

Chapter 16

Robot Ethics: Ethical Design Considerations



Dylan Cawthorne

16.1 Learning Objectives

- Identify some of the ethical design considerations in the robotics field.
- Become familiar with the three main normative ethical theories.
- Understand why technology is not ethically neutral.
- Appreciate that technology exists in a context.
- Utilize the PPPP model to identify and design for a preferable future.
- Identify some human values relevant in robot design.
- Apply the value-sensitive design methodology.
- Identify impacted stakeholders.
- Utilize ethics checklists, standards, design principles, and frameworks.
- Identify specific examples of design features which support human values.
- Learn about the AIRR framework for responsible innovation.
- Be able to answer the questions: “should I build this robot? If so, why? If not, why not?”
- Apply theories and tools in the chapter to design your own ethically informed robot.

16.2 Introduction

In May of 2019, a twelve-kilogram medical delivery drone crashed in Switzerland, only 50 m from kindergarten children (Ackerman, 2019) (Fig. 16.1). No one was hurt

D. Cawthorne (✉)
Unmanned Aerial Systems Center, University of Southern Denmark, Odense, Denmark
e-mail: dymca@sdu.dk



Fig. 16.1 Matternet (2021) drone which crashed in a wooded area of Switzerland near a kindergarten. The emergency parachute had been deployed, but the chord connecting it to the drone was cut allowing the 12 kg drone to freefall to the ground. Image used with permission, from SUST (2019)

or killed in the incident, but the failure of the drone and its parachute system caused the Swiss Post to immediately suspend operations of the large quadcopters made by the Silicon Valley company Matternet (2021). The system, which was designed to quickly transport up to 2 kg of urgently needed medical samples between hospitals and save lives, could have inadvertently caused the death of a small child. Should the drone still be used if its benefits outweigh the risks? Is it fair to subject people in the cities below the drone's flightpath to risk of injury or death? What about the sick people who need urgent medical care—isn't the drone helping them? And what responsibility do we as robot engineers and builders have for our creations? In this chapter, some of the ethical considerations in the field are presented, and theories and tools are offered for designing ethically informed robotic systems.

An Industry Perspective

Nolwenn Briquet, Control Team Lead, Kortex.

Kinova inc.



During my studies, I was always interested in Mechatronics courses. By combining the disciplines of automation, mechanics, and electronics, I felt that these subjects were the closest to cutting-edge systems. During all my studies and even after, I made all my choices little by little to work in this direction. In 2014, I started working as a robotics engineer in a robotics research laboratory on collaborative robotics applications before doing a Ph.D. in the field of human–robot interaction. What I particularly retain from this experience is that it allowed me not only to develop my expertise but also to ask myself questions about the place of the human being in automation through robotics and our responsibility in this evolution as engineers designing these systems. I finally joined Kinova in 2019 as a robotics control developer and I recently became the team lead of the control team.

When you work in robotics, the question of the societal impact of our actions is unavoidable. In some countries, robotization is seen as a way to overcome labor shortages, while in others it is seen as a real threat. When I was offered the opportunity to work in collaborative robotics, I was attracted by the concept of exploiting the complementarity of the human and the robot and thus keeping the human in the loop. The common discourse is that the human takes care of the tasks that require decision-making skills, critical thinking, and sometimes even dexterity, while the repetitive and non-value added tasks are left to the robot. However, in many situations, we can see that the real interest of these robots is that they are easy to integrate into a workspace as they do not need to be placed in a cage. In some cases, the only limitation to taking humans out of the loop is the technological limit, which I contribute to pushing back in my job. Therefore, to this day, I am faced with the following dilemma: Are these the real reasons for the success of collaborative robotics, or is it more a matter of politicization of public discourse and cobot-washing?

The field of ethics in robotics has been a hot topic in recent years. The question of responsibility can be found today at several levels: at the level of the researcher who shares his or her contributions in the public domain,

at the level of the robot manufacturer who designs the functionalities of the robotic system that will be solved on the market, or at the level of the integrator who will be responsible for the security of the robotic application. At the level of the research community, there is an awareness among researchers on the integrity of the data and results that are shared. In the industrial field, this issue is covered by standards that have evolved in recent years to take into account safety concerns. However, further steps need to be taken to address the ethical issues related to the progress of artificial intelligence, whose results appear extremely promising but also less well-controlled and predictable.

16.3 Ethics

Ethics is the branch of philosophy that deals with questions about right and wrong, and how best to live and act in the world (“Ethics,” 2021). But what does ethics have to do with robot design? According to philosopher of technology Peter-Paul Verbeek, “most scholars in the field agree that technologies actively help to shape culture and society, rather than being neutral means for realizing human ends” (Verbeek, 2008). This means that the robots we design and build won’t just perform a task, their capabilities will allow some actions to be easier to perform and others more difficult—which has moral consequences.

For example, a healthcare drone may make it faster to deliver urgent medical samples between hospitals but make it more difficult to ensure the security of the samples during the trip. An industrial robot arm designed to weld a car’s frame together could make assembly faster, but it might make it more difficult for factory workers to cultivate their welding skills. The complexity of the task becomes even more challenging—and crucial—as we consider the impacts if these technologies are scaled up. Will jobs as medical couriers and welders disappear completely? What will be the long-term impact on peoples’ physical, psychological, and material welfare? Technologies always exist in a context, so we might ask where will this robot be implemented? Is it in a country where workers are likely to be retrained to build or collaborate with robots? Or will these people become redundant? Thus, we as responsible technologists need to be aware of the ethical considerations relevant in the domain, the context of use, and the potential long-term impacts—and make well-reasoned choices about the capabilities our new robots should have.

16.3.1 Normative Ethics

Luckily, ethics has been an area of study for thousands of years, and there are a lot of theories and tools we can apply when designing robots. Ethical questions can be approached at the level of normative ethics (“Ethics,” 2021). Normative ethical theories can be useful to robot designers as they provide guidance on ways to view what is a morally good technology—or at least, technologies that support actions which are morally good. It is important to note that the way in which morally good consequences, actions, and behaviors are defined can vary depending on the context and across cultures.

16.3.2 Consequentialism

One type of normative ethics puts focus on the results or consequences of one’s actions and is called consequentialism (“Consequentialism,” 2003). This includes utilitarianism which states that we should act in a way that the consequences of our actions result in the most good for the most people (“Consequentialism,” 2003). For example, heart disease is the leading cause of death globally at 16% (World_Health_Organization, 2020)—if we could design a mobile robot or drone which encouraged people to exercise we could potentially help a lot of people lead longer, healthier lives. However, the benefits of the drone would have to outweigh the negative outcomes such as injury, privacy violations, and environmental impact caused during production and at the end of the drone’s useful life. And the context will matter too; in some countries, heart disease may be much less prevalent than others, and the way privacy is exercised could vary across cultures.

16.3.3 Deontology

Another normative ethics approach is deontology which puts focus on the rightness or wrongness of an action, rather than the outcome or consequence of that action (“Deontological Ethics,” 2020). Deontology is a rule-based approach where actions that conform to moral norms—“the Right”—are allowed, and those that do not should not be undertaken. For example, if one were to follow the rule not to kill innocent people, then we should not design a weaponized robot that targets innocent people. And if one should save lives, then we ought to design a mobile robot or drone which encouraged people to exercise. With this last example, we can see that different normative ethical theories may suggest we perform the same actions in a given situation, but perhaps for different reasons.

However, there are sometimes important differences between normative theories. In consequentialism, it would be acceptable to perform a wrong action if it leads to an overall positive outcome for more people. This would not be accepted from a deontological standpoint, where “the Right is said to have priority over the Good” (“Deontological Ethics,” 2020).

16.3.4 *Virtue Ethics*

A third type of normative theory is virtue ethics; here the focus is on the moral character of a person, and the theory aims to guide one in what type of person to be or become (“Virtue Ethics,” 2016). Examples of virtues to strive for and cultivate over a long period of time include honesty, courage, care, and wisdom (Vallor, 2016). Designing an industrial robot arm that reduces workplace injuries would be a way of (indirectly) cultivating the virtue of care for other people. Developing a drone that provides rapid medical care in a context where this increased efficiency allows medical staff more time with patients would be another way to support the cultivation of care. Again, how different virtues are manifested could vary depending on context and across cultures. The three normative theories are summarized in Table 16.1.

All three normative theories share an emphasis on human values—values are what a person or group of people consider important in life (Friedman et al., 2013). Human values relevant in technology design include human welfare, privacy, freedom, calmness, and environmental sustainability (Friedman et al., 2013). Later in this chapter, we will see how human values can be utilized throughout the robot design process to enhance human flourishing.

16.4 The Non-Neutrality of Technology

Technology interacts with and impacts people and society in complex ways—but never in an ethically neutral way. Designing and building something alters the range of capabilities and possible actions available to people (Verbeek, 2008). Sometimes, technologies are described as “platforms”—this is the case with social media applications such as Facebook and YouTube (Gillespie, 2010), as well as drones (Cawthorne & Devos, 2020) and robots. These claims represent an older concept in the philosophy of technology called technological neutrality (Vermaas et al. 2007). Technological neutrality allows companies, governments, and engineers to distance themselves from responsibility for the uses of their products. For example, if someone uses a drone to transport illegal drugs the drone manufacturer could claim that it was

Table 16.1 Three main normative ethical theories

Consequentialism	The result/consequence of actions matters most
	Guides actions
Deontology	Includes utilitarianism—the most “good” for the most people/lifeforms
	Rule-based approach
Virtue ethics	Guides actions
	Focus on virtues/moral character
	Guides the type of person one should be/aim to become

the user that *mis*-used their product. However, the drone does clearly play a role in the crime, and as good robot designers, we should be aware of lots of different possible uses of our systems and design them so they prevent—or at least make it more difficult—to do unethical things with.

The concept of technological neutrality has since been replaced by a contextually situated and interactional model. This means that the context, the user, and the technology itself all play a role in the resulting mediation and possible resulting action. As stated before, this means that technology plays a key role in human actions, and since human actions are morally relevant then technology design is also morally relevant.

16.4.1 *Dual-Use*

Creating ethically informed robots may not be easy, especially given the nature of drone and robotic systems as dual-use technologies. Dual-use refers to a system's capability to be used within civilian contexts as well as in military contexts (Novitzky et al., 2018). Many normative approaches allow for the use of military technologies, especially to protect oneself or as part of a “just war” (as in a “justifiable” war) (Lin et al., 2008), but a person that develops a drone to map fields to help farmers may not have intended for their system to be used for military reconnaissance. Or someone developing a robot to carry injured soldiers out of harm's way may not expect the system to be used to deliver packages in crowded cities.

In practice, there is a lot of technology transfer that takes place both from civil contexts to military and from military to civil contexts. Drones were initially developed in university research labs, then proliferated in the military context, and now are seeing rapid growth in civil contexts (Choi-Fitzpatrick et al., 2016). As well, it may be difficult to determine if a technology is civil or military—take for example robots and drones that perform border patrol or those that are used in private security. Still, there are certain capabilities that are relevant in one context over another, and we can design for the intended context. For example, in a military context where the enemy will be trying to destroy the system the survivability of a drone will be a highly relevant capability, while this capability is much less relevant in a civil context (Van Wynsberghe & Nagenborg, 2016). How to design for relevant capabilities will be explained in more detail in the section on value sensitive design.

16.5 Technological Determinism and Multiple Futures

It is sometimes claimed that technology moves in certain ways, and that we are powerless to stop it. For example, that it is inevitable that in the future there will be more drones and robots. This idea is called technological determinism (Verbeek, 2008). In the philosophy of technology, this conception has mostly been superseded by the idea

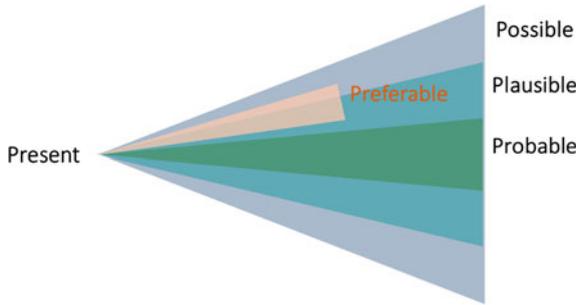


Fig. 16.2 The “PPPP”—or “probable, plausible, possible, preferable”—model shows the future as contingent on what we do in the present, and that we can choose to design our robotic systems for a preferable future rather than for the most likely (probable) future. Graphic by the author, based on Dunne and Raby (2013)

of multiple possible futures—such as that shown in the PPPP—or “probable, plausible, possible, preferable”- model in Fig. 16.2—with an emphasis on human agency and the role we play in shaping technological development. Clearly, if everyone for some reason decided to stop developing drones and robots, companies decided to stop producing them, governments outlawed them, and people stopped buying them then the future would not contain more drones and robots. There are lots of economic forces such as profitability and human forces such as curiosity which make it likely that robots will proliferate in the future, but this is not inevitable. Therefore, we as robot designers hold a lot of power when it comes to the trajectory of future technological developments and need to act in a responsible manner in doing so; this topic is explored in more detail in the section about **Responsibility**.

So in our case as designers, some critical first questions might be to ask ourselves “should we build this drone or robot at all?” “What are some of the possible opportunities, risks, and changes that it will support?” “Who will benefit the most, and who will be at the greatest risk?” “And if we should design the system, what capabilities and characteristics should it have—and which should it not have?” Later in the section on **Value Sensitive Design**, we will look at specific ways to address these questions.

16.6 Human Values in Design

Many human values that are relevant to technology design in general—and for us in robotics design—have been proposed and are shown in Table 16.2. Here, values refer to those things which humans find important and meaningful in life (Friedman et al., 2013). Values are different than preferences; preferences are opinions that individuals hold while values are more universal and are held by most people (Van de Poel, 2009). For example, I might like the color blue (it is my preference) while you might like the color green (your preference). But we both deeply value our own physical safety and the safety of others. The importance and universality of human values makes them critical to a flourishing life, and why they are so relevant

Table 16.2 Twelve human values that are considered relevant in technology design. Based on Table 4.1 in Friedman et al. (2013)

Human welfare	Includes physical, psychological, and material welfare Physical welfare deals with bodily well-being such as physical health Psychological welfare concerns mental health such as stress Material welfare refers to physical circumstances and is related to economics and employment
Ownership and property	The right to possess an object
Privacy	The ability to determine what information about one's self can be communicated to others
Freedom from bias	Without systematic unfairness toward individuals or groups
Universal usability	Technology that can be successfully used by all people
Trust	The expectation to experience goodwill from others
Human autonomy	The ability to decide, plan, and act in ways which allow one to achieve one's goals
Informed consent	Garnering voluntary agreement, such as in the use of information systems
Accountability	Ensuring that actions may be traced uniquely to the person, people, or institution responsible
Calmness	A peaceful and composed psychological state
Identity	The understanding of who one is over time
Environmental sustainability	Sustaining ecosystems such that they meet the needs of the present without compromising future generations

to designers and engineers since our technologies can support (or diminish) these values.

16.7 Value Sensitive Design

Value sensitive design, or VSD, is a way to systematically incorporate the ethical and social impacts of technologies early in the design process (Friedman et al., 2013). The process is shown in Fig. 16.3 and includes three phases: (1) conceptual, (2) empirical, and (3) technological.

16.7.1 Conceptual Phase

In the conceptual phase of VSD, the ethical considerations are identified, as well as the impacted stakeholders. If we consider the example of a cobot in a factory, this includes direct stakeholders such as those working alongside the robot as well as

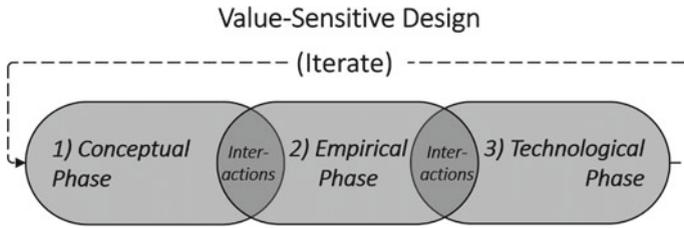


Fig. 16.3 The value sensitive design process consists of three phases: (1) conceptual, (2) empirical, and (3) technological. There are interactions between all phases, and the process itself is iterated many times throughout the design process as the technology is developed. Graphic by the author based on Friedman et al. (2013)

indirect stakeholders such as customers who buy products produced at the factory. Philosophers and technology ethicists are particularly well-suited to perform work on the ethical considerations in the conceptual phase, and social scientists can identify stakeholders and help to understand their values.

16.7.2 *Empirical Phase*

In the empirical phase of VSD, the interactions between the technology and people are investigated. Human–robot interaction (HRI) studies are a good example here. Continuing with the cobot example, how do workers expect a cobot to behave? Will they trust it and work in close proximity to it, or will they be afraid of it and stay away? (Read more in the previous chapter on social robots.) The empirical phase includes both the interaction of technology with individuals, but also society more broadly. Taking a drone example, will their more widespread use lead to a “chilling effect” where people assume that they do not have privacy anywhere because of surveillance from drone cameras? (Cawthorne & Cenci, 2019) Can we design drones so it is more obvious what their function is and who is controlling them? (Cawthorne & Frederiksen, 2020) Social scientists and HRI experts have a lot to offer in the empirical phase, and they can use a wide variety of quantitative and qualitative methods to better understand human–technology interactions such as by using surveys, interviews, and focus groups.

16.7.3 *Technological Phase*

In the technological phase of VSD, these inputs from the conceptual and empirical phases are used to design a technology—such as a cobot or drone—that supports the beneficial human values and positive social impacts identified earlier. “The technical phase is dedicated to understanding the artifact (i.e., technology, robot) in context

and how it manifests values or fails to do so” (Van Wynsberghe & Nagenborg, 2016). Alternatively, a technology can be chosen first and then the social and ethical implications can be assessed, or a social phenomenon can provide inspiration for a new technology—the VSD process can be started at any phase (see the section “Practical suggestions for using value sensitive design” in Friedman et al., (2013)).

16.7.4 Contextual Design

VSD is an example of a contextual and embedded design approach—each individual technology is considered within the location of its eventual use and in relation to the people and systems that will be impacted by its uptake. And VSD is an inherently multidisciplinary design approach, since experts from fields such as philosophy and ethics of technology can contribute to the conceptual phase, social scientists to the empirical phase, and engineers and computer scientists to the technological phase. Therefore, it is useful for us as robot systems designers to at least be aware of some of the relevant issues with regard to ethical and social impacts, and collaborate with experts in these fields when developing technology responsibly. Of course, we cannot all become philosophers or social scientists overnight, but taking into consideration philosophical and human interaction issues is part of a holistic, contextually aware, and responsible design practice.

16.8 Ethics Tools

Although it is a developing field, there are already many tools available to make it easier to incorporate ethics into the design of robotic systems.

16.8.1 Checklists

Perhaps the easiest to use is an ethical checklist such as the one utilized in European Union Horizon 2020 projects (European_Union, 2019). The checklist asks yes or no questions about the project, and the questions should identify relevant ethical issues. For example, “does your research involve human participants?” and are they volunteers, vulnerable individuals, or children? “Does your research involve the processing of personal data?”, and does this involve the processing of special categories of personal data such as genetics, sexual lifestyle, ethnicity, religion, etc.? A limitation of such checklists is that they are typically self-administered, and researchers with limited experience working with ethics may not see the potential risks of their technologies. In addition, most ethical issues do not easily resolve themselves to

simple yes or no questions and involve complex reasoning and justification. And it is possible that the checklist may simply omit a relevant ethical issue.

16.8.2 Standards

Another source for ethics guidelines is industry standards. Within robotics, the Institute of Electrical and Electronics Engineers (IEEE) is the “world’s largest technical professional organization for the advancement of technology” (IEEE, 2021). They have just released the 7000 series of standards to address ethical concerns during system design. The standard utilizes human values in design (see the earlier section on human values in design) and contains many elements of value sensitive design (see the earlier section on VSD). Industry standards help engineers design to similar requirements and promote an approach that allows companies to compare their technology to others’. However, standards can be expensive to buy, which can prevent individuals and small businesses from being able to access them.

16.8.3 Design Principles

Design principles and guidelines developed by researchers and organizations can also be helpful. For example, the “privacy by design” guidelines for drones were proposed in 2012 by the Canadian Information and Privacy commissioner (Cavoukian, 2012). These guidelines include proactively designing for privacy preservation (rather than reacting after privacy violations have occurred), privacy as the default setting, and visible and transparent operation—see Table 16.3. These design principles have since been utilized to improve privacy in drones compared to traditional approaches (Cawthorne & Devos, 2020).

The visible and transparent operation of robotic systems can be challenging—how does the robot work? What capabilities does it have? and who is controlling it? These considerations are called “explicability,” and they describe to what extent a system is transparent in its operation, and if its actions can be attributed to a person

Table 16.3 Seven privacy by design guidelines. Any robotic system that uses a camera to sense the world will need to consider privacy issues. Table from Cawthorne and Devos (2020) based on Cavoukian (2012)

Taking a proactive rather than reactive approach
Privacy as the default setting
Embedding privacy in the design
Aiming for full functionality while maintaining privacy
Ensuring full life-cycle protection of sensitive data
Visible and transparent (i.e., explicable) operations
Taking a stakeholder-inclusive approach

or organization that is responsible for it. Design for explicability principles has been proposed within artificial intelligence (AI) (Floridi et al., 2018) and drone design (Cawthorne & Frederiksen, 2020) as both can appear from the outside as “black boxes.” A series of questions to consider in designing drones for explicability have been developed, including “how can the drone be designed to convey the organization and person responsible for it?” and “how can the purpose of the drone (e.g., health care) be easily identified from a distance?” Another example of a design guidelines intended to limit the mis-use or risks drones is the five capability caution principles which ask the designer to consider aspects such as the context of use, the impact on jobs and human skills, and long-term impacts on society and the environment (Cawthorne & Devos, 2020).

Design principles and guidelines can be useful for designers since they pose open-ended questions or offer suggestions which allows room for creativity and context-specific solutions. However, they can be more difficult to apply than a checklist since they are more abstract and require ethically informed critical thinking.

16.8.4 Ethical Frameworks

A final category of tools at our disposal is ethical frameworks. Ethical frameworks are often high level which makes them useful for assessing the overall direction technologies should take and in determining what risks and opportunities may be ahead in the development of a new technology. One ethical framework concerns AI for the good of society (Floridi et al., 2018). It utilizes the four bioethics principles as its foundation—beneficence (do good), non-maleficence (do not do harm), human autonomy, justice, and a new enabling principle for AI—explicability. This framework has subsequently been translated into a drone context, producing an ethical framework for the development of drones in public healthcare (Cawthorne & Robbins-van Wynsberghe, 2020). The framework has been used to develop a prototype fixed-wing drone for rapid delivery of blood samples (Cawthorne & Robbins-van Wynsberghe, 2019); this case study is examined in the next section.

16.9 Case Study: VSD of a Danish Healthcare Drone

Can ethics and value sensitive design help us to design robotic systems that enhance human flourishing? Can the technology be designed so we can avoid some of the risks that we read about at the beginning of the chapter, such as the risk of injuring small children? In this section, we will look at a case study of a Danish healthcare drone developed using VSD (Cawthorne & Robbins-van Wynsberghe, 2019) and an ethical framework (Cawthorne & Robbins-van Wynsberghe, 2020) as a practical example of how these approaches can be used to enhance the design of real robotic systems. The prototype drone is shown in Fig. 16.4.



Fig. 16.4 Prototype Danish healthcare drone; it is the first known example of a drone developed using VSD methods. Image by the author, with permission granted by the subject in the image

Value sensitive design is holistic and contextual, so first it is important to understand the place the drone will operate and the process it could replace. The small, affluent country of Denmark consists of two large islands and a peninsula at the northern tip of Germany, along with many smaller islands. Healthcare services at these small communities may be limited since they are remote, and the small population makes it hard to justify very expensive testing equipment such as those used to analyze blood samples for certain ailments. In addition, the Danish Ministry of Health has been undergoing a process of centralizing healthcare services and has closed several regional clinics while upgrading hospitals in the larger cities into “superhospitals”—instead of 41 hospitals with 24 h care, there will soon only be 20 (Danish_Municipalities, 2015).

For this case study, we focus on the small island of Ærø, located about 25 km south of the central Danish island of Fyn. Currently, an average of 32 blood samples per day are generated at the regional hospital at Ærø and are transported twice a day on weekdays and once a day on weekends by a courier (Sand, 2019). The courier loads the samples into an insulated box and drives them to the port. There, the ferry is used to cross the 25 km of ocean to Fyn. Then, the courier drives a few kilometers to the larger hospital at Svendborg where the samples are analyzed. The infrequent deliveries and dependence on the ferry schedule mean that it can take several hours to get test results (Health_Drone, 2021) meaning some patients could be quarantined unnecessarily or go without proper treatment for some time. Several stakeholders are

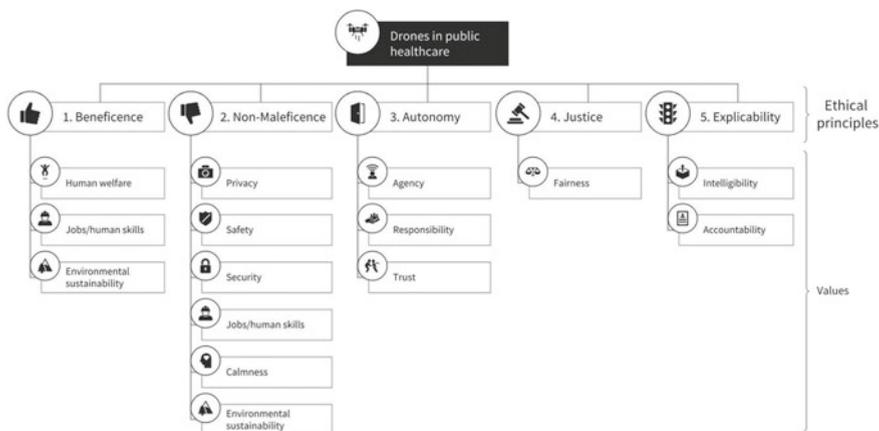


Fig. 16.5 Ethical framework for the design of drones in public health care based on the bioethics principles and AI ethics principles. The high-level ethical principles are made more specific in the second level of the framework which highlights relevant human values such as human welfare, jobs, safety, privacy, and fairness (Cawthorne & Robbins-van Wynsberghe, 2020)

relevant: citizens and sick people living on Ærø, healthcare workers, couriers, ferry operators, and Danish taxpayers among many others.

An ethical framework, shown in Fig. 16.5, was developed for drones used in public healthcare in collaboration with a robot ethicist (Cawthorne & Robbins-van Wynsberghe, 2020). The framework is designed to help the drone designer translate human values into design requirements (Van de Poel, 2013) and is based on bioethics principles since the drone could become part of the healthcare system. This values hierarchy includes four levels: at the top are ethical principles such as beneficence and non-maleficence, next are human values such as human welfare and privacy. The next lower level (not shown) is about contextual norms—specific considerations that pertain to the use-case in question. For example, we could compare the safety of the drone system to that of the current process of driving and taking the ferry. At the base of the hierarchy (not shown), we need to determine the design requirements that will support the ethical principles, human values, and norms—and enhance human flourishing. A detailed account of the development of the drone can be found in the references (Cawthorne, 2020; Cawthorne & Robbins-van Wynsberghe, 2019, 2020) along with its specifications and performance (i.e., fulfillment of the design requirements coming from the ethical framework).

The Danish health care drone is a fixed-wing aircraft which means it can be much smaller and lighter weight than a multirotor drone since flying on wings is more efficient than flying with powered rotors. The drone is so lightweight that it would not cause a fatality even if it were to hit a person on the ground—it is safe by design. The payload of the Danish drone is small, making it useful in urgent cases but not for routine transportation. The drone’s cargo compartment includes a security system making it more difficult to carry unauthorized cargo. The drone is controlled manually

by a pilot using a privacy-preserving camera system which cultivates drone piloting skills and makes responsibility more direct than with an automated system. And the drone is painted bright yellow with dark green checkers like a Danish ambulance, making it clearer what its purpose is and who is responsible for it. If we compare this drone to the one in the opening paragraph of the chapter, we see key differences. The contexts of use are not the same so we should not compare them directly, but the Danish drone exhibits a high level of safety, privacy, security, responsibility, and explicability which could provide health benefits—while protecting those on the ground.

As you can see from this example, developing robots in a holistic and value sensitive way is complex, and there are many impacted stakeholders—some who will benefit from the technology, and some who may be harmed. What is our responsibility as robot developers in this complex process? We will explore the topic in the next section on **Responsible research and innovation**.

16.10 Responsible Research and Innovation

In the previous sections, we saw how technology is not ethically neutral (**the non-neutrality of technology**), which means we need to consider ethics when we design robotic systems. This is the first step in responsible research and innovation—accepting that the things we design have ethical importance. Then, the question turns to how we actually design using ethics as a design input. Here, we could utilize normative ethical theories in Table 16.1, consider human values listed in Table 16.2, and utilize **value sensitive design, checklists, standards, design principles, and ethical frameworks**. These theories and tools can help us to combat moral de-skilling—the process where we become less adept at making ethically informed decisions (Vallor, 2015). Ideally, our moral progress should keep pace with our technological progress.

16.10.1 AIRR Framework

An often-cited framework for responsible research and innovation is called AIRR: anticipation, inclusion, reflexivity, and responsiveness (Stilgoe et al., 2013). “Anticipation prompts researchers and organizations to ask ‘what if...?’ questions, to consider contingency, what is known, what is likely, what is plausible, and what is possible” in the future (Stilgoe et al., 2013)—as we saw earlier in the PPPP model in Fig. 16.2. Inclusion means considering not just powerful stakeholders, but all those that will be impacted—directly or indirectly—by our robots (**the conceptual phase of VSD**). Reflexivity “means holding a mirror up to one’s own activities, commitments, and assumptions, being aware of the limits of knowledge and being mindful that a particular framing of an issue may not be universally held”... “reflexivity means rethinking prevailing conceptions about the moral division of labor within

science and innovation” (Stilgoe et al., 2013)—as exemplified in VSD interdisciplinary approach. Responsiveness “requires a capacity to change shape or direction in response to stakeholder and public values and changing circumstances” (Stilgoe et al., 2013). Responsiveness can be seen in the iterative nature of the VSD process (Fig. 16.3)—as circumstances change, we must adapt our robots to the new situation.

16.11 Chapter Summary

In summary, interdisciplinary collaboration, a holistic perspective, and the ethically informed design of robotic systems give us the best chance to perform responsible research and innovation—and ultimately enhance human flourishing.

16.12 Revision Questions

- What are three normative ethical theories, and what do they say?
- List some human values that are relevant in robot design.
- What is dual-use and what are some implications to your robot design?
- List the three phases of value sensitive design; what activities take place in each phase? Which research areas are most relevant in each phase?
- What are some benefits and limitations of ethical checklists? What about ethical frameworks?
- Identify industry standards related to ethics in technology design?
- Which design principles would be useful in the design of your robot?
- What does AIRR stand for? How could the four phases of the framework be applied to your robot?
- Consider an existing or proposed robot or drone system:
 - Who are the direct and indirect stakeholders?
 - What is the context of use?
 - What might be some social impacts of the system?
 - Should this robot be built? If so, why? If not, why not?
 - How will this robot enhance human flourishing?

References

Ackerman, E. (2019). *Swiss post suspends drone delivery service after second crash: An emergency parachute failure raises questions about the safety of urban delivery drones.* <https://spectrum.ieee.org/swiss-post-suspends-drone-delivery-service-after-second-crash>

- Cavoukian, A. (2012). *Privacy and drones: Unmanned aerial vehicles*. Information and Privacy Commissioner of Ontario, Canada Ontario.
- Cawthorne, D. (2020). *Value sensitive design of unmanned aerial systems*. (PhD Thesis). University of Southern Denmark.
- Cawthorne, D., & Cenci, A. (2019). *Value sensitive design of a humanitarian cargo drone*. Paper presented at the 2019 International conference on unmanned aircraft systems (ICUAS).
- Cawthorne, D., & Devos, A. (2020). *Capability caution in UAV design*. Paper presented at the 2020 International Conference on Unmanned Aircraft Systems (ICUAS).
- Cawthorne, D., & Frederiksen, M. H. (2020). *Using the public perception of drones to design for explicability*. Paper presented at the International Conference on Robot Ethics and Standards (ICRES), Taipei, Taiwan.
- Cawthorne, D., & Robbins-van Wynsberghe, A. (2019). *From HealthDrone to FrugalDrone: Value-sensitive design of a blood sample transportation drone*. Paper presented at the 2019 IEEE International Symposium on Technology and Society (ISTAS).
- Cawthorne, D., & Robbins-van Wynsberghe, A. (2020). An ethical framework for the design, development, implementation, and assessment of drones used in public healthcare. *Science and Engineering Ethics*, 26(5), 2867–2891.
- Choi-Fitzpatrick, A., Chavarria, D., Cychosz, E., Dingens, J. P., Duffey, M., Koebel, K., Siripanh, S., Yurika Tulen, M., Watanabe, H., Holland, J., & Juskauskas, T. (2016). *Up in the air: A global estimate of non-violent drone use 2009–2015*.
- Consequentialism. (2003). *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/consequentialism/>
- Danish_Municipalities. (2015). *Here is where your superhospital is located*. <http://www.danskekommuner.dk/Global/Artikelbilleder/2015/DK-3/DK-3-side-26-27.pdf>
- Deontological Ethics. (2020). *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/ethics-deontological/>
- Dunne, A., & Raby, F. (2013). *Speculative everything: Design, fiction, and social dreaming*. MIT press.
- Ethics. (2021). *Wikipedia*. <https://en.wikipedia.org/wiki/Ethics>
- European_Union. (2019). *How to complete your ethics self-assessment*. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf
- Floridi, L., Cowsls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., & Schafer, B. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.
- Friedman, B., Kahn, P. H., Borning, A., & Huldgren, A. (2013). Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). Springer.
- Gillespie, T. (2010). The politics of ‘platforms.’ *New Media Society*, 12(3), 347–364.
- Health_Drone. (2021). Health Drone: The project. <https://sundhedsdroner.dk/index.php?page=the-project>
- IEEE. (2021). *7000-2021 standard addressing ethical concerns during systems design*.
- Lin, P., Bekey, G., & Abney, K. (2008). *Autonomous military robotics: Risk, ethics, and design*. <https://apps.dtic.mil/sti/pdfs/ADA534697.pdf>
- Matternet. (2021). *Matternet website*. <https://mtr.net/>
- Novitzky, P., Kokkeler, B., & Verbeek, P.-P. (2018). The dual-use of drones. *Tijdschrift Voor Veiligheid*, 17(1–2), 79–95.
- Sand, P. S. (2019). Director after flight in medical drone: I felt completely safe. *Fyns*. <https://www.fyns.dk/erhverv/Direktoer-efterflyvetur-i-laege-drone-Jeg-foelte-mig-helt-tryk/artikel/3318057>
- Stilgoe, J., Owen, R., & MacNaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580.

- SUST. (2019). *Interim report of the Swiss Safety investigation agency SUST about the accident involving the M2 V9 drone*. https://www.sust.admin.ch/inhalte/AV-berichte/ZB_SUI-9903.pdf
- Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy and Technology*, 28(1), 107–124.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Van de Poel, I. (2009). Values in engineering design. In *Philosophy of technology and engineering sciences* (pp. 973–1006). Elsevier.
- Van de Poel, I. (2013). Translating values into design requirements. In *Philosophy and engineering: Reflections on practice, principles and process* (pp. 253–266). Springer.
- Van Wynsberghe, A., & Nagenborg, M. (2016). Civilizing drones by design. In *Drones and responsibility* (pp. 148–165). Routledge.
- Verbeek, P.-P. (2008). Morality in design: Design ethics and the morality of technological artifacts. In *Philosophy and design* (pp. 91–103). Springer.
- Vermaas, P. E., Kroes, P., Light, A., & Moore, S. (2007). *Philosophy and design: From engineering to architecture*. Springer.
- Virtue Ethics. (2016). *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/ethics-virtue/>
- World_Health_Organization. (2020). *The top 10 causes of death*. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>

Dylan Cawthorne is an Associate Professor at the Unmanned Aerial Systems Center at the University of Southern Denmark. His aim is to support human flourishing through the development of ethically informed technologies. His main area of research is using value sensitive design methods and ethical principles to develop and build prototype drones for humanitarian, public health care, and search and rescue operations. He works across disciplines, utilizing art, craft, and creativity to enhance engineering practice. A common theme in his work is the use of holistic and contextually situated technological development. He considers himself an activist engineer, and in his free time, he teaches and performs repairs with the local nonprofit organization Repair Café Odense, and volunteers with Engineers Without Borders Denmark to develop low-cost cloth masks and mapping drones for use in Africa.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits any noncommercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if you modified the licensed material. You do not have permission under this license to share adapted material derived from this chapter or parts of it.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

