Future Intelligent Autonomous Robots, Ethical by Design. Lessons Learned from Autonomous Cars Ethics

Gordana DODIG-CRNKOVIC
Division of Computer Science and Software Engineering, Mälardalen University
Västerås, Sweden
Department of Computer Science and Engineering, Chalmers University of Technology
Gothenburg, Sweden

Tobias HOLSTEIN
Division of Computer Science and Software Engineering, Mälardalen University
Västerås, Sweden

Patrizio PELLICCIONE
Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila,
L'Aquila, Italy

Jathoosh THAVARASA
Software Engineering, Efrei Paris
Villejuif, France
Division of Computer Science and Software Engineering, Mälardalen University
Västerås, Sweden

ABSTRACT

The "ethical by design" approach involves examining all stages of a lifecycle of technology to ensure that they are ethically justifiable and socially sustainable. Building on our work on the ethics of autonomous intelligent robocars, and studies of the literature on the ethics of robotics, we propose for robot applications a set of values and ethical principles including safety, security, privacy, transparency, and explainability, accountability, fairness, human control, well-being, autonomy and freedom, and sustainability. This may help stakeholders in the field of intelligent autonomous robotics to connect ethical principles with their applications. Most ethical considerations we identified in our work on autonomous cars are relevant to all AI-powered robots, but robots require additional examination depending on their application domain, such as social robots (care robots, personal companions, robots used in education, health care, elderly care, education, entertainment, chat-bots), industrial robots, etcetera. Thus, existing ethical frameworks need to be applied in a context-sensitive way, by assessments in interdisciplinary, multi-competent teams through multi-criteria analysis. Furthermore, we argue for the need for continuous development of ethical principles, guidelines, and regulations, informed by the progress of technologies and involving relevant stakeholders. This implies designing the socio-technical system as an intelligent learning ecology.

Keywords: Ethics, Artificial Intelligence, Autonomous Robots, Intelligent Robots, Roboethics, Autonomous cars, Emerging Technologies, ELSA

1. INTRODUCTION

This article builds on the findings of our book chapter [1] on ethical and social aspects of self-driving cars, which are robots classified as "mobile service robots" [2]. Those are vehicles capable of perceiving their environment and driving without (or with little) human intervention. They combine advanced sensing, controlling, and artificial intelligence with autonomous safety-critical decision-making. Ethical aspects of autonomous cars (also called self-driving cars, autonomous vehicles, driverless cars, automated cars, or robocars) have lately got a lot of attention from the general public, ethicists, researchers, industry, and decision-makers [3].

Our studies on autonomous car ethics led us to the insight that the same approach may be applied to intelligent autonomous robots in general, having in mind that autonomous cars are a special type of intelligent autonomous robots. The research question was: how much of our recommendations for the ethical analysis of autonomous cars can apply to the ethical analysis of other types of robots that present an important emerging technology?

"As a game-changing technology, robotics naturally will create ripple effects through society.", according to Lin, Abney, and Bekey, [3]. The impact of robotics technology on society is significant and far-reaching, potentially leading to major changes in everyday life, business, and culture. Therefore, it is crucial to examine the effects on ethics, law, and policy within ELSA (Ethical, Legal, and Social Aspects) studies, through the fields known as robot ethics, roboethics, and ethics of robotics, which are closely related to AI ethics, machine ethics, technology ethics, and ethical technology.

Peter Asaro posed the fundamental question: "What should we want from a robot ethic?" [4]. His answer is that we need to develop robots that progressively acquire stronger ethical abilities. The primary focus of robot ethics should be avoiding the harm caused by robots. The assignment of responsibility in complex socio-technical systems should be governed by legal theory, [4].

In addition to the ethical dilemmas arising from the use of robots, Vincent Müller [5] also addresses the topic of the "ethical status of robots" and "machine ethics" as a means of making robots ethical. Some authors [6]–[8] suggest incorporating ethics in the design of robots, referred to as "ethical by design".

Other notable references on the ethics of intelligent robotics include [9]–[12]. Rafael Capurro [10] provides an overview while emphasizing the intercultural perspective on ethics and robotics. Gianmarco Veruggio and Fiorella Operto [12] examine the ethical and social implications of robotics in their chapter in the Springer Handbook of Robotics. Asaro raises important questions about the role of a robot ethic, and the authors of this work agree with his perspectives on the three main topics of "the ethical systems built into robots, the ethics of people who design and use robots, and the ethics of how people treat robots." This is achieved by approaching robots as socio-technical systems in their different contexts.

2. LESSONS LEARNED FROM AUTONOMOUS CARS ETHICS. ETHICAL ANALYSIS WITH REQUIREMENTS, CHALLENGES, AND POSSIBLE APPROACHES

Our book chapter [1] based on the literature on value-based design and current guidelines [13]–[18], extracted the list of topics/requirements/values of relevance for real-world automated/self-driving cars that we presented in two tables, for technical and social ethical challenges, respectively. A similar convergence of ethics requirements has been observed in AI-ethics literature globally [19].

In our earlier work [20]–[22], we developed the approach to the practice-oriented, real-world ethics for self-driving cars. We started from the most important technological and societal requirements based on extensive literature studies and identified challenges. With practical applications in mind, we searched for a consensus among international studies about the most important ethical issues as requirements for intelligent cars. After producing such a list summarized in table form, we identified challenges and approaches to addressing them. We contacted leading experts in the field, as well as colleagues researching interaction design, software engineering, and the philosophy of technology. In the series of seminars and a dialogue with the general public, through the discussions in the public lectures, we tested and concretized our ideas. The details of this work are given in [1].

In the present work, starting with the ethical aspects in [1], ethical aspects of technical and social challenges presented in Table 1, and Table 2 are modified from self-driving cars to correspond to the case of robotics systems in general and reproduced from [1] with permission.
In this transition, we considered the ethics requirements found in research on robots ethical by design, medical robots, industrial robots, and other intelligent robots to which we delegate responsibilities [8], [23]–[28]. We have also taken into account the findings of works of [3], [5], [9]–[12], regarding ethical concerns and requirements for robotics.

Furthermore, we have consulted the whitepaper from the UK Robotics and Autonomous Systems Network [29] that proposes seven ethical concerns of robotics (Bias, Deception, Employment, Opacity, Safety, Oversight, and Privacy). All except for deception are part of our original framework, [1], [20], [21]. In the present work, we added deception as an important concern important for social robotics, particularly for intelligent robotic companions in case of vulnerable users [30].

The comparison between Tables 1 and 2 and the corresponding tables for autonomous cars [1] suggests that self-driving robotic cars provide a good baseline of ethics requirements since they have almost all of the requirements that apply to other autonomous intelligent robotic systems. However, particular application domains, such as social/ companion robots, medical robots, transport, entertainment, or industry applications bring particular ethical challenges and require specific analysis.

Table 1. Summary of ethical aspects of technical challenges of intelligent autonomous robots, grouped by the requirement.

Requirements	Challenges	Approaches
Safety	Hardware and software adequacy. Vulnerabilities of machine-learning algorithms. Control of trade-offs between safety and other factors (like economic) in the design, manufacturing and operation. Possibility of intervention in case of major failure of the system and graceful degradation. Systemic solutions to guarantee safety in organizations (regulations, authorities, safety culture).	Setting safety as the first priority. Learning from the history of automation. Learning from experience of use. Specification of how a system will behave in cases when autonomous operation is disabled (safe mode). Preparedness for handling "loss of control" situations-autonomous systems running amok. Regulations, guidelines, standards being developed as the technology develops.
Security	Minimal necessary security requirements for deployment of the system. Security in the context and connections. Deployment of software updates. Storing and using received and generated data in a secure way.	Technical solutions to guarantee minimum security under all foreseeable circumstances. Anticipation and prevention of the worst-case scenarios. Accessibility of all data, even in the case of accidents, learning from experience.
Privacy	Trade-offs between privacy and data collection/recording and storage/sharing.	Following/applying legal frameworks to protect personal data, such as GDPR.
Transparency	Information disclosure, what and to whom. Transparency of algorithmic decision making. Transparency in the techno-social ecosystem.	Assurance of transparency and insight into decision making. Active sharing of knowledge to ensure the interoperability of systems and services.
Algorithmic Fairness	Algorithmic decision making is required to be fair and not to discriminate on the grounds of race, gender, age, wealth, social status etc.	This requirement is related to transparency of decision making and expectation of explainability of the ground for decision making (e.g., right of explanation is enforced by GDPR as stated in Recital 71 [31]).

Requirements	Challenges	Approaches
Reliability	Reliability of hardware, sensors and software and need for redundancy. Reliability of required networks and solution for the case when the network is unavailable.	Definition of different levels for reliability, such as diagnostics, hardware, sensors, software, and external services, set the ground for reliability measures of the system and its components. The standardized process required to shift from fail-safe to fail-operational architecture.
Environmental Sustainability	Environmental sustainability ethics refers to new ways of production, use, and recycling for robotic systems.	Production, use, and disposal/recycling of technology rise sustainability issues (materials, processes, energy) that must be addressed.
Intelligent Behavior Control	Intelligent behavior may lead to unpredictable situations resulting from learning and autonomous decision making.	Development of self-explaining capability and other features ensuring desired behavior in intelligent software.
Transdisciplinary -Systemic Approach	Ethics in/for/by/through/of design. Requirements engineering, software-hardware development, learning, legal and social aspects, software-hardware interplay.	Adoption of transdisciplinarity and system approaches is increasing and should be strengthened even more.
Quality	Quality of components. Quality of decision making. Lifetime and maintenance. QA process. Adherence to ethical principles/guidelines.	Ethical deliberations included in the whole process starting with design and development. Ethics-aware decision making to ensure ethically justified decisions.

Table 2. Summary of ethical aspects of social and individual challenges of intelligent autonomous robots, grouped by the requirement.

Requirements	Challenges	Approaches
Non-maleficence	Risk of technology causing harm. Disruptive changes in the labor market. Transformation of related businesses, markets and business models (manufacturers, insurances, etc.).	Partly covered by technical solutions. Preparation of strategic solutions for people losing jobs. Learning from historic parallels to industrialization and automatization.
Stakeholders' involvement	Participation of different stakeholders – from professionals in designing, developing, maintenance and recycling, to users, and the general public.	Active involvement of stakeholders in the process of design and requirements specification as well as decisions of their use.
Beneficence	Establishment of values and priorities: Ensuring that shared public values will be embodied in the technology, with interests of minorities taken into account.	Initiatives such as "AI for good" exemplify this expectation that new technologies not only do not cause harm but actively do good for its stakeholders.
Responsibility and Accountability	Assignment and distribution of responsibility and accountability as some of central regulative mechanisms for the development of new technology. They should follow ethical principles.	The Accountability, Responsibility, and Transparency (ART) principle based on a Design for Values approach includes human values and ethical principles in the design processes [32].
Freedom and Autonomy	Freedom of choice for a human hindered or disabled by the system.	The freedom of choice determined by regulations. Determination and communication of the amount of control that humans have.
Social Sustainability	In the domain of business, social sustainability is about identifying and managing business impacts on people. In the case of social robotics, the impact of social robots on society is central.	Pursuing social equity, community development, social support, human rights, labor rights, social responsibility, social justice, etc.
Social Fairness	Ascertaining fairness of the socio-technological system.	Fairness of the decision-making. Related to inclusiveness, transparency and explainability.
Dignity and Solidarity	Challenges come from the lack of a common wholistic view.	This requirement should apply to the entire socio-technological system.
Social Trust	Establishing trust between humans and robots as well as within the social system involving robots.	Further research on how to implement trust across multiple systems. Provision of trusted connections between components as well as external services.
Justice: legislation, standards, norms, policies, and guidelines	Keeping legislation up to date with the current level of technology, and proactively meeting emergent developments. Creating and defining global legislation frameworks. Including ethical guidelines in design and development processes.	Legislative support and contribution to global frameworks. Ethics training for involved stakeholders. Establishment and maintenance of a functioning socio-technological system in addition to functional safety standards.
Cognitive and psychological effects of social/ companion robots on humans	Research and frequent evaluation of Personal Integrity, Cognitive Load, Deception, and Human Control, throughout the lifetime of the system in real social context or companionships.	Further research on how social robots and especially increasingly intelligent and human-like robot companions affect users. Solid understanding of effects, after stakeholders' interests are taken into consideration should be followed by regulation/legislation. Humanoid or zoomorphic robots may cause emotional attachment to some users. "Robots should not be designed in a deceptive way to exploit vulnerable users" [30].

3. COMPARATIVE ANALYSIS OF THE APPLICATION OF THE AUTONOMOUS CARS ETHICS FRAMEWORK EXTENDED TO OTHER CLASSES OF INTELLIGENT ROBOTS

In order to explore major ethical challenges for the development of future intelligent autonomous robotic systems, we have chosen the different classes of robots, representative of various kinds of ethical aspects, and compared their ethical challenges with our approach developed for autonomous robotic cars. We covered intelligent autonomous robots used in the following applications: transport, social, industrial, medical, military, and entertainment.

The given ethical requirement might be more or less relevant for a certain application area. For example, safety is highly relevant for transport but less relevant for entertainment. In addition to the general relevance, requirements can be prioritized differently within an organization, i.e., by project and development teams.

Three of the authors of this article ranked the different requirements independently to see similarities and differences in assessment, which resulted in Figure 1. The results showed that even among the authors, requirements are regarded as more or less important for a certain robot class, depending on the background, experiences, and perspective, of the assessor. It is not an uncommon situation that expert opinions in real life differ between experts, so one of the challenges to be met is reaching a consensus on the prioritization of requirements.

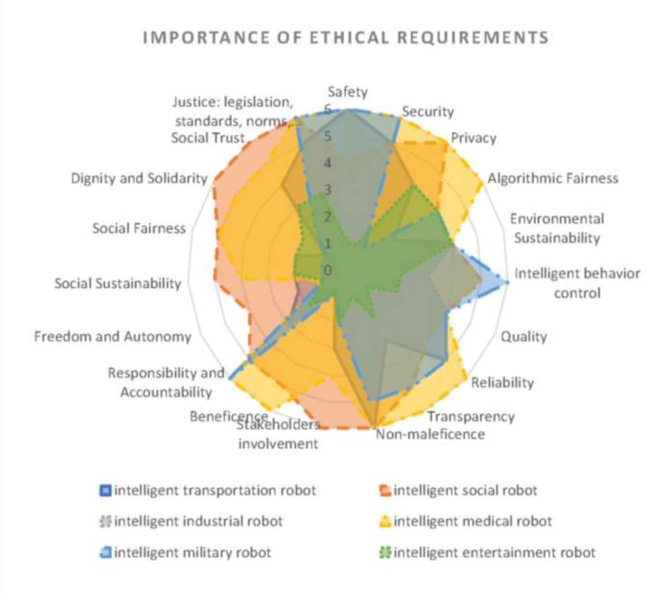

Figure 1. Spider web diagram illustrating the importance of different ethical requirements in different types of robots based on the independent ranking by the three authors.

Ethical aspects for any class of robots will require evaluation by interdisciplinary, multi-competent teams through multi-criteria analysis. In addition, it is imperative to monitor the prioritization and ranking of ethical requirements throughout the development to avoid disregarding certain requirements over others.

Applications of the framework, priorities, and actual processes are also context-sensitive and must be decided for a given application. They will evolve over time and must keep pace with technology development.

One approach to monitor ongoing development is a supervision group. Waymo has established an interdisciplinary group called the Waymo Safety Board, to maintain and improve their safety [33]. The group consists of executive leaders from Safety, Engineering, and Product teams.

Applying an ethical approach such as ours for bridging the "from principles to practice" gap, [19] points out in 'A way forward' section: "there is a need for a more coordinated effort, from multi-disciplinary researchers, innovators, policymakers, citizens, developers, and designers, to create and evaluate new tools and methodologies, in order to ensure that there is a 'how' for every 'what'". That is exactly what is needed at this stage of development.

4. CONTEXT-SPECIFIC APPLICATION OF THE PROPOSED ETHICAL APPROACH FOR INTELLIGENT ROBOTIC SYSTEMS

As argued in [19], we need not only principles but also the connection to the practices. Tables 1 and 2 summarize ethical challenges related to the technical and social aspects of robotics systems. We distinguish between three contexts in which robots appear:

1. the context of application of existing technology
2. the context of design/production/maintenance of new technology, and
3. the context of oversight of the whole socio-technical ecology.

4.1. **The context of application of already existing technology**

In a systemic view of emergent technology, we cannot analyze software in separation from hardware, and we cannot disconnect engineering from human, social, and organizational factors [34], [35]. We have shown how this applies to self-driving cars. The way robots interact with the real world can be presented in the same abstract phases of "sense, decide, act", which form an iterative process, supported and extended by AI. The context and perspective towards an emergent technology set the boundaries for each challenge in each of the phases. E.g., privacy concerns for a robot that does not "sense" (with limited perception, recording, etc.) or due to limited context of use (e.g., industrial robot) are less important than privacy concerns for a self-driving car with more capabilities to sense/ record information and a less restricted context of use. Privacy for self-driving cars is less important than privacy for social robots which relate closely to humans and possess capabilities of sensing, recording, local processing, and transmitting information.

4.2. **The context of design/production/maintenance**

For different types of robots, detailed analysis and concretization of design/production/maintenance for their specific contexts must be done. While [36] presents the value-based view of ethical IT innovation, [19] provides useful advice helping to bridge the conceptual gap between ethical principles and engineering practice. What we found to be most important in this context is collaboration, communication, systems thinking, and solving real-world problems guided by the needs of stakeholders. Globally, priorities and values are constantly changing and are negotiated in light of new developments. Ethical aspects are present in the technology from research to practice, design, development, implementation, testing, verification, and management of the entire system lifecycle, in the iterative process of continuous improvement [1], [8], [21], [34], [35].

Engineers concretize the steps for every requirement for each specific context. Frameworks provide guidelines for the approach and require contextual domain knowledge for the specific application. E.g., safety requirements for medical robots are clearly different from the safety requirements for nano-robots which are examples of robotic systems that pose new ethical challenges because of their microscopic size, the ability of self-replication, mutation, and the possibility to easily escape out of control.

For example, in the military application area, there is current demand among ethicists and roboticists to completely ban intelligent autonomous weapons because of the unacceptability of their consequences. So, the right question about certain classes of intelligent robots will be whether they are acceptable at all, rather than how to build them in an ethical way when there is no ethical way. See the arguments provided by prominent AI researcher Stuart Russel [37] and roboticist Noel Sharkey [38].

4.3. **The context of oversight of the socio-technical ecology**

Finally, it is essential to point out the importance of the oversight of “the total ecology of the socio-technological system, where ethics is ensured through education, constant information sharing and negotiation of priorities in the value system”, [1]. The development of technologies is followed by the interest of the general public and other stakeholders, which is followed by legislation, rules, and guidelines [30], [39]. As argued in [8], this process is a recursive socio-technological learning process, where experiences and new developments lead to improved oversight/regulation/legislation. It is necessary to ensure the transparency of those processes to enable independent evaluations and efficient learning.

When applying an ethical framework, such as ours [1] or [18] for a good technology-based society, in complex and often unforeseen circumstances of real-world applications, it is important to know how to interpret such general requirements to make ethical choices under uncertainty [40].

The approach described in this article recognizes that the ethical implications of robots and AI systems depend on the context in which they are used. Obviously, an autonomous vehicle that operates on public roads raises different ethical considerations than a robot used in a manufacturing plant. Therefore, it is crucial to consider the specific context in which technology is used when applying ethical principles and frameworks.

In relation to the context-dependence of the ethical approaches, we can distinguish between technology ethics and ethical technology as interrelated concepts that pertain to the ethical principles and values that guide the design, development, and use of technology. Technology ethics involves the study of ethical considerations involved in the creation and implementation of technology, while ethical technology refers to the use of technology in compliance with ethical principles.

Moreover, the approach emphasizes the need for ongoing development and adaptation of ethical principles, guidelines, and regulations to keep pace with the advancement of technologies. As new technologies emerge and existing technologies evolve, new ethical challenges may arise, and ethical principles must be updated to address these challenges. This process should involve all relevant stakeholders, including technologists, ethicists, policymakers, and the public, to ensure that ethical considerations are incorporated into the development and use of new technologies.

In short, we highlight the importance of context sensitivity and ongoing ethical development in the application of ethical principles and frameworks to robots and AI systems. By considering the specific context in which technology is used and continually updating ethical principles to reflect technological advancements, we can ensure that these technologies are developed and used in a responsible and ethical manner.

5. ANTICIPATORY ETHICS FOR EMERGING TECHNOLOGIES

Having in mind the emergent character of intelligent robotic technologies, one important aspect is their continuous and rapid development related to uncertainties and anticipation. Studying anticipatory ethics for emerging technologies, Philip Brey [41] identified the main current approaches that utilize forecasting to evaluate the ethics of emerging technologies: ethical technology assessment (eTA), the techno-ethical scenarios approach, the ETICA approach, and he proposed anticipatory technology ethics (ATE). These methods aim to identify and evaluate the potential ethical implications of new emerging technologies and to develop strategies for addressing them. By engaging diverse stakeholders and utilizing foresight exercises, these approaches help ensure that emerging technologies are developed and deployed in an ethical, responsible, and sustainable manner.

The approaches studied are:

1. Ethical Technology Assessment (eTA): eTA proposed by Palm and Hansson, based on technical assessment (TA) involves a multidisciplinary team of experts who use foresight exercises and scenario analysis to identify and evaluate the ethical, social, and environmental impacts of emerging technologies. It is focusing on continual assessment more than predictions of a far future.

2. Techno-Ethical Scenarios Approach: The Techno-Ethical Scenarios approach of Boenink, Swierstra and Stemerding involves the creation of hypothetical scenarios that explore the possible outcomes and consequences of a given technology. These scenarios are then used to identify potential ethical challenges and to develop strategies that can help policymakers to anticipate ethical controversies for addressing them.

3. ETICA Approach: The ETICA approach proposed by Stahl is based on the principles of Responsible Research and Innovation (RRI) and involves engaging a wide range of stakeholders in the development and assessment of emerging technologies. It aims to identify and address the ethical implications of emerging technologies in a way that is transparent, inclusive, and responsive to societal needs and values.

4. Anticipatory Technology Ethics (ATE): ATE proposed by Brey involves using foresight methods and ethical analysis to anticipate and proactively address the ethical challenges that may arise from emerging technologies. It distinguishes three levels of ethical analysis (the technology, artifact, and application level) and three objects of ethical analysis (things, properties, or processes) that ATE focuses on. Different forecasting methods are proposed for different levels and objects of analysis.

Overall, the above four approaches provide frameworks for anticipatory ethics of emerging technologies taking into account uncertainties and ensuring that they are developed and deployed in a way that is responsible, ethical, and sustainable.

6. CONCLUSIONS

This article discusses the necessity of considering ethical issues in relation to robots, as a prerequisite for building trust in their future use. It is based on our previous research findings on the ethical implications of autonomous cars, where we identified the following ethical aspects in technical challenges of intelligent autonomous robots: safety, security, privacy, transparency, algorithmic fairness, reliability, environmental sustainability, intelligent behavior control, transdisciplinary and systemic approach, and quality. Equally important, ethical aspects in social and individual challenges recognized are non-maleficence, stakeholders' involvement, beneficence, responsibility and accountability, freedom and autonomy, social sustainability, social fairness, dignity and solidarity, social trust, justice: legislation, standards, norms, policies, and guidelines, as well as cognitive and psychological effects of social/ companion robots on humans.

Before the question of how to build ethical technology in an ethical way comes the question if it is possible. For example, the open question of intelligent autonomous weapons currently prompted ethicists and roboticists to propose a complete ban on intelligent autonomous weapons. Thus, the first question to ask is whether certain intelligent autonomous robotic technology is acceptable at all, rather than how to build it in an ethical way if there is no ethical way, [37], [38].

When technology can be made beneficial for society and individuals, the next step is to understand how its ethics can be secured. We argue that the ethics of intelligent autonomous robots must permeate application, design, production, and/or maintenance and oversight within the corresponding techno-social system, and must be based on learning from experience [1], [8], [21], [42].

Both studies from the literature [35] and our own study emphasize the need for a system-level approach involving the entire software-hardware system as well as human, organizational, and social factors.

With the constantly evolving, emergent nature of intelligent robotic technologies, a crucial aspect is their development includes anticipation and consideration of uncertainties. Brey [41] studied anticipatory ethics for emerging technologies and identified approaches that provide frameworks for anticipatory ethics ensuring that they are developed and deployed in an ethical way.

At present, there is a gap between general principles and their specific, context-dependent implementations when making multi-criteria decisions and identifying key ethical considerations, [43]. This issue can only be resolved through the collaboration of multidisciplinary teams with the appropriate expertise, working within the specific context in question. [19].

AI policies and their implementation can be monitored globally through resources such as the OECD Policy Observatory [44] and the "Ethical, Legal and Socio-Economic Issues" directory of the EU Robotics Topics Group, [45] which contains relevant policy documents.

We argue that ethical principles, guidelines, and assessments, as well as regulatory documents, must be continually updated and developed in line with technological advancements and must involve input from all relevant stakeholders. Incorporating ethical considerations into the development and use of intelligent autonomous robots is essential for building trustworthy future technology systems.

There is still much work to be done to ensure ethical considerations are integrated throughout the entire lifecycle of technology and in all aspects of its development and deployment in society. This process should involve ongoing learning and adaptation.

7. REFERENCES

[1] T. Holstein, G. Dodig-Crnkovic, and P. Pelliccione, “Steps Toward Real-World Ethics for Self-Driving Cars,” in *Machine Law, Ethics, and Morality in the Age of Artificial Intelligence.*, Steven John Thompson, Ed. IGI Global, 2021, pp. 85–107.

[2] M. Ben-Ari and F. Mondada, “Robots and Their Applications,” in *Elements of Robotics*, 2018, pp. 1–20.

[3] P. Lin, K. Abney, and G. A. Bekey, *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press, 2011.

[4] P. M. Asaro, “What should we want from a robot ethic?,” in *Machine Ethics and Robot Ethics*, 2020.

[5] V. C. Müller, “Ethics of Artificial Intelligence and Robotics,” in *The Stanford Encyclopedia of Philosophy*, Winter 202., E. N. Zalta, Ed. Stanford University, 2020.

[6] W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press, 2009.

[7] F. Operto, “Ethics in Advanced Robotics,” *IEEE Robot. Autom. Mag.*, vol. 18, no. 1, pp. 72–78, Mar. 2011, doi: 10.1109/MRA.2010.940155.

[8] G. Dodig-Crnkovic and B. Çürüklü, “Robots: Ethical by design,” *Ethics Inf. Technol.*, vol. 14, no. 1, pp. 61–71, 2012, doi: 10.1007/s10676-011-9278-2.

[9] S. G. Tzafestas, *Roboethics - A Navigating Overview*, vol. 79. Springer International Publishing, 2016.

[10] R. Capurro, *Ethics and Robotics An Intercultural Perspective*. Akademische Verlagsgesellschaft, 2009.

[11] P. M. Asaro, “What Should We Want From a Robot Ethic,” *Int. Rev. Inf. Ethics*, vol. 6, no. 12, pp. 9–16, 2006.

[12] G. Veruggio and F. Operto, “Roboethics: Social and Ethical Implications of Robotics,” in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1499–1524.

[13] B. Friedman and P. H. Kahn Jr., “The Human-computer Interaction Handbook,” J. A. Jacko and A. Sears, Eds. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 2003, pp. 1177–1201.

[14] B. Friedman, P. H. Kahn, A. Borning, and A. Huldtgren, “Value Sensitive Design and Information Systems,” *Early engagement and new technologies: Opening up the laboratory*. Springer Netherlands, pp. 55–95, 2013, doi:

10.1007/978-94-007-7844-3_4.
[15] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, “Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,” 2019.
[16] E. G. on Ethics in Science, N. Technologies, and others, “Artificial Intelligence, Robotics and ‘Autonomous’ Systems.” Brussels, 2018.
[17] E. C. H.-L. E. G. on Artificial Intelligence (AI HLEG), “Ethics Guidelines for Trustworthy AI,” *High-Level Expert Gr. Artif. Intell.*, 2019.
[18] L. Floridi *et al.*, “AI4People---An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,” *Minds Mach.*, vol. 28, no. 4, pp. 689–707, Dec. 2018, doi: 10.1007/s11023-018-9482-5.
[19] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices,” *Sci. Eng. Ethics*, 2020, doi: 10.1007/s11948-019-00165-5.
[20] T. Holstein, G. Dodig-Crnkovic, and P. Pelliccione, “Real-world ethics for self-driving cars,” 2020, doi: 10.1145/3377812.3390801.
[21] T. Holstein, G. Dodig-Crnkovic, and P. Pelliccione, “Ethical and social aspects of self-driving cars,” *arXiv*. 2018.
[22] T. Holstein and G. Dodig-Crnkovic, “Avoiding the intrinsic unfairness of the trolley problem,” 2018, doi: 10.1145/3194770.3194772.
[23] R. Giovagnoli, A. Crucitti, and G. Dodig-Crnkovic, “ROBOTICS AND THE QUALITY OF LIFE: THE CASE OF ROBOTICS ASSISTED SURGERY,” *Philos. Inq.*, vol. 7, no. 1, pp. 77–87, Mar. 2019.
[24] M. Georgieva and G. Dodig-Crnkovic, “Who Will Have Irresponsible, Untrustworthy, Immoral Intelligent Robot?,” 2011.
[25] B. Çürüklü, G. Dodig-Crnkovic, and B. Akan, “Towards industrial robots with human-like moral responsibilities,” 2010, doi: 10.1145/1734454.1734484.
[26] G. Dodig-Crnkovic and D. Persson, “Sharing moral responsibility with robots: A pragmatic approach,” in *10th Scandinavian Conference on Artificial Intelligence, SCAI 2008*, 2008, vol. 173, pp. 165–168, [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84867516739&partnerID=40&md5=7b8ba0d6d997abcf6a958778854b3105.
[27] G. Dodig-Crnkovic, “Delegating Responsibilities to Intelligent Robots,” in *Workshop on Roboethics*, 2009, p. 17.
[28] G. Dodig-Crnkovic, “Information Ethics for Robotic Agents,” in *IACAP 2010. VIII European Conference on Computing and Philosophy*, Oct. 2010, pp. 4–6.
[29] A. F. Winfield, J. McDermid, V. C. Müller, Z. Porter, and T. Pipe, “Ethical Issues for Robotics and Autonomous Systems,” in *UK-RAS White papers*, 2019, pp. 1–18.
[30] M. Boden *et al.*, “Principles of robotics: regulating robots in the real world,” *Conn. Sci.*, 2017, doi: 10.1080/09540091.2016.1271400.
[31] E.U., “Regulation (E.U.) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General ,” 2016.
[32] V. Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer, 2019.
[33] N. Webb *et al.*, “Waymo’s Safety Methodologies and Safety Readiness Determinations.” 2020.
[34] N. Leveson, *Engineering a safer world: Systems thinking applied to safety*. MIT press, 2011.
[35] N. Leveson, “Are You Sure Your Software Will Not Kill Anyone?,” *Commun. ACM*, vol. 63, no. 2, pp. 25–28, Jan. 2020, doi: 10.1145/3376127.
[36] S. Spiekermann, *Ethical IT Innovation*. New York: Auerbach Publications, 2015.
[37] S. Russell, “AI weapons: Russia’s war in Ukraine shows why the world must enact a ban,” *Nature*, vol. 614, pp. 620–623, 2023, doi: 10.1038/d41586-023-00511-5.
[38] N. Sharkey, “The Automation and Proliferation of Military Drones and the Protection of Civilians,” *Law, Innov. Technol.*, vol. 3, no. 2, pp. 229–240, Dec. 2011, doi: 10.5235/175799611798204914.
[39] A. F. Winfield, K. Michael, J. Pitt, and V. Evers, “Machine ethics: The design and governance of ethical ai and autonomous systems,” *Proc. IEEE*, 2019, doi: 10.1109/JPROC.2019.2900622.
[40] L. Dennis, M. Fisher, M. Slavkovik, and M. Webster, “Ethical choice in unforeseen circumstances,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8069 LNAI, pp. 433–445, doi: 10.1007/978-3-662-43645-5_45.
[41] P. Brey, “Anticipatory Ethics for Emerging Technologies,” *Nanoethics*, vol. 6, no. 1, pp. 1–13, 2012.
[42] V. Charisi *et al.*, “Towards Moral Autonomous Systems,” *CoRR*, vol. abs/1703.0, 2017.
[43] S. Langman, N. Capicotto, Y. Maddahi, and K. Zareinia, “Roboethics principles and policies in Europe and North America,” *SN Appl. Sci.*, vol. 3, p. 857, 2021, doi: 10.1007/s42452-021-04853-5.
[44] “OECD. National AI policies & strategies.,” 2022. https://www.oecd.org/science/an-overview-of-national-ai-strategies-and-policies-c05140d9-en.htm%0A.
[45] “euRobotics Topics Group ‘Ethical Legal and Socio-Economic Issues.’ Policy Documents & Institutions - ethical, legal, and socio-economic issues of robotics and artificial intelligence,” 2022. https://www.pt-ai.org/TG-ELS%0A.