*European Journal of Risk Regulation* (2025), **16**, 1584–1590
doi:10.1017/err.2025.10056

CAMBRIDGE UNIVERSITY PRESS

BOOK REVIEW

# *Robot Law: Volume II*

Martin Holohan Jr

School of Law, Trinity College Dublin, Dublin, Ireland
Email: holohama@tcd.ie

## Introduction

*Robot Law: Volume II* (Edward Elgar 2025), edited by Ryan Calo, A. Michael Froomkin and Kristen Thomasen, provides an interdisciplinary examination of contemporary issues in robotics and law, with contributions from a wide array of industry experts and academics. Whereas the first volume of *Robot Law* grappled with some of the more foundational issues concerning robotics and law, this second volume explores the pressing legal, social and ethical concerns arising from robot systems becoming increasingly embedded within society and being able to (in some respects) operate independently of overt human decision-making. This second volume has four sections which are termed "Beyond neutrality," "Presentation and manipulation", "Responsibility and enforcement" and "Individual and relational rights: creativity, expression and privacy." While these sections tackle different issues pertaining to robotics and law, each section seeks to build upon the previous one by illustrating the interaction between the law and automated systems.

## Section I – Beyond neutrality

The first section engages in a philosophical investigation concerning the way in which automated systems ought to be understood in relation to humans, along with a consideration of moral and ethical questions surrounding such systems. In Chapter 1, Birhane and van Dijk raise interesting questions surrounding the "robot rights" debate. One school of thought claims that, if automated systems one day reach a level of intelligence comparable to humans, robots should be viewed as possessing similar rights as humans. Another school of thought claims it is better to view robots as "slaves," asserting that the dehumanising connotations associated with the term allow us to not become distracted by technological advancements of robotics and that human wellbeing should be the chief concern. While Birhane and van Dijk agree that human wellbeing should be the chief concern, they argue that neither school of thought correctly captures the central issue surrounding the status of robotic systems, as it is not possible to "dehumanise" something which is itself not human. They argue that robot rights are a fictitious concept and that the surrounding debate incorrectly focuses on a supposed oppression of AI. Birhane and van Dijk correctly highlight that much of

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

https://doi.org/10.1017/err.2025.10056 Published online by Cambridge University Press

the discourse in this area implicitly anthropomorphises AI. This is said to distract from the more pressing ethical issue of AI being used as an instrument by humans to oppress other humans.[1]

While robotic systems are a tool which can be of immense benefit to society, Birhane and van Dijk are right to note that AI can unfortunately be used unjustly to violate human rights of equality and dignity through bias and discrimination. It is clear that technological innovation must never be elevated to such a level to where human concerns are thrown to the wayside. While technological progression seems to be an inevitable part of human existence, Birhane and van Dijk convincingly assert that human wellbeing should be the chief concern when examining the legal, social and ethical issues arising from advancements in automated intelligence. Beginning the book with the discussion of whether robot rights should exist is effective, as it allows the reader to contemplate how automated systems should be viewed in relation to human beings. Furthermore, it successfully sets up the book's focus on human welfare in relation to autonomous systems.

In Chapter 2, Everett Jaques explores the importance and moral implications of the algorithmic policies chosen to inform the decisions of automated systems through an examination of the online platform "Moral Machine." The author asserts that the Moral Machine, which utilises the "trolley problem" to gather human observations on moral decisions made by automated systems, is in fact a "monster" and is the wrong tool for considering the morally complex scenarios which automated systems such as self-driving cars may face. The individualistic and transactional framing of the Moral Machine is successfully identified as an issue inherent in the trolley problem upon which the online platform is based. While describing the Moral Machine as a "monster" seems a tad harsh, Jaques convincingly argues that a structural approach should be adopted. Rather than incentivising an "all or nothing" approach in determining the correct course of action for an autonomous vehicle faced with a morally challenging situation, a structural approach considers a multitude of varying but still relevant factors. The effectiveness of this approach can be seen in its consideration of the relationship between deterrence and proportionality in relation to imprudent behaviours such as jaywalking. While Everett Jaques admits that a structural approach is not always going to provide an immediately correct or easy answer, its advantages over Moral Machine's trolley problem approach is compellingly demonstrated.

## Section 2 – Presentation and manipulation

In Chapter 3, Hwang and Levy observe that corporations slowly introduce automated technology into products to evaluate the technology, while simultaneously guarding against consumer concerns that new technology is lessening their control over those products. Hwang and Levy highlight the decision of some companies to retain the presence of a steering wheel in self-driving cars as an example of how design and appearance can be used to affect the reception of automated systems in traditionally human-controlled products.[2] They also convincingly argue that such design can amount to consumer manipulation or deception. Yet, a nuanced perspective is still offered, with it being recognised that such manipulation or deception may sometimes be socially beneficial. Hwang and Levy identify one such socially beneficial but deceptive design as being the inclusion of artificial engine noise in electric cars. As electric cars are comparatively much quieter than traditional internal-combustion engine cars and since

[1] Abebe Birhane and Jelle van Dijk, "Robot Rights? Let's Talk about Human Welfare Instead" 2.
[2] Tim Hwang and Karen Levy, "The Presentation of Machine in Everyday Life" 35.

https://doi.org/10.1017/err.2025.10056 Published online by Cambridge University Press

society has been conditioned to expect cars to be at least somewhat loud, the inclusion of artificial noise helps alert pedestrians (particularly those with visual impairments) of the oncoming car. Hwang and Levy convincingly raise the ethical concerns associated with this apparent deception. They do well to highlight this conceptual tension in robotic design; on one hand, there may be legitimate social reasons to "deceive" consumers but on the other hand, the classic adage of the ends not justifying the means may apply. While Hwang and Levy differentiate between benevolent deception and *mala fides* manipulation, the key question seems to be whether consumers ought to view transparency or utility as being more valuable in respect of automated systems. Such a question is not easily answered, especially given the financial motivations of commercial product providers and the palpable power imbalance between companies and consumers.

In Chapter 4, Hartzog examines the manner in which robot design may take advantage of human vulnerabilities. There is a particular focus on the consumer protection issues which arise as a consequence of robots being increasingly prevalent in day-to-day life and on the role that state authorities, in this instance, the Federal Trade Commission (FTC) of the United States, ought to play in responding to commercial manipulation resulting from unfair and deceptive robots.[3] Hartzog is right to remind readers that one of the main reasons why humans are so susceptible in respect of deceptive robots is because a sense of trust exists. Interestingly, this sense of trust seemingly stems from the fact that robotic systems are *not* human, as there is a sentiment that robots do not possess human defects and thus are more reliable and trustworthy. Hartzog outlines the existence of "scambots" and "nudgebots" and emphasises the potential which these robots have to manipulate human behaviour, which becomes increasingly worrying as humans continue to trust automated systems with their personal information. Hartzog argues persuasively that the FTC should take the lead in protecting consumers from such deceptive robots.

Together, these two chapters highlight how the design and appearance of automated systems affects the acceptance of such technology in commonplace products, but also how deceptive robotic designs can manipulate and even take advantage of consumers. While some manipulation may be justifiable at times, measures are clearly required to address these risks.

## Section 3 – Responsibility and enforcement

The third section focuses on issues surrounding the bearing of ethical or legal responsibility in relation to automated systems, along with the potential of automated rule enforcement. In Chapter 5, Elish explores the idea of the "moral crumple zone," a concept proposed by the author to signify the way in which responsibility for a certain action may be wrongly assigned to a human actor who had minimal control over the behaviour of the autonomous system. Elish argues that, while the crumple zone of a car aims to protect the human driving the vehicle, the moral crumple zone protects the supposed integrity of the automated systems at the expense of the human operator.[4] The concept of the moral crumple zone is fascinating, as it provides a thought-provoking term to describe the "human in the loop" being held responsible for automated systems which malfunction. Elish highlights that the promise of automated systems (such as autopilot functionality in planes) was that robotic systems were going to be capable enough to make up for any mistakes made by the human operator. Yet, when an accident does occur due to

[3] Woodrow Hartzog, "Unfair and Deceptive Robots" 43.
[4] Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human–Robot Interaction" 83.

https://doi.org/10.1017/err.2025.10056 Published online by Cambridge University Press

a system malfunction, the author showcases that there is commonly an attribution of blame onto the very same human operator who the automated system was supposed to assist.

In Chapter 6, Khoo examines automated systems through the lens of tort law, with it being argued that Canadian private law ought to provide a legal basis for claiming compensation caused by discriminatory machine-involved harms. Particularly, Khoo argues that a tort of negligent discrimination could be recognised, to act as an additional avenue for combatting systemic harm suffered by marginalised communities on online platforms.[5] Khoo defines systemic harm as damaging consequences which are generally discriminatory in nature and exist on a level beyond that of merely individual harms. The proposition that the tort of negligence be extended to cover discrimination is not an uncontroversial one. The Supreme Court of Canada rejected the idea of a common law tort of discrimination and noted that such a tort was not needed due to the relevant human rights protection against discrimination being present in Canadian law.[6] While there is some academic support for the idea of a tort of negligent discrimination,[7] courts have generally not recognised the standalone existence of such a tort. Furthermore, the legal basis for claims of discrimination have typically been provided for by statute.[8]

There seem to be two main reasons for the rejection of a tort of negligent discrimination. First, the existence of a tort of negligent discrimination might be deemed to contravene the classic liberal harm principle which distinguishes between offence and harm,[9] with Gardner stating that it is "seriously problematic for liberals to explain the basis upon which citizens may be required to abstain from discriminating against each other."[10] While society may be content with setting aside the concerns of contravening the harm principle when a democratically elected legislature enacts anti-discrimination laws pursuant to the will of the people, judges may be less willing to do so, particularly in the realm of private law. Second, even if discrimination were deemed by the courts to be a type of harm which could be covered by tort law, systemic harm would still seem not to be covered. As systemic harm is focused on harm to groups of persons and not individuals, it is better categorised as a societal harm and not as an individual personal harm, the latter of which being what tort law is typically concerned with. The lack of a consensus regarding the existence of such a tort seems to demonstrate that tort law might not be the correct avenue to combat systemic harm, with a breach of fundamental rights claim or a claim under statute being more suitable.

Chapter 7 sees Froomkin, Kerr and Pineau examine the use of automated systems in the medical field with respect to the principles underpinning medical negligence in tort law. They provide a disconcertingly persuasive argument that the application of tort law in the context of these systems may actually lead to a loss of human diagnostic expertise

[5] Cynthia Khoo, "The Sum of All (un)intentions: Reasonable Foreseeability, Platform Algorithms, and Emergent Systemic Harm to Marginalized Communities" 106.

[6] *The Board of Governors of the Seneca College of Applied Arts and Technology v Pushpa Bhadauria* [1981] 2 SCR 181, 194–5.

[7] See for example Sophia Moreau, "Beyond Discrimination Law: Realizing Equality Through Other Laws, Such as Tort Law" (2024) 4 American Journal of Law and Equality 427; Ruparelia Rakhi, "I Didn't Mean it That Way!: Racial Discrimination as Negligence" (2009) 44 Supreme Court Law Review 81.

[8] Milton R. Konvitz, *A Century of Civil Rights: With a Study of State Law Against Discrimination by Theodore Leskes* (Columbia University Press 1961) p 155. See also for example Civil Rights Act 1964 (United States); Canadian Human Rights Act 1977; Equality Act 2010 (United Kingdom); Equal Status Act 2000 (Republic of Ireland).

[9] John Gardner, "Liberals and Unlawful Discrimination" (1989) 9(1) Oxford Journal of Legal Studies 1.

[10] Ibid 3.

https://doi.org/10.1017/err.2025.10056 Published online by Cambridge University Press

and medical knowledge induced by tort law, rather than by the technology itself.[11] Froomkin, Kerr and Pineau assert that machine learning based diagnostic tools may one day exhibit higher success rates for diagnosing medical issues than human diagnosticians. The standard of care in medical negligence might therefore require the use of these more accurate machines in place of human physicians. Not only would this arguably lead to the loss of human diagnostic expertise and medical knowledge, but also a loss of human innovation, thus weakening the medical field as a whole. However, Froomkin, Kerr and Pineau also offer several convincing solutions, including that the standard of care could be renewed to require the use of machine learning based diagnostic tools *along with* a meaningful review by a human physician. This is quite an apt solution, as it successfully recognises the effectiveness of machine learning tools in the medical field, while also tempering the principles of medical negligence so as to not fully replace human expertise. Such a solution also takes into consideration that machine learning based diagnostic tools may also be inaccurate and are not guaranteed to take the most appropriate decision each time. By retaining both the automated system and the human physician as active diagnosticians, it would provide both machines and humans with a type of learning experience and check for each diagnosis made in respect of a patient.

In Chapter 8, Jones and Levy provide a fascinating analysis of the use of automated rule enforcement in the realm of professional sports as a replacement for human referees/rule-keepers. They highlight the potential effects which automated rule enforcement may have on the "soul" of sports, i.e., those highly regarded sports principles such as custom, tradition and sentiment, while also raising points concerning automated rule enforcement which may be applied more generally.[12] Jones and Levy correctly note that automated rule enforcement would undermine elements such as the concept of a "sporting chance" and the extra leeway occasionally given to a weaker or losing team/opponent, along with the desire to keep the flow of the game intact in line with the interests of the players and fans of the sport. Furthermore, they also note that opponents of automated rule enforcement are concerned about its potential to minimise the compelling human interactions between players, managers and referees, which are one of the hallmarks of entertaining sports games. While there is much debate surrounding whether it is proper to liken law to a game such as sports,[13] Jones and Levy successfully utilise the presence of automation systems in sports to raise compelling points more generally regarding the possibility of automated rule enforcement in society at large.

## Section 4 – Individual and relational rights: creativity, expression and privacy

The fourth and final section explores certain human rights which may be engaged by machines, particularly the rights of expression and privacy. This section begins with Craig and Kerr examining the idea of "AI authorship" in Chapter 9, where they argue that AI authorship itself is an oxymoronic idea and that seeking to assign authorship to robotic systems stems from an error regarding the ontology of authorship, since authorship is something which is intrinsically human.[14] This rejection of AI authorship largely rests upon the idea that anthropomorphising AI relies on incorrect presuppositions surrounding the nature of automated systems, drawing the reader back to those same critiques encountered in the book's first section. Craig and Kerr

[11] A. Michael Froomkin, Ian Kerr and Joelle Pineau, "When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning" 198.

[12] Meg Leta Jones and Karen Levy, "Sporting Chances: Robot Referees and the Automation of Enforcement" 230.

[13] John Gardner, "Legal Positivism: 5 ½ Myths" (2001) 46 American Journal of Jurisprudence 199, 227.

[14] Carys Craig and Ian Kerr, "The Death of the AI Author" 250.

https://doi.org/10.1017/err.2025.10056 Published online by Cambridge University Press

persuasively argue that AI cannot be an author, as authorship is fundamentally human, since it has at its core human communication. Thus, to attribute such a human behaviour to robotic systems would be to misunderstand both human nature and the nature of machine intelligence.

In Chapter 10, Massaro, Norton and Kaminski showcase a slightly different perspective in respect of AI and expression in examining the nature of the speech protections contained within the First Amendment of the United States (US) Constitution. They argue that, since the First Amendment is interpreted by the US Supreme Court to protect non-human speech with respect to companies,[15] the outputs of AI (such as tweets by Twitter's former chatbot "MS Tay") could conceivably enjoy such speech protection.[16] While the argument that AI outputs should receive speech protection might itself be unconvincing (particularly in light of the US Court of Appeals decision of *Thaler v Perlmutter* which held that AI outputs without a human author cannot be copyrighted),[17] the authors do well to demonstrate that it is not fantastical to construe the First Amendment in such a way.

Lastly, Chapter 11 contains an examination by Thomasen of the relationship between privacy, law and technology in the form of a feminist perspective on drone privacy regulation. Thomasen asserts that state approaches to drone regulation ought to be reconsidered, while also calling for a re-conceptualisation of privacy law more generally.[18] The inclusion of a feminist perspective on drone privacy regulation represents a fascinating and thought-provoking human-focused examination of privacy rights. Thomasen persuasively argues that a purported concern regarding women's safety may lead to an increase in drone surveillance, while simultaneously undermining women's ability to assert their privacy rights in public. In addition, although it is often stated that a tension exists between public safety and privacy, the author also contends that privacy concerns can be *included* in safety concerns, thus creating an opportunity for more publicly beneficial innovation. While this argument is well made, it does seem that safety and privacy represent competing concerns, with it being more accurate to view privacy as needing to be balanced against the community's interest of detecting and combatting wrongdoing. Nevertheless, the issues which Thomasen raises are pertinent and the claims surrounding the need for a re-evaluation of drone regulation (especially in respect of women's privacy) are quite cogent.

## Conclusion

*Robot Law: Volume II* effectively engages in a human-focused analysis of some pressing issues associated with the emergence of automated systems in society, marking a notable contribution to the academic literature in this area. The thematic coherence of *Robot Law: Volume II* is one of its major strengths, with a human-centric approach to robotic systems permeating throughout the entirety of this volume. The overriding claim that human welfare should be the main concern is convincing and successfully underpins the entirety of the volume, although, the commitment to a human-centric analysis may sometimes lead to the suggestion of well-meaning and principled, but unfeasible, solutions. Nevertheless, the book correctly recognises that human beings ought to be our focus as robots become increasingly present in society, while also

[15] See, for example, *Consolidated Edison Co v Public Service Commission* 447 U.S. 530 (1980); *Citizens United v Federal Election Commission* 558 U.S. 310 (2010).

[16] Toni M. Massaro, Helen Norton and Margot E. Kaminski, "SIRI-OUSLY 2.0: What Artificial Intelligence Reveals about the First Amendment" 286.

[17] 130 F.4th 1039 (D.C. Cir. 2025).

[18] Kristen Thomasen, "Beyond Airspace Safety: A Feminist Perspective on Drone Privacy Regulation" 322.

https://doi.org/10.1017/err.2025.10056 Published online by Cambridge University Press

maintaining a nuanced perspective in respect of this reality. There is not an assumption that the emergence of automated systems is inherently negative, nor is there a rush to come to such a conclusion. Rather, the editors successfully showcase nuanced examinations of the current and potential legal, social and ethical issues facing humanity in respect of machine intelligence. By engaging in a balanced analysis of these issues, the book does well to not fall into unnecessary or irrational techno-optimism or techno-pessimism.

**Competing interests.** The author has no conflicts of interest to declare.

https://doi.org/10.1017/err.2025.10056 Published online by Cambridge University Press