ISSN 2616-5767 Vol.7, Issue 2: 9-14, DOI: 10.25236/AJETS.2024.070202

Research on the Design of Value-sensitive Artificial Intelligence

Zehua Dai*

School of Marxism, Yunnan Normal University, Kunming, Yunnan, 650500, China
**Corresponding author*

Abstract: *Artificial Intelligence (AI) technology has a relatively short history, emerging in the mid-1950s. Despite its youth, AI research and applications have rapidly developed and expanded into various fields, including intelligent robots, language recognition, natural language processing, and expert systems. These advancements have had far-reaching impacts on human society, economy, culture, ecology, and other aspects. However, AI technology, like any new technology, brings both benefits and ethical issues. To promote the integration of technology and human ethical values, it is necessary to study the value-sensitive design of AI. This will enable AI to benefit society and guide its development.*

Keywords: *Artificial intelligence, value-sensitive design, science and technology*

1. Introduction

The rapid development and widespread application of artificial intelligence have demonstrated the power of science and technology. Society is constantly evolving and progressing, and promoting the advancement of science and technology is a necessary choice for modern society. The development of AI technology will impact people's work, study, and life, providing benefits and conveniences while also presenting challenges and risks to human values[1]. It is important to reflect on and address the ethical issues of AI technology promptly and implement appropriate coping strategies to ensure that AI technology benefits human beings and avoids harm. Exploring the sensitivity of AI design to solve ethical and social problems has become a research field in AI.

2. The conceptual origins of value sensitive design

In 1996, the journal Interaction published a paper titled “Value-Sensitive Design”, which introduced the value-sensitive design theory created by Batya Friedman and her colleagues. However, the earliest research results found in the open literature are in the paper titled “Human Subjects and Responsible Computing: Implications for the Design of Computer Systems”, published in 1992. In this paper, Friedman and Peter H. Kahn, Jr. proposed the idea of responsible computing through the design of computer systems to avoid the negative consequences of their work. Friedman argues that the concept of value-sensitive design originated from this article. Value-sensitive design has been in development for almost thirty years since 1992 [2]. Value-sensitive design differs from technology development methods that solely prioritize technological rationality and economic efficiency. Instead, it focuses on examining ethical values during the initial design phase of a technology. This approach integrates engineering innovation with ethical values. The book 'Value-Sensitive Design - Shaping Technology through Moral Imagination' by Friedman and David G. Hendry, et al. provides a comprehensive summary of value-sensitive design, including its theoretical foundations, practical methods, and technical applications[3].

Value-sensitive design is a technical design methodology that employs a principled and holistic approach to articulating human values throughout the design process. In this sense, ethical technology design is not only a forward-looking behavior, but should be carried out throughout the design process to effectively embed human moral values. Simultaneously, it can be interpreted as incorporating human values into the design of machine technology and integrating ethical and moral values into the initial research and development of robots. This ensures the prevention of adverse effects of technological artifacts and adherence to human value requirements. When discussing unmanned driving, the topic of whether artificial intelligence has moral agency often arises. Hypothetical situations are often used to determine responsibility in the event of a car accident, as well as how to compensate for it and how the

ISSN 2616-5767 Vol.7, Issue 2: 9-14, DOI: 10.25236/AJETS.2024.070202

law recognizes such issues. This is known as value-sensitive design. In his 1942 science fiction novel "Runaround", Asimov introduced the three laws of robotics in order of priority: the first law states that a robot may not injure a human being or, through inaction, allow a human being to come to harm. The three laws of robotics are as follows: First, robots must not harm humans or allow humans to come to harm. Second, robots must obey human orders unless doing so would violate the first law. Third, robots must protect their own existence as long as it does not conflict with the first two laws. To overcome the limitations of the first law, he proposed a higher-priority zero law of robotics: a robot must not endanger the human race as a whole or allow the human race as a whole to be endangered. These four laws outline the premise of robot manufacturing. They can also represent the premise of current artificial intelligence manufacturing, which takes the human race as a whole as the basis of value formation. With this major premise, it triggers the further refinement of the minor premise, such as the issue of human rights ethics. Human rights ethics are embodied in the subjective initiative of human beings and are mainly classified into the following aspects: Firstly, respect for the value of human life and dignity; secondly, respect for human freedom and equality; thirdly, focus on the spirit of democracy and mutual love; and fourthly, the promotion of comprehensive development of human beings. Additionally, the ethics of responsibility must be considered. As intelligent beings, we must take responsibility for any irreparable losses caused to human society. For instance, in cases where an expert system used in the medical industry causes malpractice, or when robots harm each other or humans, who should be held responsible? Should it be the robots themselves, the researchers who created them, or the users of the robots? Furthermore, what responsibilities should they each bear? For instance, the issue of environmental ethics has become increasingly pressing with the rapid development of artificial intelligence technology. This has led to the frequent updating and iteration of products across various industries. From intelligent laboratories to smart homes, a wide range of innovative products are available, but some of them may have unusual or even harmful features that need to be eliminated. Most of the materials used in the production of obsolete technology products are synthetic, metal, or non-metal materials. If these components are exposed to nature, they are difficult to decompose, which increases the burden on the environment and can lead to ecosystem collapse. This burden can only be borne by humans themselves. From this, it is evident that technology is not value-neutral as traditionally understood, but rather embedded with the values of the designer. Therefore, ethical issues must be resolved, making value-sensitive design urgent and necessary.

3. Principles of value-sensitive design

Value-sensitive design is a methodology led by David Hendry, David Friedman, and others at the University of Washington's School of Information. It advocates for a series of methods to account for human values in technology design. The methodology includes conceptual, empirical, and technological approaches to elucidate the meaning of human values in a given context and feed into technological systems. The construction of artificial moral intelligences aims to embed human values in autonomous machines, based on current technological capabilities, and to test them in the context of users' interactions. According to Ibo van de Poel and Peter Kroes, this involves a progression from "intended value" to "embodied value" to "realized value".

To achieve value-sensitive design for AI, it is essential to understand the following principles.

1) Designers must have a deep understanding of the values held by different user groups. This includes analyzing factors such as users' culture, social background, moral values, and personal beliefs to better meet their needs and expectations. For instance, numerous brands have released smart speakers that can be programmed to meet the needs of various individuals who require different commands on a daily basis. AI designers must consider the values of diverse human beings. For example, children may require children's songs or storytelling, while adults may need to control basic home appliances, such as turning the lights on and off, adjusting the water heater, or setting reminders for their to-do lists. These differences among users must be researched and investigated by the designer in advance.

2) In the design process, it is crucial to identify the potential impacts that technology products or systems may have on users and society. This includes both positive effects and negative risks, as well as possible ethical conflicts. Designers must consider these impacts holistically to avoid or minimize potential negative consequences in the design. For instance, the teaching profession, which is closely related to intelligence, will undergo redefinition as AI becomes more prevalent in education. This will have a positive effect on balancing the level of education for students in different regions and making resources more equitable. However, it also poses negative risks, such as hindering the expression of

ISSN 2616-5767 Vol.7, Issue 2: 9-14, DOI: 10.25236/AJETS.2024.070202

students' personalities by providing the same education to all. At the same time, we cannot ignore the fact that AI lectures on moral values may lead to moral conflicts, as they may not align with human moral cognition.

3) Design Decisions Based on User Values: Designers must consider user values and potential impacts when making design decisions. This involves taking values into account throughout the design process and striving to respect and realize them. Design decisions may require weighing different values and making informed choices in potentially conflicting situations. There are numerous fundamental values, including acceptance, achievement, ambition, attractiveness, caring, compassion, bravery, helpfulness, honesty, imagination, independence, equality, sensibility, and comfort. These values must be prioritized differently depending on the user to better serve human life.

4) Value-sensitive design highlights the importance of engaging with stakeholders closely. This involves involving users, designers, developers, policy makers, communities, and other parties. By collaborating with stakeholders, we can better understand and meet the needs and values of all parties, while minimizing potential conflicts and misunderstandings.

5) After completing the design, evaluation and feedback are necessary to understand its actual effectiveness and identify potential areas for improvement. Collect user feedback regularly, evaluate the design's impact, and make necessary adjustments based on the evaluation results.

4. Advantages of value-sensitive design

Value-sensitive design (VSD) is an effective engineering practice that advocates for the systematic consideration of human values during the technology design stage and their embedding in the technology system. Friedman et al. argue that VSD has several advantages.

(1) Value-sensitive design aims to proactively influence technology design from the beginning and throughout the process.

(2) It can address value generation in various environments, including the workplace, home, business, and online communities.

(3) It explores a broad range of human values, such as collaborative work and democratization.

(4) It employs a survey research methodology that involves conceptual, empirical, and technical approaches.

(5) This perspective emphasizes the interaction and shaping of technology with individual human beings and social systems.

(6) It draws on moral epistemology, which gives moral status to particular values.

(7) Value-sensitive design also takes into account specific values that are commonly held in different times and cultures.

Value-sensitive design is a comprehensive and systematic approach to analyzing context and technology. It recognizes the process of constructing technology as an iterative, nonlinear process of constant revision and refinement. The functions of artificial moral intelligences and human values are always in a continuous game, and it is impossible to realize all values in an autonomous technological system.

According to value-sensitive design theory, a technology system is always suited to certain activities and values, making it difficult to realize other activities and values. Technical usability and human value often present a dilemma:

(1) Good design promotes good usability without affecting human value;

(2) Design can improve usability but may sacrifice some value;

(3) Design can align the technical system with ethical standards and human values but at the cost of usability;

(4) Technology can be designed to conform to a specific range of human values. The development of an artificial moral intelligence involves a repeated process of analysis, verification, and refinement that balances technical usability with value suitability.

Value-sensitive design clarifies the relationship between values, technical requirements, and design

ISSN 2616-5767 Vol.7, Issue 2: 9-14, DOI: 10.25236/AJETS.2024.070202

behavior, and enhances the traceability of values throughout the technology development process.

In summary, value-sensitive design offers a comprehensive and practical perspective, utilizing a holistic and systematic approach. This integrated approach, which includes qualitative, quantitative, and technical practices, outperforms approaches that rely solely on isolated ethical and moral frameworks with ex post remedial measures. Albrecht Lund extends the value-sensitive design methodology to include more than just technological practices, placing technological systems in a multistable and relationally rich constructed scenario. It is important to consider whether value-sensitive design can solve all dilemmas in the construction of artificial moral intelligences. There are still questions to be answered. Value-sensitive design is an evolving concept that requires continuous harmonization with specific cultural contexts, technological heterogeneity, and other factors. Umbreono suggests that further research is necessary to determine the long-term applicability of value-sensitive design to artificial moral intelligences. Secondly, Moore, Wallach, and Allen, among others, have proposed the idea of fully artificial moral intelligences. However, the question remains: do fully artificial moral intelligences actually exist? Furthermore, does their existence pose a threat to human subjectivity? This is no longer a question of how to construct artificial moral intelligences, but rather a complex issue of human destiny, which is beyond the scope of value-sensitive design research. The primary objective of value-sensitive design is to create a machine system that fulfills human ethical requirements. As demonstrated by Winsberg, there are numerous justifications for developing an artificial moral intelligence, all of which are rooted in the specific value expectations of humans for the dependability and safety of machine systems. Therefore, it is crucial to express and integrate ethical principles into a value-sensitive design framework for artificial moral intelligence.

5. Practical applications of value-sensitive design

Value-sensitive design for AI involves identifying and eliminating bias and discrimination. AI systems may exhibit bias based on gender, race, age, and other factors. Designers can use data analytics and machine learning techniques to review and analyze datasets and algorithms to identify and eliminate potential bias and discrimination. This fragment of text discusses measures that can be taken to prevent bias and discrimination in AI systems, such as data cleansing, algorithm tuning, and model training.

Additionally, the text mentions the possibility of customizing AI systems to meet individual user needs and preferences. Designers can gather information about users' values and needs through user feedback, surveys, and research. This information can be used to personalize the system's design and algorithms, providing a better user experience. It is important to consider social and cultural differences when designing AI systems to meet the needs of diverse contexts. Designers can work with and study different groups of users to understand their cultural values, social habits, behavioral characteristics, etc., in order to fully consider and respect these differences in their designs. This can include measures such as multilingual support, cross-cultural design and user participation to ensure that the system is designed to be inclusive and accessible to different social and cultural groups. During the design process, ethical review and assessment is required to ensure that the design is consistent with moral and ethical principles. This includes assessing and addressing potential ethical issues with the design, as well as monitoring and evaluating the actual impact of the design. Designers can use ethical review methods and tools, such as ethical review boards, ethical assessment frameworks, etc., to identify and address ethical issues in a design, as well as to understand the actual impacts of the design and room for improvement through user feedback and evaluation methods.

The "Trolley Paradox" is also often used as a measure of whether AI is of the highest ethical standard. It's a classic thought experiment in ethics, asking whether a driver would crash into five people to save another. The question is, how can a machine be expected to make a more reasonable choice when humans themselves have failed to resolve this moral dilemma? Zhang Xiaoyu, a researcher at the Center for the Study of World Politics at East China Normal University, evaluates such a tendency in his book Technology and Civilization. He argues that most of the time people talk about what is actually a fantasized advanced intelligence, and that these virtual characters face moral choices and confusion just as people do. Talking about AI from the perspective of fantasized advanced intelligence is certainly a lot of whimsical and interesting ideas, but it doesn't help much in understanding the nature of this technology and its real impact on human civilization. This viewpoint coincides with what Shen Xincheng and Huang Qingqiao discuss in their new book, Artificial Intelligence and Values, in which they both hope that the academic community and the public will not lose sight of how the technology itself is shaping human society. Shen Xincheng, an assistant professor

ISSN 2616-5767 Vol.7, Issue 2: 9-14, DOI: 10.25236/AJETS.2024.070202

at the Institute of History and Culture of Science at Shanghai Jiaotong University, often tells his students a story. The invention of the zipper had a major impact on human society in the 20th century, but the technology was once resisted because it was seen as “morally corrupt”, because if a person chose to use a zipper, it meant that he was so lazy that he didn't even bother to fasten his buttons. Such incredible values are unimaginable for people living in this day and age. "But this just goes to show that the crystallization of any value in human society cannot be separated from the technological foundation of the era in which it is located.” Shen Xincheng points out. “The history of science abounds with such examples, because there has never been a two-way interaction between technology and values.” Huang Qingqiao, associate professor of the Institute of History of Science and Culture of Science at Shanghai Jiaotong University, said that AI provides new technological support for the reshaping of the value system of human beings, which is bound to produce an unavoidable collision with the original value system of society. “It needs to be emphasized that values are at the core of social culture, and geographical differences in culture will make social values both universal and geographically specific. Therefore, the way AI develops in each particular country will largely follow the established value system, and we still need to uphold the principle of harmony and difference to avoid going to extremes.” Huang Qingqiao said. At the same time, value-sensitive design assumes that technology is not value-neutral, that it embodies the values of the developers, and that technological artifacts can have an impact on the environment in which humans live, on a number of levels, including moral and political. In other words, the decisions researchers make during the design process have value implications. Therefore, it is extremely important to fully consider human values before designing. In terms of the specific ethical values that value-sensitive design focuses on, it covers human welfare, ownership and property, privacy, unbiasedness, universal usability, trust, sexuality, informed consent, and accountability, etc.; it also involves pragmatic values that focus on the user's experience in the process of use (e.g., simplicity of the system's operation), conventions (e.g., standardization of protocols), and personal tastes (e.g., color preferences in graphical user interfaces), etc.

6. Challenges of value-sensitive design and ways to address them

Value-sensitive design for AI faces a number of challenges and constraints, including the following: first, technical constraints and feasibility issues, in practice, the design and implementation of AI may be affected by technical constraints and feasibility issues. For example, certain values may be difficult to model and implement in an algorithmic form, or may not satisfy the needs of all users due to technical constraints. In addition, due to the complexity and uncertainty of AI systems, it may be difficult for designers to accurately predict and control the behavior and outcomes of the system. Secondly, there are also conflicts and trade-offs between different values: the design of AI systems often requires trade-offs between the different values of different user groups. Different users may have different expectations and preferences for certain features, decisions, or behaviors, which may lead to conflicts between values. Designers need to weigh the different values and find a balanced solution to try to satisfy the needs of multiple user groups. Finally there are legal and policy constraints and uncertainties: the design and use of AI is subject to legal and policy constraints and uncertainties. For example, legal requirements on personal privacy, data protection, and discrimination prohibitions may have an impact on the design and implementation of AI systems. In addition, due to the rapid development of AI technologies, relevant laws and policies may evolve and adapt, and designers need to keep abreast of and comply with the latest legal and policy requirements.

In the face of these challenges and constraints, value-sensitive design for AI needs to synthesize multiple factors, including technical, ethical, social, and legal considerations. Designers can address and cope with these challenges and ensure that the design and implementation of AI systems are consistent with human values and ethical principles through close collaboration with users, interdisciplinary teamwork, and ethical review and evaluation.

Value-sensitive design of AI is a growing and evolving field, and future research directions and development perspectives include the following:

1) Ethical decision-making and ethical principles: researchers will continue to explore ways to incorporate ethical principles and ethical decision-making into the design and decision-making process of AI systems. This includes developing more refined ethical assessment methods and tools, as well as investigating how to enable AI systems to actively consider and follow ethical principles.

2) Human-computer collaboration and shared decision-making: researchers will further investigate

ISSN 2616-5767 Vol.7, Issue 2: 9-14, DOI: 10.25236/AJETS.2024.070202

how to achieve human-computer collaboration and shared decision-making to ensure that the behaviors and decisions of AI systems can effectively interact and collaborate with human users. This includes the development of more intelligent and transparent decision-making processes, as well as the provision of user-friendly interfaces and tools that enable users to understand and participate in the system's decision-making process.

3) Fair and inclusive design: researchers will continue to focus on ways to ensure that the design and implementation of AI systems is fair and inclusive. This includes addressing bias and discrimination in the system, as well as taking into account the needs and differences of different social and cultural groups. The researcher will further explore diversity and inclusion design methods and tools to ensure that systems have fair and equal treatment for all users.

4) Interpretability and Transparency: researchers will continue to work to improve the interpretability and transparency of AI systems. This includes investigating how to explain and interpret the decisions and behaviors of AI systems, as well as providing user-friendly explanations and feedback mechanisms. Researchers will also explore ways in which AI systems can transparently demonstrate their workings and algorithms to enhance users' trust and understanding of the system.

7. Conclusion

In conclusion, value-sensitive design of AI is an area full of challenges and opportunities. Future research will continue to explore how to ensure that the design and implementation of AI systems are consistent with human values and ethical principles in order to achieve smarter, fairer, and more trustworthy applications of AI technology. Regardless of how AI develops, the premise of its development remains inseparable from human values, and under such provisions, AI can better develop in response to human needs. Value-sensitive design provides a systematic and holistic view of the methodology for constructing artificial moral intelligences under the premise of the interaction between technology and human beings, which has a certain degree of universality. However, it is more important to pay attention to the differences in specific cultural environments, social policies and other factors, which also requires in-depth research under the perspective of different disciplines, continuous improvement and development.

References

[1] Zhang H P, Xia B H. Value-sensitive design perspectives: background, current status, problems and future[J]. Natural dialectics research,2023,39(04):77-83.
[2] Liu B J. Exploration of value-sensitive design methods[J]. Natural Dialectics Newsletter, 2015, 37(02):94-98.
[3] Friedman B. Value-Sensitive Design[J]. Interactions,1996,3(6):16-23.