

## THE CAMBRIDGE HANDBOOK OF RESPONSIBLE ARTIFICIAL INTELLIGENCE

In the past decade, artificial intelligence (AI) has become a disruptive force around the world, offering enormous potential for innovation but also creating hazards and risks for individuals and the societies in which they live. This volume addresses the most pressing philosophical, ethical, legal, and societal challenges posed by AI. Contributors from different disciplines and sectors explore the foundational and normative aspects of responsible AI and provide a basis for a transdisciplinary approach to responsible AI. This work, which is designed to foster future discussions to develop proportional approaches to AI governance, will enable scholars, scientists, and other actors to identify normative frameworks for AI to allow societies, states, and the international community to unlock the potential for responsible innovation in this critical field. This book is also available as Open Access on Cambridge Core.

SILJA VOENEKY is a leading scholar in the field of the interdependence of ethics and public international law, and the governance of emerging technologies. She is Professor of Public International Law, and Comparative Law at the University of Freiburg and was a Fellow at Harvard Law School. She previously served as a member of the German Ethics Council. Since 2001, Voeneke has been, inter alia, a legal advisor to the German Federal Foreign Office and the German Federal Ministry of Environment.

PHILIPP KELLMAYER is a neurologist and neuroscientist at the Medical Center – University of Freiburg. In his neuroscientific work, he investigates language processing in the brain and the clinical application of brain–computer interfaces, virtual reality, and other digital technologies. Through his contributions to neuroethics, he has become an international expert on ethical aspects of neuroscience, neurotechnology, and AI.

OLIVER MUELLER is Professor of Philosophy with a focus on technology and on contemporary philosophy, University of Freiburg, and a Senior Fellow at the 2018–2021 FRIAS Saltus Research Group ‘Responsible AI’.

WOLFRAM BURGARD is a leading researcher in Robotics and Artificial Intelligence. He co-authored more than 400 publications including the famous book ‘Probabilistic Robotics’. He received numerous awards including the Gottfried Wilhelm Leibniz Prize, the most prestigious German research award, and is Fellow of several societies.

Cambridge University Press & Assessment

978-1-009-20786-7 — The Cambridge Handbook of Responsible Artificial Intelligence

Edited by Silja Voeneke , Philipp Kellmeyer , Oliver Mueller , Wolfram Burgard

Frontmatter

[More Information](#)

---

# The Cambridge Handbook of Responsible Artificial Intelligence

INTERDISCIPLINARY PERSPECTIVES

Edited by

**SILJA VOENEKY**

University of Freiburg

**PHILIPP KELLMEYER**

Medical Center – University of Freiburg

**OLIVER MUELLER**

University of Freiburg

**WOLFRAM BURGARD**

University of Technology Nuremberg



**CAMBRIDGE**  
UNIVERSITY PRESS

Cambridge University Press & Assessment

978-1-009-20786-7 — The Cambridge Handbook of Responsible Artificial Intelligence

Edited by Silja Voeneke, Philipp Kellmeyer, Oliver Mueller, Wolfram Burgard

Frontmatter

[More Information](#)



Shaftesbury Road, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781009207867](http://www.cambridge.org/9781009207867)

DOI: 10.1017/9781009207898

© Cambridge University Press & Assessment 2022

This work is in copyright. It is subject to statutory exceptions and to the provisions of relevant licensing agreements; with the exception of the Creative Commons version the link for which is provided below, no reproduction of any part of this work may take place without the written permission of Cambridge University Press.

An online version of this work is published at [doi.org/10.1017/9781009207898](https://doi.org/10.1017/9781009207898) under a Creative Commons Open Access license CC-BY-NC-ND 4.0 which permits re-use, distribution and reproduction in any medium for non-commercial purposes providing appropriate credit to the original work is given. You may not distribute derivative works without permission. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0>

All versions of this work may contain content reproduced under license from third parties.

Permission to reproduce this third-party content must be obtained from these third-parties directly.

When citing this work, please include a reference to the DOI 10.1017/9781009207898

First published 2022

*A catalogue record for this publication is available from the British Library.*

*Library of Congress Cataloging-in-Publication Data*

NAMES: Voeneke, Silja, 1969– editor. | Kellmeyer, Philipp, 1979– editor. | Mueller, Oliver, 1972– editor. | Burgard, Wolfram, editor.

TITLE: The Cambridge handbook of responsible artificial intelligence : interdisciplinary perspectives / edited by Silja Voeneke, University of Freiburg; Philipp Kellmeyer, Medical Center – University of Freiburg; Oliver Mueller, University of Freiburg; Wolfram Burgard, University of Technology Nuremberg.

DESCRIPTION: Cambridge, United Kingdom; New York, NY: Cambridge University Press, 2022. |

Series: Cambridge law handbooks | Includes index.

IDENTIFIERS: LCCN 2022022867 (print) | LCCN 2022022868 (ebook) | ISBN 9781009207867 (hardback) |

ISBN 9781009207898 (epub)

SUBJECTS: LCSH: Artificial intelligence–Law and legislation. | Artificial intelligence–Social aspects. | BISAC: LAW / General

CLASSIFICATION: LCC K564.C6 C3595 2022 (print) | LCC K564.C6 (ebook) | DDC 343.09/99–dc23/eng/20220630

LC record available at <https://lccn.loc.gov/2022022867>

LC ebook record available at <https://lccn.loc.gov/2022022868>

ISBN 978-1-009-20786-7 Hardback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

## Contents

<i>List of Figures</i>	<i>page</i> ix
<i>List of Contributors</i>	xi
<i>Acknowledgements</i>	xix
<b>Introduction</b>	1
Silja Voeneke, Philipp Kellmeyer, Oliver Mueller, and Wolfram Burgard	
<b>PART I FOUNDATIONS OF RESPONSIBLE AI</b>	
<b>1 Artificial Intelligence: Key Technologies and Opportunities</b>	11
Wolfram Burgard	
<b>2 Automating Supervision of AI Delegates</b>	19
Jaan Tallinn and Richard Ngo	
<b>3 Artificial Moral Agents: Conceptual Issues and Ethical Controversy</b>	31
Catrin Misselhorn	
<b>4 Risk Imposition by Artificial Agents: The Moral Proxy Problem</b>	50
Johanna Thoma	
<b>5 Artificial Intelligence and Its Integration into the Human Lifeworld</b>	67
Christoph Durt	
<b>PART II CURRENT AND FUTURE APPROACHES TO AI GOVERNANCE</b>	
<b>6 Artificial Intelligence and the Past, Present, and Future of Democracy</b>	85
Mathias Risse	
<b>7 The New Regulation of the European Union on Artificial Intelligence: Fuzzy Ethics Diffuse into Domestic Law and Sideline International Law</b>	104
Thomas Burri	

8	<b>Fostering the Common Good: An Adaptive Approach Regulating High-Risk AI-Driven Products and Services</b>	123
	Thorsten Schmidt and Silja Voeneke	
9	<b>China's Normative Systems for Responsible AI: From Soft Law to Hard Law</b>	150
	Weixing Shen and Yun Liu	
10	<b>Towards a Global Artificial Intelligence Charter</b>	167
	Thomas Metzinger	
11	<b>Intellectual Debt: With Great Power Comes Great Ignorance</b>	176
	Jonathan Zittrain	
<b>PART III RESPONSIBLE AI LIABILITY SCHEMES</b>		
12	<b>Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks</b>	187
	Christiane Wendehorst	
13	<b>Forward to the Past: A Critical Evaluation of the European Approach to Artificial Intelligence in Private International Law</b>	210
	Jan von Hein	
<b>PART IV FAIRNESS AND NONDISCRIMINATION IN AI SYSTEMS</b>		
14	<b>Differences That Make a Difference: Computational Profiling and Fairness to Individuals</b>	229
	Wilfried Hinsch	
15	<b>Discriminatory AI and the Law: Legal Standards for Algorithmic Profiling</b>	252
	Antje von Ungem-Sternberg	
<b>PART V RESPONSIBLE DATA GOVERNANCE</b>		
16	<b>Artificial Intelligence and the Right to Data Protection</b>	281
	Ralf Poscher	
17	<b>Artificial Intelligence as a Challenge for Data Protection Law: And Vice Versa</b>	290
	Boris P. Paal	
18	<b>Data Governance and Trust: Lessons from South Korean Experiences Coping with COVID-19</b>	309
	Sangchul Park, Yong Lim, and Haksoo Ko	
<b>PART VI RESPONSIBLE CORPORATE GOVERNANCE OF AI SYSTEMS</b>		
19	<b>From Corporate Governance to Algorithm Governance: Artificial Intelligence as a Challenge for Corporations and Their Executives</b>	331
	Jan Lieder	

*Contents*

vii

- 20 **Autonomization and Antitrust: On the Construal of the Cartel Prohibition in the Light of Algorithmic Collusion** 347  
 Stefan Thomas
- 21 **Artificial Intelligence in Financial Services: New Risks and the Need for More Regulation?** 359  
 Matthias Paul
- PART VII RESPONSIBLE AI HEALTHCARE AND NEUROTECHNOLOGY GOVERNANCE**
- 22 **Medical AI: Key Elements at the International Level** 379  
 Fruzsina Molnár-Gábor and Johanne Giesecke
- 23 **“Hey Siri, How Am I Doing?”: Legal Challenges for Artificial Intelligence Alter Egos in Healthcare** 397  
 Christoph Krönke
- 24 **‘Neurorights’: A Human Rights–Based Approach for Governing Neurotechnologies** 412  
 Philipp Kellmeyer
- 25 **AI-Supported Brain–Computer Interfaces and the Emergence of ‘Cyberbilities’** 427  
 Boris Essmann and Oliver Mueller
- PART VIII RESPONSIBLE AI FOR SECURITY APPLICATIONS AND IN ARMED CONFLICT**
- 26 **Artificial Intelligence, Law, and National Security** 447  
 Ebrahim Afsah
- 27 **Morally Repugnant Weaponry? Ethical Responses to the Prospect of Autonomous Weapons** 475  
 Alex Leveringhaus
- 28 **On ‘Responsible AI’ in War: Exploring Preconditions for Respecting International Law in Armed Conflict** 488  
 Dustin A. Lewis

Cambridge University Press & Assessment

978-1-009-20786-7 — The Cambridge Handbook of Responsible Artificial Intelligence

Edited by Silja Voeneke , Philipp Kellmeyer , Oliver Mueller , Wolfram Burgard

Frontmatter

[More Information](#)

---



## Figures

5.1 The Turing Test	<i>page</i> 74
5.2 The fundamental relations between humans, AI, data, and the lifeworld	81
9.1 A development process of AI application	160
12.1 The ‘physical’ and the ‘social’ dimensions of risks associated with AI	189
14.1 Fairness-index for statistical profiling based on measures for the under- and over-inclusiveness of profiles	249
18.1 Daily newly confirmed COVID-19 cases	310
18.2 The COVID-19 Epidemic Investigation Support System	317
18.3 The KI-Pass, a QR code-based electronic visitor booking system	318
18.4 User interface of the Self-Quarantine App	319

Cambridge University Press & Assessment

978-1-009-20786-7 — The Cambridge Handbook of Responsible Artificial Intelligence

Edited by Silja Voeneke , Philipp Kellmeyer , Oliver Mueller , Wolfram Burgard

Frontmatter

[More Information](#)

---

## Contributors

**Ebrahim Afsah** is an associate professor for public international law at the University of Copenhagen (Denmark); he was a professor of Islamic law at the University of Vienna (Austria), and a visiting professor of international relations at the University of Brest (France). He was trained at the School of Oriental and African Studies, London, Trinity College Dublin, the Kennedy School of Government at Harvard University, and the Max Planck Institute for Comparative Public and International Law in Heidelberg (Germany). In recent years, *Ebrahim Afsah* has won fellowships to the European University Institute in Florence (Fernand Braudel), the Norwegian Academy of Science (Nordic Civil Wars), Harvard Law School (Islamic Legal Studies Program), and the National University of Singapore (Centre for Asian Legal Studies).

**Wolfram Burgard** is a professor of computer science at the University of Technology Nuremberg (UTN), and was a professor of computer science at the University of Freiburg (Germany), where he headed the research lab for Autonomous Intelligent Systems. Besides, he was a Senior Fellow at the 2018–2021 FRIAS Saltus Research Group ‘Responsible AI - Emerging ethical, legal, philosophical and social aspects of the interaction between humans and autonomous intelligent systems.’ His areas of interest lie in AI and mobile robots. His research focuses on the development of robust and adaptive techniques for state estimation and control. Over the past years, his group and he have developed a series of innovative probabilistic techniques for robot navigation and control. They cover different aspects such as localization, map-building, SLAM, path-planning, and exploration. *Wolfram Burgard* has published more than 400 papers and articles in robotic and AI conferences and journals. He is coauthor of two books, *Principles of Robot Motion: Theory, Algorithms, and Implementations* and *Probabilistic Robotics*. He is a Fellow of the European Association for Artificial Intelligence (EurAI), the Association for the Advancement of Artificial Intelligence (AAAI), and the Institute of Electrical and Electronics Engineers (IEEE). He is, furthermore, a member of the German Academy of Sciences Leopoldina as well as of the Heidelberg Academy of Sciences and Humanities.

**Thomas Burri** is a professor of international law and European law at the University of St. Gallen (Switzerland). His research has been published in numerous international outlets. He has published three books, *The International Court of Justice and Decolonisation* (ed. with Jamie Trinidad, Cambridge University Press 2021), *The Greatest Possible Freedom* (Nomos 2015), and *Models of Autonomy?* (Schulthess 2010). His research covers traditional international law and EU law, but *Thomas Burri* has also investigated AI and autonomous systems with empirical methods for more than a decade.

**Christoph Durt** is currently a postdoctoral researcher at the Freiburg Institute of Advanced Studies (FRIAS, Germany). The core topic of his research is the interrelation of computational technology with human experience and thought. He combines systematic philosophy with interdisciplinary collaboration and conceptual investigation of the developments that underlie the digital age. Before coming to Freiburg, he was the scientific head of several interdisciplinary projects on consciousness, self, and AI, funded by the European Union's Horizon 2020 research and innovation program and the Volkswagen Foundation. He has taught on a wide range of topics and authors from Ancient Philosophy to the present at Munich University, the University of California, Santa Cruz, the University of California, Berkeley, and the University of Vienna. He has received several research and teaching awards, including one for the essay 'The Computation of Bodily, Embodied, and Virtual Reality' and another for the essay 'How the Digitization of Our World Changes Our Orientation' (links and more on: [www.durt.de](http://www.durt.de)).

**Boris Essmann** studied philosophy, cognitive science, and anthropology at Freiburg University (Germany) and is currently finishing his PhD project. He has worked on several research projects in the field of neurophilosophy and neuroethics, including the Bernstein Focus Neurotechnology Freiburg/Tübingen (2010–2012) and a neurophilosophical Junior Research Group in BrainLinks-BrainTools at the University of Freiburg (2013–2018). He currently works at the EU project FUTUREBODY (2018–2020).

**Johanne Giesecke** is an articulated clerk in the Dresden Higher Regional Court district (Germany). She studied law at the University of Heidelberg (Germany) and at the Université de Montpellier (France), the latter with an Erasmus scholarship. She completed her first state exam in 2021 in Heidelberg. From 2018 until 2021, she was a student assistant in the research group of Prof. Dr. *Fruzsina Molnár-Gábor* in various projects in the field of medical and health law at the Heidelberg Academy of Sciences and Humanities.

**Jan von Hein** is a Director at the Institute for Comparative and Private International Law at the University of Freiburg (Germany). He is the chairman of the Second Commission of the German Council for Private International Law, a member of the Board of the International Law Association's German branch and an associate member of the International Academy of Comparative Law. He is the author of numerous books and articles on private international and comparative law, which have been honoured by the Max Planck Society and the German Stock Corporation Institute.

**Wilfried Hinsch** is a full university professor and holds the chair for Practical Philosophy at the University of Cologne (Germany). His most recent book publications are *Die gerechte Gesellschaft* (Stuttgart 2016), *Die Moral des Krieges* (München 2018) and *Öffentliche Vernunft? Die Wissenschaft in der Demokratie* (co-edited with Daniel Eggers, Berlin/Boston 2019). Currently, he is working on a book manuscript with the title *Legitimacy Beyond Procedural Justice*.

**Philipp Kellmeyer**, is a neurologist at the Medical Center – University of Freiburg (Germany), where he heads the Human-Technology Interaction Lab at the Department of Neurosurgery. He was a Senior Fellow at the 2018–2021 FRIAS Saltus Research Group 'Responsible AI – Emerging ethical, legal, philosophical, and social aspects of the interaction between humans and autonomous intelligent systems'. *Philipp Kellmeyer* studied human medicine in Heidelberg and Zurich and received a Master of Philosophy from the University of Cambridge (UK). As a neuroscientist, he works in the fields of neuroimaging and translational neurotechnology, in particular the clinical application of AI-based brain–computer interfaces. He is a scientific

member of the BrainLinks-BrainTools centre at the University of Freiburg. In his neuroethical research, he works on ethical, legal, social, and political challenges of neurotechnologies, big data, and AI in medicine and research. He is also a research affiliate of the Institute for Biomedical Ethics and History of Medicine at the University of Zurich, where he also teaches biomedical ethics.

**Haksoo Ko** is a professor of law at Seoul National University School of Law in Seoul (Korea). He primarily teaches areas in law and economics as well as in data privacy and AI law. He regularly sits on various advisory committees for the Korean government and other public and private institutions. Prior to joining academia, he practiced law with law firms in New York and in Seoul. He currently serves as President of Asian Law and Economics Association; President of Korean Association for AI and Law; Co-director of SNU AI Policy Initiative; and Associate Director of SNU AI Institute. He had visitor appointments at UC Berkeley, University of Hamburg, Vrije Universiteit Brussel, University of Freiburg, and National University of Singapore. He holds a B.A. in Economics from Seoul National University and received both J.D. and Ph.D. (Economics) degrees from Columbia University in New York, USA.

**Christoph Krönke** is a full university professor of public law at Vienna University of Economics and Business (WU, Austria). His research interests include the law of digitalisation, in particular the law of the digital economy. He studied law at Ruprecht Karl University in Heidelberg (2003–2005) and Ludwig Maximilian University Munich (2005–2009). He received his doctorate in law in 2013 with a dissertation on European and administrative law. In his habilitation thesis of 2020, he developed the foundations of the ‘Regulation of the Digital Economy’, with a focus on the regulation of digital platforms and networks as well as intelligent systems.

**Alex Leveringhaus** is a lecturer on political theory at the University of Surrey (UK). Furthermore, he is the Co-director of the Centre for International Intervention. Beforehand, *Alex Leveringhaus* was a Leverhulme Early Career Research Fellow in the Centre for Political Theory at the University of Manchester (UK). Prior to this, he was a postdoctoral research Fellow at the Oxford Institute for Ethics, Law and Armed Conflict (UK). Generally, his research focuses on the ethical and political repercussions of the widespread introduction and use of AI. In 2016 he published his monograph, *Ethics and Autonomous Weapons*.

**Dustin A. Lewis** is the Research Director for the Harvard Law School Program on International Law and Armed Conflict (HLS PILAC) (USA). With a focus on public international law sources and methods, he leads research into several wide-ranging contemporary challenges concerning armed conflict. Among his recent areas of focus, *Dustin Lewis* led the HLS PILAC project on ‘International Legal and Policy Dimensions of War Algorithms: Enduring and Emerging Concerns’. Alongside a team of faculty and research assistants, he explored how international law governs the development and use in war of AI and certain other advanced algorithmic and data-reliant socio-technical systems. As part of that project and its forerunner, he presented the program’s research on AI, international law, and armed conflicts in Beijing, Geneva, Moscow, New Delhi, Oxford, Shanghai, Stockholm, and Washington, D.C. He is an honours graduate of Harvard College (A.B.) and Utrecht University School of Law (LL.M.).

**Jan Lieder** is a full professor, holds the chair for Civil Law, Commercial Law and Business Law and is the Director of the Department of Business Law at the Institute for Business Law, Labor and Social Law at the University of Freiburg (Germany). Moreover, he is a judge at the Higher Regional Court Schleswig. In 2017, *Jan Lieder* was a Stanford Law Senior Visiting Scholar. He

attained an LLM from Harvard Law School with a concentration in corporate law. His main areas of research include German, European, and international commercial law, corporate law, and business law. In these fields of law, he is an author of more than 200 publications and articles.

**Yong Lim** is an associate professor at Seoul National University, School of Law (Korea). He is also the Co-director of the SNU AI Policy Initiative at SNU's Center for Law and Economics. He has graduated from Seoul National University, College of Law, and obtained his S.J.D. at Harvard Law School. Prior to joining academia, *Yong Lim* practiced law at Korea's leading law firm specializing in antitrust and competition law. His areas of specialty also include consumer protection, information technology law, and privacy and data protection law.

**Yun Liu** is a Postdoc at the Tsinghua University School of Law and Assistant Director of the Institute for Studies on Artificial Intelligence and Law, Tsinghua University (China). In 2017, he received his Ph.D. at the China University of Political Science and Law. From August 2016 to August 2017, he was a visiting scholar at the Ohio State University. His research interests include computational law, standardization law, and civil law. He has worked as an expert in the legislation on technical standards and data governance. Since 2018, he has served in a LegalTech research program, organizing computer scientists and jurists to develop intelligence tools for judicial trials on civil and commercial cases in Tsinghua University. He has published more than twenty papers in Chinese on AI governance, data governance, technical standards, and civil law.

**Thomas Metzinger** was a full professor of theoretical philosophy at the Johannes Gutenberg-Universität Mainz until 2019 (Germany); from 2019 to 2021 he was awarded a Senior Forschungsprofessur by the Ministry of Science, Education and Culture. He is past president of the German Cognitive Science Society (2005–2007) and of the Association for the Scientific Study of Consciousness (2009–2011). As of 2011, he is an Adjunct Fellow at the Frankfurt Institute for Advanced Studies; from 2008 to 2009 he served as a Fellow at the Berlin Institute for Advanced Study; from 2014 to 2019, he was a Fellow at the Gutenberg Research College. From 2018 to 2020 *Thomas Metzinger* worked as a member of the European Commission's High-Level Expert Group on AI.

**Catrin Misselhorn** holds a chair for philosophy at the University of Göttingen (Germany). From 2012 until 2019 she was chair for the philosophy of science and technology at the University of Stuttgart. Prior to that she was visiting professor at the Humboldt University in Berlin, the University of Zurich, and the University of Tübingen. In 2003 she received her Ph.D. at the University of Tübingen and in 2010 she finished her habilitation. In the years 2007–2008 she was Feodor Lynen research Fellow at the Center of Affective Sciences in Geneva, at the Collège de France, and the Institut Jean Nicod for cognitive science in Paris. Her main research areas are philosophical problems in AI, robot and machine ethics, and human–machine interaction. She is leading a number of third party funded projects on the ethical assessment of assistive systems in different areas, for instance, in care, at the workplace, and in education.

**Fruzsina Molnár-Gábor** is a professor at the Faculty of Law, University of Heidelberg (Germany) and a member of the Akademie Kolleg of the Heidelberg Academy of Sciences and Humanities. Her research focuses on the regulation of biomedicine and biotechnology, including the fields of data law, medical law, international law, and the law of the European Union. She is a member of the Junge Akademie of the Berlin-Brandenburg Academy of

Sciences and Humanities, the Leopoldina, and the European Group on Ethics in Science and New Technologies (EGE). She has received the Manfred Fuchs Prize, the promotion award of the VG Wort for her Ph.D., the Young Scholars Award of the Research Network on EU Administrative Law, and the Heinz Maier Leibnitz Prize 2020 for her interdisciplinary research.

**Oliver Mueller** is a professor of philosophy with a focus on technology and on contemporary philosophy, University of Freiburg (Germany) and was a Senior Fellow at the 2018–2021 FRIAS Saltus Research Group ‘Responsible AI – Emerging ethical, legal, philosophical, and social aspects of the interaction between humans and autonomous intelligent systems’. His areas of research are *inter alia* philosophy of technology (AI, bio-/neurotechnologies, human–machine interactions, digitalization), and philosophical anthropology, ethics. Since 2008, *Oliver Mueller* has been Principal Investigator in several interdisciplinary projects on current technologies. In 2015 he obtained the Heisenberg Grant (DFG) and a deputy professorship at the University of Koblenz-Landau, and between 2015 and 2017 he held a visiting professorship at the ETH, Zurich (Switzerland).

**Richard Ngo** is a student at the University of Cambridge. His dissertation focuses on researching the philosophy of machine learning, especially comparing AI development to the evolution of human intelligence. In 2017, he graduated with a first-class honours degree Bachelor in Computer Science and Philosophy from the University of Oxford (UK). He graduated with distinction, earning a Master of Philosophy from the University of Cambridge (UK). Prior to beginning his dissertation, *Richard Ngo* worked as a research engineer at DeepMind. There, he extensively investigated the foundations of AI safety research. His work experience also includes a summer internship at the Future of Humanity Institute.

**Boris P. Paal** is a professor of civil law and information law, data and media law, and a Director of the Institute for Information Law, Media Law, and Law of Digitalization at the University of Leipzig (Germany). Before this, *Boris Paal* was Director of the Institute for Media and Information Law, Dept. I (Civil Law), University of Freiburg (Germany). He researches and teaches, advises, and publishes in all areas of Private and Business Law with a special focus on data, competition, media, and information law as well as on compliance topics. As an author and editor (among others of Gersdorf/Paal on *Information and Media Law*, and Paal/Pauly on the *GDPR*), he is responsible for more than 180 publications in the aforementioned fields. He regularly works as an expert advisor for both state institutions and private entities.

**Sangchul Park** is an assistant professor at the Seoul National University School of Law (Korea). He completed his JSD at the University of Chicago and his undergraduate studies at Seoul National University. His main research area is the application of machine learning to legal studies. Prior to beginning his academic career, he spent more than 13 years in private practice specialising in technology, media, and telecommunications.

**Matthias Paul** was appointed as a professor for Digital Business Management at the Baden-Wuerttemberg Corporate State University (DHBW) in Loerrach (Germany) in 2019, after teaching for a couple of years at other business schools. Before his academic tenure, he served in senior management roles in the financial service industry, as CEO of Ned Davis Research in Boston (USA), an international investment research firm for the asset management industry, and in various managing director roles for financial market data vendors and information providers like IDC (today Factset) and Dow Jones, both in Germany and the USA. He started his business career as a Strategy Consultant at the Boston Consulting Group, working on digitalisation

projects across different industries, embracing the first internet revolution. His academic background is in AI, Cognitive Science, and analytical philosophy. He completed his Ph.D. at the University of Edinburgh at the Centre for Cognitive Science with a thesis on natural language understanding. His dissertation *Success in Referential Communication* was published in 1999 (reprint 2010) by Kluwer as part of the Philosophical Studies Series (Vol. 80).

**Ralf Poscher** is a Director at the Max Planck Institute for the Study of Crime, Security and Law in Freiburg and an Honorary Professor at the University of Freiburg (Germany). His research focuses on constitutional law, national security law, and legal theory.

**Mathias Risse** is Berthold Beitz Professor in Human Rights, Global Affairs, and Philosophy and Director of the Carr Center for Human Rights Policy at Harvard University (USA). His work primarily addresses questions of global justice ranging from human rights, inequality, taxation, trade and immigration to climate change, obligations to future generations, and the future of technology, especially the impact of AI on a range of normative issues. He has also worked on questions in ethics, decision theory, and nineteenth century German philosophy, especially *Nietzsche*. *Mathias Risse* is the author of *On Global Justice* (Princeton University Press) and *Global Political Philosophy* (Palgrave Macmillan), as well as *On Trade Justice: A Philosophical Plea for a New Global Deal* (Oxford University Press, with Gabriel Wollner) and *On Justice: Philosophy, History, Foundations* (Cambridge University Press).

**Thorsten Schmidt** is a professor of Mathematical Stochastics at the University of Freiburg (Germany). His research combines financial mathematics with the area of stochastic processes and statistics. In his career, he met interesting problems, both from statistics and financial mathematics, which inspire deeper mathematical understanding by their surprising complexity. In Freiburg, he and his research team are working on tackling these challenges with an improved mathematical model and targeting various applied areas where this can be useful. Besides finance, this includes medicine, robotics, and, in general, all areas where stochastic modelling is used.

**Weixing Shen** is a professor and Dean of the School of Law, Tsinghua University, Beijing (China), awarded with the 'National Outstanding Young Jurist' in 2014. He also serves as Executive Council Member of China Civil Law Society, Vice President of China Cyber and Information Law Society, etc. His research interests include civil law, computational law, and health law. He also hosts the National Key R&D Program 'Research on Intelligent Assistive Technology in the Context of Judicial Process Involving Concerned Civil and Commercial Cases' and the Major Project of National Social Science Foundation 'Research on the Legal-based Governance of Internet Economy'.

**Jaan Tallinn** is a founding engineer of Skype and Kazaa as well as a co-founder of the Cambridge Centre for the Study of Existential Risk (UK) and Future of Life Institute (Boston, USA). He is on the Board of Sponsors of the Bulletin of the Atomic Scientists, and has served on the High-Level Expert Group on AI at the European Commission, as well as on the Estonian President's Academic Advisory Board. *Jaan Tallinn* is also an active angel investor, a partner at Ambient Sound Investments, and a former Investor Director of the AI company DeepMind.

**Johanna Thoma** is an associate professor in the Department of Philosophy, Logic and Scientific Method at the London School of Economics and Political Science (UK). Before that, she completed her Ph.D. in Philosophy at the University of Toronto and was a visiting researcher at Stanford University and LMU Munich. *Johanna Thoma* has published widely in the areas of



practical rationality and decision theory, economic methodology, and ethics and public policy. She has a special interest in decision making under uncertainty and over time, by policymakers, individuals, and artificial agents.

**Stefan Thomas** is a professor at the Law Faculty of the Eberhard Karls University of Tübingen (Germany), where he holds the chair in Private Law, Commercial Law, Competition and Insurance Law since 2009. He is Director of the Tübingen Research Institute on the Determinants of Economic Activity (TRIDEA), and a Member of the International Advisory Board of the Institute for Global Law, Economics and Finance, Queen Mary University of London. His principal areas of research are German, European, and international antitrust- and competition law and regulation. His most recent work on the intersection between AI and antitrust is *Harmful Signals: Cartel Prohibition and Oligopoly Theory in the Age of Machine Learning* (2019), which was nominated for the antitrust writing awards 2020.

**Antje von Ungern-Sternberg** is a professor of Comparative Public Law and International Law at Trier University (Germany). She is a Director of the Institute for Legal Policy at the University of Trier and a Director of the Institute for Digital Law Trier. Her research focuses on comparative constitutional law, European and public international law, and the law of digitalisation.

**Silja Voeneke** is a professor of public international law, comparative law, and ethics of law at the University of Freiburg (Germany). She was a Fellow at Harvard Law School (2015–2016) and a Senior Fellow at the 2018–2021 FRIAS Saltus Research Group ‘Responsible AI – Emerging ethical, legal, philosophical, and social aspects of the interaction between humans and autonomous intelligent systems’. Her areas of research include environmental law, laws of war, human rights, and the governance of disruptive research and technologies. She previously served as a Director of a Max Planck Research Group at the Max Planck Institute for Comparative Public and International Law in Heidelberg (2005–2010), and a member of the German Ethics Council appointed by the Federal Government. For many years, *Silja Voeneke* has been a legal advisor to the German Federal Foreign Office, and the German Federal Ministry of Environment. She is an author and (co-)editor of several articles and books in the fields of international law and the governance of emerging technologies, as for instance Voeneke/Neuman (eds), *Human Rights, Democracy, and Legitimacy in a World of Disorder* (Cambridge University Press 2018).

**Christiane Wendehorst** has been professor of civil law at the University of Vienna (Austria) since 2008. Amongst other functions, she is a founding member and Scientific Director of the European Law Institute (ELI), chair of the Academy Council of the Austrian Academy of Sciences (ÖAW) and co-head of the Department of Innovation and Digitalisation in Law. She is a member of the Bioethics Commission at the Austrian Federal Chancellery; a member of the Managing Board of the Austrian Jurists’ Association (ÖJT); a member of the Academia Europea (AE), the International Academy for Comparative Law (IACL), and the American Law Institute (ALI). Currently, her research focuses on legal aspects of digitalisation and she has been working as an expert on topics such as digital content, the Internet of Things, AI, and data economy, for the European Commission, the European Parliament, the German Federal Government, the ELI and the ALI. *Christiane Wendehorst* is currently leading the transatlantic project ‘Principles for a Data Economy’ as well as various projects in the area of digitalisation, on topics such as safety- and liability-related aspects of software and biometric techniques. She is also involved in a number of projects on algorithmic fairness. Prior to moving to Vienna, she was a professor in Göttingen (1999–2008) and Greifswald (1998–1999) and was Managing Director of the Sino-German Institute of Legal Studies (2000–2008).

**Jonathan Zittrain** is the George Bemis Professor of International Law at Harvard Law School (USA). He is also a professor at the Harvard Kennedy School of Government, a professor of computer science at the Harvard School of Engineering and Applied Sciences, Director of the Harvard Law School Library, and co-founder and Director of Harvard's Berkman Klein Center for Internet & Society. His research interests include the ethics and governance of AI; battles for control of digital property; the regulation of cryptography; new privacy frameworks for loyalty to users of online services; the roles of intermediaries within Internet architecture; and the useful and unobtrusive deployment of technology in education. *Jonathan Zittrain* established and co-leads the Institute for Rebooting Social Media, a three-year 'pop-up' research initiative that is bringing together participants from across industry, governments, and academia in a focused, time-bound collaboration. He also championed the development of the Caselaw Access Project, which has expanded free public access to US case law. His book, *The Future of the Internet – And How to Stop It*, predicted the end of general purpose client computing and the corresponding rise of new gatekeepers.

## Acknowledgements

We want to start by thanking the Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, for the immense support of our Saltus Responsible AI Research Group 2018–2021.\* Without the FRIAS, its directors, especially Bernd Kortmann, and the team, this Handbook would not have been possible.

We would also like to thank the researchers at our Responsible AI Research Group and at the chair of Silja Voeneke, Alisa Pojtinger, Jonatan Klaedtke, and Daniel Feuerstack, who helped us edit this book, for their important work. Furthermore, we are very grateful to the research assistants Alena Huenermund (Responsible AI Research Group), Hannah Weber, and Heidi He (University of Cambridge) for their valuable support.

We would also like to acknowledge the Baden-Württemberg Stiftung for their financial support in the context of our research project AI Trust. Finally, we would like to express our thanks to the Klaus Tschira Foundation for supporting the work of editor Philipp Kellmeyer in the project “Neuroethics and AI Ethics Lab” (grant no. 00.001.2019).”

\* UP 14/1 (2), DFG-ProjektID: 422010107.

Cambridge University Press & Assessment

978-1-009-20786-7 — The Cambridge Handbook of Responsible Artificial Intelligence

Edited by Silja Voeneke , Philipp Kellmeyer , Oliver Mueller , Wolfram Burgard

Frontmatter

[More Information](#)

---