

Verity Harding, *AI Needs You: How We Can Change AI's Future and Save Our Own*, Princeton and Oxford: Princeton University Press, 2024. 274 pages.

This book is an accessible read on artificial intelligence (AI) aiming for a broad audience. It is the newest work from Verity Harding of the University of Cambridge, who is a British politico turned AI expert. This book intends to provide a clear sense of what is at stake for everyone through revisiting the past examples of transformative technologies such as space exploration, in vitro fertilization (IVF), and the internet. Per personal experience and expertise, the author acknowledges the geographical limits of this study by focusing mainly on the United States and the United Kingdom. The discussion is limited to relevant Western experiences on potential societal structures and political processes while mentioning what is happening in other parts of the world, such as China. In addition to complex technical details, the author recognizes the need and importance of finding proper principles, governance, and shared beliefs for the future in order to guide the development of AI technology. In conclusion, this book is to encourage people to raise their voices in this transforming experience of AI, no matter their level of technical background, to ensure the future of AI can reflect the best of us.

This book is concise in length and has a total of six chapters. It starts with “Shadow Self,” highlighting the author’s own experience working at DeepMind for Google and reflecting on personal observations of AI technology, which follows a mentality of pure tool development without much social considerations. The author argues that technology should be built on societal trust and serve the public good. From Chapter 2 to Chapter 5, the author examines three major socially impactful tech innovations. Space exploration is the first one discussed (Chapter 2). The 1967 United Nations Treaty on the Peaceful Uses of Outer Space declares space as the province of all mankind. Despite geopolitical strife about space to this day, the space program manages to symbolize the aspirations of peace. The author articulates that the secret to keeping

the program away from an act of war lies in the legal and diplomatic innovation, not in the technology of going to the moon. In 2020, the Global Partnership on AI (GPAI) was launched, and it has twenty-five nation-state members plus the European Union.

Following space exploration, the author visits IVF, another socially impactful tech example (Chapter 3). She compares the similarity of the biotechnology debates of the 1970s and 1980s with today's AI discourse and applauds the position reached for IVF through steady and thoughtful democratic process supported by the Warnock Commission in the UK. The ethical and moral issues for IVF are like some kind of precursors for Generative AI. The author emphasizes the urgency of the need for a regulatory landscape with the leadership to assert limitations on the capabilities of science in order for the innovation to thrive. The last example used in the book is the internet and it is separated into two parts, i.e., purpose and profit (the internet before 9/11) (Chapter 4) and trust and terror (the internet post 9/11) (Chapter 5). It is impossible to understand today's technology industry without understanding its unique culture. It is the existing tech industry that built much of today's AI.

The author gives credit of thriving internet to a great deal of social movement in the past several decades. She praises the role of governing bodies such as the Internet Corporation for Assigned Names and Numbers (ICANN), launched in 1998. In the post 9/11 world, there was increasing surveillance of the internet from both governments and companies for security purposes. However, the author thinks that "we have missed something and we have lost something," such as preventing and identifying misinformation and fake stories online. There is no parallel system in AI like ICANN for the internet to ensure representative participation to assess the acceptable limits despite the fact that ICANN has its own challenges. At the end of the book (Chapter 6), the author reiterates her position on AI's future that it is not value-neutral and

it should be developed with intention and purpose. The objective should be peaceful in intent and for social good while embracing its limitations.

The strength of this book is its readability and opens such a difficult topic like AI for everyone without the fear of technical complexity. Readers who are curious about this topic can benefit from this book through those common comparable technology examples. The extensive account from the author as a former politician, technology executive, and academic researcher presents rare behind-the-scene information for the audience. It is precious for the author to acknowledge the challenges we are facing and the limits we may not be able to overcome over a short period of time. Still, the author repeatedly reminds the readers about the danger of turning a blind eye to the development of AI. This book tells us the importance of participating in the debate and raising our voices if we want a future society for all to enjoy. The discussions in this book inspire us to think not only about the convenience brought by AI but also about future risks from ignoring its impact and the danger of losing human control. Maybe a topic for the future volume of this book is to discuss the further details of AI security, e.g., “the need of a brake should unexpected danger arise.”

In summary, this book provides a good starting point for us all to think about AI security and development. The emphasis is on the governance of AI. It should be an interesting read for anyone who cares about AI and its future. In addition, it can serve as a good reference book for social scientists in the fields of economics, history, political science, sociology, and technology studies.

Yongsheng Wang*, Ph.D.
Former Pi Gamma Mu Board Member
Pittsburgh, Pennsylvania

*The opinion in this review is solely of the author's and does not represent any institution.