The Challenges for Socially Responsible AI for Well-being

Takashi Kido [1] and Keiki Takadama [2]

[1] *Teikyo University, Advanced Comprehensive Research Organization, Japan*
[2] *The University of Electro-Communications, Japan*

Abstract
In this AAAI Spring Symposium 2023, we discuss Socially Responsible AI for Well-being. For AI to truly benefit society, it must go beyond mere productivity and economic advantages; embrace social responsibility; and emphasize fairness, transparency, safety, and other key principles. For instance, AI diagnostic systems should not only be accurate but also free from bias, ensuring equitable data representation across races and locations. This highlights the need for ongoing discussions on the nature of "social responsibility" in AI applications. There are two main perspectives: (1) Individually Responsible AI: Focuses on designing AI systems that consider individual well-being, such as understanding how digital experiences influence emotions and quality of life. (2) Socially Responsible AI: Emphasizes broader societal impacts, striving for decisions that are fair and beneficial for all. Addressing biases in AI is crucial to achieving fairness. Additionally, the knowledge produced by AI, such as health advice, should be universally applicable and not only beneficial to a subset of individuals.
This paper outlines the underlying motivations, key terms, areas of focus, and research inquiries for this symposium.

Keywords
Socially Responsible AI, Well-being

1. Motivation

For our well-being, AI is insufficient to achieve rapid growth or to dominate the economy. AI must adhere to principles of fairness, transparency, accountability, reliability, safety, privacy, and security. Taking an AI diagnostic system as an example, although it is essential to provide accurate results with clear explanations, it is equally crucial that such results are derived from unbiased data. This ensures an equitable representation across different races and regions. As AI decisions influence the quality of life, it is vital to define "social responsibility" in the upcoming AI era.

The first viewpoint focuses on "Responsible AI at the Individual Level." This perspective seeks to understand the mechanisms necessary to shape an AI that considers personal well-being. With daily fluctuations in health conditions, AI should recognize how digital interactions impact emotions and the quality of life.

The second viewpoint centers on "AI with Societal Responsibility." This perspective aims to understand the considerations for integrating societal values with AI to improve overall well-being. An integral aspect of this responsibility is ensuring fairness in AI decisions so that they benefit everyone. Addressing AI bias is vital. Furthermore, AI-generated health insights such as sleep tips for one individual might not be relevant to another, emphasizing the importance of

universally beneficial knowledge. To address these concerns, it is essential to develop methods that prevent AI from inheriting human bias and that ensure fairness and societal responsibility.

We invite both technical and philosophical conversations centered on "Socially Responsible AI for Well-being" as it relates to the ethical design and deployment of machine learning, robotics, and social media, among other areas. Topics such as clear predictions, trustworthy social media practices, beneficial robotics, using AI/VR to combat loneliness, and advocating health are integral to our discussion.

2. Our Scope of Interest

We plan to explore key interdisciplinary issues that will shape the future direction of Socially Responsible AI for Well-being. The following scope of interest is the focal point of our symposium:

2.1. Individually Responsible AI

1. Interpretable AI

Interpretable AI seeks to comprehend AI's decisions and actions. This involves understanding how AI results are responsible for well-being. To address this, it is essential to develop advanced tools that can elucidate the operation of deep neural networks and various analytical techniques. We advocate both theoretical and empirical studies to understand the strengths and weaknesses of present AI/ML technologies in the context of interpretable AI for well-being. Areas of interest encompass but are not restricted to interpretable AI for precision medicine, accountability of black-box prediction models, interpretability of machine learning systems, interpretability in human/robot communications, establishing trust in AI, and using social computing to foster trust in human-centric computational systems.

2. How can human well-being be defined and measured?

To ensure that the outcomes genuinely enhance well-being, we must first establish and quantify the meaning of well-being. This pursuit paves the way for new success metrics for individually tailored responsive AI. We welcome contributions from various fields, including positive psychology, positive computing, predictive medicine, studies on human well-being, neuroscience behind happiness, cultural algorithms, studies on thriving environments, and cross-cultural assessments of well-being.

3. Dynamical change in well-being

To address the dynamic shifts in human health, it is essential to delve into cutting-edge machine learning techniques and incorporate them into individually tailored Responsible AI. We covered areas such as deep learning, data mining, wellness knowledge modeling, monitoring shifts in health accuracy and efficiency, collective intelligence, life log analytics (vital data and Twitter-based insights), data representation, human-centered computing, biomedical data management, and tailored medical care. We invite dialogue to assess the potential and constraints of the existing technological approaches.

2.2. Socially Responsible AI

1. How can fairness be defined and measured?

From the perspective of well-being, AI should deliver equitable and unbiased outcomes to all its users. To address this, we need to initially delineate what "Fairness" means in the context of well-being, guiding us toward novel benchmarks for Socially Responsible AI. We welcome interdisciplinary studies that cover but are not confined to standards and metrics for fairness, equity in robotics, machine learning, social platforms, systems with human involvement, collective frameworks, causal inference for fairness comprehension, simulations involving multiple agents for fairness insights, game-theoretical fairness evaluations, contrasts between human prejudices and algorithmic biases, biased evaluations of social platforms, and analyses of political orientations.

2. Knowledge applicability for well-being

To explore health-centric information that can benefit a wide audience, we invite both empirical and technical studies. The areas of interest encompass but are not restricted to the analysis of social data and design of social relations, mood tracking, healthcare communication

infrastructures, conversational AI systems, insights into individual behaviors, explorations into 'Kansei', creativity zones, compassion, calming technology, the principles of Kansei engineering, gamification, support tools, technologies such as Ambient Assisted Living (AAL), medical recommendation systems, elderly care systems, web services for personal wellness, games for health and happiness, digital health diaries, trials for disease amelioration (such as metabolic issues or diabetes), sleep enhancement studies, support systems for disabled people, and community computing platform.

2.3. Ethical Issues on "AI and Humanity"

It is imperative to cultivate effective human–AI collaborations to foster trust and acceptance of AI outcomes. We encourage contemplative discussions centered on ethics and philosophy related to this. Key subjects of interest comprise the juxtaposition of machine intelligence and human cognition, the ramifications of AI on societal constructs and human thought processes, challenges posed by misinformation (or 'infodemics') via social platforms, and notions of individual identity, among others.

3. Conclusion

In this paper, we discuss the inspiration, technical aspects, and philosophical concerns associated with "Socially Responsible AI for well-being." As the planners and coordinators of the AAAI23 symposium, our goal is to disseminate the most recent advancements, existing obstacles, and the prospective benefits of AI implementations. Evaluating digital interactions and gaining insights into human well-being are central themes.

4. Acknowledgements

5. References

[1] Kido, T. and Takadama, K., AAAI 23 Spring Symposium Report on "Socially Responsible AI for Well-Bing", AI Magazine, Volume 44, Issue 2, Summer 2023, pp. 211-212, https://onlinelibrary.wiley.com/doi/full/10.1002/aaai.12092

[2] Kido, T. and Takadama, K.,The Challenges for Fairness and Well-being: How Fair is Fair? Achieving Well-being AI, The AAAI 2022 Spring Symposia, How Fair is Fair? Achieving Wellbeing AI, Stanford, CA, March 21-23, 2022, pp. 1-3, https://ceur-ws.org/Vol-3276/SSS-22_opening1.pdf

[3] Kido, T. and Takadama, K., The Challenges for Interpretable AI for Well-being - Understanding Cognitive Bias and Social Embeddedness- AAAI 2019 Spring symposia, Interpretable AI for Well-being: Understanding Cognitive Bias and Social Embeddedness, Stanford, CA, March 25-27, 2019, http://ceur-ws.org/Vol-2448/SSS19_Paper_Upload_210.pdf

[4] Kido, T., Oono, K. and Swan, M., The Challenges for Machine Learning and Subjective Computing in Well-being AI, The AAAI2017 Spring symposia, Wellbeing AI: From Machine Learning to Subjectivity Oriented Computing, Stanford, CA, March 27-29, 2017, pp. 751, https://cdn.aaai.org/ocs/15767/15767-68317-1-PB.pdf