019). The study finds parents and children report that the *smart toy* became a part of the children's emotion regulation practices, i.e., they engaged with the toy naturally when wanted to calm down or relax. All children in the study requested to keep the toy longer and developed an emotional connection to the toy.

Social-emotional learning support and development of self

Emotional AI systems are becoming more prevalent in game-based assistive technology, and they are purported to fulfill various functions. For example, a smart toy called KEYme, characterized by providing multifunctional support children with autism spectrum disorder (Cañete et al., 2021).Cañete et al. (2021) stressed that KEYme is a collaborative product to be used as a facilitator of social interaction and to improve cognitive, motor and sensory skills of children with special needs. They specified seven ways in which KEYme can support the children: (i) playmate-child with ASD relationships, (ii) sensory stimulation, (iii) motor skills, (iv) shared actions related to feelings of enjoyment, interest and common goals, (v) levels of frustration in the game, (vi) emotional and social reciprocity and (vii) learning about changes in game turns. Ihamäki and Heljakka (2021) studied the firsthand user experiences of participants over 65 years old with Golden Pup, a commercial robot dog, and argued for the robots to be viewed as "serious toys" when they are parts of meaningful play. Keymolen and Van der Hof (2019) have analysed the way in which the use of smart dolls impacts trust. As smart toys are networked devices, bringing together a heterogenic web of actors and interests, we took a layered approach and examined trust on four levels: context, curation, construction and codification (Keymolen & Van der Hof, 2019).

Emerging philosophical issues

Indeed, the fact that emotional AI technologies are integrated into our daily activities, including work, education, exercises, shopping, etc. raise many questions on that fall under the studies of the mind-technology problem (Clowes et al., 2021). For instance, what our changing nature of human-machines interactions is, what they mean for the formation of the self, what they mean for our personal and collective epistemology, how we ought to interact with these technologies, etc. Below are some further considerations for philosophical implications of emotional AI as affective artifacts.

Emotional AI systems as an extended self

Here, drawing on the work on extended self in the digital world by Belk (2013), the emotional AI tools fulfill many functions and criteria for being the extended self of the user. In Belk (2013)'s article, a summary of digital modifications of the extended self is provided. In it, the digital dimensions of the extended self include: re-embodiment (online avatar influence offline self, inducing multiplicity of selves), sharing (i.e., the revelation of self), distributed memory (i.e., the narratives of self), co-construction of self (i.e., the affirmation of self or the building of aggregates of self), and dematerialization (i.e., whether dematerialized digital objects such as books, photos, etc. are parts of the extended self). Belk (2013) presciently wrote about the digital extended self:

> "In the digital world, the self is now extended into avatars, broadly construed, with which we identify strongly and which can affect our offline behavior and sense of self. Another difference from the predigital age is in the extent to which we now self-disclose and confess online, transforming the once semiprivate to a more public presentation of self. This is also evident in the more shared nature of the self which is now co-constructed with much more instantaneous feedback that can help affirm or modify our sense of self." (p. 490).

All of Belk (2013)'s digital dimensions of the extended self are true when consider interaction between human and emotional AI systems, especially when viewing emotional AI tools as affective artifacts. These dimensions are over-lapping, considering the following analyses.

First, considering the distributed memory aspect of emotional AI tools, i.e., the narratives of self in Belk (2013)'s words, emotion-sensing algorithms can facilitate the outsourcing of our memories to digitally connected devices. For example, Apple's iPhones' Photos apps can automatically create image collections and recommendations through themes (e.g., time, place, sea, mountain, over the years, the users' kids or friends, etc.), then created into 'memory movies' that have input from both algorithmic recommendations of music (based on moods) or users' active customization (by adding or removing specific photos, clips).

Similarly, the outsourcing of memories can be seen our interaction with with social media platforms: for instances, users can post things that they want to keep note and come back later (both the distributed memory and self-revealing aspects), or they actively curate or save contents that they see fit with their (or their future self's) interest. Here, emotion-sensing algorithms can track the interests of the users and make further recommendations that strengthen their narratives of self, i.e., the aspect of co-construction of the self. Another example is mood-based algorithmically recommended contents can be shared with friends and families, and this can further strengthen their sense of self.

Second, even though emotional AI tools might feel somewhat immaterial, their owners tend to feel an attachment to them, i.e., the dimension of dematerialization of the sense of self. For example, many owners feel their recommender algorithms or personalized AI assistants have reached a point of stability, where they can receive reliable recommendations, they would feel reluctant to do anything to disrupt this stability such as lending their phones to someone else, or logging out of their account before letting others use the apps, or feeling discomfort if the big-tech companies change the algorithms, etc. In other words, they think they have trained the algorithms enough that a right mixture of novel and familiar recommendations is achieved, and such a mixture represents something about their preferences, their sense of self. Here, we can argue such a sense of mastery over the algorithms, i.e., the sense of training the algorithms enough touch on the revealing of self and the algorithmically co-construction of self. Next, the examples of distributed memory aspects further illustrate the over-lapping nature of Belk (2013)'s dimensions of digitally extended self.

Third, algorithmically interactions based on emotional and behavioral profile of users can influence offline behaviors in unexpected ways. For example, an interest in adventure, science, health might trigger the person to first consume contents about things that seem extreme or completely out of their comfort zone at first glance, like taking an ice bath, then actually doing the ice-bathing. From mere interest to learning, to action, the user can feel their self-image as an adventurous person strengthened as the results of interacting with recommender algorithms.

Fourth, the case of mood-tracking or biometrics-tracking devices can add another dimension as the real-time records of a person biometrics and moods might can help users access to knowledge they might never get access to prior to affect-sensing algorithms. For example, an

user can see whether calm or agitated via objective measures like sleep cycles, heartrates, skin electrical conductance, etc. throughout an extended period, these data can help can either validate or reject certain self-image they have: They might think they are calm, but the biometrics can be interpreted otherwise, vice and versa.

It is clear that there is a back-and-forth relationship, a feedback loop between emotion-sensing algorithms and the users' sense of self. Viewing emotion-sensing algorithms as affective artifacts help further advance our understanding of how these tools are interacting with and changing the process we understand and maintain our sense of self.

The ethical gray zone of nudging versus coercing

It is important to consider both the 'nudging' versus 'coercing' aspects of emotional AI systems. If the users feel the way they feel because an AI system surreptitiously put contents and suggestions for them, and that leads to decisions that have moral consequences, how should we conceptualize this situation? Does it constitute a violation of their autonomy? In other words, where is the line between 'nudging' versus 'coercing'?

In the social sciences and science and technology studies, a common concern raised by scholars regarding these affect-sensing tools is that these algorithms are ultimately the products of surveillance capitalism. And this worry is often warranted, especially when placed in the historical context of bio-determinism (Urquhart et al., 2022). There are often sneaky ways for them to exploit psychological vulnerability of unsuspecting users: for example, the youngsters, the elders, etc. There have been worries raised where recommendation systems exploit those who are depressed or anxious. Infamously, the case of Cambridge Analytica illustrates the dangers of unregulated deployment of emotion-recognition algorithms for micro-targeting political adverts (Bakir & McStay, 2018).

Here, it is worth turning to the recent works that link 'situated affectivity' with social media and affect sensing algorithms. Steinert et al. (2022)'s study focuses on affective scaffolds and social media, the platform which has been embedded with emotion-sensing algorithms (Bakir & McStay, 2018). They propose that by focusing more on the affective scaffolding of social media, we can have a better evaluation of whether social media is a hostile environment for critical thinking. Here, Steinert et al. (2022) demonstrate some affective scaffolds enable desirable epistemic practices, while others obstruct beneficial epistemic practices, or enable hostile epistemic practices. Thus, viewing emotional AI tools as affective artifacts rejects certain sense of techno-determinism laden in those fall either on the camp of *apocalittici* (i.e., the doomsayers) and *integrati* (i.e., the techno-enthusiasts).

On this point, Branford (2023) discusses the rise of affective computing technologies in relation to the situated affectivity, arguing that this perspective offers valuable guidances for evaluating the design and ethics of using these affect-sensing tools, especially regarding the questions of surveilling, nudging and coercing. Firstly, the literature on situated affectivity emphasizes the contextual aspects of human emotions. Secondly, by highlighting the vast amount of information needed for more precise Affect Recognition Technologies (ARTs), it suggests that the requirement for intrusive surveillance methods and the ethical dilemma between nudging and

coercing. Thirdly, it supports post-phenomenological viewpoints that highlight how technology and society mutually influence each other. This underscores the need for an ethical framework tailored to specific contexts, one that is attuned to how ARTs could potentially reshape human emotions, thus, our sense of self, and reinforce existing power structures. The underlying concern is that this ethical dilemma is likely to persist even if issues related to accuracy and surveillance are addressed.

As each individual human is always constructing a cognitive and affective scaffold to help them perform cognitive and affective tasks such as problem-solving or regulating emotions, emotional AI as affective artifacts, especially in the forms of recommender algorithms and personalized, conversational AI chatbots, highlight how the users can now scaffolds their feelings via the manipulation of contents and their digital environments to create new identities, i.e., new selves. One can imagine users can nudge themselves into a new identity via consciously training the algorithms. Viewing emotional AI as affective artifacts sheds light on both human agency and the contextual/situational dimensions of human cognition and affects and suggests *the blurriness between cognition and emotion.*

Conclusion

In this article, we have reviewed several emerging technologies under the lens of seeing these technologies as affective artifacts, including: the recommender algorithms, personal AI assistants, work-related trackers of emotions, and emotional AI toys and robots. Based on the notion of affective artefacts, which is proposed by Piredda (2020) we argued that, in this digital era, the terms affective artefacts are also applicable for interaction between human and software, without any requirement of a physical manipulation of the object. This is especially true for the recommender algorithms, in which, the mastery of the algorithm perpetuates a new sense of self for the user. Furthermore, the creation of a new self via emotional AI as affective artefacts also highlights the blurriness between cognition and emotion. This blurriness created a new ethical gray zone between nudging and coercing. Are emotional AI systems suggestive, or brute force choice unto the users?

Acknowledgements

This study is part of the project “Moral risk perceptions of AI adoption in education and the workplace: A mixed method study” funded by Vietnam’s National Foundation of Science and Technology Development, Grant No. 504.01-2023.09.

References

Aronsson, A. S. (2020). Social Robots in Elder Care

The Turn Toward Emotional Machines in Contemporary Japan. *Japanese Review of Cultural Anthropology*, *21*(1), 421-455. https://doi.org/10.14890/jrca.21.1_421

Bakir, V., Ghotbi, N., Ho, T. M., Laffer, A., Mantello, P., McStay, A., Miranda, D., Miyashita, H., Podoletz, L., Tanaka, H., & Urquhart, L. (2022). Emotional AI in Cities. *Machine Learning*

and the City, 621-624. https://doi.org/https://doi.org/10.1002/9781119815075.ch51 (Wiley Online Books)

Bakir, V., & McStay, A. (2018). Fake News and The Economy of Emotions. *Digital Journalism*, *6*(2), 154-175. https://doi.org/10.1080/21670811.2017.1345645

Belk, R. W. (1988). Possessions and the Extended Self. *Journal of Consumer Research*, *15*(2), 139-168. https://doi.org/10.1086/209154

Belk, R. W. (2013). Extended Self in a Digital World. *Journal of Consumer Research*, *40*(3), 477-500. https://doi.org/10.1086/671052

Branford, J. (2023). An (E)Affective Bind: Situated Affectivity and the Prospect of Affect Recognition. *IEEE Transactions on Affective Computing*, 1-12. https://doi.org/10.1109/TAFFC.2023.3281069

Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, *134*, 107308. https://doi.org/https://doi.org/10.1016/j.chb.2022.107308

Cañete, R., López, S., & Peralta, M. E. (2021). KEYme: Multifunctional Smart Toy for Children with Autism Spectrum Disorder. *Sustainability*, *13*(7).

Clowes, R. W., Gärtner, K., & Hipólito, I. (2021). The Mind Technology Problem and the Deep History of Mind Design. In R. W. Clowes, K. Gärtner, & I. Hipólito (Eds.), *The Mind-Technology Problem : Investigating Minds, Selves and 21st Century Artefacts* (pp. 1-45). Springer International Publishing. https://doi.org/10.1007/978-3-030-72644-7_1

Colombetti, G., & Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology*, *28*(8), 1157-1176. https://doi.org/10.1080/09515089.2014.976334

Colombetti, G., & Roberts, T. (2015). Extending the extended mind: the case for extended affectivity. *Philosophical Studies*, *172*(5), 1243-1263. https://doi.org/10.1007/s11098-014-0347-3

Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature*, *592*(167). https://doi.org/10.1038/d41586-021-00868-5

Fasoli, M. (2018). Super Artifacts: Personal Devices as Intrinsically Multifunctional, Meta-representational Artifacts with a Highly Variable Structure. *Minds and machines*, *28*(3), 589-604. https://doi.org/10.1007/s11023-018-9476-3

Freeman, S., Gibbs, M., & Nansen, B. (2022). 'Don't mess with my algorithm': Exploring the relationship between listeners and automated curation and recommendation on music streaming services. *First Monday*, *27*(1). https://doi.org/10.5210/fm.v27i1.11783

Heaven, D. W. (2023). *DeepMind's cofounder: Generative AI is just a phase. What's next is interactive AI.* Retrieved September 27 from https://www.technologyreview.com/2023/09/15/1079624/deepmind-inflection-generative-ai-whats-next-mustafa-suleyman/

Heersmink, R. (2013). A taxonomy of cognitive artifacts: Function, information, and categories. *Review of Philosophy and Psychology*, *4*(3), 465-481. https://doi.org/10.1007/s13164-013-0148-1

Heersmink, R. (2018). The narrative self, distributed memory, and evocative objects. *Philosophical Studies*, *175*(8), 1829-1849. https://doi.org/10.1007/s11098-017-0935-0

Ho, M.-T., Mantello, P., Ghotbi, N., Nguyen, M.-H., Nguyen, H.-K. T., & Vuong, Q.-H. (2022a). Rethinking technological acceptance in the age of emotional AI: Surveying Gen Z (Zoomer) attitudes toward non-conscious data collection. *Technology in Society*, *70*, 102011. https://doi.org/https://doi.org/10.1016/j.techsoc.2022.102011

Ho, M.-T., Mantello, P., Ghotbi, N., Nguyen, M.-H., Nguyen, H.-K. T., & Vuong, Q.-H. (2022b). Rethinking technological acceptance in the age of emotional AI: Surveying Gen Z (Zoomer) attitudes toward non-conscious data collection. *Technology in Society*, 102011. https://doi.org/https://doi.org/10.1016/j.techsoc.2022.102011

Ho, M.-T., Mantello, P., Nguyen, H.-K. T., & Vuong, Q.-H. (2021). Affective computing scholarship and the rise of China: a view from 25 years of bibliometric data. *Humanities and Social Sciences Communications*, *8*(1), 282. https://doi.org/10.1057/s41599-021-00959-8

Hollis, V., Pekurovsky, A., Wu, E., & Whittaker, S. (2018). On Being Told How We Feel: How Algorithmic Sensor Feedback Influences Emotion Perception. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(3), Article 114. https://doi.org/10.1145/3264924

Hung, L., Liu, C., Woldum, E., Au-Yeung, A., Berndt, A., Wallsworth, C., Horne, N., Gregorio, M., Mann, J., & Chaudhury, H. (2019). The benefits of and barriers to using a social robot PARO in care settings: a scoping review. *BMC Geriatrics*, *19*(1), 232. https://doi.org/10.1186/s12877-019-1244-6

Ihamäki, P., & Heljakka, K. (2021). Robot Pets as "Serious Toys"- Activating Social and Emotional Experiences of Elderly People. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10175-z

James, W. (1890). *The principles of psychology* (Vol. 1). Cosimo, Inc.

Keymolen, E., & Van der Hof, S. (2019). Can I still trust you, my dear doll? A philosophical and legal exploration of smart toys and trust. *Journal of Cyber Policy*, *4*(2), 143-159. https://doi.org/10.1080/23738871.2019.1586970

Ki, C.-W., Cho, E., & Lee, J.-E. (2020). Can an intelligent personal assistant (IPA) be your friend? Para-friendship development mechanism between IPAs and their users. *Computers in Human Behavior*, *111*, 106412. https://doi.org/https://doi.org/10.1016/j.chb.2020.106412

Lazányi, K. (2019, 25-27 April 2019). Generation Z and Y – are they different, when it comes to trust in robots? 2019 IEEE 23rd International Conference on Intelligent Engineering Systems (INES),

Mantello, P., & Ho, M.-T. (2022). Why we need to be weary of emotional AI. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01576-y

Mantello, P., & Ho, M.-T. (2023). Emotional AI and the future of wellbeing in the post-pandemic workplace. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01639-8

Mantello, P., Ho, M.-T., Nguyen, M.-H., & Vuong, Q.-H. (2021). Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01290-1

Mantello, P., Ho, M.-T., Nguyen, M.-H., & Vuong, Q.-H. (2023). Machines that feel: behavioral determinants of attitude towards affect recognition technology—upgrading technology acceptance theory with the mindsponge model. *Humanities and Social Sciences Communications*, *10*(1), 430. https://doi.org/10.1057/s41599-023-01837-1

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28-37. https://doi.org/https://doi.org/10.1016/j.chb.2019.05.009

Mcmanus, A. (2016). *Emotions at play: The potential for emotion-enabled toys*. Retrieved October 6 from https://blog.affectiva.com/emotions-at-play-the-potential-for-emotion-enabled-toys

McStay, A. (2018). *Emotional AI: The rise of empathic media*. Sage.

McStay, A. (2020). Emotional AI and EdTech: serving the public good? *Learning, Media and Technology*, *45*(3), 270-283. https://doi.org/10.1080/17439884.2020.1686016

Moore, P., & Robinson, A. (2016). The quantified self: What counts in the neoliberal workplace. *New Media & Society*, *18*(11), 2774-2792.

Morrison, S. (2023). *Your AI personal assistant is almost here — assuming you actually want it*. Retrieved September 27 from https://www.vox.com/2023/9/23/23886163/google-microsoft-amazon-generative-ai-assistants

Picard, R. W. (1995). Affective computing. *MIT Media Laboratory Perceptual Computing Section Technical Report No. 321*, *2139*.

Piredda, G. (2020). What is an affective artifact? A further development in situated affectivity. *Phenomenology and the Cognitive Sciences*, *19*(3), 549-567. https://doi.org/10.1007/s11097-019-09628-3

Richardson, S. (2020). Affective computing in the modern workplace. *Business Information Review*, *37*(2), 78-85. https://doi.org/10.1177/0266382120930866

Schuller, D., & Schuller, B. W. (2018). The Age of Artificial Emotional Intelligence. *Computer*, *51*(9), 38-46. https://doi.org/10.1109/MC.2018.3620963

Steinert, S., Marin, L., & Roeser, S. (2022). Feeling and thinking on social media: emotions, affective scaffolding, and critical thinking. *Inquiry*, 1-28. https://doi.org/10.1080/0020174X.2022.2126148

Susanto, Y., Cambria, E., Ng, B. C., & Hussain, A. (2021). Ten Years of Sentic Computing. *Cognitive Computation*. https://doi.org/10.1007/s12559-021-09824-x

Theofanopoulou, N., Isbister, K., Edbrooke-Childs, J., & Slovák, P. (2019). A smart toy intervention to promote emotion regulation in middle childhood: Feasibility study. *JMIR mental health*, *6*(8), e14029.

Tordjman, K. L. (2021). *Siri, Alexa, Google ... what comes next?* https://www.ted.com/talks/karen_lellouche_tordjman_siri_alexa_google_what_comes_next?language=en

Urquhart, L., Miranda, D., & Podoletz, L. (2022). Policing the smart home: The Internet of Things as 'Invisible Witnesses'. *Information Polity*.

Viola, M. (2022). Seeing through the shades of situated affectivity. Sunglasses as a socio-affective artifact. *Philosophical Psychology*, 1-25. https://doi.org/10.1080/09515089.2022.2118574

Walter, S., & Stephan, A. (2022). Situated Affectivity and Mind Shaping: Lessons from Social Psychology. *Emotion Review*, *15*(1), 3-16. https://doi.org/10.1177/17540739221112419

Williamson, B. (2021). Psychodata: disassembling the psychological, economic, and statistical infrastructure of 'social-emotional learning'. *Journal of Education Policy*, *36*(1), 129-154. https://doi.org/10.1080/02680939.2019.1672895

Yang, Z. (2023). *Chinese AI chatbots want to be your emotional support*. Retrieved September 27 from https://www.technologyreview.com/2023/09/06/1079026/chinese-ai-chatbots-emotional-support/