Education for Information 40 (2024) 229–231
DOI 10.3233/EFI-240033
IOS Press

Book Review

The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities. Floridi, Luciano (2023). Oxford University Press. 272pp. ISBN 9780198883098

Artificial intelligence (AI) has become a larger and larger focus of scholarly discussion across the information field in recent years. Work on AI dates back to the 1940s and has gone through multiple cycles of hype and wintering over the years. We are currently in a period of hype, but this time may be different.

Advances in machine learning (ML) over the past decade have made AI a household term, and AI has been incorporated into everyday products used by millions, from mapping applications to social media and email – to say nothing of the recent efflorescence of generative AI products such as ChatGPT. While some forecast another impending winter, many expect the latest trends in AI technology to produce changes – if not wholesale revolution – across society.

It is in this context that the philosopher Luciano Floridi has published the latest installment in his Principia Philosophiae Informationis, *The Ethics of Artificial Intelligence*. This is the fourth volume in the series that began with *The Philosophy of Information* (Oxford University Press, 2011).

The Ethics of Artificial Intelligence is a work of digital ethics, which Floridi defines as "the branch of ethics that studies and evaluates moral problems relating to *data and information*, including generation, recording, curation, processing, dissemination, sharing, and use; *algorithms*, including AI, [artificial agents], ML and robots; and corresponding *practices and infrastructures*, including responsible innovation, programming, hacking, codes of ethics, standards" (p. 81, emphasis original). This definition makes obvious the relevance of this work for researchers, educators and practitioners in information studies, which is also concerned with many of these issues.

Overall, the book provides an accessible overview of conceptual issues regarding AI as it finds application across society. The first part of the book paints the conceptual backdrop, and these are the most heavily philosophical chapters of the book; then the second part, the major substance of the book, discusses various ethical issues regarding AI, from the risks of AI ethics itself (ch. 5), to crimes that AI enables (ch. 8), to how AI can be used to aid in sustainable development (ch. 12).

In my view, the early part of the book is also the strongest. There, Floridi conceptualizes the "cleaving power" of the digital (ch. 1): how digital technologies couple things that were separate in the pre-digital environment (e.g., identity and personal data) and decouple ones that were previously together (e.g., presence and location). Floridi makes the point that AI is not actually a form of intelligence (despite its name), but rather a new form of agency (ch. 2). Finally in this section, he explains that a key frontier in AI is synthetic data (ch. 3); certain systems can generate their own data

to teach themselves. Chess is a prime example, but this technique has many further applications.

In the second part of the book, Floridi discusses AI ethics head-on, presenting numerous frameworks for understanding and intervening in this space. As Floridi writes, AI introduces numerous ethical issues and risks (ch. 7): inconclusive, inscrutable or misguided evidence may lead to unjustified actions, opacity or bias; unfair outcomes may lead to discrimination; and transformative effects may challenge human autonomy and informational privacy. AI also introduces a greater capacity for crime, particularly in the realms of finance, smuggling, personal offenses, and fraud (ch. 8).

In response to these risks, ethicists have been working to establish key ethical principles for AI; Floridi's favored framework (ch. 4) draws on the established principles of bioethics, with a final addition specific to AI: beneficence (doing good), nonmaleficence (avoiding harm), autonomy (enabling human agency), justice (ensuring fairness) and explicability (ensuring accountability). But even trying to do AI ethics introduces risks and loopholes for exploitation, such as "ethics washing," which must be guarded against (ch. 5).

On the positive side, Floridi is convinced that AI has clear possibilities for good, particularly under the monikers of AI4SG ("AI for Social Good," ch. 9) and AIxSDG ("AI and the UN Sustainable Development Goals," ch. 11–12). In particular, Floridi foresees a future of AI helping us navigate the climate crisis, assuming policymakers and voters can keep his recommendations in mind (ch. 11, 13). Floridi calls this "a marriage of green and blue" (p. 201) – green for the environment, blue for computing.

Helping us visualize this future, in one of the final chapters (ch. 10), Floridi illustrates some characteristics of a society that has integrated AI well:

- Enabling human self-realization without devaluing human abilities
- Enhancing human agency without removing human responsibility
- Increasing societal capabilities without reducing human control
- Cultivating societal cohesion without eroding human self-determination

The book is welcome and well worth reading. Several chapters would make excellent course readings for students in programs across the information field.

However, the book has some limitations. Most of the latter chapters are formatted as "listicles" that would benefit from some further analysis; in particular, they do not quite feel philosophical, an oddity for a philosophy book. Some of the chapters are a bit outdated; the book was published in late 2023, but several chapters (particularly those presenting literature reviews) are reprinted from papers published before 2020. In some cases Floridi claims that the work has been "updated" for this book, but it is not clear if or how that is really the case. Relatedly, the organization of the final five chapters is haphazard; the content feels out of order and repetitive. For example, the topics of AI4SG and AIxSDG overlap substantially, but they are treated (without explanation) in separate chapters.

While the book could have been better, it is still a vital contribution as we continue to move forward in the digital age, now with AI as – a tool? a partner? a nemesis?

– and as we continue to navigate the possibilities for humans to live good lives and understand what that means. As Floridi writes, "we are mere analogue organisms trying to adapt" in a new, onlife environment that blends the analogue and digital (p. 204).

The next, and perhaps final, volume of Floridi's magnum opus is slated to be *The Politics of Information*, which I expect will also be of great interest to those of us in the information field. Fortunately for the impatient among us, Floridi has already been publishing articles that will likely compose parts of that book.

Tim Gorichanaz
Drexel University
E-mail: gorichanaz@drexel.edu