

# An Ethical Framework for Smart Robots

Mika Westerlund

*“Never underestimate a droid.”*

Leia Organa

Star Wars: The Rise of Skywalker

This article focuses on “roboethics” in the age of growing adoption of smart robots, which can now be seen as a new robotic “species”. As autonomous AI systems, they can collaborate with humans and are capable of learning from their operating environment, experiences, and human behaviour feedback in human-machine interaction. This enables smart robots to improve their performance and capabilities. This conceptual article reviews key perspectives to roboethics, as well as establishes a framework to illustrate its main ideas and features. Building on previous literature, roboethics has four major types of implications for smart robots: 1) smart robots as amoral and passive tools, 2) smart robots as recipients of ethical behaviour in society, 3) smart robots as moral and active agents, and 4) smart robots as ethical impact-makers in society. The study contributes to current literature by suggesting that there are two underlying ethical and moral dimensions behind these perspectives, namely the “ethical agency of smart robots” and “object of moral judgment”, as well as what this could look like as smart robots become more widespread in society. The article concludes by suggesting how scientists and smart robot designers can benefit from a framework, discussing the limitations of the present study, and proposing avenues for future research.

## Introduction

Robots are becoming increasingly prevalent in our daily, social, and professional lives, performing various work and household tasks, as well as operating driverless vehicles and public transportation systems (Leenes et al., 2017). However, given that the field of robotics has grown to become interconnected with other technologies, it seems more and more difficult to provide a commonly accepted definition of a robot (Leenes et al., 2017). According to Ishihara and Fukushi (2010), the word “robot” was first introduced in Karel Capek’s 1921 play that dealt with conflict between human beings and robots, that is, artificial persons molded out of chemical matter. Belanche et al. (2019) add that the word “robot” originates from the Czech word “robota”, which means “forced labor;” or, put another way, “slavery”. Thus, robots are often seen as mechanical devices programmed to perform specific physical tasks for human beings. That said, many of today’s robots are no longer mere slaves - unpaid labor that respond only to human requests - but increasingly embody autonomy and progressive “decision making”

capabilities (Lichocki et al., 2011; Petersen, 2007). Hence, Lin et al. (2011) define a “robot” as an engineered machine that senses, thinks, and acts, thus being able to process information from sensors and other sources, such as an internal set of rules, either programmed or learned, that enables the machine to make some “decisions” autonomously. The degree of autonomy, we will see, is a crucial indicator of how “smart” a robot is or is not. Nevertheless, the notion of anthropomorphizing robots, or treating them “as persons”, is not under consideration in this paper.

Advancements in robotics have led to the emergence of “smart robots”, which are defined as autonomous artificial intelligence (AI) systems that can collaborate with humans. They are capable of “learning” from their operating environment, experiences, and human behaviour feedback in human-machine interaction (HMI), in order to improve their performance and capabilities. The smart robot market was valued at USD 4.5 billion in 2017, and is expected to reach USD 15 billion by 2023 (*Market Research Future*, 2019). Among robotics engineers, the increased focus on HMI and use

# An Ethical Framework for Smart Robots

*Mika Westerlund*

of AI components has shifted the attention from “mechanoids”, that is, robots with a machine-like appearance, towards the development of human-shaped (“humanoid”) and animal shaped (“animaloid”) smart robots (Kumari et al., forthcoming; Mushiaki, 2013/2014). Belanche et al. (2019) note that while humanoids may only have stylized human features, “droids” (android if male, gynoid if female) have an appearance and behaviour closer to a real human being, at least on the technical level. However, robots’ appearances are less important than how easy they are to communicate with, to train to do what we want, and how well they solve tasks. Thus, design and usability matter significantly when choosing what types of smart robots we will want in our home or work (Torresen, 2018).

There are multiple ways to categorize robots, including conceptual typologies based on a robot’s function and application area (Lin et al., 2011), the degree of a robot’s anthropomorphism (that is, human characteristics of a robot), the purpose or task of its operation (Leminen et al., 2017), its ability to adapt to the environment (Bertolini & Aiello, 2018), and a robot’s level of “cognitive” computing and affective resources (Čaić et al., 2019). Leenes et al. (2017) argue that robots can be categorized by their autonomy, task, operative environment, and HMI relationships.

Nonetheless, as the number of different types of robots and their uses increase in our daily lives, there will unarguably be more and more ethical challenges and questions arising with new robotic achievements and applications (Demir, 2017). Although concern about ethical issues in robotics is actually older than the field of robotics itself, “roboethics” has only recently emerged as a discipline dealing with ethical issues related to robotics (Ishihara & Fukushi, 2010; Veruggio & Operto, 2006). In fact, the study of social and ethical issues related to robotics is still in its infancy and calls for more research, although attention to the theme is increasing rapidly (van der Plas et al., 2010). In particular, there is a need for coherent ethical frameworks in order to frame and discuss new types of robots, and contribute to the virtuous development and adoption of such robots (Demir, 2017). Hence, this conceptual article aims at reviewing previous literature on roboethics in order to discuss the main roboethics perspectives, and at the same time use those perspectives to create an ethical framework for “smart robots” as a rapidly emerging new robotic “species”.

The article is structured as follows. After this introductory section, the study reviews previous literature on roboethics and discusses the main perspectives on ethics in robotics. It then makes use of the perspectives identified in order to establish an ethical framework for smart robots. Upon establishing and elaborating the framework, the paper identifies two underlying dimensions based on key concepts in ethical and moral theory. Finally, the article concludes by discussing key tenets from the study and highlighting avenues for future research on roboethics in light of the surge coming with ever smarter robots.

## Roboethics as an Emerging Discipline

Ethical issues in regard to robots and their impacts on our society are the subject of “roboethics” (Demir, 2017). Research in robotics and discussions about roboethics are currently being promoted globally by several organizations, including universities and technology companies, as well as online and open-source maker communities dedicated to robotics development (Prescott & Szollosy, 2017). Hence, roboethics has mainly addressed the “human ethics” of robot designers, manufacturers, and users (Mushiaki 2013/2014). However, “machine ethics” indicates ethics relating to forms and codes of conduct implemented *in the AI of robots*. The aim of this research field is to guarantee that autonomous robots will exhibit ethically acceptable behaviour during their interactions with human beings. The risk that the actions of robots may have negative consequences on human beings or the environment is a growing area of study in roboethics (Lichocki et al., 2011; Veruggio et al., 2011). In fact, recent research (for example, Beltramini, 2019) uses the term “roboethics” as a synonym for “machine ethics”, thus acknowledging that the ethical behaviour of machines is determined by the way their systems have been designed. Nevertheless, both the discourse and application of roboethics remain poorly understood, lacking a clear explanation of basic principles regarding the present and potential consequences of what we can now call “smart robots” on society (Alsegier, 2016).

Fundamental issues in roboethics include the dual use problem of robots (robots can be used or misused), the anthropomorphisation of robots (the illusion that machines have internal states that correspond to the emotions they express, like a “ghost in the machine”), and the equal accessibility to technology challenge, such as for care robots (Bertolini & Aiello, 2018;

## An Ethical Framework for Smart Robots

*Mika Westerlund*

Veruggio & Operto, 2008). Further, many engineering projects lean toward trying to develop more humanized robots, partly due to the increased use of AI components and a focus on developing HMI.

However, a note of caution is expressed that there is an ethically significant distinction between human-human interaction and human-robot interaction (Borenstein & Pearson, 2013). Engineers should therefore be highly sensitive to the potential impacts of their creations on human thinking and emotions, as people interact with robots (Steinert, 2014). The humanoid appearance of a robot might deceive users into believing that the robot has capabilities it does not actually have. The more “intelligently” a robot acts, the more people are inclined to attribute “liveliness” or “life” to it, thus leading them to at least in some ways treat that machine as they would treat other living beings (Steinert, 2014). Lumbrenas (2018) suggests discarding the ongoing current efforts at humanization of robots, and instead distinguishing HMI from inter-personal interaction with human beings, by avoiding the practice of giving names to technology. As it turns out, however, technology manufacturers seem to be navigating in an entirely opposite direction for their AI-driven technologies (for example, Apple’s “hey Siri” call).

The unfolding scenarios made possible by smart robotics technology are both fascinating and unsettling at the same time. The increasing adoption of smart robots will raise new ethical, legal, and social issues (Alsegier, 2016; Veruggio et al., 2011). Advanced robotics can be very harmful if it is applied to people’s lives without understanding the potential issues that may arise from introducing ever “smarter” technology (Alsegier, 2016). Hence, it is crucial that everyone in a society, especially the creators of smart robots, knows that there are ethical principles that govern the field. They may then in a practical sense try to apply those principles in real life (Alsegier, 2016). As a major branch of philosophy, ethics may be simply described as “the intrinsic control of good behaviour”, which is in contrast to “law” that acts as the “extrinsic control of good behaviour” (Majeed, 2017). The main ethical concern involving robotics is the conflict between basic human rights and the responsibilities of scientists and engineers. Accordingly, people have the right to be safe, while at the same time, corporations have the right to attempt to profit from the development of robotic technology (Alsegier, 2016). Hence, addressing key tenets in roboethics as they are likely to arise is a fundamental, market sensitive requirement for assuring a sustainable, ethical, and beneficial human-robot

symbiosis (Tsafestas, 2018) in digitized social ecosystems.

### Key Ethical Perspectives for Smart Robots

Building on suggestions by Steinert (2014), roboethics provides four key ethical perspectives on smart robots. These are, 1) smart robots as amoral and passive tools, 2) smart robots as recipients of ethical behaviour in society, 3) smart robots as moral and active agents, and 4) smart robots as ethical impact-makers in society. The following sections provide an in-depth elaboration on these perspectives.

#### *Smart robots as amoral and passive tools*

According to the instrumental perspective, robots are mere extensions of human capabilities, and can be used as tools to alter a situation according to human desires (Steinert, 2014). A robot can also be part of larger systems that have some control over its actions (Coeckelbergh, 2011). Solis and Takanishi (2010) point out that while robots are viewed as tools that humans use to perform hazardous or dull tasks (for example, robot vacuums), humanoids are increasingly designed to engage people through communications strategies, in order to achieve social or emotional goals. Whether or not such robots are capable of making ethical decisions, thus has become a non-trivial point of contention (Borenstein & Pearson, 2013). Robots are still seen as amoral instruments, because technology is supposed to be neutral concerning the purpose of its usage. For example, a robot can be used to perform a life-saving surgery, while the very same robot could also be used to hurt or kill someone, as a result of human will (Steinert, 2014). In fact, along with the increasing intelligence, speed, and interactivity of robotics technology (Kumari et al., forthcoming), smart robots can potentially be used as “killer robots” by militaries, that is, as offensive semi-autonomous weapons (Demir, 2017). Yet, even if a robotic weapon is built as an intelligent, autonomous or semi-autonomous system, the ethical concerns that arise from its usage nevertheless remain entirely focused on the human designing or using them (Steinert, 2014).

Kelley et al. (2010) note that robots are analogous to domesticated animals in disputes about liability. If a robot is involved in an accident, the robot’s owner should be liable, unless the robot is defective in manufacture or design, or has an inadequate warning label, in which case the robot’s manufacturer may be held liable for damages (Kelley et al., 2010). Further, either owners or users can be held liable if a robot under their custody harms someone, or if they made the robot

## An Ethical Framework for Smart Robots

*Mika Westerlund*

unsafe through modifications to display features not intended by the robot's manufacturer (Bertolini & Aiello, 2018). Smart robots cannot also be held liable in case of privacy issues. Advanced social robots such as robot companions and care robots can record sensitive information about customers and patients, even without them being aware of having disclosed that information (Bertolini & Aiello, 2018). The instrumental view argues that machines are unlikely in the foreseeable future to be able to undertake the same or similar reasoning processes of handling sensitive information as human beings can do (Borenstein & Pearson, 2013). Nevertheless, only strong autonomy considered as a robot's full ability to freely determine its own will and course of action would justify treating the robot as a "subject" that (who) can be held liable for its actions. Instead, the instrumental perspective holds that a robot is not an active agent, but merely a passive object to an active human agent's will (Bertolini & Aiello, 2018).

### *Smart robots as recipients of ethical behaviour in society*

Another perspective in roboethics views smart robots as recipients of human ethical behaviour in society. Nowadays, it is unimaginable for civilized societies to hold slaves. As ethical sensibilities concerning our behaviour towards animals has recently advanced, there is also need to contemplate whether the moral realm should also encompass intelligent technology such as smart robots (Steinert, 2014). For example, a scenario arises where it could be considered wrong to be "inhumane" to a homecare robot that is no longer of use to a household, even though that robot has no real autonomy or personality (Petersen, 2007). Similarly, Anderson et al. (2010) argue that roboethics should put more emphasis on developing ethical research guidelines for experimentation on robots, along the lines of rules for experimentation and testing on animals. Although one might argue that robots do not possess "personality", societies actually make "persons" by producing them partly through a process of personification, that is, attributing human qualities to non-human objects, which is conferring the status of a "person" to something non-human (Steinert, 2014). Another issue arises if robots gain an ability to learn to reason themselves out of a "desire" for doing their designed task. Thus, forcing an autonomous smart robot to stick with its designed task, in such a situation, could be deemed unethical, perhaps upheld by law even if the "owner" of the robot paid for the robot to do the designed task (Petersen, 2007). Thus, future work in roboethics needs to discuss more about the potential domain of "robots' rights" (Anderson et al., 2010), alongside of whether rights only exist for human beings

as owners of robots, the latter which by definition have no "rights" at all.

Smart robots become part of the "social-relational whole", that is, members of an interactive network of human beings and intelligent machines (Coeckelbergh, 2015). Whatever capacity and understanding of how to interact with human beings a robot is built with, designers have to consider its ethical consequences in HMI (Coeckelbergh, 2015; Solis & Takanishi, 2010). Programming social values and norms into robots that are designed to interact with humans requires input from several types of experts (Weng, 2010), such as engineers, scientists, legal advisors, sociologists, and psychologists. That said, experts working on areas characterized by complexity and controversy, such as AI and smart robotics, cannot assume their technical qualifications will be enough to satisfy questions involving the human condition in HMI (Prescott & Szollosy, 2017).

This partly relates to advancements in robotics, leading to a shift from the ability to execute "simple" navigational tasks, to being able to perform "complex" social interaction with human beings (Campa, 2016). Nonetheless, one issue that arises from people interacting with social robots is that they may show indifference and even cruel behaviour in HMI, knowing that the robot's displayed emotions are not real (Wirtz et al., 2018). On the other hand, there is a danger that children or other groups may interpret the behaviour of robots as controlled by internal cognitive or emotional states (for example, the robot moved or said something because it "wanted" to), as opposed to externally regulated by human control (for example, a programmatic response based on information about the environment gathered through sensors) (Melson et al., 2009). Thus, interacting with a smart robot may spark empathy toward the robot for its "good" behaviour, or, alternatively, aggressive behaviour such as punching or kicking the robot by children simply because of its occasionally irrational, uncanny or "wrong" behaviour (Darling, 2015). Likewise, a robot's right to self-defense against potential abusive behaviour in HMI is an under-researched area that needs further study.

### *Smart robots as moral and active agents*

The third perspective views robots as moral agents in themselves, that is, as active subjects in their own right, rather than as objects and passive instruments of human beings. Sophisticated trading robots and autonomous vehicles can be considered as non-human "decision-makers", because the actions they "choose" to

## An Ethical Framework for Smart Robots

*Mika Westerlund*

take can have pervasive real-world consequences (Steinert, 2014). Decision-making capacities come inevitably with the question of ethics. At its simplest, ethics signifies conduct a balanced assessment of the harms and benefits of any actions (Iphofen & Kritikos, forthcoming). However, a robot's inability to have human emotions and feelings has raised concerns about their capabilities to act respectfully or in a "moral" way towards human beings (Leenes et al., 2017). The more autonomous a robot is, the more it would seem necessary to be both sensible and responsible to legal and social values and norms, as well as to perceive and interpret its present situation, including to identify what is demanded, forbidden or tolerated (Steinert, 2014). For instance, robotic street cleaners and driverless cars will have to observe traffic regulations (Leenes et al., 2017), and care robots in hospitals need to be able to monitor and perform analyses and operate courses of action that are consistent with established codes of ethics during their interaction with patients (Luxton, 2014). Whereas a robot's simple "decision making" needs to be founded on case-based reasoning, rather than on generic moral principles (Iphofen & Kritikos, forthcoming), at the same time a pre-programmed understanding of the use context will be crucial in order to adjust a robot's design to accommodate ethics based on context and practice (Van Wynsberghe, 2013).

Coeckelbergh (2011) argues that engineers should not implement roboethics in a top-down fashion, but rather design robots that have the capacity to learn, develop and even eventually reproduce themselves over time. According to Vetrò et al. (2019), an overly deterministic approach to a robot's algorithmic operations might affect the machine's behaviour in a way that produces negative social effects. Rather, they suggest it would be better if a robot learned to autonomously perform human tasks and behaviour, by mimicking the demonstration of human subject performances (Solis & Takanishi, 2010). While this technology is still in exploratory territory, it is noteworthy that algorithmic operations involving individuals can result in harmful discrimination, even in the case of robotic learning.

Attempts by robots to reproduce observed human behaviour, may lead to under- or overestimation of certain human beings and representatives of human groups, because of disproportionate historical datasets and learning methods in these different "species" (Iphofen & Kritikos, forthcoming; Vetrò et al., 2019). Although a robot might not be held morally or legally responsible for its operations, or liable for the damage it

causes because technology has no intentionality (Bertolino & Aiello, 2018; Lichoocki et al., 2011), the "robots as moral and active agents" perspective maintains that an autonomous smart robot capable of learning to perform tasks should have at least "limited liability". This argument is even more crucial if a robot were to show emergent behaviours that were not explicitly programmed, and which only became observable with time (Trentesaux & Rault, 2017).

### *Smart robots as ethical impact-makers in society*

Finally, smart robots can be seen as impact-makers. This view holds that robots can be ethical-impact agents that influence for social norms and values (Steinert, 2014). For example, the spread of smart social robots could alter the structure of the societies globally, influencing humanity and our relationship with technology (Ishihara & Fukushi, 2010). Futuristic visions about a coming "Ubiquitous Robot Society" and "Neo Mechatronic Society" are frequently to be found in public discussions (van der Plas et al., 2010). Thus, this perspective on roboethics stresses the potential constructive and beneficial relationship between humans and robots, focusing on questions involving if, when, and how we can potentially learn to flourish with robots (Coeckelbergh, 2011).

Social norms regarding receptiveness to technology vary in time and place. There are differences, for example, between Japanese and Western cultures about robots. Whereas Japanese culture generally views robots as helpmates, in contrast, Western cultures have tended to lean toward the idea that machines created by humans will ultimately turn against their makers (Leenes et al., 2017). Similarly, while Japanese robot developers are now actively pursuing the creation of smart care home robots for their aging population, Majeed (2017) argues that the provision of widespread robotic care in one culture, may turn out to impose a societal stigma on it from other cultures. Borenstein and Pearson (2013) submit that as the adoption of social robots in some cultures increases, especially children may grow to prefer robots over humans. In this vein, some people may develop a tendency to retreat from social interaction with others, and even start competing with other people for a robot companion's attention, which may bring attendant harmful social consequences.

Also, smart robots are already capable of taking over a steadily increasing number of human tasks (Leenes et al., 2017). Although robotics is often associated with the "three Ds", that is, robots perform jobs that are "dull,

# An Ethical Framework for Smart Robots

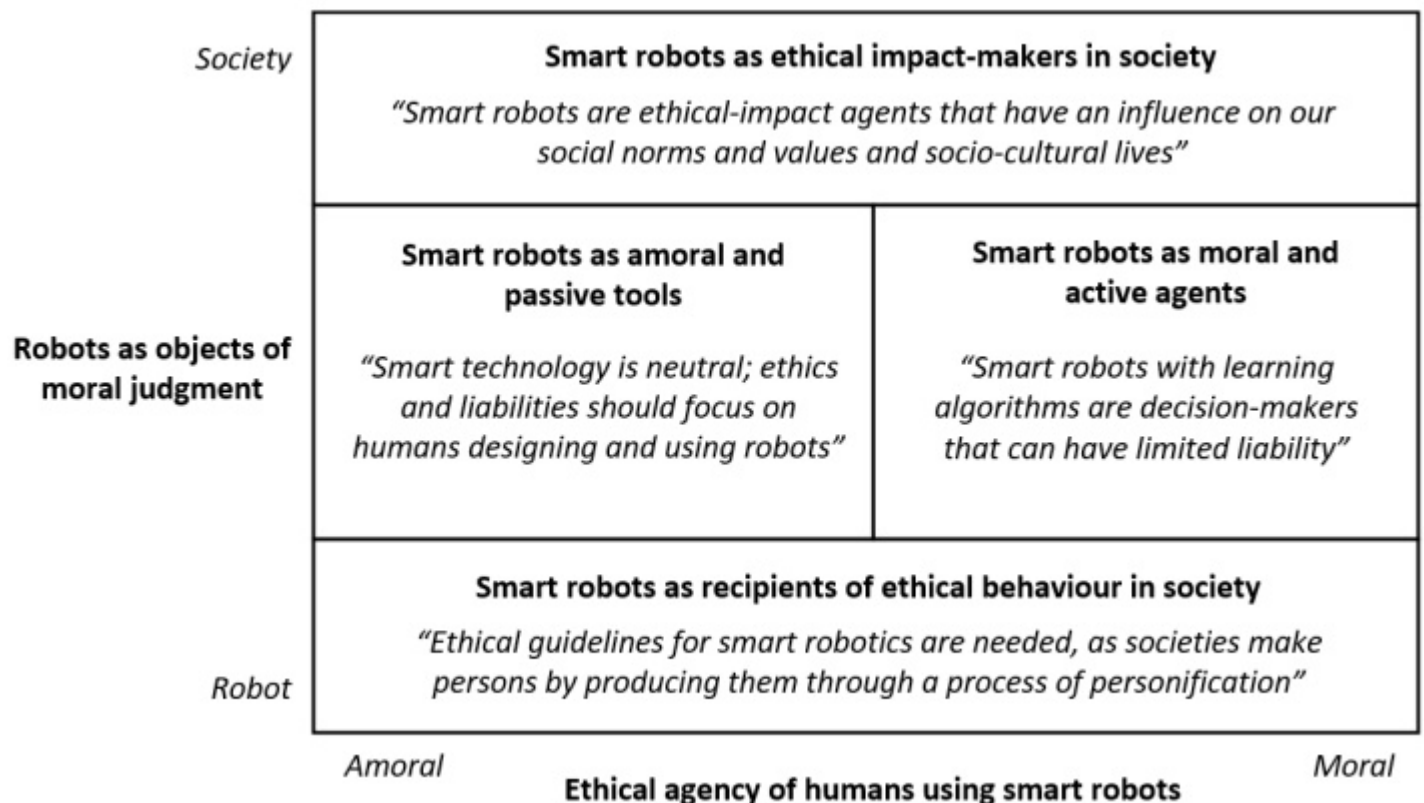
*Mika Westerlund*

dirty, or dangerous”, meanwhile advanced robots can now perform increasingly delicate and difficult jobs, such as medical surgeries, with more precision and accuracy than human hands (Lin et al., 2011). Indeed, intelligent robotics technology is coming more and more to replace human labour for performing complicated tasks in domains ranging from manufacturing and economy to finance and health (Beltrami, 2019). Although such robotic “servitude” is perceived quite differently from human “slavery”, the growth in robots as unpaid labor brings with it the issue of human “replaceability” changing the composition of the workforce (Petersen, 2007). This begs a question of who or what would be to blame if a large-scale labour force replacement of human workers due to robots were to occur; robots, their designers, or the society and people who pay to use them (Steinert, 2014). After all, humanity has deliberately built automated tools to increase its power and foster economic progress by eliminating manual labour and needless drudgery (Veruggio & Operto, 2008). Thus, in the meantime we have become highly reliant on technology (Anderson et al., 2010). On the other hand, robots do not only cause job losses, but also create jobs. However, the kinds of available jobs for

humans will change, with low-skilled jobs being replaced by higher-skilled jobs. This development may exacerbate social inequality in the labour market (Leenes et al., 2017).

## An Ethical Framework for Smart Robots

Summing up the discussion on diverse approaches to roboethics, we can establish a conceptual framework that distinguishes four major ethical perspectives regarding smart robots, based on the work of Steinert (2014). Steinert (ibid.) recommends that robotics developers treat all four ethical perspectives simultaneously and, further, that ethical, social, cultural, and technical considerations should be combined. Moreover, Steinert (ibid.) suggests that roboethics taxonomies should incorporate more than one dimension, although one is all that is often used in current roboethical categorizations. Along with advancements in AI and robot technologies, some popular dimensions, such as a robot’s autonomy (Wallach & Allen, 2010) are becoming obsolete, as increasingly smarter robots are becoming autonomous or semi-autonomous *de facto*. This means that robots



**Figure 1.** A framework of ethical perspectives to smart robots

## An Ethical Framework for Smart Robots

*Mika Westerlund*

are nowadays capable of making what more and more look like “decisions”, and of performing complicated actions in HMI. Similarly, other dimensions such as a robot’s area of usage (Steinert, 2014) are increasingly difficult to define in an accurate manner. New smart robots, such as Samsung’s “Ballie”, can perform tasks in multiple areas, being a life companion, personal assistant, robotic pet, fitness assistant, personal care robot, manager, and coordinator for a number of other home robots in a household, at the same time (Hitti, 2020).

Lin et al. (2011) note that although smart robots may seem to jump out of the pages of science fiction, technological progress nevertheless continues, and we therefore need to consider the ethical issues that are coming along with advancing robotics. In accordance with Steinert’s notion (2014) on the need to use key concepts in ethics as dimensions for categorizing roboethics, our framework identifies two underlying dimensions behind the four ethical perspectives to smart robots: 1) ethical agency of human beings using smart robots (in terms of smart robots as amoral tools vis-à-vis moral human agents) and 2) robots as objects of moral judgment in themselves (in terms of smart robots being objects of ethical behaviour vis-à-vis ethical changes in society due to the introduction of smart robots) (see Figure 1). The underlying approach to each of the perspectives is summarized below the label of the perspective.

Ethical and moral theory (see for example Craig, 1993) put forward many important and relevant concepts. The two dimensions chosen for the purpose of this study have been previously suggested in literature on roboethics, yet they have not been extensively discussed, nor connected together. “Roboethical agency”, that is, the ability of a smart robot to commit ethical or unethical actions, is discussed as a dimension by Moor (2006) and Dyrkolbotn et al. (2017). “Robots as objects of moral judgment”, that is, whether the consequences of ethical or unethical actions affect a smart robot or human society, is discussed by Davenport (2014). The dimensions are not exclusive; whether smart robots are considered amoral tools or as autonomous moral agents, or even as both at the same time, can be the case irrespective of the object of moral judgment. That is, ethical actions can impact either robots, or society at large, or both. This is accords with Steinert’s (2014) argument that various roboethical perspectives have blurry boundaries. The features in each of the perspectives are summarized in Figure 1 below the label of the perspective.

### Discussion and Conclusion

This article has aimed at creating and discussing an ethical framework for smart robots based on previous scholarly literature on roboethics. Smart robots were defined as autonomous AI systems that can collaborate with humans and are capable of learning from their operating environment, experiences, and human behaviour feedback in HMI, in order to improve their capabilities. Upon reviewing previous literature on roboethics, the study discussed and elaborated on four perspectives to roboethics, as originally suggested by Steinert (2014). Then it established a conceptual framework to illustrate these perspectives, as well as a general robotics strategy suitable for near future HMI with smart robots. In so doing, the study argued that the dimensions of a framework should be based on key concepts in ethical and moral theory, and identified two dimensions underlying Steinert’s four ethical perspectives: 1) ethical agency of humans using smart robots (amoral tools vis-à-vis moral agents), and 2) robots as objects of moral judgment (smart robots as objects of ethical behaviour vis-à-vis the ethical consequences of smart robots in human societies).

The study contributes to extant literature on roboethics in several ways. First, it updates Steinert’s (2014) discussion on roboethics by specifying how smart robots, as a kind of new robotic “species” that is being increasingly adopted by users at all levels of society, may serve to affect our ethical outlook regarding both robots and robotics. For example, the study points out that some popular dimensions in roboethics categorizations, such as a robot’s autonomy (see for example Wallach & Allen, 2010), are becoming obsolete, as increasingly smarter robots are becoming “semi-autonomous” or “autonomous” *de facto*. Similarly, a robot’s technical features, or area of usage (Lin et al., 2011; Steinert, 2014), are currently becoming increasingly difficult to define, as new smart robots emerge that are capable of performing tasks in multiple areas (Hitti, 2020). Second, the study establishes a conceptual framework that presents Steinert’s four perspectives on roboethics, and summarizes the ethical approach to smart robots from each perspective in a descriptive sentence. Third, the framework contributes to extant literature on roboethics by identifying two dimensions underlying the four perspectives. These dimensions are based on ethical and moral theory (Craig, 1993), and have been suggested in prior studies on roboethics (Moor, 2006; ; Davenport, 2014; Dyrkolbotn et al., 2017), but have not been discussed extensively, nor simultaneously in the roboethics literature. Fourth, the study suggests that

## An Ethical Framework for Smart Robots

*Mika Westerlund*

these two dimensions are not mutually exclusive, but rather can occur at least in part together at the same time. In this vein, the study both accepts and confirms Steinert's (2014) argument that the four ethical perspectives should be considered simultaneously because their boundaries are blurry.

Both researchers and practitioners such as smart robot designers can benefit from the study. First, scientists can use the framework and its dimensions to better focus their area of research in regard to the emergence of new types of robots, including the ethical challenges those robots may impose. Second, the majority of people in technologically advanced nations want robots to contribute to a better and more ethical world. Yet, there is nonetheless still disagreement in regard to how to bring this goal about (Lichocki et al., 2011). According to Alseguer (2016), designers must consider how their robots will impact peoples' behaviours, and continually review their robotics applications, including both the technological and psychological aspects, as safety measures to ensure that their robots do not cause harmful effects on a person or society. However, the present study reiterates the recommendation that engineers should not implement roboethics in a top-down manner, but rather design robots that can learn from mimicking the demonstration of performing human subjects, in order to avoid the negative effects of overly deterministic algorithmic decision-making (Coeckelbergh, 2011; Solis & Takanishi, 2010; Vetrò et al., 2019). This means that in order for smart robots to function ethically with human beings and to exercise context-awareness, they must be able to absorb the necessary legal, social, and even cultural norms and standards from their environment (Steinert, 2014). This, of course, is no small challenge. Third, smart robot engineers can use this framework on roboethics as an aid to assess the potential consequences and risks of AI-driven robotics technologies for people and societies.

Regarding limitations and future research opportunities in the field, Tsafestas (2018) argues that robotic behaviour, behavioural expectations, and related ethical questions vary significantly by the type of smart robot, for example, assistive robots, social robots, and military robots. Although it is acknowledged that context-awareness is important for ethical robotic decisions, the variety of smart robots was only covered briefly at once by discussing generic perspectives about roboethics in the smart robot context at an abstract level. Thus, future research should examine if the roboethics perspective has a relationship with newer "smart robots". Further, previous research (Tuisku et al., 2019) argues that public

opinion about the widespread use of robots in society continues to be mainly negative. The paper thus discussed ethical issues in regard to smart robots on largely an abstract level, and did not address opinions about robots and their possible relationship with roboethics as adopted by any specific party such as robot engineers or the general public. Future research should investigate whether negative public opinion about robots can be explained by the particular types of perspective that many people have adopted. Overall, the study concludes that the spread of ever smarter robots will cause numerous ethical challenges in societies around the world.

### References

- Alseguer, R. A. 2016. Roboethics: Sharing our world with humanlike robots. *IEEE Potentials*, 35(1): 24-28.  
<http://dx.doi.org/10.1109/MPOT.2014.2364491>
- Anderson, M., Ishiguro, H., & Fukushi, T. 2010. "Involving Interface": An Extended Mind Theoretical Approach to Roboethics. *Accountability in Research*, 17(6): 316-329.  
<http://dx.doi.org/10.1080/08989621.2010.524082>
- Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. 2019. Service robot implementation: a theoretical framework and research agenda. *The Service Industries Journal*.  
<http://dx.doi.org/10.1080/02642069.2019.1672666>
- Beltramini, E. 2019. Evil and roboethics in management studies. *AI & Society*, 34: 921-929.  
<http://dx.doi.org/10.1007/s00146-017-0772-x>
- Bertolini, A., & Aiello, G. 2018. Robot companions: A legal and ethical analysis. *The Information Society*, 34(3): 130-140.  
<http://dx.doi.org/10.1080/01972243.2018.1444249>
- Borenstein, J., & Pearson, Y. 2013. Companion Robots and the Emotional Development of Children. *Law, Innovation and Technology*, 5(2): 172-189.  
<http://dx.doi.org/10.5235/17579961.5.2.172>
- Čaić, M., Mahr, D., & Odekerken-Schröder, G. 2019. Value of social robots in services: social cognition perspective. *Journal of Services Marketing*, 33(4): 463-478.  
<http://dx.doi.org/10.1108/JSM-02-2018-0080>
- Campa, R. 2016. The Rise of Social Robots: A Review of the Recent Literature. *Journal of Evolution and Technology*, 26(1): 106-113.
- Coeckelbergh, M. 2011. Is Ethics of Robotics about Robots? Philosophy of Robotics Beyond Realism and Individualism. *Law, Innovation and Technology*, 3(2): 241-250.  
<http://dx.doi.org/10.5235/175799611798204950>
- Craig, R. P. 1993. Ethical and Moral Theory and Public



# An Ethical Framework for Smart Robots

Mika Westerlund

- School Administration. *Journal of School Leadership*, 3(1): 21–29.  
<https://doi.org/10.1177/105268469300300103>
- Darling, K. (2015). Children Beating Up Robot Inspires New Escape Maneuver System. *IEEE Spectrum*, August 6, 2015. Retrieved from  
<https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/children-beating-up-robot>
- Davenport, D. 2014. Moral Mechanisms. *Philosophy & Technology*, 27(1): 47-60.  
<http://dx.doi.org/10.1007/s13347-013-0147-2>
- Demir, K. A. 2017. Research questions in roboethics. *Mugla Journal of Science and Technology*, 3(2): 160-165.  
<http://dx.doi.org/10.22531/muglajsci.359648>
- Dyrkolbotn, S. K., Pedersen, T., & Slavkovik, M. 2017. Classifying the Autonomy and Morality of Artificial Agents. *CARe-MAS@PRIMA 2017*: 67-83.
- Hitti, N. 2020. Ballie the rolling robot is Samsung's near-future vision of personal care  
<https://www.dezeen.com/2020/01/08/samsung-ballie-robot-ces-2020/>
- Iphofen, R., & Kritikos, M. forthcoming. Regulating artificial intelligence and robotics: ethics by design in a digital society. *Contemporary Social Science*.  
<http://dx.doi.org/10.1080/21582041.2018.1563803>
- Ishihara, K., & Fukushi, T. 2010. Introduction: Roboethics as an Emerging Field of Ethics of Technology. *Accountability in Research*, 17(6): 273-277.  
<http://dx.doi.org/10.1080/08989621.2010.523672>
- Kelley, R., Schaerer, E., Gomez, M., & Nicolescu, M. 2010. Liability in Robotics: An International Perspective on Robots as Animals. *Advanced Robotics*, 24(13): 1861-1871.  
<http://dx.doi.org/10.1163/016918610X527194>
- Kumari, R., Jeong, J. Y., Lee, B.-H., Choi, K.-N., & Choi, K. forthcoming. Topic modelling and social network analysis of publications and patents in humanoid robot technology. *Journal of Information Science*.  
<http://doi.org/10.1177/0165551519887878>
- Leenes, R., Palmerini, E., Koops, B.-J., Bertolini, A., Salvini, P., & Lucivero, F. 2017. Regulatory challenges of robotics: some guidelines for addressing legal and ethical issues. *Law, Innovation and Technology*, 9(1): 1-44.  
<http://dx.doi.org/10.1080/17579961.2017.1304921>
- Leminen, S., Westerlund, M., & Rajahonka, M. 2017. Innovating with service robots in health and welfare living labs. *International Journal of Innovation Management*, 21(8): 1740013.  
<http://dx.doi.org/10.1142/S1363919617400138>
- Lichocki, P., Kahn Jr., P. H., & Billard, A. 2011. The Ethical Landscape of Robotics. *IEEE Robotics and Automation Magazine*. 18(1): 39-50.  
<http://dx.doi.org/10.1109/MRA.2011.940275>
- Lin, P., Abney, K., & Bekey, G. 2011. Robot ethics: Mapping the issues for a mechanized world. *Artificial Intelligence*, 175(5/6): 942-949.  
<http://doi.org/10.1016/j.artint.2010.11.026>
- Lumbrenas, S. 2018. Getting Ready for the Next Step: Merging Information Ethics and Roboethics - A Project in the Context of Marketing Ethics. *Information*, 9(8), 195.  
<http://dx.doi.org/10.3390/info9080195>
- Majeed, A. B. A. 2017. Roboethics - Making Sense of Ethical Conundrums. *Procedia Computer Science*, 105: 310–315.  
<http://dx.doi.org/10.1016/j.procs.2017.01.227>
- Market Research Future. 2019. *Smart Robot Market Research Report - Global Forecast Till 2023*. December 2019.  
<https://www.marketresearchfuture.com/reports/smart-robot-market-6622>
- Melson, G. F., Kahn, Jr. P.H., Beck, A., & Friedman, B. 2009. Robotic Pets in Human Lives: Implications for the Human-Animal Bond and for Human Relationships with Personified Technologies. *Journal of Social Issues*, 65(3): 545-567.  
<https://doi.org/10.1111/j.1540-4560.2009.01613.x>
- Moor, J. H. 2006. The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4): 18-21.  
<http://dx.doi.org/10.1109/MIS.2006.80>
- Mushiaki, S. 2013/2014. Chapter 1. Ethica Ex Machina: Issues in Roboethics. *Journal International de Bioéthique*, 24(4): 17-26.
- Petersen, S. 2007. The ethics of robot servitude. *Journal of Experimental & Theoretical Artificial Intelligence*, 19(1): 43-54.  
<http://dx.doi.org/10.1080/09528130601116139>
- Prescott, T., & Szollosy, M. 2017. Ethical principles of robotics. *Connection Science*, 29(2): 119-123.  
<http://dx.doi.org/10.1080/09540091.2017.1312800>
- Solis, J., & Takanishi, A. 2010. Recent Trends in Humanoid Robotics Research: Scientific Background, Applications, and Implications. *Accountability in Research*, 17(6): 278-298.  
<http://dx.doi.org/10.1080/08989621.2010.523673>
- Steinert, S. 2014. The Five Robots - A Taxonomy for Roboethics. *International Journal of Social Robotics*, 6: 249-260.  
<http://dx.doi.org/10.1007/s12369-013-0221-z>
- Torresen, J. 2018. A Review of Future and Ethical Perspectives of Robotics and AI. *Frontiers in Robotics and AI*, 4: 75.  
<http://dx.doi.org/10.3389/frobt.2017.00075>
- Trentesaux, D., & Rault, R. 2017. Designing Ethical Cyber-Physical Industrial System. *IFAC PapersOnLine*, 50(1): 14934-14939.  
<http://dx.doi.org/10.1016/j.ifacol.2017.08.2543>

## An Ethical Framework for Smart Robots

Mika Westerlund

- Tsafestas, S. G. 2018. Roboethics: Fundamental Concepts and Future Prospects. *Information*, 9(6), 148.  
<http://dx.doi.org/10.3390/info9060148>
- Tuisku, O., Pekkarinen, S., Hennala, L., & Melkas, H. 2019. "Robots do not replace a nurse with a beating heart" - The publicity around a robotic innovation in elderly care. *Information Technology & People*, 32(1): 47-67.  
<http://dx.doi.org/10.1108/ITP-06-2018-0277>
- van der Plas, A., Smits, M., & Wehrmann, C. 2010. Beyond Speculative Robot Ethics: A Vision Assessment Study on the Future of the Robotic Caretaker. *Accountability in Research*, 17(6): 299-315.  
<http://dx.doi.org/10.1080/08989621.2010.524078>
- van Wynsberghe, A. 2013. A method for integrating ethics into the design of robots. *Industrial Robot: An International Journal*, 40(5): 433-440.  
<http://dx.doi.org/10.1108/IR-12-2012-451>
- Veruggio, G., & Operto, F. 2006. Roboethics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics. *International Review of Information Ethics*, 6: 2-8.
- Veruggio, G., & Operto, F. 2008. Roboethics: Social and Ethical Implications of Robotics. In B. Siciliano, & O. Khatib (Eds.). *Springer Handbook of Robotics*: 1499-1524. Berlin: Springer.
- Veruggio, G., Solis, J., & Van der Loos, M. 2011. Roboethics: Ethics Applied to Robotics. *IEEE Robotics & Automation Magazine*, 18(1): 21-22.  
<http://dx.doi.org/10.1109/MRA.2010.940149>
- Vetrò, A., Santangelo, A., Beretta, E., & De Martin, J. C. 2019. AI: from rational agents to socially responsible agents. *Digital Policy, Regulation and Governance*, 21(3): 291-304.  
<http://dx.doi.org/10.1108/DPRG-08-2018-0049>
- Wallach, W., & Allen, C. 2010. *Moral machines: teaching robots right from wrong*. New York: Oxford University Press.
- Weng, Y.-H. 2010. Beyond Robot Ethics: On a Legislative Consortium for Social Robotics. *Advanced Robotics*, 24(13): 1919-1926.  
<https://doi.org/10.1163/016918610X527220>
- Wirtz, J., Patterson, P., Kunz, W., Gruber, T., Lu, V., Paluch, S., & Martins, A. 2018. Brave new world: service robots in the frontline. *Journal of Service Management*, 29(5): 907-931.  
<https://doi.org/10.1108/JOSM-04-2018-0119>

### About the Author

Mika Westerlund, DSc (Econ), is an Associate Professor at Carleton University in Ottawa, Canada. He previously held positions as a Postdoctoral Scholar in the Haas School of Business at the University of California Berkeley and in the School of Economics at Aalto University in Helsinki, Finland. Mika earned his doctoral degree in Marketing from the Helsinki School of Economics in Finland. His research interests include open and user innovation, the Internet of Things, business strategy, and management models in high-tech and service-intensive industries.

Citation: Westerlund, M. 2020. An Ethical Framework for Smart Robots. *Technology Innovation Management Review*, 10(1): 35-44.  
<http://doi.org/10.22215/timreview/1312>



Keywords: Smart robot, Artificial intelligence, AI, Roboethics, Ethics