Chapter 2
Robot: Multiuse Tool and Ethical Agent

Brigitte Krenn

Abstract In the last decade, research has increasingly focused on robots as autonomous agents that should be capable of adapting to open and changing environments. Developing, building and finally deploying technology of this kind require a broad range of ethical and legal considerations, including aspects regarding the robots' autonomy, their display of human-like communicative and collaborative behaviour, their characteristics of being socio-technical systems designed for the support of people in need, their characteristics of being devices or tools with different grades of technical maturity, the range and reliability of sensor data and the criteria and accuracy guiding sensor data integration, interpretation and subsequent robot actions. Some of the relevant aspects must be regulated by societal and legal discussion; others may be better cared for by conceiving robots as ethically aware agents. All of this must be considered against steadily changing levels of technical maturity of the available system components. To meet this broad range of goals, results are taken up from three recent initiatives discussing the ethics of artificial systems: the EPSRC Principles of Robotics, the policy recommendations from the STOA project *Making Perfect Life* and the MEESTAR instrument. While the EPSRC Principles focus on the tool characteristics of robots from a producer, user and societal/legal point of view, STOA *Making Perfect Life* addresses the pervasiveness, connectedness and increasing imperceptibility of new technology. MEESTAR, in addition, takes an application-centric perspective focusing on assistive systems for people in need.

Keywords Application-centric perspective • Connectedness and increasing imperceptibility of new technology • Ethics for robots as autonomous agents • Human-like communicative and collaborative behaviour • Initiatives discussing the ethics of artificial systems • Pervasiveness • Robots as ethically aware agents • Socio-technical systems • Tool characteristics of robots

B. Krenn (✉)
Austrian Research Institute for Artificial Intelligence (OFAI), Freyung 6/6, 1010 Vienna, Austria
e-mail: brigitte.krenn@ofai.at

R. Trappl (ed.), *A Construction Manual for Robots' Ethical Systems*, Cognitive Technologies, DOI 10.1007/978-3-319-21548-8_2

2.1 Introduction

Robots as we knew them in the past were fully controlled technical devices that are either controlled by a computer programme or a human operator. As regards the former, the robots need to operate in closed, non-changing environments, as it is the case for classical industry robots which can, for instance, be found in the automotive, the chemical, the electrical and electronics, the rubber and plastics or the food industries. In classical industry robotics, all possible events and robot actions are known beforehand, and the robot is programmed accordingly. However, there is a strong demand in industry robotics for robots that are flexible enough to easily adapt to new processes and to collaborate in human–robot teams; cf. [1]. Tele-operated robots are a different kind of controlled robots. They can operate in open environments, because human operators interpret the robot's sensory data and steer the robot's actions. These types of robots are typically employed for operation in conditions that are dangerous for humans, such as underwater, in fire incidents and chemical accidents, warfare, and medical operations, e.g. in minimally invasive surgery [2].

In the last decade, research has increasingly focused on the robot as an autonomous agent that knows its goals, interprets sensory data from the environment, makes decisions, acts in accordance with its goals and learns within an action–perception loop. Thus, the robot becomes more apt to autonomously act in open and changing environments. These developments are of interest for both industry and service robotics. Autonomous robots come in different forms. A prominent example in current times are robots as socio-technical systems assisting people in need. The development of robot companions or robot caretakers that support the elderly is of particular interest from a societal point of view. Europe, especially, has to face a growing share of people aged over 65. According to a Eurostat projection from 2013 to 2080, the population aged 65 years or above will account for 28.7 % of the European population (EU-28) by 2080, as compared with 18.2 % in 2013 [3].

Overall, a broad range of aspects must be considered when discussing robot ethics, including the robots' autonomy, their display of human-like communicative and collaborative behaviour, their characteristics of being socio-technical systems designed for the support of people in need, their characteristics of being technical devices or tools with different grades of technical maturity, including the range and reliability of sensor data, the criteria and accuracy guiding their integration and the quality of the thus resulting actions. A different kind of discussion is needed in the context of basic and applied research, regarding the implementation of a policy to create awareness of potential ethical and legal mishaps a certain research or engineering endeavour may lead to and the countermeasures that need to be taken. To be effective, interdisciplinary contexts must be created where technology development will systematically be intertwined with research on ethical (and psychological) impacts of intelligent, life-like artefacts in general and, even more important, in the light of specific application contexts the technology will be developed for. Some of the relevant aspects must be regulated by societal and legal

discussion; others may be better cared for by conceiving robots as ethically aware agents. All of this must be considered against steadily changing levels of technical maturity of individual system components.

The chapter is organised as follows: In Sect. 2.2, three recent initiatives/instruments are presented which discuss legal and ethical aspects of intelligent artificial systems from complimentary perspectives, including ethical guidelines for robots as technical devices (Sect. 2.2.1), legal and ethical requirements of human–computer interfaces (Sect. 2.2.2) and guidelines for the ethical assessment of socio-technical applications (Sect. 2.2.3). In the remainder of the chapter, these three perspectives are taken up and applied to a broader discussion of robot ethics, taking into account robots as multiuse tools (Sect. 2.3), the special case of care robots (Sect. 2.4), robot ethics and system functionality (Sect. 2.5). The discussions are concluded in Sect. 2.6.

2.2 Ethics: Setting the Context

The last few years have already demonstrated increased awareness regarding the necessity for regulating legal and ethical issues related to new technologies which act autonomously, which are likely to blur boundaries between life-likeness or human-likeness and technology, and which are used as assistive systems for people in need. Three examples for recent results of discussion are (1) the EPSRC Principles of Robotics (UK, 2011), addressing ethical issues of robots viewed as technical tools rather than autonomous, self-learning systems; (2) the policy recommendations from the STOA project *Making Perfect Life* (EU, 2012), more generally addressing the ethics of "intelligent" computer interfaces; and (3) the MEESTAR model (Germany, 2013) which is an analysis instrument for structuring and guiding the ethical evaluation of socio-technical systems, i.e. systems that interact with and support their human users in everyday life. Whereas each of the initiatives has its specific views on the ethical assessment of such systems, all three taken together support a broader discussion of legal and ethical requirements of socio-technical systems.

2.2.1 EPSRC Principles: Ethical Guidelines for Robots as Multiuse Tools

The UK Engineering and Physical Sciences Research Council published the so-called EPSRC Principles of Robotics in 2011. The principles quoted below are the result of a workshop bringing together researchers from different areas including technology, industry, the arts, law and social sciences. The principles 1 to 5 are quoted from [4].

Principles:

1. Robots are multiuse tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.
2. Humans, not robots, are responsible agents. Robots should be designed and operated as far as is practicable to comply with existing laws and fundamental rights and freedoms, including privacy.
3. Robots are products. They should be designed using processes which assure their safety and security.
4. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.
5. The person(s) with legal responsibility for a robot should be attributed.

The EPSRC Principles strongly focus on robots as technical devices, as tools which are used by someone. It is in the nature of tools that they may be used in more than one way and that they are used under human responsibility. For instance, a hammer may be used to nail a picture on the wall, but also to smash somebody's head. How a tool is used is under the responsibility of its user as well as of politics and society providing legal and ethical frames for the uses of the specific kind of tool. In this view on robotic systems, the agent-like aspects of robots, their autonomy, self-learning and adaptive capabilities are not further assessed. However, these are key features of a new generation of robots, which must be addressed, too.

2.2.2 *STOA Project* Making Perfect Life: *Ethical Requirements of Human–Computer Interfaces*

Another workshop, held in 2011, was initiated by the European STOA project *Making Perfect Life*: *Human–Computer Interfaces*. It brought together experts from law, behavioural science, artificial intelligence, computer science, medicine and philosophy. "Implanted Smart Technologies: What Counts as 'Normal' in the 21st Century?" was discussed as overall topic. Results are published in [5]. STOA is the European Parliament's Science and Technology Options Assessment (http://www.europarl.europa.eu/stoa/). The project *Making Perfect Life* (2009–2011) looked into selected fields of engineering artefacts and resulting consequences for policymaking. As for human–computer interfaces, the study distinguishes three types of systems and makes related high-level policy recommendations. Systems are grouped into:

1. **Computers as human-like communication partners**: The computer takes on several roles such as teacher, nurse and friend and acts and communicates accordingly.
2. **Computers as devices for surveillance and alert**: The computer monitors, measures and intervenes with human states such as attention, fatigue, etc.

3. **Ambient intelligence and ubiquitous computing**: The computer becomes more and more imperceptible.

For details, see [5], p. 130f.
The resulting policy recommendations address:

1. **Data protection**, specifically for pervasive and highly connected IT systems.
2. **Privacy, transparency and user control** must be embedded in systems design.
3. An **external regulating body** is required to monitor technology developments and issue warnings with respect to ethical, legal and societal challenges.

See [5], p. 131.

Robots and in particular care robots are realisations of the first two types of systems, i.e. "computers as human-like communication partners" and "computers as devices for surveillance and alert", and they feed data into the third type of systems ("ambient intelligence and ubiquitous computing"), for instance, when they transmit information to applications of telemedicine. They simulate human-like communicative behaviour. They survey and measure their human fosterlings' states, and apart from merely transmitting these data to external services, they are designed to intervene when something goes wrong or moves into an undesirable direction. This immediately leads into ethical discussion of what is (un)desirable under which circumstances, who determines it and according to which criteria. Here the MEESTAR analysis instrument [6] comes into play. It is an attempt to guide the ethical assessment of socio-technical systems. These are systems that interact with and support their human users in everyday life. MEESTAR was developed having in mind assistant systems for the elderly; however, the instrument as such can be applied to assess any socio-technical system.

2.2.3 MEESTAR: Ethical Assessment of Socio-Technical Applications

The major characteristics of MEESTAR are as follows: (1) It is geared to model a specific application scenario, i.e. the specific assistant needs of a concrete person in her/his social context. It does not aim at universal validity. On the contrary, MEESTAR is an instrument to identify at any time the ethical objectionability of concrete applications. The focus lies on identifying and solving ethically problematic effects of the socio-technical system under assessment. (2) The model takes into account the perspectives of different groups of persons including those who use the system such as the elderly, professional caretakers as well as family and friends, the system providers and its developers. A minimal requirement is that the socio-technical system must not do any harm or only a minimum of harm, given the benefit of the system clearly exceeds the harm it may cause. This must be transparent and in consent with the persons concerned.

MEESTAR assessments focus on ethically negative aspects of a socio-technical application. They are guided by questions regarding three levels of analysis:

1. **Ethical dimensions**: care, autonomy/self-determination, safety, equity, privacy, participation and self-conception. At this level, the content of the ethical questions is formulated.
2. **Ethical objectionabilities related to a specific ethical question given a particular application scenario**: A specific socio-technical application may be uncritical, (b) ethically sensitive but can be handled in practic, (c) ethically highly sensitive with need to be constantly monitored, or (d) the application must be rejected because of severe objections.
3. **Perspectives under which 1. and 2. are assessed**: individual, organisational and societal.

While the EPSRC Principles focus on the tool characteristics of robots from a producer, user and societal/legal point of view, STOA *Making Perfect Life* addresses the pervasiveness, connectedness and increasing imperceptibility of new technology. MEESTAR, in addition, takes an application-centric perspective focusing on assistive systems for the elderly. As MEESTAR has been designed for assessing the ethical objectionability of a concrete application for a specific person in her or his social context, the model provides explicit questions for guiding the ethical assessment. MEESTAR assessments are complex qualitative decision processes which cannot be directly implemented on a computer system. However, thinking of robots as autonomous agents with ethical responsibility, the MEESTAR model can be seen as a starting point for deriving capabilities an ethically aware artificial agent should be equipped with. What the EPSRC Principles, the STOA *Making Perfect Life* and MEESTAR can do for developing ethically aware artificial agents will be explored in the following sections.

2.3 Robot Ethics Under the Perspective of Robots as Multiuse Tools

Under the assumption of robots as multiuse tools, the manufacturers and users are responsible for their robots. In this respect, the main discussion in robot ethics must concentrate on the societal and legal frame of robot use. A transparent and broad societal and political discussion of technology is required, in particular of technology which is part of devices which are already in the market or soon to be launched. This is an interdisciplinary endeavour including experts from various fields such as computer science, engineering, AI, ethics, philosophy and law, as well as the general public, especially after expert discussions have reached a certain level of maturity.

In this respect, the formulation of robot ethics requires first of all the articulation of good habits and standards a society and their members should adhere to in the

development and use of intelligent, (semi-) autonomous, agentive artificial systems. It is the task of *normative ethics* to devise moral standards that regulate right and wrong conduct; cf. [7].

Understanding a robot as a multifunctional technical device also suggests that robots should be conceived as implicit ethical agents. Therefore, in a first step, we should strive at developing artificial agents whose actions are constrained in such a way that unethical outcome can be avoided. To achieve this, strategies are required to systematically assess the ethical implications of an application, and this is where the MEESTAR framework comes into play. Even though the MEESTAR instrument was developed with focus on caregiving for the elderly, the questions guiding the ethical assessment can be generalised to any socio-technical system. Following is the adapted list of guiding questions. For the original formulation of the questions (in German), see Appendix 1:

1. Is the use of a specific type of assistant system ethically questionable or not?
2. What are the specific ethical challenges?
3. Given the use of a specific kind of assistant systems, is it possible to attenuate or even resolve related ethical problems? If yes, what would be potential solutions?
4. Are there (potential) situations in the use of the system which are ethically so alarming that the system should not be installed and used?
5. Did the use of the system lead to novel and unexpected ethical problems which were not anticipated during the design of the system?
6. What are the specific aspects and functionalities of the system under investigation which require specific ethical care?

Summing up, in a first stage of the development of robot ethics, the following issues must be dealt with:

1. Robots, including sociable robots, are technical devices/multifunctional tools and should be treated as such. This also holds for ethic requirements imposed on robots. Therefore, measures to be taken to implement robot ethics at technology level must accord with the ethical and legal framework devised at societal and political levels. This framework however still needs to be defined.
2. When we talk about robot ethics, we should talk about normative ethics for the use of robots, i.e. right and wrong conduct of robots is the responsibility of the robot users and not of the robots themselves.
3. Following from claim 2, a robot should not be ethical by itself; it should be ethically used. Therefore, robots should be conceived as implicit ethical agents.
4. The discussion about robot ethics should be divided into ethical and legal issues concerning smart and (semi-) autonomous technology (a) that is already integrated or on the verge of being integrated into commercial applications and (b) that is a matter of basic research. While for the former a broad societal consensus and clear legal regulations are required, for the latter, the discussions will be on a more explorative level, together with round tables of groups of experts from various fields, including technology, AI, philosophy, medicine, law, etc.

2.4 The Special Case of Care Robots for the Elderly: Ethical Dimensions Under Assessment in the MEESTAR Model

In Table 2.1, a summary is provided of the ethical dimensions and related questions investigated by MEESTAR, and it is assessed what they mean in terms of intelligent agents. What are the relevant questions for their assessment, and what would be required for their implementation in a robot?

Summing up, the preceding discussion of potential realisations of MEESTAR ethical dimensions within an artificial agent provides input to requirements on modelling mind components for explicit ethical agents.

2.5 Robot Ethics and System Functionality

Robots are a specific type of human–computer interfaces; thus, the considerations from both EPSRC and STOA *Making Perfect Life* hold for robots and determine robot ethical requirements. On the one hand, robots are artefacts, tools and manufactured products for which the human manufacturers and users have legal responsibility. On the other hand, robots are human–computer interfaces that may be designed to simulate human communication and social interaction, to function as devices for surveillance and alert and to operate on data from virtual as well as real-world contexts. They may be equipped with technology that allows them to connect to the internet and to technical devices in their vicinity including smartphones, tablets, sensors and actuators of smart homes. Being computers and hooked up to other computers on which virtually any programme may run, robots do not only have physical presences with specific object/body features but also may create a broad range of virtual presences. This broad potential is constrained by the specific realisation of a particular robot and by its application scenario. Both condition the requirements for the robot to be ethically and legally compliant.

From a point of view of technical realisation, there exists a broad range of mechanisms that may be built into a robot in order to facilitate its ethically compliant use and behaviour. To achieve this, however, we need to know what should constitute ethically compliant behaviour of a specific robot in a concrete application scenario. The definition and formalisation of what is ethical under which conditions are by far harder than their technical implementation. The following is a checklist of technical dimensions that should be considered in order to devise an artificial (implicit or explicit) ethical agent.

Table 2.2 contains a checklist for creating an ethical artificial agent. Guiding questions are posed from a perspective of robots as situated perceptors and actors.

Different constraints for ethical and legal use apply, depending on what can be perceived, which actuators a robot has in use, what the application scenario is and who the users are. Conceiving robots as multifunctional tools also implies the idea of flexible assembly of different functionalities on an individual robot. This

Table 2.1 MEESTAR ethical dimensions, related questions and their potential for realisation in an artificial agent

Ethical dimension	Related questions
Care (Ge.: Fürsorge) To support the ability of a person in need to conduct a self-governed life	Q1: At which point does the technically supported care for a person in need become problematic, because it changes the person's self-esteem and her/his relationship to the world in a way which is not desirable for the person—from her/his own point of view, as well as from an external perspective? Q2: What kind of dependencies in caregiving structures are still acceptable and desirable, and at which point does the positive intention of care turn into paternalism which may be supported or caused by the technical system?
Potential for realisation in an artificial agent/robot	
Needed are intelligent systems that assess, monitor (over time) and foster the user's, i.e. care receiver's, self-esteem and avoid paternalistic behaviours in caregiving. This requires first of all models for the assessment of self-esteem which must be informed by psychology and nursing science and strategies to avoid paternalism in caregiving which also must be informed by nursing science. These models must be implemented in such a way that they are intertwined with the agent's long-term memory LTM and its dialogue system. These are preconditions for making the agent capable of asking questions regarding the assessment of the user's self-esteem and for initiating supportive dialogue	
Autonomy/self-determination(Ge.: Selbstbestimmung) To support freedom of choice and action for the individual To foster social inclusion	Q1: How can people be supported in their right to exercise self-determination? Q2: How can people be supported in their self-determination, for whom the "normal" criteria of self-determined decisions and actions have become questionable or obsolete? Q3: How do we handle the discrepancy that the ascription of self-determination can be in conflict with the demand for care and support?

(continued)

Table 2.1 (continued)

Ethical dimension	Related questions
Potential for realisation in an artificial agent	
Criteria may be implemented to provide levels of choice for people without or with different levels of impairment. This may be achieved by making use of a computer system's capability to constantly monitor its user and environment and to assess the resulting data according to criteria derived from the agent's theory of mind (TOM) of the user and its theory of the user's physical condition. While there is work on modelling TOM in virtual agents and robots [8, 9], computational models of physical condition still need to be developed, taking into account insights from nursing science	
Ethical dimension	Related questions
Safety (Ge: Sicherheit) To prevent the patient to be harmed To ensure immediate service/support in health critical situations To ensure operating safety in the intelligent home To increase objective safety and the subjective feeling of being safe for the person concerned and the caregivers	Q1: How to deal with the effect that the creation of safety (by the socio-technical system) may decrease the existing capabilities of the human (i.e. when people start to rely on technology, they may stop taking care of things themselves)? Q2: How should it be assessed that the assistant system increases the subjective feeling of safety without objectively increasing safety? Q3: How can conflicts be solved between safety and privacy or privacy and self-determination?
Potential for realisation in an artificial agent	
While Q2 and Q3 are subject for ethical discussion at societal level, solutions to Q1 are well suited to be realised as part of the agent's ethical system: A well-funded TOM and theory of the user's physical condition integrated with the agent's LTM and dialogue system allow the agent to encourage the user to do things on her or his own. The artificial agent monitors and assesses the situation and takes initiative only when absolutely needed	

Ethical dimension	Related questions
Privacy (Ge.: Privatheit) On the one hand, age-adequate assistant systems should do their work as discreet and invisible as possible; on the other hand, almost always assistant systems are based on collecting, processing and evaluating sensitive personal data. Both aspects together may be in conflict with the ethically motivated postulation of informed consent	Q1: How can privacy—above informational self-determination—be assured as a moral right for the individual when designing age-adequate assistant systems? Q2: How can privacy be protected for cognitively impaired people? Q3: How to deal with cultural differences in the assessment of the private and the public sphere, e.g. when introducing age-appropriate assistant systems for people with migration background?
Potential for realisation in an artificial agent	
The ethical assessment of privacy comprises the following issues: data protection, protection of privacy in general, protection of privacy for cognitively impaired people as well as cultural differences in what is considered as private and what as public As stated in the policy recommendations resulting from STOA *Making Perfect Life*, specific data protection regulations are needed for artificial systems that are increasingly pervasive, distributed and connected. Apart from the necessary societal and legal developments, different levels of data protection and security should be implemented in the respective socio-technical application, so that it is hard to (nearly) impossible for unauthorised persons to access the data collected by the agent. This addresses data stored in the agent's memory as well as data that are transmitted by the agent to external servers. While data protection and security at the agent level is a matter of low-level technical solution, the protection of the user's privacy lends itself to be modelled as part of the agent's cognitive system, combining LTM, dialogue system, a model of what is considered to be private in the concrete area of application of the given socio-technical system and the TOMs of its users (e.g. the person cared for, the caregivers) augmented with a cultural dimension of discretion. The goal is to steer the agent's dialogue and action strategies. In addition, the agent's culturally augmented user model of the person cared for may also help the caregivers to better understand the individual needs for privacy of the person cared for. As regards the dialogue capabilities, the agent should be able to articulate which security levels apply to which functionality. Moreover, it should be able to issue warnings in the dialogue with the user, for instance: "Are you sure you want to put this information on Facebook?" or "Did you know that someone from outsight might be able to listen to our conversation?" Equally, the user should be able to tell the agent that some information is strictly confidential and must not be shared with anybody else or only with a certain restricted set of people. Accordingly, the agent needs to be aware to whom it is transmitting what kind of information. As far as communication with other computer systems is concerned, this requires an elaborate concept of data security and its implementation. In addition, it requires the implementation of methods for user identification when it comes to face-to-face communication with humans. This may be done by voice-based speaker identification or other biometric identification methods as face scan, finger or palm print, and iris and retina scan [10–12]	

(continued)

Table 2.1 (continued)

Ethical dimension	Related questions
Equity (Ge.: Gerechtigkeit) To ensure social justice, inter- and intragenerational equity and access to age-appropriate assistant systems	Q1: Who is granted access to age-appropriate assistant systems? Q2: How should age-appropriate assistant systems be financed (how pays how much)? Q3: What is the understanding of intra- and intergenerational justice?
Potential for realisation in an artificial agent	
The ethical dimension equity addresses aspects of socio-technical systems which are outside of the agent and must be regulated by societal and political discussion	
Ethical dimension	Related questions
Participation (Ge.: Teilhabe) To support a self-governed life and equal participation in societal life	Q1: What is the participation of elderly people in societal life, who are not or cannot be part of the labour force anymore? What kind of participation do they wish for themselves? Q2: What kind of participation is (a) aimed at with age-appropriate assistant systems and (b) which one is actually fostered?
Potential realisation in an artificial agent	
Artificial agents have a high potential to foster participation, because of their ability to connect with and monitor the activities in social networks and to influence the group dynamics in virtual communities [13]. The artificial personal assistant may help its user to select appropriate social networks and monitor network activity. In addition, it may support people with special needs in their communication, e.g. in case of typing impairment by making use of intelligent, personalised auto-completion [14], or for vision impaired by making use of text-to-speech technology [15, 16]	

Ethical dimension	Related questions
Self-conception (Ge.: Selbstverständnis)	Q1: How do socio-technical systems account for the question of meaning which may be of particular interest in old age? Q2: In how far changes the tendency of medicalising life cycles the attitude towards age and ageing? Q3: What are the direct or indirect social restraints of dominant views regarding medicalised and technically supported ageing? Q4: In how far establishes age-appropriate technology routines of standardisation?
Potential realisation in an artificial agent	
While questions Q1 to Q3 are a matter of societal discussion, intelligent technology has the potential to counteract routinely grinding-in of treatments and instead support a broader bandwidth of caregiving strategies and behaviours	

Table 2.2 Checklist for creating an ethical artificial agent

What is the agent's perception space? What are the agent's perceptors/sensors?	Does the agent perceive data from digital worlds, from the physical world or from both? Perceptors can be computer software that collects data from virtual environments (e.g. e-mail, social media, telephone links, queries to search engines, etc.) as well as software that collects data in the physical world (e.g. vision data, audio, biofeedback, etc.)
This gives information about which data can be gathered by the system and thus about the necessities for data protection and data security	
What is the agent's action space? What are the agent's actuators?	Does the agent act in digital and/or physical environments? Actuators can be computer software that triggers individual actions in the agent's virtual or physical environments. In this context, it is important to assess what is the (potential) outcome of each single action the agent is capable of
These considerations help to assess the potential of each agent action to do harm in the virtual and/or physical world. Becoming clear about this is a precondition to devise respective control mechanisms, internal and external to the agent	
What are the dimensions of autonomy built into the agent?	This question mainly addresses the working mechanisms of the agent's interpretation and control layers, i.e. those aspects of the system that interpret the sensor data and decide upon which actions will be triggered. This leads to further questions, including: Who is the actor and who has control over the action—the human, the robot or both? To which extent and related to which aspects are learning mechanisms employed in the system components?
With increasing autonomy of the system, agenthood is shifted from the human to the robot, and the more it is necessary to build mechanisms into the artificial agent that allow the agent itself to be aware of its actions and their potential effects and to be transparent about the reasons for the respective actions	
What is the degree of human-likeness in the agent's appearance and behaviour?	Does the system aim at an illusion of human-likeness, e.g. does it engage in natural language communication, to which extent has the agent features of a human body, does the agent emit socio-emotional signals and does it engage in social interaction?
These questions help to assess how likely it is that the agent's simulation of human-likeness will deceive its human user. Depending on the application scenario, the user's knowledge about the technology and her or his mental condition, the assessment of one and the same technical device may lead to different results	

requires certain mechanisms that allow for flexible connection and disconnection of functionalities on the robot at perception and action levels as well as their integration into the robot's control mechanisms (mind), also including mechanisms that support ethically compliant robot action. This requires:

- An action–perception architecture that allows to connect and disconnect action and perception components, i.e. the agent's tools and senses to interact with the outside world, be it a virtual or a physical one
- Models and mechanisms to structure the agent's knowledge of self, others and the environment it is acting in
- Mechanisms that generate natural language utterances based on the agent's memory content and its various models of self, others and the environment

For initial work in this direction, see, for instance, [17–19].

2.6 Conclusion

The formulation of ethical principles for robots has different facets and is a moving target, especially as the technical developments in modelling self-learning, autonomy and natural language faculty are successively improving. Depending on the technical realisation of a robot and its area of application, different requirements regarding robot ethics apply, including question of legal liability, data collection and privacy as well as the rights of those people who are given care by assistive robots. In this chapter, three recent initiatives debating aspects of the above-mentioned requirements are discussed, including the EPSRC Principles of Robot Ethics, the STOA project *Making Perfect Life*: Human–Computer Interfaces and the MEESTAR instrument for assessing the ethical implications of socio-technical systems. While the EPSRC Principles focus on robots as multifunctional technical devices their human producers and users are responsible and liable for, the STOA project *Making Perfect Life* defines policy recommendations for computer systems that act as human-like communication partners and surveillants, and the MEESTAR model is devised to guide the ethical assessment of socio-technical systems in concrete application scenarios.

Understanding a robot as a multifunctional technical device also suggests that the robot should be conceived as implicit ethical agent. In this respect, it is argued in the chapter that, first of all, developers should strive at creating artificial agents whose actions are constrained in such a way that unethical outcome can be avoided. In this respect, creating an implicit ethical agent is an issue of robot design. To find out about relevant design criteria, strategies are required to systematically assess the ethical implications of concrete applications, and MEESTAR provides a framework for this kind of assessment. Furthermore, in this chapter, the MEESTAR ethical dimensions and related questions are assessed with respect to their potential for realisation in an artificial agent's mind. For instance, while data protection and security at agent level is a matter of low-level technical solution suitable to be

realised in an implicit ethical agent, the protection of the user's privacy lends itself to be modelled as part of the agent's cognitive system, combining long-term memory, dialogue system, a model of what is considered to be private in a concrete area of application of a given socio-technical system and respective theories of mind (TOM) of the agent's users (e.g. the person cared for and the caregivers) augmented with a cultural dimension of discretion. This already requires the realisation of explicit ethical agents capable of identifying and interpreting relevant information and deriving ethically sound behaviours. For a distinction of implicit and explicit ethical agents, see, for instance, [20].

Overall, two bodies of questions arise for the development of ethically aware agents: (1) How to determine what we expect from an ethical agent? This includes questions such as: In which sense an artificial agent should be ethical? What are the ethical requirements we pose on robots in specific application scenarios? How do we determine these requirements? Instruments such as MEESTAR help to further assess these questions. (2) What are the preconditions to be modelled and technically implemented in order to create ethically aware artificial agents? This implies questions such as: What kind of ethically aware artificial agent can be realised given the state-of-the-art in technical as well as in model development? For instance, well-funded TOM models and theories of users' mental and physical condition are required for health care and assistant robots. Accordingly, developing ethical agents not only requires close collaboration between technicians such as computer scientists and AI researchers, philosophers and lawyers but also must include experts from the specific application domains an artificial agent is going to be developed/deployed for.

Appendix 1: MEESTAR Guiding Questions Original Formulation (German)

1. Ist der Einsatz eines bestimmten altersgerechten Assistenzsystems ethisch bedenklich oder unbedenklich?
2. Welche spezifisch ethischen Herausforderungen ergeben sich durch den Einsatz eines oder mehrerer altersgerechter Assistenzsysteme?
3. Lassen sich ethische Probleme, die sich beim Einsatz von altersgerechten Assistenzsystemen ergeben, abmildern oder gar ganz auflösen? Wenn ja, wie sehen potenzielle Lösungsansätze aus?
4. Gibt es bestimmte Momente beim Einsatz eines altersgerechten Assistenzsystems, die ethisch so bedenklich sind, dass das ganze System nicht installiert und genutzt werden sollte?
5. Haben sich bei der Nutzung des Systems neue, unerwartete ethische Problempunkte ergeben, die vorher – bei der Planung oder Konzeption des Systems – noch nicht absehbar waren?

6. Auf welche Aspekte und Funktionalitäten des untersuchten altersgerechten Assistenzsystems muss aus ethischer Sicht besonders geachtet werden?

   Quoted from [6], p. 14.

Appendix 2: Ethical Dimensions Assessed in MEESTAR, Original Formulation (German)

All quotes [6], pp. 16–20.

Ethical dimension	Related questions
Care (Ge.: Fürsorge)	Q1: "An welchem Punkt wird eine technisch unterstützte Sorge für hilfebedürftige Menschen problematisch, weil sie das Selbstverhältnis und das Weltverhältnis dieser Menschen auf eine Weise verändert, die diese selbst nicht wünschen bzw. die wir Anderen im Blick auf diese Menschen nicht wünschen sollen?" p. 16 Q2: "Welche Grade der Abhängigkeit in Fürsorgestrukturen sind noch akzeptabel bzw. gewünscht und ab welchem Punkt wird aus positiv gemeinter Fürsorgehaltung eine Bevormundung bzw. eine negativ bewertete paternalistische Einstellung, die unter Umständen technisch unterstützt bzw. hergestellt werden kann?" p. 16
Autonomy/self-determination (Ge.: Selbstbestimmung)	Q1: "Wie können – in Anlehnung an eine konsequent am Selbstbestimmungsrecht des Einzelnen orientierte Praxis – Menschen bei der Ausübung ihrer Selbstbestimmung unterstützt werden?" p. 16 Q2: "Wie können Menschen in ihrer Selbstbestimmung unterstützt werden, bei denen die 'normalen' Kriterien selbstbestimmten Entscheidens und Handelns fraglich oder gar hinfällig geworden sind?" p. 16 Q3: "Wie gehen wir damit um, dass die Zuschreibung von Selbstbestimmung mit dem Anspruch auf Fürsorge und Unterstützung in Konflikt treten kann?" p. 16
Safety (Ge: Sicherheit)	Q1: "Wie ist dem zu begegnen, das die Herstellung von Sicherheit unter Umständen zur Verringerung vorhandener Fähigkeiten führt, d.h. wenn Menschen beginnen, sich auf Technik zu verlassen, hören sie vielleicht auf, sich selbst um bestimmte Dinge – in einem produktiven Sinn – zu sorgen?" p. 17 Q2: "Wie ist es zu bewerten, wenn durch ein Assistenzsystem das subjektive Sicherheitsgefühl steigt, ohne dass objektiv die Sicherheit erhöht wurde?" p. 17 Q3: "Wie können Konflikte zwischen Sicherheit und Privatheit oder Sicherheit und Selbstbestimmung (Freiheit) gelöst werden?" p. 17

(continued)

Ethical dimension	Related questions
Privacy (Ge.: Privatheit)	Q1: "Wie kann die Privatsphäre des Einzelnen über die informationelle Selbstbestimmung hinaus als moralischer Anspruch bei der Gestaltung altersgerechter Assistenzsysteme zur Geltung gebracht werden?" p. 18 Q2: "Wie kann die Privatheit kognitiv eingeschränkter Menschen geschützt werden?" p. 18 Q3: "Wie ist mit kulturellen Unterschieden in der Bewertung von privater und öffentlicher Sphäre umzugehen – z.B. bei Einführung von altersgerechten Assistenzsystemen bei Menschen mit Migrationshintergrund?" p. 18
Equity (Ge.: Gerechtigkeit)	Q1: "Wer bekommt Zugang zu altersgerechten Assistenzsystemen?" p. 18 Q2: "Wie soll die Finanzierung von altersgerechten Assistenzsystemen gestaltet werden (wer zahlt wie viel)?" p. 18 Q3: "Welches Verständnis von intragenerationeller und intergenerationeller Gerechtigkeit liegt vor?" p. 18
Participation (Ge.: Teilhabe)	Q1: "Welche Teilhabe besteht für ältere Menschen, die nicht mehr in das Arbeitsleben integriert werden (sollen)? Welche Teilhabe wünschen sie sich?" p. 18 Q2: "Welche Art und Weise der Teilhabe wird durch altersgerechte Assistenzsysteme a) anvisiert und b) tatsächlich gefördert? Inwiefern werden durch technische Assistenzsysteme bestimmte Teilhabevarianten be- oder verhindert?" p. 18
Self-conception (Ge.: Selbstverständnis)	Q1: "Wie wird der Sinnfrage, die im Alter verstärkt auftreten mag, Raum und Perspektive in sozio-technischen Arrangements geboten?" p. 19f Q2: "Inwiefern verändert die Tendenz zur Medikalisierung des Lebens auch die Haltung zum Alter und Altern?" p. 19f Q3: "Welche (direkten oder auch indirekten) sozialen Zwänge entstehen durch dominante Bilder des medikalisierten bzw. technisch unterstützten Alter(n)s?" p. 19f Q4: "Inwiefern werden durch altersgerechte Technik Normierungsroutinen etabliert?" p. 19f

References

1. World Robotics – Industrial Robots 2014. Statistics, Market Analysis, Forecasts and Case Studies. IFR Statistical Department, Frankfurt, Germany. http://www.worldrobotics.org/uploads/media/Executive_Summary_WR_2014_02.pdf http://www.worldrobotics.org/uploads/media/Foreword_2014_01.pdf
2. Lichiardopol, S.: A Survey on Teleoperation. DCT Report, Technische Universiteit Eindhoven (2007)
3. Eurostat, http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing
4. EPSRC Principles of Robotics, http://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/
5. van Est, R., Stemerding, D. (eds.): Making Perfect Life. European Governance Challenges in 21st Century Bio-engineering. European Parliament STOA – Science and Technology Options Assessment (2012) http://www.rathenau.nl/uploads/tx_tferathenau/Making_Perfect_Life_Final_Report_2012_01.pdf

6. Manzeschke, A., Weber, K., Rother, E., Fangerau, H.: Studie "Ethische Fragen im Bereich Altersgerechter Assistenzsysteme". VDI/VDE Innovation + Technik GmbH, Januar (2013)
7. Internet Encyclopedia of Philosophy, http://www.iep.utm.edu/ethics/
8. Hiatt, L.M., Harrison, A.M., Trafton, G.J.: Accommodating human variability in human–robot teams through theory of mind. In: Walsh, T. (ed.) Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI'11), vol. 3, pp. 2066–2071. AAAI Press (2011)
9. Pynadath, D.V., Si, M., Marsella, S.C.: Modeling theory of mind and cognitive appraisal with decision-theoretic agents. In: Gratch, J., Marsella, S. (eds.) Social Emotions in Nature and Artifact, pp. 70–87. Oxford University Press, Oxford (2013)
10. Orság, F.: Speaker recognition in the biometric security system. Comput. Inform. **25**, 369–391 (2006)
11. Chakraborty, S., Bhattacharya, I., Chatterjee, A.: A palmprint based biometric authentication system using dual tree complex wavelet transform. Measurement **46**(10), 4179–4188 (2013)
12. Dehghani, A., Ghassabi, Z., Moghddam, H.A., Moin, M.S.: Human recognition based on retinal images and using new similarity function. EURASIP J. Image Video Process. **2013**, 58 (2013)
13. Skowron, M., Rank, S.: Interacting with collective emotions in e-Communities. In: Von Scheve, C., Salmela, M. (eds.) Collective Emotions, Perspectives from Psychology, Philosophy, and Sociology. Oxford University Press, Oxford (2014)
14. Matiasek, J., Baroni, M., Trost, H.: FASTY – a multi-lingual approach to text prediction. In: Miesenberger, K., et al. (eds.) Computers Helping People with Special Needs, pp. 243–250. Springer, Heidelberg (2002)
15. King, S., Karaiskos, V.: The blizzard challenge 2009. In: Proceedings of the International Blizzard Challenge TTS Workshop (2009)
16. Laghari, K., et al.: Auditory BCIs for visually impaired users: should developers worry about the quality of text-to-speech readers? In: International BCI Meeting 3–7 June, Pacific Grove, CA (2013)
17. Eis, C., Skowron, M., Krenn, B.: Virtual agent modeling in the RASCALLI platform. In: PerMIS'08 – Performance Metrics for Intelligent Systems Workshop, August 19–21, pp. 70–76. Gaithersburg, MD, USA (2008)
18. Skowron, M., Irran, J., Krenn, B.: Computational framework for and the realization of cognitive agents providing intelligent assistance capabilities. In: The 18th European Conference on Artificial Intelligence Proceedings 6th International Cognitive Robotics Workshop, July 21–22, pp. 88–96. Patras, Greece (2008)
19. Gregor Sieber, G., Krenn, B.: Towards an episodic memory for companion dialogue. In: Allbeck, J., et al. (eds.) Intelligent Virtual Agents, LNAI 6356, pp. 322–328. Springer, Heidelberg (2010)
20. Moor, J.H.: The nature, importance, and difficulty of machine ethics. In: Anderson, M., Anderson, S.L. (eds.) Machine Ethics, pp. 13–20. Cambridge University Press, New York (2011)