Bernd Carsten Stahl, *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies* (Switzerland: Springer Nature, 2021).

Reviewed by Maheen Shafeeq Researcher, Centre for Aerospace & Security Studies, Islamabad, Pakistan

Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies is authored by Professor Bernd Carsten Stahl who is serving as Director of the Centre for Computing and Social Responsibility at De Montfort University, United Kingdom. It was written based on the work undertaken in the SHERPA (Securing a Hybrid Environment for Research Preservation and Access) project.

As the title elaborates, the book offers parallels between Artificial Intelligence (AI) and a biological ecosystem. This conceptualisation hinges on the understanding that AI uses various types of algorithms such as machine learning (narrow AI), general AI (strong AI) and sociotechnical AI systems. For this reason, AI algorithms cannot be clubbed under one generic term rather AI and its distinct applications form a complete ecosystem with different functions and operations. The author uses the term 'ecosystem' as a metaphor to elaborate this concept. The central theme of this

book dictates that AI ethics should be employed in a manner that promotes ethical human flourishing. Professor Stahl draws the idea of human flourishing from Aristotelian roots whose key tenets dictate that human flourishing is central to ethics. He argues that the aim of 'human flourishing' is to promote emancipation.

What is most interesting is that the ethical theories explained in the book (consequentialism/utilitarianism-outcome, deontology-motivation, and virtue-ought to do) hold great value. Every ethical theory pulls in a manner that one may end up aligning oneself with the theory. Unfortunately, the theories of ethics mentioned in the book, if put into application, may end up justifying acts that seem unethical. If, for example, the 'motivation' (under the deontology theory) behind an unethical act was right; the unethical act can become ethical. Similarly, some ethical theories may end up justifying the use of autonomous weapons on the battlefield. For instance, by stating that the 'intention' (under the virtue theory) was to protect the soldiers from volatile situations. Professor Stahl also realises these limitations and proposes a method to promote ethics. He proposes that whatever act AI applications perform, they must promote 'human flourishing.' He defines use of AI for human flourishing by using a quote by Virginia Dignum (professor in social and ethical AI, and Wallenberg Chair on Responsible Artificial Intelligence), 'Responsible Artificial Intelligence is about human responsibility for the development of intelligent systems along fundamental human principles and values, to ensure human flourishing and well-being in a sustainable world' (p. 25). From this, the author extracted that 'the explicit aim to do the ethically right thing with AI can be described with reference to human flourishing' (p. 25).

Furthermore, Professor Stahl attempts to analyse the ethical issues that stem from AI and draws analysis from the SHERPA project case study. According to this study, the most immediate ethical issues can arise from machine learning, and he specifically indicates that privacy and data protection, reliability, transparency, and safety need immediate attention (p. 49). He also mentions that

there are ethical issues of living in a digital world such as issues related to economy, justice and fairness, freedom, lack of human contact, loss of human decision-making and so on. However, he feels that these require only 'medium attention.' He also provides an interesting list of metaphysical issues, such as what would happen if machines displayed human reasoning abilities (p. 46). The author does not seem too worried about such metaphysical issues as he is convinced that such a form of AI does not exist and states that even if research is being conducted in this domain, it is slow. He believes that matters such as job loss due to AI and discrimination are more important worries as compared to conscious machines (p. 46).

The book offers three options in which ethical issues of AI can be mitigated - policy level, organisational level and through guidance mechanisms. The options offer a wealth of ideas about policy initiatives, regulations, regulators, governance, strategic initiatives, frameworks, tools and methodologies that can be adopted to address ethical AI issues. Professor Stahl also points to stakeholders who could play a special role in these areas. However, he highlights that those various stakeholders have different interests. For example, he mentions that 'there are corporate end users of AI, but these tend to have different interests and motivations from individual end users' (p. 73). Due to conflicting interests, he suggests that there is a need to look at various aspects with a 'higher-level view' (p. 74).

The 'higher-level view', that the author recommends, is to develop ethics for an AI ecosystem. In an AI ecosystem, however, the relation between ethics and the ecosystem is not straightforward (p. 87) as it inculcates a 'system of systems of systems' (p. 92). The author explains that interventions designed to address ethical challenges of AI can be developed in a multitude of ways (p. 96) and recommends that in order to develop ethical AI, there is a need to clearly define boundaries around the AI ecosystem and learn and evolve the concept. For example, the governance of an AI ecosystem should be sensitive to motivations and incentives of the

members of such an ecosystem. This needs to be carefully balanced with possible as well as expected benefits and downsides as only then can developments in AI technology promote human flourishing. Professor Stahl confesses that implementation of ethics in an AI ecosystem that promotes human flourishing would require more thought and research.

The best part of the book is that it delves into the concept of AI from theories to their practical implications. Moreover, it cites a number of renowned authors working in this area. Unfortunately, while the author talks about different forms of AI, he does not dig into specific ethical AI issues. Rather, he offers a generic overview such as what ethical issues may arise in the field of business, military, or education. Hence, *Artificial Intelligence for a Better Future* would be useful for those who have some prior knowledge of AI and the debate that surrounds it to help them better understand the author's point of view. In comparison to other books on the subject, this one offers a unique perspective on the nexus between ethics and AI.