https://doi.org/10.48417/technolang.2024.04.10
Research article

Ethical Reflections on Persuasive Technology

Sumei Cheng (✉)
Institute of Philosophy, Shanghai Academy of Social Sciences, Shunchang Road, No. 622, Shanghai, 200025, China
csm@sass.org.cn

Abstract

Persuasive technology arose as a new kind of interdisciplinary field of arts and science. Techniques and technologies of persuasion traditionally involved oral or written language, be it for the presentation of arguments or for rhetorical strategies and seductive slogans. In contrast, the term now refers to a human-computer interaction technology that has the ability to influence or even to change people's perceptions, attitudes, or behaviors. There is a wide range of applications with significant social impact. However, the novelty, concealment, polymorphism, and other characteristics of AI products with persuasive functions will conceal their intentions, limit the free choice of their users, put their users in a disadvantaged position, and might even prove to be addictive. In order to avoid or mitigate these problems, it is necessary to conduct an ethical examination of the development and application of persuasive technology. At the same time, the indeterminacy or uncertainty of data-driven algorithmic systems and the multiple moral agents associated with computing products have made traditional assessments difficult. We can't cope with these challenges, until we have gone beyond the dichotomy between theoretical and applied ethics, expanding the semantic and pragmatic scope of the concept of responsibility. Regarding the ethics of technology we need to effect a shift from an emphasis on the responsibility for passively conceived users to their actively taking responsibility, and establish a new conceptual framework of ethics for the human future.

Keywords: Human-computer interaction; Data-driven algorithmic systems; Accountability; Ethics about the future of humanity

Citation: Cheng, S. (2024). Ethical Reflections on Persuasive Technology. *Technology and Language, 5*(4), 143-157. https://doi.org/10.48417/technolang.2024.04.10

УДК 17:62
https://doi.org/10.48417/technolang.2024.04.10
Научная статья

Этические размышления о технологии убеждения

Сумэй Чэн (✉)
Институт философии Шанхайской академии социальных наук, Шуньчан-роуд, 622, Шанхай, 200025, Китай
csm@sass.org.cn

Аннотация
Технология убеждения возникла как новый вид междисциплинарной области искусства и науки. Методы и технологии убеждения традиционно включали устную или письменную речь, будь то для представления аргументов или для риторических стратегий и соблазнительных лозунгов. Напротив, в настоящее время этот термин относится к технологии взаимодействия человека и компьютера, которая способна влиять или даже изменять восприятие, установки или поведение людей. Существует широкий спектр приложений, оказывающих значительное социальное воздействие. Однако новизна, скрытность, полиморфизм и другие характеристики продуктов искусственного интеллекта с функциями убеждения будут скрывать их намерения, ограничивать свободу выбора пользователей, ставить их в невыгодное положение и даже могут вызывать привыкание. Чтобы избежать или смягчить эти проблемы, необходимо провести этическую экспертизу разработки и применения технологий убеждения. В то же время неопределенность алгоритмических систем, управляемых данными, и множество моральных агентов, связанных с компьютерными системами, затрудняют традиционную оценку. Мы не сможем справиться с этими вызовами, пока не выйдем за рамки дихотомии между теоретической и прикладной этикой, расширив семантический и прагматический охват концепции ответственности. Что касается этики технологий, то нам необходимо сместить акцент с ответственности за пассивно мыслящих пользователей на активное принятие ответственности и создать новую концептуальную основу этики для будущего человечества.

Ключевые слова: Взаимодействие человека и компьютера; Алгоритмические системы, управляемые данными; Ответственность; Этика будущего человечества

Для цитирования: Cheng, S. Ethical Reflections on Persuasive Technology // Technology and Language. 2024. № 5(4). P. 143-157. https://doi.org/10.48417/technolang.2024.04.10

INTRODUCTION

Many computing products or AI systems are now developed on the basis of a deep integration of computer technology and persuasion theory. They are increasingly fostering new possibilities for our life and new breakthroughs in societies driven by science and technology. These computing products, due to their persuasive capabilities, have been widely applied to many fields such as education, sports, gaming, advertising, finance, social networking, social governance, healthcare, e-commerce, environmental protection, disease management, platform management, and personal self-management – as such achieving significant social impact. The problem is that persuasive activities used to consist in one person or group persuading and inducing another person or group to change their attitude or behaviors, whereas these applications now enable such persuasive activities to be carried out through human-computer interaction systems. In view of this, one urgently needs to strengthen ethical governance and conduct an ethical examination of the development, implementation, and use of persuasive technology, so that developers and designers are guided to choose what is good in their design. This paper attempts to expound briefly the significance of persuasive technology, to reveal the ethical challenges and ethical problems caused by application of persuasive technology, and to explore the corresponding ethical governance principles, so as to deepen our ethical understanding of human-computer interaction systems.

THE SIGNIFICANCE OF PERSUASIVE TECHNOLOGY AND THE NECESSITY OF ITS ETHICAL EXAMINATION

The concept “persuasive technology” was proposed in the 1990s by B. J. Fogg who was a social scientist and the founder of the Behavior Design Lab at Stanford University. It refers to a human-computer interaction technology that can influence or even change people’s perceptions, attitudes, values, or behaviors. In other words, it refers to human-computer interaction technology that can influence and guide users' perceptions, attitudes, or behaviors, orienting them towards specific goals desired by designers, businesses or institutions. “Human-computer interaction” here refers to the interaction between people and computer systems or AI systems, rather than to the manual operation of conventional machines by people. In his book *Persuasive Technology: Using Computers to Change What We Think and Do*, published in 2003, Fogg further codified the term “Captology” to describe a new interdisciplinary field in which traditional persuasive design and computer-based techniques overlap where the word “Captology” is an abbreviation for “Computer as Persuasive Technology.” The goal of this book is to study how computer-based technology can be made more persuasive and better at changing users’ attitudes or behaviors (Fig. 1).

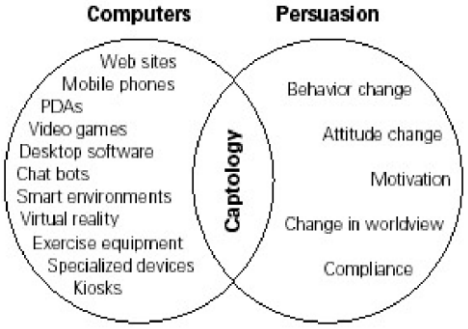

Fig. 1. Captology describes the area where computing technology and persuasion overlap (Fogg, 2003, p. 11)

Fogg entered Stanford University in 1993 as PhD student. He used the way of experimental psychology to prove that computers could change people's perceptions and behaviors in predictable ways during his doctoral studies. The title of his doctoral thesis was *Charismatic Computers: Creating More Likable and Persuasive Interactive Technologies by Leveraging principles from Social Psychology* (Fogg, 1998). Fogg founded the Persuasive Technology Lab (later renamed the Behavioral Design Lab) at Stanford University after obtaining his Ph.D. degree in 1997. He prospected in detail the research findings regarding human behavior and persuasion, and how they can be combined with computers to create a new field of persuasive technology (Fogg, 1998). His book, *Persuasive Technology: Using Computers to Change What We Think and Do*, shares experiences and provides a theoretical summary of the first decade of his laboratory research. In 2005, he received a grant from the US National Science Foundation for a project "Experimental Work Investigating How Mobile Phones can Motivate and Persuade People."[1]

Researchers from the United States, the Netherlands, Denmark, Finland, Italy, Austria, Canada. and other countries held an annual international conference about the development of persuasive technologies since 2016. The 18th conference took place in 2023. Obviously, these conferences not only extended the influence of Fogg's work, they also deepened the scope of invention and application of persuasive technologies. After attending the second conference held by Stanford University in 2007, Aaron Marcus was inspired to apply the concept design of persuasive design theory with information design/visualization theory to the software and hardware development of mobile devices. Two years later, he started a five-year (2009-2014) project to develop mobile devices in his company. Each sub-project was set a clear goal to change users' attitudes or behaviors for specific problems and application scenarios. Designers used the design techniques of "user-centered" and persuasive design. The specific design process and operational details of this project, as well as the experience of designers constitute the main content of the book *Mobile Persuasion Design: Changing Behaviour by Combining Persuasion Design with Information Design* (Marcus, 2015).

[1] https://peoplepill.com/people/b-j-fogg

Persuasive theory, which can be traced back to Aristotle's rhetoric, describes how one person influences or even changes another person's mind. Captology is the study of how these technology developers, designers, and researchers embed the findings of rhetoric, psychology, cognitive science, behavioral science, and social dynamics, such as persuasion, information extraction, behavior change, user experience, and the way of behavioral incentive, into human-computer interaction systems. This approach enables computing products to influence subtly and guide their users' behaviors to change toward specific goals favored by designers or businesses during the process of serving their users more proactively and enhancing the convenience of interaction. In fact, the various social media recommendation systems we are using currently, such as website recommendation systems, e-commerce systems, smartwatches, smartphones and other computing products, have persuasive functions.

In terms of the significance of persuasion, persuasion means that one person accepts voluntarily the opinions of another person. This voluntary acceptance stems from one's internal motivation. Persuasion is different from coercion, deception, brainwashing, etc. Coercion means exerting external pressure on a person in order to force them to change their attitudes or behaviors. Coercion may fall within the realm of education or it may be a crime. For example, the coercion of parents who force their children to change bad habits belongs to education, while criminals who force children to steal act immorally and even illegally. Also, all forms of deception are immoral or illegal. By brainwashing one indoctrinates and imposes one's values on. Rather than persuade someone to understand the truth from a standpoint of justice, brainwashing makes people change their believes of values for cultural or political purposes. Fogg defines persuasion as "an attempt to change attitudes or behavior or both (without any coercion or deception)" (Fogg, 2003, p. 16). According to Fogg, this means that the designers of persuasive technology have good intentions for the sake of their users and then embed persuasive intentions into human-computer interaction computing products in order to induce users to change their attitudes or behaviors. This kind of persuasion is internal to the products.

There are many different design ideas and approaches to developing computing products or AI systems that incorporate persuasion technology. Such as, first of all, simplification or making products simpler, that is, breaking down or simplifying complex tasks or activities, improving the benefit-to-cost ratio of users' behaviors, so that the usage of products or features becomes easier and more convenient. For example, the step counting and sorting functions of the social media platform WeChat is supposed to better motivate its users to exercise. Secondly, catering to users' preferences, that is, making computing systems to automatically predict and meet users' needs. These include, for example, algorithmic systems which automatically provide relevant information based on users' interest, consumption habits, and even geographical location. Thirdly, persuasive technologies can simulate experiences, that is, modify and improv existing design plans through vivid and visible simulation effects, such as simulation experiments for urban planning. Fourthly, there is interactive experience, that is, making users enjoy the best experience by improving their environmental perception at least in interactive computing environments. Here, for example, the computer system may always play games at a level comparable to the player so as to attract players to continue playing in e-sports games.

Finally, there is the principle of similarity, that is, increasing the user's sense of identification with a product by making the function of interactive products adapt to the user's personality This is to "persuade" users to more willing keep buying these products which can also be utilized, however, for online education tools that are designed to meet the psychological characteristics of different groups of people (Fogg, 2003, see Chapter 4 and Chapter 5).

In his book *Tiny Habits: The Small Changes That Change Everything*, published in 2019, Fogg argued the viewpoint that "behavior designs can change everything." He offered a large number of concrete examples, and elaborated on the "Fogg behavior model" for people to develop permanent habits. This model is represented by the formula B=MAP, where B represents Behavior, M represents Motivation, A represents Ability, and P represents Prompt (Fogg, 2019). This formula indicates that changes in human behavior depend on the convergence of motivation, ability, and prompt. In 2011, for example, the World Economic Forum Alliance for Occupational Health chose the "Fogg Behavior Model" as the framework for health behavior change. With the development in recent years of embedded algorithms or data-driven machine learning algorithms, the persuasive functions of persuasive technologies have become more diverse and embedded in our daily lives in a more hidden way.

Although Fogg emphasized that the initial intention of developing persuasive technology is to make human life healthier, more environmentally friendly, convenient, and enjoyable, etc., and although its overall goal is to improve human experience in every aspect and to meet people's needs, these intentions and goals imply, epistemologically speaking, a paternalistic way of thinking. They are based on the belief that users generally lack the ability to make correct choices and handle affairs, and need to be guided, reminded, or even controlled. Methodologically, persuasive technology presupposes a worldview of techno-solutionism, which gives priority or assigns special importance to the use of technology when it comes to solving human problems. At he same time, persuasive technologies are not always objective, transparent, impartial, just and so on. There have been many worrying social consequences such as racial hatred, gender discrimination, recruitment of terrorist organizations, cybercrime, privacy violations, and increased social inequality in practical applications (Noble, 2018).

These circumstances have required us to be vigilant and evaluate critically the development and applications of computing products or AI systems which include persuasive technologies. Therefore, we need to monitor and exam in the values optimized or disseminated by automated decision-making systems, so as to guide developers or vendors to lay a solid ethical foundation for the development and application of persuasive technology, and especially to prevent misuse and abuse.

Although there is a lot of discussion about the ethics of technology and governments have introduced corresponding governance principles, there are still many gaps between theory and practice. On the one hand, the stakeholders or moral agents are not trained in ethics and even do not have any systematic ethical knowledge. On the other hand, extant ethical training and ethical examination are mostly conducted in fields related to medicine, and largely neglected in technical fields based on AI. Most designers and engineers still believe that the question of value is a topic of discussion for

philosophers, sociologists, political scientists, or policy makers, and that the goodness or badness of technology lies with the users, not the inventor. According to the standard example, the invention of knives and guns does not involve value judgments, only the ways of using them does involve such judgments. Accordingly. many ethicists consider ethics of technology as applied ethics.

However, persuasive technology challenges not only this view of the neutrality of technology and the dichotomy between theoretical and applied ethics. It also challenges the corresponding conceptual framework, because the persuasion of persuasive technology products is active and context-sensitive or intrinsic: The good or evil of persuasive intention is related not only to the motivation of the designers and their methods of persuasion, but also to the technical limitations of the algorithmic systems and their path-dependencies. This concerns, for example, the question of information interaction in the human-computer interaction process, but also the addiction or gamification effects of programs that override human nature by changing users' attitudes or behaviors through automatic rewards. Therefore, revealing the ethical challenges and ethical problems brought by persuasive technology or AI systems has become a necessary aspect to regulate its development.

ETHICAL PROBLEMS CAUSED BY PERSUASIVE TECHNOLOGY AND THE ETHICAL PRINCIPLES FOLLOWED BY DESIGNERS

From the perspective of ethics, what is related to ethics is not Captology but the developmental motivations of designers, the marketing of applications and the social consequences of persuasive technologies. This is akin to the study of nuclear physics having nothing to do with ethics, in contrast to the study of atomic bomb technology. In terms of content Captology is a theoretical study, including the software architecture of persuasive systems, technical infrastructure, the design of persuasive systems, visualized interactions between human and the persuasive systems, tailored personalized persuasion and gamified persuasion, a digital marketplace which carries persuasive functions, the creation of smart environments (e.g., internet of things), and so on. In a nutshell, Captology is a specific discipline that studies how to make computer-based technology better perform its persuasive function; "Captology focuses on planned persuasive effects of technology, not on side effects" (Fogg, 2003, p. 18). In contrast persuasive technologies are to develop specific products with persuasive functions. Therefore, the relationship between Captology and persuasive technology is one in between theory and practice. The relationship among ethics, persuasion, technology, computer-based technology, Captology, and persuasive technology can be illustrated as follows (Fig. 2):

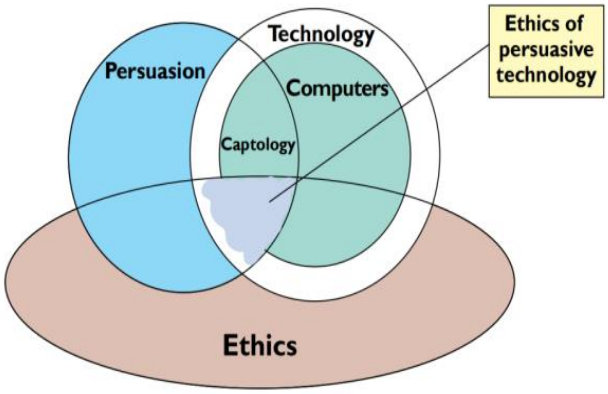

Figure 2: Convergence of ethics, persuasion, and technology. Ethical concerns extend beyond persuasive computers to all forms of persuasive technology – from the simply structural to the complex and cybernetic (Berdichevsky & Neuenschwander, 1999, p. 53).

The figure above indicates that persuasion activity based on human-computer interaction is still a value-laden activity, although persuasive technology transforms persuasion from an interpersonal relation to an interaction between human and computer or AI systems. The persuasive ways of AI system is so variable that users without specialized knowledge may not feel the persuasion activities that are carried out by increasingly intelligent algorithmic systems driven by data. The persuasive intention of the AI systems is enduring enough to make their users resonate emotionally. The persuasive process of AI systems is so adaptable as to dominate the choices of their users. The problem is, however, that these features of persuasive technology are its advantages, but at the same time they also raise some unprecedented ethical questions.

Firstly, persuasive technology may disguisedly hide or weaken its persuasive intention. The AI products that perform the task of persuasion are both a provider of method and an executor of method in the human-computer interaction system. On the one hand, this dual identity, leads to synergies with their users by the form of graphics, audio, video, animation, simulation models, hyperlinks, etc., so as to achieve the best persuasive effects. On the other hand, because of the novelty of persuasive technology, they may hide their persuasive intention and distract the attention of their users, so as to make their users accepting the information delivered by without carefully reviewing the content. In most cases, their users have no choice but to accept many of the default settings offered by the designers when they use AI products. Therefore, in the process of human-computer interaction, the choices of users are not only affected by the information content advanced by the AI system, but also depend on the presented way of content and unconscious acceptance of implicit settings (Fogg, 2003, pp. 213-215).

Secondly, persuasive technologies may potentially limit their user's right of free choice. The AI products have controlled the process of interaction in the human-computer interaction system, because their users only have the right to choose whether or not to continue an interaction, but not the right to debate or ask the AI system for clarification or explanation. When technological persuasion is applied to a person, the success of

persuasion does not depend on his or her logical reasoning ability, but on the guidance of his or her emotion. When persuasive technology is applied to a group, it is difficult for individuals to make a free choice about the persuasive purpose of the AI system, in other words, individual actions are no longer the result of voluntary choice, but are constructed. For example, a "health code" was widely used in China during the COVID-19 epidemic, in order to control the epidemic effectively. The company encourages employees to use fitness software (such as Fitbit) or persuasive activity tracking systems in the form of company benefits, in order to pay a lower insurance for each employee.

Thirdly, persuasive technology may prove so addictive so as to become a new "opium of the people." A digital environment with persuasive features not only automatically adjusts its interactive action according to the digital information about its users, but abolishes altogether the dichotomy between passive materiality and active mind. Its user has been placed in a state of being "interpreted" and "fed." Especially, the popularity of smart phones and their many apps with persuasive functions have caused many adults to become addicted to mobile phones. For children and teenagers whose emotional controls have not fully developed, their addiction, obsession, or indulgence are no longer the consequence of being weak-willed, but rather created by the designers of persuasive technologies who are exploiting their developmental weaknesses and psychological vulnerabilities. Thus, the psychological manipulation used by persuasive technologies is likely to put their physical and mental health at risk, so that we may be caught in a new type of Opium War.

Fourthly, the emotional cues of AI systems may put people at a disadvantage. In the process of person-to-person persuasion, both will often achieve a fairer and more ethical persuasion effect due to empathy. However, in the persuasive process of human-computer interaction, the emotional cues provided by AI systems will affect people's choices and judgments, because they are context- sensitive by way of picking up cues from the users and adapting to their behaviors However, the AI systems do not have real emotional resonance, because they are a material system. This asymmetry will leave their users at a disadvantage. For example, when a social interactive toy uses emotional words – such as expressions of friendship – to communicate with children, this may affect the children's feelings and actions. Whether this kind of influence is moral or not has become a focus of debate. At present, the emotional expression of AI system is a moral gray area of human-computer interaction (see Fogg, 2003, pp. 217-222, p. 105). This urgently requires that we systematically study the ethical problems caused by smart toys.

Although the four ethical problems raised by persuasive technologies are not exhaustive, they have indicated at least that one of the effective ways to avoid these problems in practice is to proactively conduct an ethical examination of the designer's design intention, the persuasive methods, the foreseeable social consequences, and the unexpected circumstances caused by the application of AI products. That is, we need to comprehensively evaluate the ethical nature of AI products by judging whether each step or aspect of it is ethical. In particular, the designers of persuasive technology should abide by the following four ethical principles of an ethics of persuasive technologies (see Berdichevsky & Neuenschwander,1999, pp. 52-58):

1. The Principle of Dual Privacy

Designers of persuasive technology must at least ensure that the privacy of their users is respected as much as their own. When a user's personal information is transferred to a third party through persuasive technology, privacy settings must be strictly examined. Persuasive technology is able to collect personal information of its users in the process of human-computer interactions, and to make persuasion more targeted. Therefore, designers must comply with the principle of dual privacy when they design AI products or systems that collect information from their users.

2. The Principle of Disclosure

The designers of persuasive technology should disclose their motives, methods, and expected results to the public unless it would seriously undermine other moral objectives. This is because the motivations behind designing a AI systems should never be unethical, even if they employ traditional means of persuasion. No result foreseen by persuasive technology should ever be immoral, even if there are socially beneficial results independently of the means of persuasion. Designers of persuasive products must take responsibility for the consequences of their products that can reasonably be expected in practical applications.

3. The Principle of Accuracy

The designers of persuasive technology must not provide misinformation in order to achieve their persuasive goals. Most users have seen technology as something reliable and honest in the majority of cases. They cannot have any awareness of the deceptiveness of technology in their use of technologies. Therefore, the designers of persuasive products must abide by the principle of accuracy and avoid abuse in order to ensure the credibility of AI products.

4. The Golden Principle

Like the principle of dual privacy, the golden principle also invokes the idea of reciprocity which means that the designers of persuasive technology should never seek to make anyone believe or do things that even the designers themselves would not want to be persuaded to believe or to do. This principle is also supported by John Rawls's consideration of the ethical issues behind the "veil of ignorance" in his famous book *A Theory of Justice.* Rawls designs the "veil of ignorance" to ensure that the choices made by participants or stakeholders are guaranteed not to be distorted by their special interests and benefits. Therefore, the golden principle may minimize the possibilities for ethical harm caused by persuasive products.

Taken together, these four design principles provide the bottom line for the development and application of persuasive technologies, they are also the basic principles for ethical review of the entire process.

ETHICAL CHALLENGES CAUSED BY PERSUASIVE TECHNOLOGY; AND AN ETHICAL RESPONSE

The ethical principles obeyed by the designers of persuasive technology are only for the design and application of computing products. They do not concern the specific technical details. For this, one has to consider the three kinds of inevitable bias when designers develop data-driven algorithmic systems. The first one is the preexisting bias,

which is caused by the social culture and customs which form the cognitive background of the designers. This bias is similar to Heidegger's concept of pre-supposition, pre-understanding and pre-existence, because everyone is a person in a specific environment, and his or her values must imperceptibly include the background concepts of the society and culture in which they live. So the preexisting bias is also an unconscious background idea or exists in the subconscious of the designers. The second one is data bias, which is caused by the incomplete data of the algorithm systems while they are being trained and by the way the data are selected, because whether it is a completely data-driven algorithm system or a data-driven algorithm system with knowledge embedding, it needs to be trained based on a specific dataset so that it can acquire "expertise" or "advantage." The third one is emergent bias, which is caused by the fundamental features of machine learning algorithms that are currently in use or emerge from algorithmic systems in the process of human-computer interaction.

These three biases of the algorithmic systems and the particularity of the persuasive function of computing products have given rise to the ethical challenges that cannot be solved within the original ethical conceptual framework or according to traditional ways of thinking. The most obvious ethical challenge is the "attribution of responsibility"-problem. In traditional moral philosophy, accountability is the assignment of responsibility to all relevant moral agents according to the causes and effect of an event that occurred, including the moral condemnation or other punishment for moral agents who have caused harm to users due to their bad motives and intentions or negligent actions. The moral agents are those who can bear moral responsibility and have the ability to compensate. However, a persuasive technological system is a computer-based technological system and as such a novel device that intervenes between the designers and their users. The complexity and interconnectedness of AI systems makes it difficult to trace responsibility by traditional ways. It has led to the following four "dilemmas of accountability," which are distilled from the works of Cooper et al. (2022).

Firstly, there is the causal dilemma of accountability. The development and application of human-computer interactive computing products involve cooperation or collaboration among multiple moral agents, such as scientists, engineers, designers, trainers, evaluators, decision makers, managers, regulators and other diverse or decentralized experimental groups. Both hardware and software production are done in company settings. It is extremely difficult to find a moral agent to take a responsibility for all developmental decisions and every detail of the technology, because the entire computing system is composed of multiple modules. In most cases, these modules are developed by multiple engineering groups as is the case, for example, in the development of machine learning models as a multi-level process. Open-source software, databases, multi-target toolkits, and other products developed by other groups come together. Some control systems have the capacity of interoperation. Some AI products may continue to be used on the Internet and never disappear from the market, although the companies that produced them have gone out of business, or the person responsible for the project has been changed. In these cases, finding a morally accountable person among multiple interconnected groups is no longer an easy task after harm has occurred.

Secondly, there is the dilemma of accountability when there is a bug in the operating system. The data-driven algorithmic systems have necessarily relied on some specific abstract assumptions about a phenomenon. The statistical nature of algorithmic systems and the incompleteness of training data can lead to misclassifications, statistical errors, and uncertain outcomes. When machine learning experts describe these "bugs" as features of machine learning, it is possible for the developers to attribute the resulting damage to these features or "bugs" of the algorithmic system, rather than to the mistakes resulting from human negligence or insufficient ability to generalize and predict. In this way, the currently inherent statistical characteristics of the algorithmic system, – which are at the heart of persuasive technology – may become an excuse for moral agents to shirk their responsibilities, so that their users have to passively bear the resulting losses. The existence of this phenomenon implies that an intangible "treaty of inequality" was signed between technology providers and their users.

Thirdly, there is the ownership dilemma of the disclaimer. The two concepts of ownership and responsibility are important ethical and legal concepts with rich meanings and a long history. However, the trend in the computing industry is towards greater property rights and less liability. Consider, for example, the overlord clauses established by shrink-wrap and click-through licenses used for software copyright authorization; or consider the disclaimers set forth in the terms of service for web services, mobile apps, IoT devices, content moderation decisions, etc. There is also the refusal of third-party providers of algorithmic systems to submit their products to ethical review, on the ground of protecting their intellectual property or keeping their trade secrets. At the same time, manufacturers and owners of cyber-physical systems (e.g., robots, IoT devices, drones, autonomous vehicles) can shift responsibility to environmental factors or human-machine loops and so on. The strengthening of the sense of ownership and the weakening of a culture of accountability will lead to many new social challenges, because these trends grant technology companies more and more control of the rules such that users' losses appear to be just bad luck.

Fourthly, there is the dilemma of accountability caused by artificial agents. With the improvement of the degree of intelligence of algorithmic system, the capacity and agency of computational products will be increasingly similar to that of human beings and will increasingly embody a tendency to personify. Developers and critics describe these systems as intelligent. This means that AI products should be held responsible for the mistakes they make in some complex cases. However, computing systems, even if they have the ability of action, do not become a moral agent like a human being within the traditional framework of accountability. In this case, when we have to track the users' losses caused by AI products back to the human moral agents related to them – such as designers, developers, owners and trainers etc. – accountability becomes downgraded to the inspection of the quality of AI products, rather than serving as a normative concept associated with ethical responsibility.

For the purpose of taking responsibility and punishment, an algorithmic or AI system is a material system. Although the original persuasive intention of the material system is designed or provided by its designers and coaches, the ability of environmental awareness and knowledge discovery undercuts the traditional dichotomy of matter and

the environment being passive, of consciousness and mind being active. Their interactivity, autonomy, and adaptability in the process of human-computer interaction not only turn users imperceptibly to a state of being "interpreted" and "fed," also changing their "factory settings." Therefore, damages caused by complex algorithmic systems cannot necessarily be attributed to the fault of the people involved, because these damages may be related to the bias and randomness of the algorithmic system itself. In addition, the inputs of data-driven algorithmic systems are digital or discontinuous, while the causal tracing of accountability is based on assumption of continuity and linearity. As a result, the way of thinking that retraces causality to assign responsibility among human moral agents loses its applicability in a persuasive technology system. This inapplicability is also manifested in two ways. The one is that it makes no sense to punish a material system, because it is not a human moral agent. The other is that the material system itself does not have the capacity to bear liability.

Given all this, the effective approach to get out of the above four dilemmas of accountability may consist in moving beyond or abandoning the way of thinking which considers accountability only among human moral agents. This approach would expand the semantic and pragmatic scope of the concept of responsibility, it would propose a new conceptual framework for ethics and conceive a new kind of accountability which can be applied to the development and application of AI systems. It would then establish a compensation mechanism that is does not require accountability as a condition of punishment. Thus we need to distinguish between the responsibility taken by the material system and the punishment delivered by the material system by preparing a pool of funds for each complex intelligent system so that a user's loss can be compensated to a certain extent. This might involve, for example, binding human moral agents together to form a new collective personality. This discussion began more than 30 years ago (Solum, 1992) and has now become a hot topic of concern in philosophical and legal circles. It should be emphasized here that this article advocates that the material system should be held responsible, neither to reduce people's responsibility, nor to shift responsibility to the material system so as to avoid the accountability of persons, but as a proposal for an effective mechanism for victims to obtain financial compensation in cases where the responsible person cannot be found.

CONCLUSION: BUILDING A NEW CONCEPTUAL FRAMEWORK OF ETHICS ABOUT THE FUTURE OF HUMAN BEING

Techniques and technologies of persuasion traditionally involved oral or written language, be it for the presentation of arguments or for rhetorical strategies and seductive slogans. In contrast, the persuasive function of persuasive technology is based on the automated decision-making capabilities of intelligent systems. One of the reasons why users prefer to adopt automated decision-making is that they generally believe that automated decision-making based on massive data is not only faster and more reliable than human decision-making, but also able to provide decision-making suggestions beyond human imagination. This ignores, however, the biases which are intrinsic to the algorithmic system and the decision-making errors owing to the randomness of algorithm

systems. Therefore, when designers construct AI products with persuasive functions, they need to provide the choice of giving informed consent, so that the persuasive intention, methods, and social outcomes of the algorithmic system are aligned to the values and interests of users. They also need to respect the autonomy of users so that they can eliminate paternalistic persuasion assumptions. In short, there is an urgent need to raise ethical awareness, strengthen humanistic education, and coordinate the relationship between interests in maintaining security and promoting technical innovation. Ethicists need to expand their traditional conceptual framework to ensure that there can be compensation for the damages caused by intelligent systems. This includes a shift from an emphasis on being passively responsible to actively assuming responsibility, that is, a shift to a new conceptual framework of ethics for the future of humanity. Legal scholars and legislators need to create new mechanisms which can solve the problems caused by an intelligent algorithmic system or an intelligence machine. In order to ensure the healthy working of an intelligent society, regulators need to create a set of new rules to guide the process of developing and applying AI systems with persuasive functions.

REFERENCES

Berdichevsky, D., & Neuenschwander, E. (1999). Toward and Ethics of Persuasive Technology. *Communications of the ACM*, *42*(5), 53. https://doi.org/10.1145/301353.301410

Cooper, A. F., Moss, E., Laufer, B. & Nissenbaum, H. (2022). Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)* (pp. 864–876). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533150

Fogg, B. J. (1998). Charismatic Computers: Creating More Likable and Persuasive Interactive Technologies by Leveaging principles from Social Psychology. https://www.semanticscholar.org/

Fogg, B. J. (2003). *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann.

Fogg, B. J. (2019). *Tiny Habits: The Small Changes That Change Everything*. Houghton Mifflin Harcourt.

Marcus, A. (2015). *Mobile Persuasion Design: Changing Behaviour by Combining Persuasion Design with Information Design*. Springer-Verlag London Ltd. https://doi.org/10.1007/978-1-4471-4324-6

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.

Solum, L. B. (1992). Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, *10*(4), 1231-1287.

СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR

Сумэй Чэн, csm@sass.org.cn

Sumei Cheng, csm@sass.org.cn

Статья поступила 21 сентября 2024
одобрена после рецензирования 16 октября 2024
принята к публикации 30 ноября 2024

Received: 22 September 2024
Revised: 16 October 2024
Accepted: 30 November 2024