DOI: 10.62461/TMY210223

BOOK REVIEWS

Julian Nida-Rümelin and Nathalie Weidenfeld. *Digital Humanism For a Humane Transformation of Democracy, Economy and Culture in the Digital Age*. Cham: Springer, 2021. 129 pp. ISBN 978-3-031-12482-2 (eBook).

Artificial intelligence (AI), or machine intelligence, as we understand it today has been around since the invention of the programmable digital computer. In public discourse, AI has been a subject of great fascination and fear, and this has notably been expressed through numerous science fiction films depicting the dangers of an advanced and malicious AI. Interest surrounding the topic of AI has accelerated in the past two years as AI tools (generative AI, AI art, etc.) became widely available for public use. This surge in interest has led to very active social and academic discourse surrounding the ethics of AI development and use. It is in this space where Julian Nida-Rümelin and Nathalie Weidenfeld's book *Digital Humanism for a Humane Transformation of Democracy, Economy and Culture in the Digital Age* makes a timely contribution.

The book explores the philosophical questions surrounding AI and explores the prevalent tendency to assign human-like qualities when describing AI. The authors associate this tendency to what they describe as the "Silicon Valley ideology" of AI, where development of machine intelligence is measured and marked by its unbridled quest to develop AI to the point of little distinction between computer processes and human behaviour. The authors sharply critique this approach to understanding AI, stating that such discourse of AI ultimately leads to a "technicist utopia in which the human is left behind" (p. 4). In simpler terms, the book makes the argument that current narratives of AI, where it is sometimes seen as "the answer to all our economic, social, and even spiritual problems" (p. 122), eventually leads to human obsoletion.

As an alternative, the concept of "digital humanism" is proffered as an alternative approach to thinking about the relationship between humans and AI. The authors argue for a clear distinction between human traits and the affordances provided by digital technologies, or as they

put it "digital humanism argues for an instrumental attitude towards digitalization: what can be economically, socially, and culturally beneficial, and where do potential dangers lurk?" (p.122). The authors systematically address several philosophical enquiries about assigning human characteristics to AI – digital self-determinism and free will, digital slavery, digital utilitarianism, and more, and highlights how human-like functions of AI programmes do not make it human.

Methodologically, these arguments are creatively presented through the extensive use of science-fiction films that feature AI significantly, such as *I, Robot, AI: Artificial Intelligence, The Matrix*, and more. Each chapter of the book begins with a brief description of a key scene from one of these films, followed by a textual analysis of the different features of AI to generate discussion. For example, a key scene depicting the spectacle of destroying seemingly self-aware robots in *AI: Artificial Intelligence* is used to discuss the notion of assigning "human dignity" to robots, with the writers cautioning that "anyone who thinks that there can be no categorical difference between human brains and computers is denying the foundations not only of scientific practice but of the human way of life in general" (p.10).

This approach is effective and insightful for two reasons. Firstly, films – or art, in general, capture the sentiment, expression, or understanding of human culture of the time. The use of filmic narratives of AI situates the discussion in an accessible medium of understanding. Through describing the layers of interpretations of these films' representations of AI, readers are guided to an nuanced view of human-computer relations. A particularly interesting demonstration of this is the exploration of friendship and morality in the films *I, Robot* and *Ex Machina*. The authors highlight how these films, on the surface, demonstrate the fantasy that robots are able to become sufficiently aware of the moral conditions to become "friends" with humans.

Beyond the sentimentality offered in these scenes, the authors note that "AIs do not act according to their own reasons. They have no feelings, no moral sense, no intentions, and they cannot attribute these to other persons. Without these abilities, however, proper moral practice is not possible" (p. 42). AI "friendship" is ultimately rooted in their

programming rather than morality, and as demonstrated in *Ex Machina,* trust in the robot's ability to function morally has fatal consequences. This nuance is also seen in the exposition of *AI: Artificial Intelligence*. Using a scene where discarded robots are "tortured" and "destroyed" in a circus arena as a spectacle for a human audience, the authors argue that while it is easy to interpret the scenes literally as representations of robot dignity, it is more likely that it is a commentary of racism rather than a "serious assessment of the status of robots" (p.11).

Overall, this book is an enlightening intersection of contemporary issues related to AI, philosophy, and film analysis. This is a book that uses analysis of science-fiction films to bring to life the philosophical questions about AI. The "Silicon Valley ideology" of AI is sharply criticized on a general level, though a deeper analysis of these accusations with specific examples and case studies would be welcome. This would possibly allow for a clearer contrast between what is offered by tech giants and digital humanism as proposed by the authors. That being said, I recognize that this is outside the purview of this book. In its current form, this is a book that is very useful in introducing readers to the various debates surrounding digital technology and AI, and showing us why these issues are relevant in the present time.

Tan Meng Yoe
School of Arts and Social Sciences
Monash University, Malaysia