Patrick Lin, Ryan Jenkins and Keith Abney (eds.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (New York: Oxford University Press 2017).

Five years since the edition of *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney and George Bekey, its sequel, *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, is published by Patrick Lin, Ryan Jenkins and Keith Abney. While the first edition, as editors say in the Preface, was presented as the first comprehensive book on robot ethics, a lot has changed since then. After the first collection of papers was released, two interesting campaigns involving robots appeared. One was the Campaign to Stop Killer Robots in 2013, and the other was the Campaign Against Sex Robots that appeared in 2015. Both campaigns showed that there is a great deal of public concern involving robots. Editors covered these subjects already in the first edition, but they emphasise that there is a lot more going on with the new types of robots and areas of robotics that requires attention from robot ethics. First thing they mention are self-driving vehicles – a crucial case study that is use throughout the entire collection. There are also robots used in the law enforcement, e.g. when Dallas Police Department turned a robot into a mobile bomb and used it to kill dangerous suspects. Editors say that their goal with this collection was to "create a one-stop authoritative resource of the latest research in the field" (p. x) and give something more accessible to policymakers and the broader public. They decided to include in this edition more diverse researchers working on robot ethics, which was not so much the case with 2012 edition.

In the first part of the collection, devoted to "Moral and Legal Responsibility", autonomous cars occupy a central place. As much as it is expected for robot systems to be error-free programmed, it is a fact that even our less complexed everyday technology fails on a daily basis in much more controlled environments. Another issue is how should these robots be programmed and how should they make judgment calls in uncertain conditions. Because of the unpredictable behaviour caused by machine learning, the main question is: Who should be responsible if there was an accident? One way of preventing these scenarios is to programme robots with an ethical theory, but the question is which one. Vikram Bhargava and Tae Wan Kim ask themselves the same question in the chapter "Autonomous vehicles and moral uncertainty". Instead of advocating a particular theory, they present a methodology for choosing between ethical theories. Second chapter, "Ethics settings for autonomous vehicles", is by Jason Millar, who suggests that it is okay that robots do make some decisions, but also that when it comes to the important ones, users should guide the decision. Wulf Loh and Janina Sombetzky in the third chapter ("Autonomy and responsibility in hybrid systems: the example of autonomous cars") discuss responsibility in the context of self-driving cars, arguing that drivers or users remain responsible for the outcomes of ethical dilemmas revolving around them. In the fourth chapter ("Imputing driverhood: applying a reasonable driver standard to accidents caused by autonomous vehicles") Jeffrey K. Gurney refers to law and liability and holds technology developers as responsible as drivers of the self-driving car in case some harm is caused by the car. Existing legal frameworks are the main subject of the chapter by Trevor N. White and Seth D. Baum ("Liability law for present and future robotics technology"). They analyse the possibility of overdeveloping robots in the future and the insufficient existing legal framework. In the last chapter of the first part of the collection, "Skilled perception, authenticity, and the case against automation", David Zoller discusses the impact of technology on our relationship with reality, especially the question: "what's our moral responsibility to remain 'authentic' to ourselves?" (p. 3).

Second part of the collection – titled "Trust and Human-Robot Interactions" – brings articles focused on issues that can appear in close relationships between humans and robots. In their chapter "Could a robot care? It's all in the movement", Darian Meacham and Matthew Studley open this subject by analysing the question if there is a possibility for "carebots" to really care. Alexis Elder, in chapter "Robotic friends for autistic children: monopoly money or counterfeit currency?" points out that it has been shown that robots can help in treating patients with autistic spectrum. He also argues that this fake friendly behaviour can cause a moral harm to those patients and how it can be prevented with responsible design and use. Issue with *overtrust* is the main subject the chapter by Jason Borenstein, Ayanna Howard and Alan R. Wagner, "Pediatric robotics and ethics: the robot is ready to see you now, but should it be trusted?". Authors claim that there should be a responsibility in robotic community to examine tendency for children, parents, and healthcare workers to overtrust robots and they also suggest some strategies to reduce this risk. In "Trust and human-robot interactions", Jesse Kirkpatrick, Erin N. Hahn and Amy J. Haufler use multidisciplinary approach (comprising law, philosophy and neuroscience) to discuss the issue of trust in human-robot interactions. In chapter eleven, "White lies on silver tongues: why robots need to deceive (and how)", Alistair M. C. Isaac and Will Bridewell focus on deception as something that shouldn't be avoided in human-robot interaction, but as something that is rather a moral necessity. Authors suggest robots, in order to be successful social robots, should be able to use deceptive speech themselves. Kate Darling, in the chapter "'Who's Johnny?' Anthropomorphic framing in human-robot interaction, integration, and policy", considers whether we should encourage anthropomorphization of robots through framing or is this something that should be avoided it because it affects how people perceive and treat robots.

Third part of the collection – "Applications: From Love to War" – addresses the use of robots in two controversial areas: love and warfare. In chapter thirteen ("Lovotics: human-robot love and sex relationships"), Adrian David Cheok, Kasun Karunanayaka and Emma Yann Zhang consider the term "Lovotics", coined by David Levy (author of *Love and Sex with Robots*) for robot lovers that should be driven by artificial hormones making them "experience" complex human-like biological and emotional states. Authors provide a number of different perspectives on ethical permissibility of using "Lovotics" robots in sex industry. In chapter "Church-Turing lovers", Piotr Bołtuć investigates the future of robot lovers and considers issues from philosophy of mind and ethics revolving around sexbot he calls Church-Turing Lover, "a sex robot that could attain every functionality of a human lover" (p. 190). He claims that Church-Turing Lover could be like human being, but without inner life. Adam Henschke – in "The Internet of Things and dual layers of ethical concern" – suggests that human-robot interaction is not only sexual but has two layers that are both ethically problematic. One is the "physical layer" which raises questions of safety and risk issues. Another is the "information layer" which is about controlling information. Henschke discusses which of these layers should have priority when it comes to ethics of robots. How can we create a robot with moral reasoning? Michał Klincewicz tries to answer that question in his chapter "Challenges to engineering moral reasoners: time and context". He discusses the idea of an algorithm that combines philosophical moral theories and analogical reasoning programmed into a piece of software that could potentially engage in moral reasoning. The question that still needs to be answered, of course, is which moral theories should be programmed into this software. That is the question Brian Talbot, Ryan Jenkins and Duncan Purves are concerned with in chapter "When robots should do the wrong thing". Authors are wondering if we should allow the use moral views that we believe to be false and should they be consequentialist or deontological. They argue "that deontological evaluations do not apply to the actions of robots, for without phenomenal consciousness they lack the mental capacities required for agency" (p. 191). Last chapter of this part of the collection – Leonard Kahn's "Military robots and the likelihood of armed combat" – deals with the problem of possible increase of military activity under the influence of robotics in armed battles.

Last part of the collection is dedicated to "AI and the future of robot ethics", that is, to exploring possible implications of not only creating but also living alongside artificial beings possibly similar us. There is a danger of neglecting possible mental life of robots and a question of moral status of robots, which is the main subject of the chapter by Michael LaBossiere ("Testing the moral status of artificial beings, or 'I'm going to ask you some questions"). LaBossiere explores various tests that could determine if some artificial being has moral status, and to what being is it comparable to (the author argues in favour of presumption of such status. In chapter "Artificial identity", James DiGiovanna explores personal identity of artificial beings over time, especially of the so-called *para-persons* – the term he introduces for "beings that are able to change all of their person-making qualities instantaneously" (p. 290). In his chapter on "Superintelligence as superethical" Steve Petersen deals with the morality of the so-called "Superintelligence", originally introduced by Nick Bostrom. Petersen suggests that coherent reasoning is the most important thing that artificial intelligence should be equipped with because he considers it necessary for both intelligence and ethical behaviour. Shannon Vallor and George A. Bekey – in "Artificial intelligence and the ethics of self-learning robots" – consider potential risks of excessive use of AI and artificial intelligence and the effect it can have on people and society in general. They take as an example possibility of machine learning in autonomous cars where vehicle is learning by interacting with human driver which can be dangerous for the driver. Popular subject of displacing human workers by robots is also one of the main concerns for Vallor and Bekey. There are areas, however, where robotics is desirable. One of those areas is space exploration, discussed in Keith Abney's chapter "Robots and space ethics". A worry of the future world filled with technology is the subject of the last chapter of the collection: Jai Galliott's "On the Unabomber and robots: the need for a philosophy of technology geared toward human ends". Galliott deals with concerns by the "Unabomber", Ted Kaczynski, who ended up on FBI's "Most Wanted" list in America by his nationwide bombing campaign against people that were involved in development and use of modern technology. Inspired by worries Kaczynski had, Galliot wonders "how can we construct a philosophy of technology that is human-centric rather than one that risks subsuming human life to an abstract machine?" (p. 291).

Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence offers a variety of approaches to robot ethics and an excellent starting point for exploring this new and propulsive field of ethics. Although most of the topics covered in this edition have been addressed in its 2012 prequel (*Robot Ethics: The Ethical and Social Implications of Robotics*), it deals with some new important issues (e.g. robots for medical and caring purposes). Robots became an important part of our everyday lives and give rise to important ethical questions, like the question of responsibility in case of an accident that includes robot or the question of using robots for treating patients with mental illnesses. By addressing questions like these, robot ethics is important because, among other things, it can provide a much needed normative framework for the rapid development of robotics and AI and warn of long-term consequences they might have if not approached with responsibility and care. In this respect is this collection highly welcomed and it will be interesting to see what its future editions will bring.

Antea Anđelić
Centre for Croatian Studies
Borongajska cesta 83d
10000 Zagreb
Croatia
antea.andelic@gmail.com