Edited by

Spyridon Stelios Kostas Theologou

THE ETHICS GAP IN THE ENGINEERING OF THE FUTURE

MORAL CHALLENGES FOR THE TECHNOLOGY OF TOMORROW

The Ethics Gap in the Engineering of the Future

This page intentionally left blank

The Ethics Gap in the Engineering of the Future: Moral Challenges for the Technology of Tomorrow

EDITED BY

SPYRIDON STELIOS

National Technical University of Athens, Greece

AND

KOSTAS THEOLOGOU

National Technical University of Athens, Greece

United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited
Emerald Publishing, Floor 5, Northspring, 21-23 Wellington Street, Leeds LS1 4DL

First edition 2025

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-83797-636-2 (Print)
ISBN: 978-1-83797-635-5 (Online)
ISBN: 978-1-83797-637-9 (Epub)

'To our students, engineers and citizens of tomorrow'

This page intentionally left blank

Contents

About the Editors

Dr Spyridon Stelios is a Teaching and Research Associate at the National Technical University of Athens (Faculty of Applied Mathematical and Physical Sciences/Department of Humanities, Social Sciences and Law). He lectures at several schools of NTUA, and his main research interests are in ethics, epistemology, and experimental philosophy. He has authored several book chapters, articles in international peer-reviewed journals and conference proceeding in these areas. He also serves as a reviewer for several academic journals.

Kostas Theologou is a Professor at the Faculty of Applied Mathematical and Physical Sciences of the National Technical University of Athens (and Director of the Department of Humanities, Social Sciences and Law); he is also the Director of the journal Signum http://aked.ntua.gr/signum.html. He is the Vice Director of Studies of European Civilization studies and the Moderator of the module 'Social Theory and Modernity" of the School of Humanities of the Hellenic Open University. His research interests focus on the History and Philosophy of Culture (technology, urban space, collective identities, bioethics, biotechnology, etc.).

This page intentionally left blank

About the Contributors

Rakhat Abylkasymova, independent researcher, Poland. Her areas of interest are linguistics, international business, and future challenges affecting humanity, with a focus on our future in space. She has published in *the International Journal of Astrobiology* and *Space Policy*, among others.

Panorios Benardos currently serves as an Assistant Professor in the School of Mechanical Engineering of the National Technical University of Athens (NTUA) in the field of 'Intelligent Manufacturing". He received his Diploma in Mechanical Engineering in 2001 and his PhD in 2009 from NTUA. He has worked as a Lecturer and an Assistant Professor in the University of Nottingham from 2014 to 2018. He has participated in 21 nationally (GSRT, EPSRC, Innovate UK) and EU-funded research projects (FP7, H2020, HE). He has published 42 papers in international scientific journals and conferences (24 and 18 respectively) and coauthored 2 book chapters. His current research interests involve the application of artificial intelligence and machine learning methods in modeling and optimization of manufacturing processes and systems.

Moses Boudourides served as a Professor of Practice at the Arizona State University School of Public Affairs and is currently a member of the faculty of the Northwestern University School of Professional Studies Data Science Online Graduate Program. Previously, he held the position of Professor of Computational Mathematics at the University of Patras in Greece and the Department of Electrical and Computer Engineering at the Democritus University of Thrace, also in Greece. His expertise lies in applied and computational mathematics, network science, and computational social science.

Maria Hedlund is an Associate Professor at the Department of Political Science, Lund University, Sweden. Her research focuses on theories on democracy and the role of experts in democratic processes and in public administration, in particular regarding questions on emerging technologies, ethics, governance, responsibility, and legitimacy. Her recent publications include 'AI and epigenetic responsibility" (in *Epigenetics and Responsibility: Ethical Perspectives*, Anna Smajdor, Daniela Cutas, Emma Moormann & Kristien Hens (eds.), Bristol University Press, 2024), 'Ethicisation and reliance on ethics expertise" (*Res Publica*, April 2023), with Erik Persson 'Expert responsibility in AI development" (*AI & Society*, June 2022), and 'När maskiner fattar beslut – vem är ansvarig?" (When machines make decisions – who is responsible?) (in *Statsvetenskaplig tidskrift* 122, 2020).

Rahulrajan Karthikeyan is a graduate student in the Computer Systems Engineering (CSE) program at the School of Computing and Augmented Intelligence, Arizona State University. He has worked as a Machine Learning Research Assistant in the Cooperative Robotics Systems Lab (CRS), collaborating with PhD students on research involving multi-modal robot communication, temporal-spatial inverse semantics, and various projects leveraging Inverse Reinforcement Learning, ROS, and Computer Vision. In addition, he co-founded the Google Developer Student Club (DSC) at ASU, where he has taught Flutter, TensorFlow, Machine Learning, and Computer Vision to over 200 Bachelor's and Master's students at each DSC event. His accomplishments at ASU include winning more than 20 hackathons and coding challenges. He has also collaborated with Professor Boudourides on several machine learning projects.

Ivica Kelam, PhD, Head of the Department of Philosophy and History at the Faculty of Education, teaches the philosophy of education, ethics of the teaching profession and bioethics. Since 2018, he has been the Head of the Osijek Unit of International Bioethics Chair. In 2014, he successfully defended his doctoral thesis entitled Genetically modified crops as a bioethical problem. Since 2016, he has been the Head of the Center for Integrative Bioethics at J.J. Strossmayer University of Osijek. Since 2020, he has taught the bioethics of sports and sociology of sports at the Faculty of Kinesiology, teaches ethics in biosciences at the doctoral study in molecular biology at the University of Osijek, and teaches bioethics at the doctoral study at the Faculty of Education. He is the President of the Organization Committee of the Osijek Days of Bioethics, an international bioethical conference held in Osijek. He has published the book Genetically Modified Crops as a Bioethical Issue, over 40 scientific papers in Croatian and foreign scientific journals, and participated in more than 80 scientific conferences in Croatia and abroad. In 2020, he was elected President of the Croatian Bioethical Society.

Evie Kendal is a Bioethicist and Public Health Scientist whose work focuses on emerging reproductive technologies. She is the Head of the Ethical, Legal, and Social Implications of Emerging Technologies (ELSIET) research group of the Iverson Health Innovation Research Institute at Swinburne University of Technology and a Senior Lecturer of health promotion in the School of Health Sciences. She is the author of 'Ectogenesis and the Ethics of New Reproductive Technologies for Space Exploration" (in *Transhumanism*, Springer 2022), 'Biological Modification as Prophylaxis: How Extreme Environments Challenge the Treatment/Enhancement Divide" (in *Human Enhancements for Space Exploration*, Springer 2020), and 'Desire, Duty and Discrimination: Is There an Ethical Way to Select Humans for Noah's Ark?" (in *Reclaiming Space*, Oxford University Press 2023).

Tony Milligan is a Research Fellow in Ethics with the Department of Theology and Religious Studies at King's College London. His publications include *The Ethics of Political Dissent* (2023), *Nobody Owns the Moon* (2015), and the coedited volume *The Ethics of Space Exploration* (2016). He is currently working on a short introduction to societal issues concerning space: *What is Space Exploration For?* with Koji Tachibana, and a SpringerBrief on *Indigenous Cosmological Visions of the Pleiades*.

Amir Muzur is a Full Professor (since 2013) and the Head of Department of Social Sciences and Medical Humanities (since 2008) at University of Rijeka – Faculty of Medicine. His major fields of research/teaching interest are history of medicine and bioethics and theoretical neuroscience. He authored/coauthored more than 30 books (local history, history of medicine, popularization of neuroscience, essays, travelogues, poetry) and about 700 articles in scientific and other periodicals and participated in more than 200 conferences (more than 40 invited lectures). Amir Muzur was the President of Croatian Bioethics Society (2012–2016) and now is the Director of the Fritz Jahr Documentary and Research Centre for European Bioethics at University of Rijeka (since 2013).

Erik Persson received his PhD in Philosophy from Lund University in 2009. He is currently an Associate Professor in Practical Philosophy at the Department of Philosophy, Lund University and Researcher at RISE – Research Institutes of Sweden. His previous employers include Umeå University, the Nordic Genetic Resource Centre and the Institute for Theological Inquiry in Princeton. His research and teaching are primarily focused on applied ethics, value theory, environmental ethics, space humanities, and the philosophy of AI and other emerging technologies. His recent publications include 'The future of AI in our hands? To what extent are we as individuals morally responsible for guiding the development of AI in a desirable direction?" (*AI & Ethics* 2: 683 – 695), co-authored with Maria Hedlund.

George Profitiliotis is a Research Scientist affiliate at Blue Marble Space Institute of Science and a speculative fiction author serving the New Weird. He is an electrical and computer engineer, further trained in interdisciplinary environmental studies, biomimetics, and strategic innovation management, and holds a PhD on the application of environmental economics to planetary protection policy. He has worked extensively in the fields of strategic foresight, futures literacy, and futures studies. During his first PostDoc, he studied the pertinence of anticipation to the search for extraterrestrial life from the viewpoints of particular subfields of the humanities, social sciences, and policy, with a focus on the proactive management of a future discovery. His second PostDoc focused on anticipating the future ethical, legal, and societal implications of quantum technologies.

Iva Rinčić graduated in 2000 in Sociology and Croatian Culture from the Department of Croatian Studies at the University of Zagreb and defended her Master's thesis in 2005 at the Faculty of Political Sciences. She defended her PhD thesis 'Theoretical Strongholds, Achievements and Perspectives of Bioethical Institutionalisation in the European Union' in 2010 at the Faculty of Philosophy in Zagreb. Since 2001, she has been working at the Department of Social Sciences and Medical Humanities of the Rijeka University Faculty of Medicine (since 2022 as Full Professor), and since 2014, also at the Department of Public Health of the University of Rijeka Faculty of Health Studies. She has participated at more than 60 Croatian and international conferences, published approximately 70 scientific and professional articles, and five books. From 2013 till 2021, she was the Director of the University of Rijeka Foundation.

Goran Rujević is an Assistant Professor at the Department of Philosophy of the Faculty of Philosophy in Novi Sad. He earned his Bachelor's and Master's degrees in Philosophy from the Faculty of Philosophy in Novi Sad, where he also defended his PhD thesis on Kant's philosophy of science in 2019. His main areas of interest are philosophy of natural sciences and argumentation theory, but he also has interest in issues regarding teaching and popularizing philosophy. Apart from scholarly work, he often translates works from English and German into Serbian; of note is his work on translating Nick Bostrom's seminal book *Superintelligence* into Serbian.

Emmanouil Stathatos is a Mechanical Engineer who graduated from the National Technical University of Athens (NTUA) in 2006. He earned his PhD in 2020, focusing on the optimization of the Selective Laser Melting (SLM) process through numerical modeling and machine learning techniques. Since 2009, he has served as a Technical Consultant to Greek industrial firms. Currently, he is a Postdoctoral Researcher at the Manufacturing Technology Lab, NTUA, where he contributes to Horizon Europe and Erasmus+ projects. His teaching roles have included Laboratory Teaching Assistant for Computer Aided Manufacturing, Additive Manufacturing, and Discrete Event Simulation for manufacturing systems, as well as Instructor for the Advanced Manufacturing Systems course in the Interdepartmental Master's Program on Automation Systems. He has published 11 papers in international scientific journals and conferences. His research is dedicated to integrating AI into manufacturing to redefine the paradigms of smart production.

Konrad Szocik, University of Information Technology and Management in Rzeszow, Poland. From 2021 to 2022, he was a Visiting Fellow at the Yale Interdisciplinary Center for Bioethics. His interests include ethics, philosophy and bioethics of space exploration, feminism, antinatalism, procreative ethics, and population ethics. He is the author of *The Bioethics of Space Exploration* (Oxford University Press, 2023) and *Feminist Bioethics in Space* (Oxford University Press, 2024).

George-Christopher Vosniakos studied Mechanical Engineering at the National Technical University of Athens and obtained an MSc in Advanced Manufacturing Technology in 1987 and a PhD in Intelligent CAD-CAM interfaces in 1991 from UMIST-Britain. He has worked in the software industry in Germany and as a Lecturer at UMIST. In 1999 he joined NTUA, where he is currently Professor of Manufacturing Systems and director of the Manufacturing Technology Laboratory. He has authored 110 Journal and over 108 Conference papers obtaining 2,850 Scopus citations. He is editorial board member of six major International Journals. His recent research interests include AI in Manufacturing.

Chieh Yi is a graduate student in the Computer Systems Engineering (CSE) program at the School of Computing and Augmented Intelligence at Arizona State University. He has collaborated with Professor Boudourides on several machine learning projects and is actively involved in the ASU EPICS program, where he contributes to innovative technical projects.

Acknowledgments

Special thanks to Katy Mathers, Lauren Kammerdiener, and Brindha Thirunavukkarasu from Emerald for their help, encouragement, and support as well as to the anonymous reviewers for their insightful comments.

This page intentionally left blank

Prolegomena: Mind the Gap! Now and for Ever

Spyridon Stelios and Kostas Theologou

National Technical University of Athens, Greece

This book discusses Technology, Engineering, and Ethics and the ways they may be related to two strictly mutually defined theoretical concepts: the "gap" and the "future". Both concepts have not occupied, as much as we think they should, the research around Engineering ethics and the Philosophy of technology. Nevertheless, they are particularly important as they express both the dizzying technological progress (see "future") and the side effects of this forward march (see "gaps").

What is the future? One could respond that it is something that (in any case) does not exist as it cannot be experienced as such – it is experienced only as present. It is a mental, immaterial, functional aspiration that we use to give meaning to our existence, to make choices and plan our actions. As a mental construct or merely a perception, the future is an extremely fruitful and therefore useful field of thought guiding our behavior. But does it have any value beyond its functionality? Or rather, what can we think of a concept which is also only a thought, which is on the same sphere as dreams and imagination and which has no tangible empirical manifestation? The answer is "a lot". The future offers fertile ground for philosophical inquiry, the preeminent mental pursuit that humans developed and which led to the birth of science itself. And it is this philosophical examination of a nonempirically existent field of reference that can transform or change the present.

The swift leaps of technology create inductively documented images of the future of humanity. Technology is threaded together with the concept of progress leading to a better future, so the ethics of technology is interwoven, above all, with future representations. We are curiously interested in what happens, e.g., with generative intelligence today, but we are more interested in what will happen tomorrow regarding that and similar technologies. The future is omnipresent in the present/here, or more precisely, the present/here is constantly in the future. This collective obsession undoubtedly awakens people to potential dangers, but at the same time prepares them to accept innovations and interventions more easily in their lifeworld. Let us not forget also that there is always the risk any discussion

The Ethics Gap in the Engineering of the Future, 1–5

doi:10.1108/978-1-83797-635-520241001

of the future makes it easier to avoid discussion and criticism of the state of the present.

Regarding technology, ethical reflection is limited by a multiplicity that derives from the object of study itself and which creates a kind of gap. But why refer to a gap? What is missing from our moral concerns, intuitive and explicit, that needs to be filled in? What might largely led to the conception and creation of this book's theme in the first place? To begin with, the moral gap is somehow fed by and related to the future dimensions of technology. No one can predict exactly what specific ethical issues will arise from tomorrow's amazing innovations and the ongoing digital transformation of our civilization. As much as we have an image of the future through the technological present, this image and sense has already contaminated, altered, and changed the course of development for every artefact. In fact, it could be argued that any assessment adds an extra boost to the final form of the "new" to come; and the new always shocks upon appearance the bystanders, either in the form of art or artefact. Simply put, what we expect, because we expect it, will be even further ahead, even more sophisticated. This does not involve a specific timeframe. When we refer to the appearance/evolution of an artefact in the future, we are not referring to temporal estimates about when it will come. It is not worth much, in philosophical terms, to say that, e.g., in 2050 there will be vehicles that can drive on the road, dive in water or fly (as, for example, in the 1960s the author Philip K. Dick predicted in his novel "Do Androids Dream of Electric Sheep?" that in 1992 there would be hovercars). The assessment of the future of any technological product associates with the product's final form, regardless of when this format will be completed. It is the form that we set as final and beyond which we cannot think of another.

Of course, one would reasonably argue, this is completely subjective. Another form I believe cars will have—keeping the essence of what it means to be a car—and another someone else. In fact, the possibilities, and forms, are endless. And not only. Their very future perspective, expressed through these forms, geometrically evolves their possible final state. In other words, the more we think about the *end* (*telos*) of a technological phenomenon, the further it draws away! This explosive expansion of future projections of the technological phenomenon creates gaps in their ethical processing. The nature of technology creates a kind of never reached imaginative infinity within our everyday lives. Against this, moral philosophers try, but rather fail, to find suitable clothes for artificial children who never stop growing.

Today's technology also seems to have gaps, especially if it is seen in its holistic social dimension. Technology is not a uniform phenomenon that satisfies the needs and desires of every culture to the same extent. Also, if we carefully examine today's innovations, we have the impression that they have not spread to all areas of our lives equally. At least not yet. For example, developments in the field of biotechnology seems fewer and to a lesser extent than those in the field of artificial and computational intelligence. Space technology is also developing less than artificial intelligence and biotechnology. Of course, the reasons for these development gaps are many (see for instance, economic) and certainly cannot be analyzed in this short introduction. Perhaps it is naive even to talk about

categorization and a common ground between different artefacts that allows their comparison. What could be argued is that the larger the network of people and organizations involved in a field, the greater will be its rate of development. In the case of space exploration technology, for instance, the players are not really that many. There are a few dozen countries in the world that have and are developing this kind of technology.

In this book, the concept of the future is examined in three general technological fields to highlight possible gaps in their moral orientation. These three thematic sections are: artificial intelligence, biotechnology and space. Of course, the multifactorial dimension of the topics under discussion inevitably leads to overlaps between the sections.

Section 1: Artificial Intelligence provides general reflections on the ethical nature of AI. This part is composed of five chapters. The first chapter, entitled "Who should obey Asimov's laws of robotics? A question of responsibility" by Maria Hedlund and Erik Persson, discusses the safety value of implementing Asimov's Laws of Robotics as a future general framework that humans should obey. Within this framework, the implementation of the law in human legislation is being considered, especially regarding people or companies that develop, build, market, or use AI, now and in the future. The second chapter, entitled "Criminal justice in the age of AI: Addressing bias in predictive algorithms used by courts" by Rahulrajan Karthikeyan, Chieh Yi, and Moses Boudourides focuses on recent studies aimed at mitigating biases in algorithmic decision-making within the realm of predictive policing algorithms employed in the criminal justice system. The authors reassess recidivism rates and implement adversarial debiasing in conjunction with fairness metrics. In "Ethical challenges in the new era of intelligent manufacturing systems", our NTUA colleagues Emmanouil Stathatos, Panorios Benardos, and George-Christopher Vosniakos explore the ethical challenges arising from the integration of advanced AI technologies into intelligent manufacturing systems. The chapter reflects on issues such as data privacy, job displacement, the impact of automation on workforce dynamics, and the psychological effects of working alongside AI-powered systems. The fourth chapter, entitled "From Croesus to computers: Logic of perverse instantiation" by Goran Rujević, analyses old and new tales of perverse instantiation arguing that is fundamentally a philosophical in nature, problem so old that even Socrates had to face it. The author aims at uncovering this fundamental problem and show its connection to the contemporary field of AI ethics, especially when such failure mode presents an existential risk (see superintelligent AI). In the last chapter of the section, entitled "The gradual unavoidable colonization of the lifeworld by technology" by Kostas Theologou and Spyridon Stelios, the technological colonization of the lifeworld is being discussed, introducing the concept of techlifeworld.

Section 2: Biotechnology focuses on specific ethical issues raised by human activity in the field of biotechnology. This part consists of two chapters. The first chapter, entitled "Ethical aspects of promises and perils of synthetic biology" by Ivica Kelam, investigates the phenomenon of synthetic biology through an ethical analysis of the unfulfilled promises and potential perils surrounding this

technology. At first, he discusses the problem of defining the inter-disciplinary field of synthetic biology and then proceeds with a brief history of systemic biology and the groundbreaking creation of Synthia, the first synthetic organism. The potential benefits of synthetic biology are then discussed together with ethical and regulatory issues. In "Adapting (bio)ethics to technology and vice versa: when to fight and when to collaborate with artificial intelligence", Iva Rinčić and Amir Muzur investigate the possibility of a systematic study of adaptations human society will have to consider guaranteeing the obeyance to the fundamental ethical values within an environment of rapid advancement of AI. This possibility emerges as particularly important if we consider recent developments such as the creation of computer-generated synthetic lifeforms displaying characteristics of living beings. Against these contemporary developments, the authors propose a new discipline, *Epharmology*, as a methodological approach that could provide a systematic ethical framework.

Section 3: Space explores ethical challenges that might emerge through our greater familiarity with extraterrestrial reality. Space offers fertile ground for the discussion of existing and new ethical conceptualizations. This part consists of four chapters. The first chapter entitled "Are space technologies untimely?" by Tony Milligan provides an overview of fears and suspicions concerning the development and expansion of space technologies, classifying them as instances of "space skepticism." It goes on to argue that the technologies in question are not untimely with respect to the major challenges facing humanity. The second chapter entitled "Moral vistas to xenic beyonds: Fostering moral imagination to pre-empt monsterization in future encounters with extraterrestrial life" by George Profitiliotis offers a theoretical exploration and a practical intervention, in line with pragmatist ethics, in the form of a novel futures literacy workshop to help pre-emptively decrease the potential for the monsterization of humans and extraterrestrial life in the case of a future discovery. The workshop is envisioned as a preparatory, complementary pedagogical approach to the traditional teaching of applied ethics to university students. The third chapter, entitled "Planning for the future in space—with or without radical biomedical human enhancement?" by Rakhat Abylkasymova and Konrad Szocik, discusses the possibility of long-term space exploration requiring extraordinary solutions, such as the possible obligation or requirement to apply radical human enhancement. Furthermore, they refer to the feminist perspective and to issues such as exclusion and power structures. The final chapter, entitled "Building better (space) babies: Bioastronautics, Bioethics and off-world ectogenesis" by Evie Kendal explores some of the ethical issues surrounding ectogenesis—the development and use of artificial womb technology—and its space applications in establishing an off-world human society.

Advances in several technological sectors have made engineering a rapidly and dynamically expanding industry. Every invention concerning robots, Bioengineering, Space, and other fields is now being developed, perfected, and implemented in a fast-track pace. Generally, there is not enough time to experiment, to test, or to approve technical innovations. This is how markets push consumers to novel risks creating gaps in their effort and process of integrating these fascinating

and not always necessary artifacts. This dynamic also gives insight on technology's future impact both on the individual and on the collective/social level. We are witnessing technology transforming itself into new potential and sometimes hypnotic applications. We do not only see it transform but we feel rather unprepared to adjust our existence, our being, within its ever-altering framework. The future is being shaped up in an imperceptible manner now and requires proper moral treatment and provisions. This process is better comprehended by looking at future scenarios and challenges. For example, imagine in the future waking up in a colony on the planet Mars and having to decide how to manage an impending biological threat using AI-based systems and biotechnological knowledge. This kind of approach is generally taken in the chapters.

This book infuses applied ethics and engineering in three technological domains to ferment and stimulate ethical reflections on their future development. The reflections and insights that the 11 chapters provide are key to understanding the change our personal, everyday life is undergoing. In particular, the fifth chapter seeks to collect all the threads spun by the different contributions to enlighten the thesis of a technological colonization of the lifeworld. This technological colonization of lifeworld serves as a link between the chapters and themes.

So, the technological development is moving so far that it is difficult to timely and properly adjust culturally-socially and personally. This fast-moving pace leads to a diversity of ethical aspects of the future visions of life and artificial–biological interactions. Based on this insightful and incisive look into tomorrow, engineers, philosophers, intellectuals and scientists as well as researchers, educators, and scholars in the fields of technology and humanities will likely benefit from reading the chapters of this collective volume finding theoretical and practical guidelines.

Section 1

Artificial Intelligence

This page intentionally left blank

Chapter 1

Who Should Obey Asimov's Laws of Robotics? A Question of Responsibility

Maria Hedlund and Erik Persson

Lund University, Sweden

Abstract

The aim of this chapter is to explore the safety value of implementing Asimov's Laws of Robotics as a future general framework that humans should obey. Asimov formulated laws to make explicit the safeguards of the robots in his stories: (1) A robot may not injure or harm a human being or, through inaction, allow a human being to come to harm; (2) A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law; (3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Law. In Asimov's stories, it is always assumed that the laws are built into the robots to govern the behaviour of the robots. As his stories clearly demonstrate, the Laws can be ambiguous. Moreover, the laws are not very specific. General rules as a guide for robot behaviour may not be a very good method to achieve robot safety – if we expect the robots to follow them. But would it work for humans? In this chapter, we ask whether it would make as much, or more, sense to implement the laws in human legislation with the purpose of governing the behaviour of people or companies that develop, build, market or use AI, embodied in robots or in the form of software, now and in the future.

Keywords: The laws of robotics; Asimov's laws; robot ethics; AI ethics; safety; responsibility; democracy

Introduction

The aim of this chapter is to explore the value of implementing Asimov's Laws of Robotics as a general framework for humans with the purpose of governing the behaviour of people or companies that develop, build, market or use artificial

The Ethics Gap in the Engineering of the Future, 9–25

doi:10.1108/978-1-83797-635-520241002