

SpringerBriefs in Ethics

Springer Briefs in Ethics envisions a series of short publications in areas such as business ethics, bioethics, science and engineering ethics, food and agricultural ethics, environmental ethics, human rights and the like. The intention is to present concise summaries of cutting-edge research and practical applications across a wide spectrum.

Springer Briefs in Ethics are seen as complementing monographs and journal articles with compact volumes of 50 to 125 pages, covering a wide range of content from professional to academic. Typical topics might include:

- Timely reports on state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- In-depth case studies or clinical examples
- Presentations of core concepts that students must understand in order to make independent contributions

More information about this series at <http://www.springer.com/series/10184>

Christoph Bartneck · Christoph Lütge ·
Alan Wagner · Sean Welsh

An Introduction to Ethics in Robotics and AI

 Springer

Christoph Bartneck
HIT Lab NZ
University of Canterbury
Christchurch, New Zealand

Christoph Lütge
Institute for Ethics in Artificial Intelligence
Technical University of Munich
München, Germany

Alan Wagner
College of Engineering
Pennsylvania State University
University Park, PA, USA

Sean Welsh
Department of Philosophy
University of Canterbury
Christchurch, New Zealand



ISSN 2211-8101

ISSN 2211-811X (electronic)

SpringerBriefs in Ethics

ISBN 978-3-030-51109-8

ISBN 978-3-030-51110-4 (eBook)

<https://doi.org/10.1007/978-3-030-51110-4>

© The Editor(s) (if applicable) and The Author(s) 2021. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland



Fig. 1 The logo of the EPIC project

This book was made possible through the European Project “Europe’s ICT Innovation Partnership With Australia, Singapore & New Zealand (EPIC)” under the European Commission grant agreement Nr 687794. The project partners in this consortium are:

- eutema GmbH
- Intersect Australia Limited (INTERSECT)
- Royal Melbourne Institute Of Technology (RMIT)
- Callaghan Innovation Research Limited (CAL)
- University Of Canterbury (UOC)
- National University Of Singapore (NUS)
- Institute For Infocomm Research (i2r)

From February 2–6, 2019 we gathered at the National University of Singapore. Under the guidance of Laia Ros from Book Sprints we wrote this book in an atmosphere of mutual respect and with great enthusiasm for our shared passion: artificial intelligence and ethics. We have backgrounds in different disciplines and the synthesis of our knowledge enabled us to cover the wide spectrum of topics relevant to AI and ethics.

This book was written using the BookSprint method (<http://www.booksprints.net>).

Contents

1	About the Book	1
1.1	Authors	1
1.2	Structure of the Book	2
2	What Is AI?	5
2.1	Introduction to AI	7
2.1.1	The Turing Test	9
2.1.2	Strong and Weak AI	10
2.1.3	Types of AI Systems	10
2.2	What Is Machine Learning?	11
2.3	What Is a Robot?	12
2.3.1	Sense-Plan-Act	12
2.3.2	System Integration. Necessary but Difficult	13
2.4	What Is Hard for AI	13
2.5	Science and Fiction of AI	14
3	What Is Ethics?	17
3.1	Descriptive Ethics	17
3.2	Normative Ethics	19
3.2.1	Deontological Ethics	19
3.2.2	Consequentialist Ethics	20
3.2.3	Virtue Ethics	20
3.3	Meta-ethics	20
3.4	Applied Ethics	21
3.5	Relationship Between Ethics and Law	22
3.6	Machine Ethics	22
3.6.1	Machine Ethics Examples	23
3.6.2	Moral Diversity and Testing	25

4	Trust and Fairness in AI Systems	27
4.1	User Acceptance and Trust	28
4.2	Functional Elements of Trust	28
4.3	Ethical Principles for Trustworthy and Fair AI	28
4.3.1	Non-maleficence	29
4.3.2	Beneficence	30
4.3.3	Autonomy	30
4.3.4	Justice	33
4.3.5	Explicability	35
4.4	Conclusion	37
5	Responsibility and Liability in the Case of AI Systems	39
5.1	Example 1: Crash of an Autonomous Vehicle	39
5.2	Example 2: Mistargeting by an Autonomous Weapon	40
5.2.1	Attribution of Responsibility and Liability	41
5.2.2	Moral Responsibility Versus Liability	41
5.3	Strict Liability	42
5.4	Complex Liability: The Problem of Many Hands	43
5.5	Consequences of Liability: Sanctions	43
6	Risks in the Business of AI	45
6.1	General Business Risks	46
6.1.1	Functional Risk	46
6.1.2	Systemic Risk	47
6.1.3	Risk of Fraud	47
6.1.4	Safety Risk	47
6.2	Ethical Risks of AI	48
6.2.1	Reputational Risk	48
6.2.2	Legal Risk	48
6.2.3	Environmental Risk	48
6.2.4	Social Risk	49
6.3	Managing Risk of AI	49
6.4	Business Ethics for AI Companies	50
6.5	Risks of AI to Workers	51
7	Psychological Aspects of AI	55
7.1	Problems of Anthropomorphisation	55
7.1.1	Misplaced Feelings Towards AI	56
7.1.2	Misplaced Trust in AI	57
7.2	Persuasive AI	57
7.3	Unidirectional Emotional Bonding with AI	58
8	Privacy Issues of AI	61
8.1	What Is Privacy?	61
8.2	Why AI Needs Data	62

- 8.3 Private Data Collection and Its Dangers 63
 - 8.3.1 Persistence Surveillance 64
 - 8.3.2 Usage of Private Data for Non-intended Purposes 67
 - 8.3.3 Auto Insurance Discrimination 69
 - 8.3.4 The Chinese Social Credit System 69
- 8.4 Future Perspectives 69
- 9 Application Areas of AI 71**
 - 9.1 Ethical Issues Related to AI Enhancement 71
 - 9.1.1 Restoration Versus Enhancement 71
 - 9.1.2 Enhancement for the Purpose of Competition 72
 - 9.2 Ethical Issues Related to Robots and Healthcare 73
 - 9.3 Robots and Telemedicine 73
 - 9.3.1 Older Adults and Social Isolation 73
 - 9.3.2 Nudging 74
 - 9.3.3 Psychological Care 75
 - 9.3.4 Exoskeletons 76
 - 9.3.5 Quality of Care 76
 - 9.4 Education 76
 - 9.4.1 AI in Educational Administrative Support 76
 - 9.4.2 Teaching 77
 - 9.4.3 Forecasting Students’ Performance 78
 - 9.5 Sex Robots 78
- 10 Autonomous Vehicles 83**
 - 10.1 Levels of Autonomous Driving 83
 - 10.2 Current Situation 84
 - 10.3 Ethical Benefits of AVs 84
 - 10.4 Accidents with AVs 85
 - 10.5 Ethical Guidelines for AVs 86
 - 10.6 Ethical Questions in AVs 86
 - 10.6.1 Accountability and Liability 87
 - 10.6.2 Situations of Unavoidable Accidents 87
 - 10.6.3 Privacy Issues 89
 - 10.6.4 Security 89
 - 10.6.5 Appropriate Design of Human-Machine Interface 90
 - 10.6.6 Machine Learning 90
 - 10.6.7 Manually Overruling the System? 90
 - 10.6.8 Possible Ethical Questions in Future Scenarios 90

11	Military Uses of AI	93
11.1	Definitions	94
11.2	The Use of Autonomous Weapons Systems	95
11.2.1	Discrimination	95
11.2.2	Proportionality	96
11.2.3	Responsibility	96
11.3	Regulations Governing an AWS	97
11.4	Ethical Arguments for and Against AI for Military Purposes.....	97
11.4.1	Arguments in Favour	97
11.4.2	Arguments Against.....	98
11.5	Conclusion	98
12	Ethics in AI and Robotics: A Strategic Challenge	101
12.1	The Role of Ethics	103
12.2	International Cooperation	103
	References	105
	Index	115

List of Figures

Fig. 2.1	Siri’s response to a not so uncommon question	6
Fig. 2.2	Alan Turing (1912–1954) (<i>Source</i> Jon Callas)	9
Fig. 3.1	Immanuel Kant (1724–1804) (<i>Source</i> Johann Gottlieb Becker)	18
Fig. 3.2	Plato (<i>Source</i> Richard Mortel)	21
Fig. 3.3	The Sophia robot (<i>Source</i> Hanson Robotics)	23
Fig. 4.1	Justitia (<i>Source</i> Waugsberg)	33
Fig. 6.1	Der Spiegel covers on mass unemployment	52
Fig. 7.1	Robot guiding people out of a building	57
Fig. 8.1	Amazon Echo Plus uses Alexa (<i>Source</i> Amazon)	64
Fig. 8.2	Hello Barbie (<i>Source</i> Mattel)	66
Fig. 8.3	Former US president Barack Obama was used to showcase the power of deep fakes (<i>Source</i> BuzzFeedVideo)	68
Fig. 9.1	da Vinci surgical system (<i>Source</i> Cmglee)	74
Fig. 9.2	Paro robot (<i>Source</i> National Institute of Advanced Industrial Science and Technology)	75
Fig. 9.3	A real doll (<i>Source</i> real doll)	79
Fig. 10.1	Waymo’s fully self-driving Chrysler Pacifica Hybrid minivan on public roads (<i>Source</i> Waymo)	85
Fig. 10.2	Example question from the Moral Machine experiment that confronted people with trolley problems (<i>Source</i> MIT)	88
Fig. 11.1	MIM-104 Patriot (<i>Source</i> Darkone)	94