The Challenges for Fairness and Well-being

- How Fair is Fair? Achieving Well-being AI -

Takashi Kido
Teikyo University,
Advanced Comprehensive Research Organization
kido.takashi@gmail.com

Keiki Takadama
The University of Electro-Communications
Department of Informatics
keiki@inf.uec.ac.jp

Abstract

In the AAAI Spring Symposium 2022, we discussed fairness and well-being in the context of well-being AI.One of the important keywords is "well-being." We define "well-being AI" as Artificial Intelligence that promotes psychological well-being *(i.e.*, happiness) and maximizes human potential ability. The well-being AI helps understand how our digital experience affects our emotions and quality of life and how to design a better well-being system that puts humans at the center. The second important keyword is "fairness." AI can potentially assist humans in making fair decisions. However, we must tackle the "bias" problem in AI (and in humans) to achieve fairness. Although statistical machine learning predicts the future based on past data, several types of data biases may lead to an AI-based system making incorrect predictions. For AI to be deployed safely, these systems must be well-understood, and we need to understand "How fair is fair" for achieving "Well-being AI." This paper describes the motivation, scope of interest, and research questions of this symposium.

Motivation

What are the ultimate goals and outcomes of AI? Although AI has incredible potential to help make humans happy, it can potentially cause unintentional harm. This symposium aims to combine humanity perspectives with technical AI issues and discover new success metrics for well-being AI instead of productive AI in exponential growth or economic/financial supremacies.

Especially in the COVID world, people's lives are transforming on an unprecedented scale. From this fact, it is important to investigate how people's mindsets are shifting and how desirable human-AI partnerships would be. COVID-19 may change human-AI collaborations by easing people's concerns about technology. For example, the number of people working from home has increased and business trips have almost disappeared. Meetings are held online, and virtual ceremonies are held using AI bots. The COVID-19 prevention measures promoted digital transformation, generating enormous amounts of data. Therefore, the need for AI has increased, as shown in the race to find a COVID-19 vaccine through global collaborations.

We call for AI-related challenges in new human-AI collaboration and discuss desirable human-AI partnerships for providing meaningful solutions to social problems from humanity's perspectives. This challenge is inspired by the "AI for social good" movements, which pursue the positive social impacts of using AI, supporting the Sustainable Development Goals (SDGs), a set of 17 objectives for the world to be more equitable, prosperous, and sustainable. In particular, we focused on two perspectives: well-being and fairness.

The first is "well-being". We define "well-being AI" as Artificial Intelligence that aims to promote psychological well-being (that is, happiness) and maximize human potential ability. Our environment escalates stress, provides unlimited caffeine, distributes nutrition-free "fast" food, and encourages unhealthy sleep behavior. To address these issues, well-being AI provides a way to understand how our digital experience affects our emotions and quality of life, and how to design a better well-being system that puts humans at the center.

The second perspective is "fairness". AI has the potential to assist humans in making fair decisions. However, we must tackle the "bias" problem in AI (and in humans) to achieve fairness. In the recent trend of big data becoming personal, AI technologies for manipulating the inherent cognitive biases have evolved, such as social media (Twitter and Facebook) and commercial recommendation systems. The "echo chamber effect" is known to make it easy for people with the same opinions in a community. Recently, there has been a movement to use cognitive biases in the political world. Advances in big data and machine learning should not overlook new threats to enlightenment thought.

This symposium called for the technical and philosophical issues of achieving well-being and fairness in the design

In T. Kido, K. Takadama (Eds.), Proceedings of the AAAI 2022 Spring Symposium "How Fair is Fair? Achieving Wellbeing AI", Stanford University, Palo Alto, California, USA, March 21–23, 2022.

and implementation of ethics, machine-learning software, robotics, and social media (but not limited to). For example, interpretable forecasts, sound social media, helpful robotics, fighting loneliness with AI/VR, and promoting good health are important aspects of our discussions.

Our Scope of Interests

This symposium discussed important interdisciplinary challenges for guiding future advances of fairness and well-being in AI. We have the following scope of interest in this symposium:

(1) How can we define and measure the well-being of humans?

To discover new success metrics for well-being AI instead of productive AI in exponential growth or economic/financial supremacies, this symposium called for basic research to define human well-being, which provides inspiration for new success metrics for well-being AI. Interdisciplinary research such as positive psychology, positive computing, predictive medicine, human well-being, economics beyond GDP, social computing for understanding AI job replacement and disparity, neuroscience of happiness and pleasure, multi-agent social simulations, cultural algorithms, a flourishing environment, and cross-cultural analyses for well-being values were the topics of this symposium.

- **Well-being AI: Machine Learning and other advanced analyses for Health & Wellness**

Advanced machine learning technologies, such as deep learning and other quantitative methods, need to be explored in the health and wellness domains. We called for theoretical and empirical research on the well-being AI. Discussions on evaluating the possibilities and limitations of current technologies were also called for.

The topics included deep learning, data mining, knowledge modeling for wellness, collective intelligence/knowledge, life log analysis (*e.g.*, vital data analyses, Twitter-based analysis), data visualization, human computation), biomedical informatics, and personalized medicine.

- **Better Well-being systems design**

To explore empirical and technical research on improving well-being system design, the topics included social data analyses and social relation design, mood analyses, human-computer interaction, health care communication system, natural language dialog system, personal behavior discovery, Kansei, zone and creativity, compassion, calming technology, Kansei engineering, gamification, assistive technologies, Ambient Assisted Living (AAL) technology, medical recommendation system, care support system for older adults, web service for personal wellness, games for health and happiness, life log applications, disease improvement experiments (*e.g.*, metabolic syndrome, diabetes), sleep improvement experiments, healthcare/disabled support systems, and community computing platforms.

(2) How can we define and measure Fairness?

To explore basic research to define the "fairness" for "human-in-the-loop computational systems," providing inspiration for new success metrics for fair AI, interdisciplinary research such as bias and fairness in machine learning, fairness criteria and metrics, responsible AI, trusting AI, social computing for trusting humans-in-the-loop computational systems, multi-agent simulations on fairness, and game theory-based analyses on fairness, were called for in this symposium.

- **Interpretable AI**

Interpretable AI is artificial intelligence whose derived results can be easily understood by humans. For example, we need to develop powerful tools to understand exactly what deep neural networks and other quantitative methods are performing. To address this issue, we called for theoretical and empirical research to understand the possibilities and limitations of current AI/ML technologies for interpretable AI. The topics included human bias vs. computational (data) bias, interpretability of machine learning systems, accountability of black box prediction models, interpretable AI for precision medicine, interpretability in human/robot communications, bias analysis on social media, political orientation analyses, accuracy and efficiency issues in health, economics, and other fields, causal inference to reason about fairness, and actionable recommendations based on causal inference.

- **Better Fairness systems design**

To explore the empirical and technical research on the design of better fairness systems, the topics included criteria and metrics for fairness in robotics, machine learning software, social media, "human-in-loop systems," collective systems, recommendation systems, and personalized search engines.

(3) Ethical Issues on "AI and Humanity": desirable human-AI partnerships.

To explore the ethical and philosophical discussions on desirable human-AI partnerships, the topics included "Machine Intelligence vs. Human Intelligence," "How AI affects our human society or way of thinking," issues on basic income, issues on infodemic (*e.g.*, fake news) with social media, and personal identity. More technically, we need to deepen our understanding of the possibilities and limitations of machine learning and other advanced analyses of health and wellness.

Conclusion

In this paper, we describe the motivation, technical, and philosophical challenges related to "AI Fairness and Well-being" as proposers and organizers of the AAAI2022 symposium. This symposium aimed to share the latest progress, current challenges, and potential well-being of AI applications and discussed the evaluation of digital experience and understanding of human well-being.

References

Kido,T., Takadama, K. 2019. The Challenges for Interpretable AI for Well-being -Understanding Cognitive Bias and Social Embeddedness- 2019 March, Stanford: http://ceur-ws.org/Vol-2448/SSS19_Paper_Upload_210.pdf

Kido,T., Takadama, K. 2018. Wellbeing AI: From Machine Learning To Subjectivity Oriented Computing, AAAI Spring symposium 2018 March, Stanford: https://aaai.org/Library/Symposia/Spring/ss17-08.php

Acknowledgments

We thank the program committees of this symposium for their valuable support.