Towards a Code of Ethics for Artificial Intelligence

By Paula Boddington

Springer International Publishing AG; 124 pages

Laurens Naudts*

As indicated by the inclusion of the word ‘towards’ in the title of this book, author Paula Boddington does not intend to present the reader with the definitive code of ethics for artificial intelligence. Rather she aims to provide a roadmap on how and under what conditions a code of ethics for AI could take form. Both short and practical in nature, the author offers a concise guide detailing the value of, and need for, ethics and its codification within the AI environment, whilst also highlighting the particular pitfalls and challenges that might impede their formulation.

Boddington’s work is pragmatic: it serves towards facilitating the creation of an end-result, i.e. a code of ethics. Hence, it should provide methods to make a difficult procedural enterprise more workable in practice. It should moreover provide the necessary tools for non-philosophers to make more value-sensitive assessments when dealing with AI. Considering the depth of both ethics and AI, a clear and comprehensive structure is needed in order to enable the reader to understand and reflect upon the breadth of issues explored. In this regard, the author rightfully notes¹ that the process of drafting a code not only requires an understanding of what ethical questions are,² but also what the purpose of codes and

regulations might be.³ Furthermore, in order for the latter to be workable and efficient, an understanding should also be created concerning who will be involved in the production, critique and application of the code in question. In addition, one should remain mindful of the contextual factors that shape the construction of both ethics and AI.⁴ Through eight chapters, the author, with success, sets out to clarify the entire life cycle of code development, and the associated questions that might arise. The chapters follow a natural progression, and even though a wide range of topics and considerations are covered, the reader is able to maintain oversight and comprehension. Particularly relevant is the comment, ‘what this means for AI’, interspersed throughout the text in which the sometimes abstract ethical principles or challenges are applied in an effort to illustrate their specific relevance for AI. Practical examples or anecdotes are highlighted, which increases the understandability and readability of the subject matter.

One could question Boddington’s ambition to tackle AI as a whole. As the author notes, though omnipresent in current day debates, the term artificial intelligence is complex and not easily defined.⁵ Moreover, AI covers a broad range of applications, which each raise their own set of particular challenges. Given this disparity, how then do we establish the building blocks towards a code of ethics for artificial intelligence? Boddington acknowledges that within the debate, a focus on distinctive characteristics of AI might be needed.⁶ In addition, she notes that, where principles often tend to favor AI in general, ‘the value issues we need to consider are much more local and contextualised.’⁷ If the AI debate could benefit from contextualisation, then perhaps instructions concerning the creation of a code of ethics should also follow a contextualized approach? Yet, even when considered in broad terms, AI does raise particular, overarching challenges. For example, Boddington challenges the focus on ‘intelligence’ as the hallmark of humanity.⁸ Admittedly, some challenges related to AI are shared with other rapidly evolving technologies, both past, present and future, such as the impact of ‘hype’ on ethical reasoning.⁹ Furthermore, ethical confrontations in AI might not be all that new.¹⁰ As the author notes: ‘The extension, enhancement and replacement of human agency and reasoning in AI serve as the loci of many of the ethical issues that arise in its use, sometimes

DOI: 10.21552/delphi/2019/2/11

* Laurens Naudts, Doctoral Researcher at the KU Leuven Centre for IT and IP Law, Belgium. Laurens’ research interests focus on the interrelationship between artificial intelligence, ethics, justice, fairness and the law. His Ph.D. research reconsiders the concepts of equality and data protection within the context of machine learning and algorithmically guided decision-making. For correspondence: <laurens.naudts@kuleuven.be>

1 Paula Boddington, *Towards a Code of Ethics for Artificial Intelligence* (Springer 2017) Preface.

2 *ibid* chs 1 - 3.

3 *ibid* ch 4.

4 *ibid* chs 5 - 7.

5 *ibid* 1.

6 See *inter alia* *ibid* 91-92, 111.

7 *ibid* 111.

8 *ibid* 86.

9 *ibid* 31-35.

10 *ibid* 86.

presenting us with vivid versions of old questions.¹¹ Of course, Boddington also provides sufficient attention to the challenges AI raises for professional codes of ethics in particular, including the difficulties of developing a code of ethics amidst fast technological changes.¹² Even though the questions raised, might often seem general, they nonetheless are pertinent.

The final chapter can be considered both a synthesis of previous chapters, and a training exercise for the reader. Looking to the future, Boddington elaborates upon, and critiques, the Asilomar AI principles, a tentative approach towards governing AI research drafted by the Future of Life Institute.¹³ Here, Boddington's commentary on a set of existing principles allows the reader to follow an applied critique of the lessons learned throughout the book, whereas the reader, at this point in time, might have gained sufficient knowledge to insert their own vision on the principles.¹⁴

Towards a Code of Ethics arose from a project focused on professional codes of ethics for artificial intelligence researchers.¹⁵ However, the book never fully explicates who the artificial intelligence researcher might be. This makes it difficult to discern in what field, context, or setting, a professional code of ethics for AI research would actually apply. Nevertheless, Boddington argues that 'given the turbulent landscape in which AI is developing, professional codes

of ethics will need the backing and support of other formalised or less formal, and institutionalised ways of addressing the ethical questions confronting us.'¹⁶ Hence, she aims for a wide field of application, which could serve both research and industry. Likewise, the broad, and perhaps general, nature of the book, expands the target audience. *Towards a Code of Ethics* not only provides insight to those tasked in drafting an actual code of ethics, it could equally be of interest for researchers dealing in AI, such as legal scholars or social scientists. Boddington has delivered a balanced product, stressing the importance of foundational underpinnings, whilst at the same time not losing track of their practical implementation. In a time where institutes and corporations are likely thinking about implementing a code of ethics for their AI research activities, her work presents a roadmap, as well as a compendium of elements to consider in AI research.

11 *ibid* 29.

12 *ibid* 67-79.

13 It should be noted that Boddington's work was funded by the Future of Life Institute. This could explain the preference given to these principles over others that currently exist.

14 Paula Boddington, *Towards a Code of Ethics for Artificial Intelligence* (Springer 2017) 104-110.

15 *ibid* xiii.

16 *ibid* 40.