Toward a social theory of Human-AI Co-creation: Bringing techno-social reproduction and situated cognition together with the following seven premises

Manh-Tung Ho[1,2]* & Quan-Hoang Vuong[2]

1. Institute of Philosophy, Vietnam Academy of Social Sciences, 59 Lang Ha, Ba Dinh, Hanoi, Vietnam

2. Centre for Interdisciplinary Social Research, Phenikaa University, Yen Nghia, Ha Dong, Hanoi, Vietnam.

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

< Draft Paper No. 20240503V1>

Abstract

This article synthesizes the current theoretical attempts to understand human-machine interactions and introduces *seven premises* to understand our emerging dynamics with increasingly competent, pervasive, and instantly accessible algorithms. The hope that these seven premises can build toward a social theory of human-AI cocreation. The focus on human-AI cocreation is intended to emphasize two factors. First, is the fact that our machine learning systems are socialized. Second, is the coevolving nature of human mind and AI systems as smart devices form an indispensable part of our cognitive scaffolds, thus, shaping how we perceive the world and ourselves. The seven premises include: primacy of social structures; human's desire for freedom and autonomy; AI systems will form *inseparable* parts of our cognitive/affective scaffolds and can change our self-understanding; philosophy and humanistic foundations on human flourishing as a guide to human-AI interaction; mindsponge information-filtering process; acculturative process of how values change and emerge from human-AI interaction; overlapping Venn diagram of the human-human, human-nature, and human-machine interactions. This article concludes with a discussion on human agency in our entanglements with socialized machines and the illusoriness of the Cartesian agent view of the mind.

Key words: artificial intelligence; human-machine interaction; human-AI cocreation; socialized machines; cognitive artifacts

Introduction

This article synthesizes the current theoretical attempts to understand human-machine interactions and introduces *seven premises* to understand our emerging dynamics with increasingly competent, pervasive, and instantly accessible algorithms. The seven premises build toward a social theory of human-AI co-creation, which emphasizes three factors: 1) the increasingly complex dynamics resulting from interactions between humans and increasingly

competent machines, 2) how these interactions are changing our social relationships and our self-understanding, 3) and these AI systems are not static and merely reproduced social meanings and boundaries, but humans and AI systems are *coevolving* as smart devices form indispensable parts of our cognitive scaffolds.

Importantly, the focus on human-AI co-creation is a constructive response toward two attempts at understanding the human-machine relationship, namely, the theory of *techno-social reproduction* of sociologist Massimo Airoldi (2021) and the *situated cognition* literature (Walter & Stephan, 2022). On the one hand, the theory of techno-social reproduction by Airoldi is also known as the theory of machine habitus, which is proposed in his book 2021, *Machine habitus: Toward a Sociology of Algorithms*. Here, Airoldi builds on the classic work of Bourdieu on the habitus to show machines and algorithms practically contribute to "the reproduction of society, with its arbitrary discourses, invisible boundaries, and structures" (p.112). This theory seeks to answer the questions 'How are algorithms socialized?' and 'How do socialized machines participate in society and, by doing so, reproduce it?' In other words, the focus of the theory of machine habitus is *socialized* machines and how these machines *reproduce* social meanings, propensities, boundaries, and structures.

Conversely, situated cognition designates a philosophy of technology literature that treats technological tools as parts of our cognitive scaffoldings that facilitate us in performing cognitive tasks and even forming our identities (Heersmink, 2013; Piredda, 2020; Walter & Stephan, 2022). This literature sheds light on the contextual/situational dimensions of human cognition and affects, thus, human agency. 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, competent 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 is purely happening in the brain. It is not brain-bound and is argued to be situated in the external world. This line of reasoning follows the famous extended mind hypothesis proposed by Clark and Chalmers (1998).

This article argues that these two pieces of literature complement each other, especially when viewing from the perspective that our increasingly smart and interactive devices, namely artificial intelligence chatbots, robots, and mobile apps, influence and change human minds and cognition in unexpected ways. The increasingly competent, pervasive, and instantly accessible algorithms We introduce seven premises that bring these two attempts at understanding human-AI interactions together and argue the focus should be on *human-AI co-creation*. The word co-creation is meant to emphasize the coevolving nature of the human mind and smart technologies due to AI technologies' entangled, networked effects on how we perceive the world, see ourselves, interact with each other, and interact with machines. Thus, a social theory of human-AI co-creation focuses on how we and our algorithms are co-creating emerging dynamics, values, ways of being in the world, and even injustice.

Socialized machines as cognitive artifacts: Bridging techno-social reproduction and situated cognition

Socialized machines

First, we will lay out some key points in Airoldi's account of techno-social production. Massimo Airoldi's *Machine Habitus: Toward a Sociology of Algorithms* (2021) draws from two parts of the sociological literature, the classic works of Bourdieu on the habitus and the recent works in Science and Technology Studies as well as Actor Network Theory, and analyzes how machines and algorithms, albeit unconscious, practically contribute to the "the reproduction of society, with its arbitrary discourses, invisible boundaries, and structures" (p.112). The author asks two main open questions that are answered throughout the book: How are algorithms socialized? How do socialized machines participate in society and, by doing so, reproduce? To answer these questions, Airoldi defines machine habitus "as the set of cultural dispositions and propensities encoded in a machine learning system through data-driven socialization processes" (p.113)," then embarks on articulating his account of techno-social reproduction.

Massimo Airoldi starts out by investigating fundamental literature about the distinction between humans and machines, the definition of algorithms, and the history of algorithms in three eras, i.e., the analogue era (-1945), the digital era (1946-1998), and the platform era (1989-). Chapters 2 and 3 explain the dynamics of culture in code and code in culture, respectively. On the one hand, machine bias, as one dimension of the cultures in the code, is socialized through machine creators: designers, developers, and computer programmers, labeled as *deus in machina*. On the other, the cultural dispositions encapsulated in machine learning systems are also created by machine trainers. Chapter 4, Theory of Machine Habitus, contains a primer on the *theory of machine habitus* and offers a treatment of the entanglements of the world of human society and the world of algorithms. In Chapter 5, the author concludes the book with a comprehensive research agenda consisting of four directions to further follow and examine: *the machine creators, users, medium, and algorithms*.

Airoldi utilizes the concept of a techno-social field, which has a parallel concept in a "social field" articulated in the earlier writings of Bourdieu. The social field has both *external/geographical* and *internal/internalized* dimensions. The external dimensions are social institutions, material inequalities, etc., outside the technologically inclined communities and in the real world. The internal/internalized dimension refers to how the social agents internalize these objective/external objects and social structures through socializing experiences. As such, the social structures are reproduced, thus being structured by the regularities in the lifestyles, life goals, and 'position takings' of individuals in the social space.

Building on these external and internal dimensions of social structures, the author provides an account of information concerning a socialized machine, which produces and reproduces objective meaning like a Bourdieusian social actor. Throughout the book, readers are introduced to many important thinkers who have shaped the development of critical thinking regarding the role of technology and society. For example, readers are introduced to how Sufiya Umoja Noble (2018) painstakingly shows the socio-cultural biases of straight white men get encoded into Google's search engine. Equipped with an understanding of the theory of machine

habitus, Massimo Airoldi hopes to show the readers that we must bring the socio-cultural, political, and historical context back to understanding our day-to-day interactions with algorithms and digital networks.

Here, Massimo Airoldi articulates human freedom and agency challenges when entangled with networks of millions of users and socialized machines. In this process, the author destroys the popular myth of marketing regarding the optimality of algorithms in knowing intimate desires and, thus, being able to provide the most 'personalized experiences' to the users. What goes on behind the feedback loop and automated Netflix or YouTube recommendations are never truly personalized, stressed Airoldi, but are made possible through digital infrastructure, social structure, and encoded culture propensities. The effect of this entanglement is conceptualized into four possible scenarios: *boundary differentiation*, *boundary fragmentation*, *boundary normalization*, and *boundary reconfiguration (see Figure 5, p.138).*

The absence of ontological interactions between an autonomous human and an autonomous machine signifies the limitation of human freedom and agency that the techno-socially reproduced society generates. The author reviews critical algorithm studies and tackles two main aspects: data mining through "digital forms of surveillance" (p.15), and the output of algorithmic computations controls and influences the users. Researchers within humanities and social sciences may instantly understand the fact that machines are socialized and that machine habitus and human habitus reciprocally interact and shape each other, establishing a structured social order. However, such a nuanced view is overlooked because of the mainstream media's zealousness that swings back and forth between the *apocalittici* (i.e., the doomsayers) and *integrati* (i.e., the techno-enthusiasts). The machines and algorithms are often sold as drivers of positive social transformations. Yet, it is likely that they'll cause a lot more displacement (if not disruption) (Wright, 2023) and chaos before their fruits are *justly distributed* to all classes of people.

Socialized machines as cognitive artifacts

The conceptual framework afforded by the theory of machine habitus complements the cognitive artifacts and situated cognition literature (Walters & Stephan, 2022), a philosophy of technology literature that treats technological tools as parts of our cognitive scaffoldings that facilitate us in performing cognitive tasks and even forming our identities.

David Krakauer, the president of the Santa Fe Institute, is known for distinguishing two kinds of cognitive artifacts (CA): *complimentary cognitive artifacts* and *competitive cognitive artifacts*. An example of a complimentary cognitive artifact is the abacus, as when one masters it, this tool will become an integral part of how his or her mind works and will augment his or her thinking process. In contrast, a calculator is an example of a competitive cognitive artifact. When a person starts using this kind of tool, his ability to calculate with his mind is diminished (Krakauer, 2016).

Krakauer took inspiration from Donald Norman (1991), the director of the Design Lab, California, San Diego, who was among the first to notice much of the scientific understanding during his time had been devoted to the unaided mind: the issues of memory, attention, action,

and thought. However, careful work in cognitive science on how artifacts can shape the mind has still 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. We do this by manipulating objects, tools (increasingly smart devices), and spaces (physical and cyber) (Fasoli, 2018; Piredda, 2020). 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 and are used to support and enhance our problem-solving skills. Essentially, augmenting human cognition and the ability to solve problems is the key feature of cognitive artifacts.

Next, having introduced the background of the theory of techno-social reproduction and situated cognitions, we will introduce seven premises that build toward a social theory of human-AI co-creation.

Seven premises of human-AI co-creation

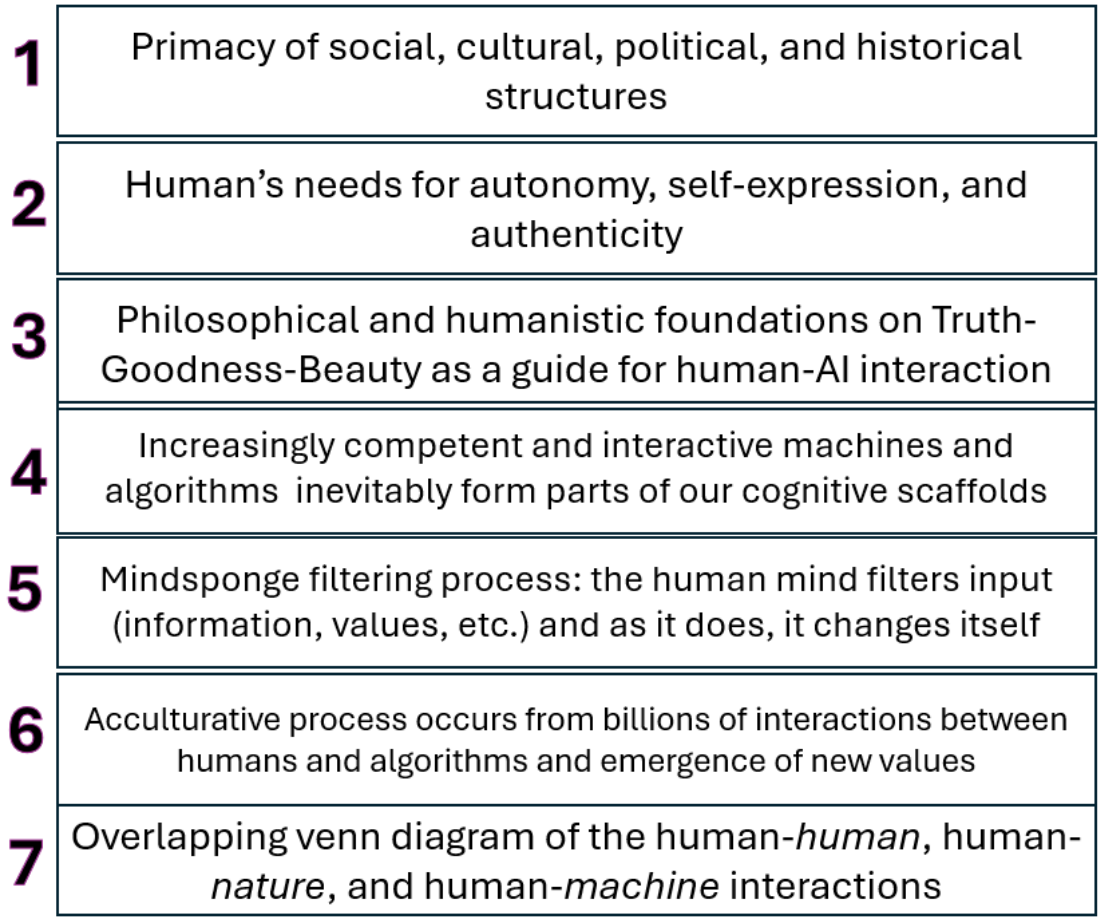

Figure 1: Seven premises of human-AI co-creation

> "Experienced and quite up-dated on new technologies, Kingfisher says (to the birds about the robot bogeyman): - You guys have met rivals of the 4.0 age. You would sure fail if you didn't study them carefully. You must do thorough research on their behaviors to find a solution."
>
> In *Bogeyman*, The Kingfisher Story Collection (Vuong, 2022a)

The primacy of social structures

The first premise is the primacy of the structures. Prior to any individual human and to any AI system, there have always been social, cultural, political, and historical structures that predispose individuals and machines to certain ways of thinking, behaving, and being. An individual is always in the process of being socialized, and he/she/they have to internalize certain socio-cultural norms and values to function as a member of society. Likewise, any algorithms that interact with humans are also trained with datasets from the preexisting social worlds. In this fundamental way, as sociologist Massimo Airoldi argues, machine learning systems such as search engines or recommender algorithms are socialized and constantly receive an influx of feedback from millions of users. The networked human-algorithms entanglements, i.e., the machine habitus, Airoldi argues, ultimately reproduce "society, with its arbitrary discourses, invisible boundaries, and structures" (p.112). Airoldi, thus, proposes that researchers systematically follow and investigate *the machine creators, users, medium, and algorithms* (Airoldi, 2021).

The premise on the primacy of social structures calls attention to another factor: we must understand how human values *propagate, reproduce, and, importantly, emerge* across and together with the machine habitus. This is also an important agenda (Ho & Nguyen, 2024; Mantello et al., 2023). This leads to point no. 2.

Human's desire for autonomy, self-expression, and authenticity

While each human desires differing levels of agency and autonomy, all do have some desire for freedom, whether in the form of self-expression, authenticity, or autonomy. Psychologist Paul Bloom explores the philosophy and psychology behind the strange appeal of perverse actions from St. Augustine's The Confessions to contemporary thinkers such as Agnes Callard. Bloom used the term "existential perversity" to refer to "perverse behavior that is deliberate, knowing, and strangely satisfying." He reasons existential perversity comes from the deep desire to show to yourself and others that you are a free, authentic, and autonomous being. This desire, thus, motivates a person to avert from the expected, the reasonable, or even the moral (Bloom, 2019).

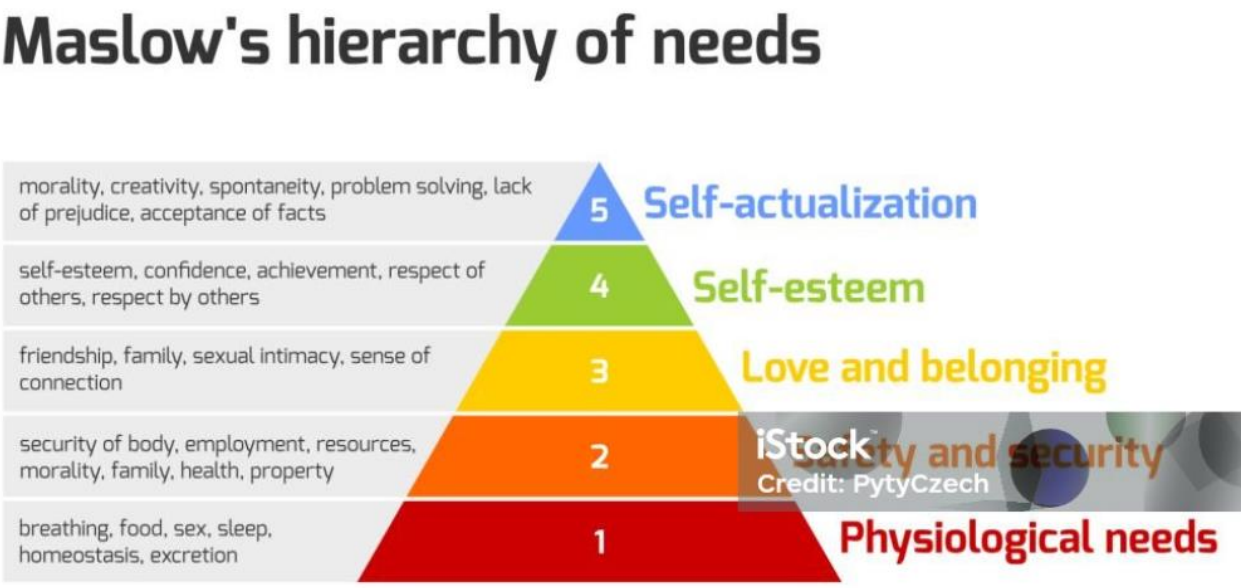

Figure 2: Maslow's hierarchy of needs. Source: Free license image from iStock

The need to feel like a free, autonomous subject also corresponds somewhat to the need for self-actualization in Abraham Maslow's famous *hierarchy of needs*, a model that appears in his scientific article *A Theory of Human Motivation*. The need for self-actualization is placed at the top of the pyramid (Maslow, 1943).

As humans always seek to assert themselves as autonomous beings, even if we have the most moral and superhuman intelligent machines, it is hard to see the future where humans merely obediently abide by what the machines instruct. This leads to point no.3.

Philosophical and humanistic foundations on Truth-Goodness-Beauty as a guide for human-AI interaction

Whether human-computer interactions serve the foundational values of Truth-Goodness-Beauty depends on the 'awakening' of humans. "What does that mean to be enlightened or awakened as a human being?" has been the question that great religious leaders and philosophers have investigated since the earliest days of human civilizations. Humans have developed a huge, rich repertoire of philosophy, humanities, and, recently, the empirical science of human flourishing. In an age where machines are becoming increasingly competent, smart devices will unlock a huge amount of cognitive surplus in many individuals and organizations. Thus, the interaction between humans and computers will bring endless possibilities and open new doors of vast opportunities and horizons.

Nevertheless, to choose wisely among these doors, to be able to open and close them at the right moment, people, both individually and collectively, need to go back, study and apply knowledge and wisdom that sometimes lies deeply within traditions, traditional practices, and cultural treasures. Here, cultural products such as poetry, ancient literature, the classics, fables

(Vuong, 2022a), folk tales (Vuong et al., 2018), and satirical cartoons (Ho et al., 2021), etc. are highly educational tools and can be a guide for human choices in the era of interaction with increasingly intelligent machines. We can contemplate this premise through the example of the virality of what many scholars call street philosophy, such as various interpretations of Stoic teachings in social media platforms (Kelsey, 2022; Rosenberg, 2020).

Increasingly competent and interactive machines and algorithms inevitably form parts of our cognitive scaffolds

We must admit that whether in the forms of recommender algorithms, dating apps, news feeds, virtual assistants, chatbots, or measurement of physical health indicators in self-tracking devices such as Fitbit, etc., increasingly competent machines form an inevitable part of our cognitive scaffolds. Smart algorithms augment our cognition (Raisamo et al., 2019) and, increasingly, our emotional lives (Ho & Ho, 2024). In other words, smart devices are now parts of the web of objects that help us perform daily cognitive and affective tasks: remembering, perceiving, thinking, regulating emotion, etc. This view is even clearer when we start to see smart devices as cognitive and affective artifacts (Fasoli, 2018; Heersmink, 2018).

For example, in *The Mind Technology Problem and the Deep History of Mind Design*, the editors Clowes et al. (2021) contrast the early view of the extended mind hypothesis that we are naturally born cyborgs with the later notion of Inforgs, i.e., "informational embodied organisms," proposed by Floridi Floridi (2015). Floridi argues our cognitive profiles remain more or less the same, yet we increasingly conceive the mind in technological and computational terms. Citing the seminal works in AI and the extended mind, Clowes et al. illustrate how our thinking about AI has been driven by the Cartesian agent view, whose underlying assumption is that a system is fully in charge of its cognitive environment and its agency is separate from artifacts which it depends on. However, the contributors to the volume point out the recent rise of ubiquitous computing and smart devices such as smartphones, social media platforms, wearables, etc., which have increasingly blurred the distinction between artificial and human intelligence, between real life and the virtual world. The authors point out several examples of how new technologies make us rethink concepts that define us humans. For example, the ubiquity of smartphones and readily accessible information have changed how we form memory, thus requiring a reconceptualization of human memory, the epistemic environment, and even personal identity. Importantly, this is how the mind-technology problem is implicated in our folk psychology, i.e., the common sense of what we think a mind is and what it does, argued Clowes et al. (2021).

Mindsponge filtering process: the human mind filters input (information, values, etc.), and as it does, it changes itself

Another important premise to consider about human-AI co-creation is how we process information. Here, analogizing the mind to a sponge (Vuong, 2022b) will be helpful intuitively. First, a sponge absorbs and depending on its material, it can absorb a lot or a little. Second, there is a point of saturation when it cannot further absorb; only by squeezing out, can it continue to absorb again. Third, a kind of filtering process happens: a sponge changes its color or becomes

hardened depending on the material of the input, e.g., pure water will have different impacts on the sponge than other liquids. These interactions of a sponge with liquids help us think about the relationship between the human mind and the infosphere, which is increasingly populated and even polluted by AI-generated and AI-recommended content (Ho & Nguyen, 2023).

Let's draw several obvious parallels with the above analogy. The human mind does filter information from the infosphere. Too much information will overwhelm the mind, and it cannot process information further. Depending on the core values of an individual mind, some information will resonate more with it than others. Putting a similar fact in front of two minds, the two minds can process a fact differently. This issue can be further exacerbated when we consider how humans process stories. Some information can become a trigger for mind change.

The changing of core values and beliefs is metabolically taxing. Thus, the human mind naturally wants to construct a predictable infosphere that aligns with its core values and beliefs. This issue is illuminated in several recent renowned works on the neural correlates of beliefs and the predictive brain (Barrett, 2017; Kaplan et al., 2016; Seth, 2021; Seth, 2013). Thus, *epistemic justice requires epistemic courage* in the form of being ready to move away from one's own comfort zone. This means we can put structures and interact with algorithms, i.e., a new, interactive, increasingly competent category of cognitive artifacts, in ways that keep us epistemically humble and grounded.

The acculturative process occurs from billions of interactions between humans and algorithms and the emergence of new values

The sixth premise is about what we might call acculturative emergence phenomena. It designates the phenomenon that values arising from billions of interactions between humans and algorithms on digital platforms, i.e., the meganets (Auerbach, 2023), once peripheral or forgotten, can become mainstream or even become the core values to society, or at least a sub-culture of society. As David Auerbach observed, while big-tech owners such as Mark Zuckerberg or Elon Musk and even each user might personally value democracy, free speech, or inclusiveness, the interactions in the hyperweb and their resulting emergent social dynamics might undermine such values in unexpected ways as (Auerbach, 2023).

An example of mainstream value change and value conflict is in American universities before and after the arrival of social networks. In the era before social networks, university lecture halls in the US were considered almost sacred cathedrals of freedom of speech; any idea can be discussed and feeling uncomfortable might not be the most important factor. However, values such as safety-ism and trigger warnings have recently become dominant values in interactions between people, individuals and organizations in organizations in the US (Haidt & Lukianoff, 2018). According to many Western psychologists and sociologists, notably Jonathan Haidt, the rise of safety*ism*, coupled with unregulated social networks, has contributed to creating a generation of students who tend to be psychologically more fragile and unstable, with higher rates of self-harm, depressive symptoms, and feelings of loneliness than previous generations (Haidt, 2024).

Observing the value changes, emergence, and conflicts in the acculturation process in everyday interactions between humans and machines makes us worry about the health of the spiritual development, physical fitness, and habits of people of different ages and social positions.

Overlapping Venn diagram of the human-human, human-nature, and human-machine interactions

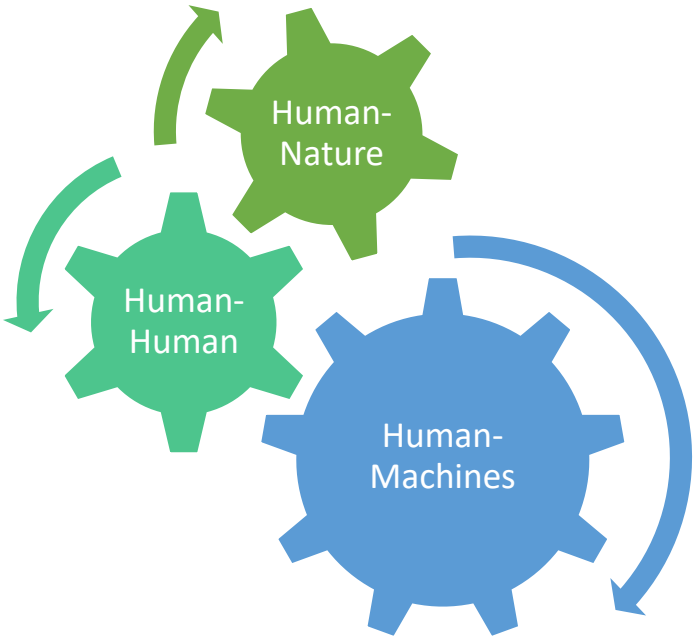

Figure 3: The interactions of humans-to-humans, humans-machines, and humans-nature are cross-cutting and influencing each other.

Finally, human-computer interactions are already and will change human-to-human and human-nature interactions. This triangle of interactions is intertwined and can influence each other. Recent evidence points to a worrying trend that increasingly technological mediation in human-to-human interaction has worsened our mental health, and its harmful effects have started to be evident in the brain. For example, a new study of more than 27,000 participants from Sapien Lab shows that smartphone ownership at a younger age increases perceptions of self-worth, lower self-esteem, motivation, and resilience—and more sadness, anxiety, and aggression—especially for girls (Jerzy et al., 2024). Recent studies also show that many Gen Z members admit to having difficulty forming deep social relationships, thus lacking trustworthy friends during difficult times (Dombrosky et al., 2018; Pichler et al., 2021). Research in Japan by Dr. Ryuta Kawashima of Brain Science (Tohoku University) on more than 70,000 children from elementary to high school age showed that long-term use of smartphones negatively affects their grades and brain development (Japan Today, 2019). Brain scan results of 214 Sendai teenagers between the ages of 5 and 18 who had brain MRI scans over three years showed that young people who regularly used the internet via smartphones showed impaired physical development. Lobar area in all major areas of their brain. This group also showed a similar deficit in the amount of white matter

that transmits data through nerve cells compared to the group that rarely used smartphones (Japan Today, 2019).

Meanwhile, there has been clear evidence of the health benefits of contact with nature. For example, a systematic (PRISMA-P) review of 51 articles shows 90% of articles ($n = 46$) demonstrated at least one positive association between nature-based recreation and mental health, which comprises improvements in affect, cognition, restoration, and well-being, and decreases in anxiety and depression symptoms (Lackey et al., 2021). It is also ironic that in the present ascendance of ever-powerful AI systems, the world is lackluster in combating climate change, the phenomenon known as climate apathy (Cripps, 2023). It is only right that we must find a way to channel the power of our most powerful AI agents toward the fight for our ecosystem, i.e., to escape climate apathy (Vuong & Ho, 2024).

We believe finding the balance among these three overlapping spheres of interactions would allow for human flourishing and undermining the negative influences of over-interacting with and over-relying on machines and algorithms. Thus, to comprehensively assess the impacts of human-computer interactions in the age of AI at both personal and collective levels, researchers need to account for the impacts it has on human-human and human-nature interactions.

Discussion: Human agency in the entanglements with socialized machines

A descriptive account of human agency in the entanglements with socialized machines

Putting together all seven premises, we can start mapping human flourishing in a time of ever-pervasive human-AI interactions. As a human strives to be an autonomous and authentic self, he or she does so in a world that has norms and values precede his/her existence. In this process, with the growth of AI technologies, this person will interact with networked, socialized machines and algorithms that could lead to unexpected emerging norms and values. These machines and algorithms inevitably form a part of his/her cognitive and affective scaffoldings (Piredda, 2020), which will influence how he/she will perceive the world and himself/herself.

Frictions and conflicts within this process and how people resolve them largely determine their well-being. Traditional cultures have different values constraining normative human-to-human and human-to-nature interactions, so respecting these values might lead to more normative human-computer interactions. Conversely, constructing new cultures that center around values and norms conducive to human well-being is also an important conclusion of these premises. The dynamic, complex, and dialectic process of understanding human-AI cocreation entails an interdisciplinary exploration of how the human mind changes as it interacts with increasingly competent AI systems.

Letting go of the Cartesian agent view: The illusoriness of our mind as a completely independent system

Equipped with the conceptual tools of machine habitus and cognitive artifacts, the binary, simplistic, reductionistic, and agentic view of algorithms and machines might start losing its grip on our collective imagination. Machines and algorithms, in and of themselves, neither diminish nor liberate human existence. The fact that we, collectively and individually, construct these

machines as parts of our cognitive scaffolding and the fact that they are encoded with our social and cultural propensities are evidence against certain strands of techno-determinism. It also highlighted the illusoriness of our mind as a completely independent system. As pointed out by philosophers Clowes et al. (2021) in a recent book titled 'The Mind-Technology Problem : Investigating Minds, Selves and 21st Century Artefacts,' our thinking about AI has been driven by the Cartesian agent view. The Cartesian agent view's underlying assumption is that a system is fully in charge of its cognitive environment and its agency is separate from artefacts which it depends on. Increasingly, the rise of ubiquitous computing devices embedded in our smartphones, wearables, etc. have demonstrably blurred the distinction between artificial and human intelligence, between real life and the virtual world. For example, as pointed out in recent works in cognitive science, the rise of smartphones which has given us readily accessible information have been found to change how we form memories. Thus philosophically, such new human-machine relationship demands a reconceptualization of human memory, our epistemic environment, and even personal identity (Clowes et al., 2021).

In our quest to resolve the tension between privacy and autonomy implicated in the age of *socialized algorithms*, perhaps, it is wise to abandon the commonsensical, intuitive dualist notion that we are a system is fully in charge of its cognitive environment and our agency is separate from artefacts which it depends on (i.e., the Cartesian agent view in the literature on the extended mind hypothesis (Clowes et al., 2021) or the notion of natural-born dualist popularized by Yale psychologist, Paul Bloom (Bloom, 2007, 2013)). It is wiser, perhaps, to start internalizing the Homo Faber view, which states the evolution of the mind depends both on the history and the pre-history of our artefacts. In other words, our mind has been shaped both by a natural evolutionary process in the pre-history of our artefacts as well as by the creation of technologies. A clear example is in Ultrasociety, where Peter Turchin (2016) applies multi-level cultural evolution dynamics and demonstrates our deep-seated preference for equality has been shaped by the creation of projectile technology (a spear, stone-throwing, bows, and arrows, etc.). Turchin argues hunter-gather society is fiercely egalitarian, not because of an innate tendency toward egalitarianism as suggested by the metaphor of a noble savage by Jean-Jacques Rousseau (Wynn et al., 2017), but because the ease of using a projectile technology to punish anyone who wants to upset the hierarchy. The technological invention of projectile technology also puts immense evolutionary pressure on humans to evolve the capacity to communicate complex thoughts and form social alliances and to detect the intentions of other human beings.

Indeed, once we relax the hold of the Cartesian agent view, and see ourselves more as interdependent, contingent beings, thus, we might appreciate the important of external intervention whether from families, closed friends, local communities, social structures, even at the state level can help us cocreate with AI systems a structure more conducive to human well-being.

Conclusion

Following the seven premises, we hope to show our choices in designing, interacting and interfacing with the cognitive artifacts (Fasoli, 2018) in our smart devices, recommender algorithms, virtual assistants, etc., matter. In the worst case, blind reliance on these technologies

could reduce people into data points of their preferences and make them more predictable and entrenched in their narrow views and prejudices. In other cases, underuse of AI systems might undermine human potential. At best, algorithms can enrich our lives by lifting individual out of their predisposed views dictated by their social classes and upbringings. Finally, we would like to conclude the paper on a positive note from Massimo Airoldi: "When used differently, artificial intelligence systems and learning machines can become amazing tools for solving socially relevant problems, cultivating social relations, preserving local cultures, reducing social inequalities and improving the life of communities" (p.155). This is the *human-centric* future we need to *proactively* strive toward if we want to avoid the bleak, dystopian future predicted by Orwell and Huxley (Ho & Mantello, 2023).

Data availability

There is no data associated with this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Airoldi, M. (2021). *Machine habitus: Toward a sociology of algorithms*. John Wiley & Sons.

Auerbach, D. (2023). *Meganets: How Digital Forces Beyond Our Control Commandeer Our Daily Lives and Inner Realities.* Public Affairs.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1-23. https://doi.org/10.1093/scan/nsw154

Bloom, P. (2019). *The Strange Appeal of Perverse Actions*. Retrieved December 26, 2023 from https://www.newyorker.com/culture/annals-of-inquiry/perverse-incentives

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7-19.

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

Cripps, E. (2023). *How to escape climate apathy*. Retrieved October 27 from https://psyche.co/guides/how-to-escape-climate-apathy-and-help-to-avert-catastrophe

Dombrosky, K., Templeton, B., & Fong, J. (2018). *An insider's guide to Generation Z and higher education 2018*. University Professional and Continuing Education Association. Retrieved March 03 from https://upcea.edu/wp-content/uploads/2018/01/An-Insiders-Guide-to-Generation-Z-and-Higher-Education-eBook.pdf

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

Floridi, L. (2015). *The onlife manifesto: Being human in a hyperconnected era*. Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-319-04093-6

Haidt, J. (2024). *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental IllnessJonathan Haidt on the 'National Crisis' of Gen Z*. Penguin Press.

Haidt, J., & Lukianoff, G. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. Penguin UK.

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., & Ho, M.-T. (2024). Bridging divides: Empathy-augmenting technologies and cultural soul-searching. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-01887-2

Ho, M.-T., & Mantello, P. (2023). Smart technologies and how they create the reality feared by Orwell and Huxley. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01652-x

Ho, M.-T., & Nguyen, T. H.-K. (2024). Bài toán về hệ thống hóa quá trình sàng lọc thông tin trước sự trỗi dậy của thuật toán thông minh *Tạp chí Thông Tin & Truyền Thông*(March/2024).

Ho, M.-T., Progler, J., & Vuong, Q.-H. (2021). An Anatomy of Satirical Cartoons in Contemporary Vietnam: Political Communication and Representations of Systemic Corruption in a One-party State. *Asian Studies Review*, *45*(4), 711-728. https://doi.org/10.1080/10357823.2021.1932743

Ho, M. T., & Nguyen, H.-K. T. (2023). From the eco-calypse to the infocalypse: The importance of building a new culture for protecting the infosphere. *AI & SOCIETY*. https://doi.org/https://doi.org/10.1007/s00146-023-01737-7

Japan Today. (2019). *Brain expert warns parents of horrors that await juvenile smartphone users*. Retrieved March 27 from https://japantoday.com/category/features/kuchikomi/brain-expert-warns-parents-of-horrors-that-await-juvenile-smartphone-users

Jerzy, B., Jennifer Jane, N., & Tara, C. T. (2024). Hierarchy of demographic and social determinants of mental health: analysis of cross-sectional survey data from the Global Mind Project. *BMJ Open*, *14*(3), e075095. https://doi.org/10.1136/bmjopen-2023-075095

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific Reports*, *6*(1), 39589. https://doi.org/10.1038/srep39589

Kelsey, D. (2022). Street Philosophy and the Stoics. In D. Kelsey (Ed.), *Storytelling and Collective Psychology: Ancient Wisdom, Modern Life and the Work of Derren Brown* (pp. 41-71). Springer International Publishing. https://doi.org/10.1007/978-3-030-93660-0_3

Lackey, N. Q., Tysor, D. A., McNay, G. D., Joyner, L., Baker, K. H., & Hodge, C. (2021). Mental health benefits of nature-based recreation: a systematic review. *Annals of Leisure Research*, *24*(3), 379-393. https://doi.org/10.1080/11745398.2019.1655459

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

Maslow, A. (1943). A theory of human motivation. *Psychological review*, *50*(4), 370–396. . https://doi.org/https://doi.org/10.1037/h0054346

Norman, D. A. (1991). Cognitive artifacts. *Designing interaction: Psychology at the human-computer interface*, *1*(1), 17-38.

Pichler, S., Kohli, C., & Granitz, N. (2021). DITTO for Gen Z: A framework for leveraging the uniqueness of the new generation. *Business Horizons*, *64*(5), 599-610. https://doi.org/https://doi.org/10.1016/j.bushor.2021.02.021

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

Raisamo, R., Rakkolainen, I., Majaranta, P., Salminen, K., Rantala, J., & Farooq, A. (2019). Human augmentation: Past, present and future. *International Journal of Human-Computer Studies*, *131*, 131-143. https://doi.org/https://doi.org/10.1016/j.ijhcs.2019.05.008

Rosenberg, J. (2020, 2020 January-February). THE GRUELING CLASS: Why Silicon Valley loves Stoicism. *Mother Jones*, *45*(1), 55+. https://link.gale.com/apps/doc/A616314490/AONE?u=anon~87943545&sid=googleScholar&xid=6496bbbe

Seth, A. (2021). *Being you: A new science of consciousness*. Penguin.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565-573. https://doi.org/https://doi.org/10.1016/j.tics.2013.09.007

Vuong, Q.-H., Bui, Q.-K., La, V.-P., Vuong, T.-T., Nguyen, V.-H. T., Ho, M.-T., Nguyen, H.-K. T., & Ho, M.-T. (2018). Cultural additivity: behavioural insights from the interaction of Confucianism, Buddhism and Taoism in folktales. *Palgrave Communications*, *4*(1), 143. https://doi.org/10.1057/s41599-018-0189-2

Vuong, Q.-H., & Ho, M.-T. (2024). Escape climate apathy by harnessing the power of generative AI. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01830-x

Vuong, Q. H. (2022a). *The Kingfisher Story Collection.* . https://www.amazon.com/dp/B0BG2NNHY6.

Vuong, Q. H. (2022b). *Mindsponge theory*. Walter de Gruyter GmbH.

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