9. Robot ethics: Assessing risk from the perspective of autonomy

Pablo Jiménez-Schlegl *0000-0003-3627-4938, pjimenez@iri.upc.edu,*

Júlia Pareto *0000-0003-4879-8800, jpareto@iri.upc.edu,*

Carme Torras *0000-0002-2933-398X, torras@iri.upc.edu*

Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
Llorens i Artigas, 4-6, 08010 Barcelona, Spain

Keywords Robots, Robotics, Embodied AI, Autonomy, Robot Applications, Ethical Concerns of Robotics, Robot Ethics

Abstract

Robots are machines and have thus a physical presence. They can also be programmed with AI software, for learning, reasoning, planning and decision making. For this reason, they can as well be considered as embodiment of AI. It is precisely the touch of AI that elevates robots over ordinary machines, enabling autonomy beyond mere automatism. Autonomy plus enhanced perception, mobility and dexterity allows robots to expand their activity in unstructured and unpredictable surroundings, pervading all the environments of human activity, in work and leisure, home and abroad, in intimacy and in public spaces. Robot users could be not only professionals, but practically anyone. The close and extended contact between humans and robots would provide a plethora of situations where important ethical questions may arise, concerning security, privacy, dignity, liberty, equality/justice, or sustainability. The possible ways to tackle such issues include normative/legal initiatives, dissemination/education of all possible stakeholders, as well as programming/learning of ethics in(to) robots.

Summary: 9.1. The robot as embodied AI: machines, automatism and autonomy. 9.2. A characterization and a categorization of robots. 9.3. Kinds of autonomy in robotics. 9.3.1. Quantifying autonomy. 9.4. Human perception of robot autonomy and user satisfaction. 9.5. Robot autonomy in application domains and the rise of ethical concerns. 9.5.1. Technological autonomy in assistive robotics: The emergence of social assistive robots. 9.5.2 The risks of technological autonomy in social assistive robots. 9.6. Conclusions.

9.1. **The robot as embodied AI: machines, automatism and autonomy.**

From the perspective of AI, robots are bodies. For an individual AI, understanding as such a particular set of cognitive programs housed in the same computer, a given robot is its body. It is the means that enables the AI to have a direct perception of its surroundings, to interact with the world. The sensors a robot is equipped with provide information of the current state of the robot itself (proprioception and interoception) and of its surroundings (exteroception). The actuators (motors), on the other hand, furnish the robot's mechanical

structure with the required motion to fulfill actions, by transforming electrical energy into movements. Such actions may be just displacements (locomotion) or some sort of manipulation, possibly including the use of a tool or device. The characteristics of the robot shape the way the AI can interact with the physical world: the body sets the limits of what the AI can do and what cannot. The link to the material world delivered by the robotic body comes at the price of subjecting that AI to the laws of physics. It is an anchor on reality, which, metaphorically, wakes up the AI from dwelling in the potentially unbounded dreamscapes of simulations[1]. Unlike virtual worlds, resetting is not possible in the physical world, the AI is confronted with the possible irreversibility of decisions translated into actions.

Robots are machines. A machine is a human-made artifact, that can be as simple as a pulley or a bellows, or as complex and sophisticated as an airplane or a supercomputer. Simple machines and tools are human powered, they are immediate extensions of human force and direct enablers of human will. Next, there are machines powered by natural forces, be it animal (e.g. the *saqiya* or animal-powered scoop-wheel), water (e.g. the water-wheel), or wind (e.g. the windmill), which do only work if the corresponding power source is present. Human autonomy in using these machines is constrained by the availability of the power source. The same can be stated about potential energy storage machines, such as clockworks (energy is stored in compressed springs), combustion engines, and electrical motors. Most of these machines are *automatic*, which etymology reveals that it comes from Ancient Greek αὐτόματος (*autómatos*) meaning, “self-moving, moving of oneself, self-acting, spontaneous”[2]. The same root is behind the word *automaton* (pl. automata), which refers to mechanical figures that seem to move and operate by their own (hidden) power, whose rise comes in hand with the development of precision mechanics in the form of sophisticated clockworks in the XVIII and XIX centuries, with the famous Digesting Duck of Jacques de Vaucanson, the Writer of Friedrich von Knauss, the Musician, Writer and Draughtsman of the Jaquet-Droz family, or the Draughtsman-Writer of Henri Maillardet. All of these automata liberate with their motions the energy stored in wind-up springs, and they are often regarded as the predecessors of today’s robots, but there is a key difference, which is quite relevant for our discussion: while robots can be reprogrammed to perform different tasks, as we will see below (their actions are conditioned by their programs), automata have to do what they are constructed for, their actions are determined by their internal mechanisms, by their bodies. Moreover, robots can be programmed to respond to external stimuli (if they are equipped with the corresponding sensors), which is not the case for automata. Thus, turning to the etymology of automatic and automaton, all of the meanings related to internal drives (strive, yearn, desire, force) have to be interpreted in terms of necessity. Self-moving and self-acting, powered by an internal force (which at the end is just the potential energy stored in the springs by a human who has been turning the wind-up key), but not able to perform otherwise. And if the ultimate root of these words is related to conceiving a thought, this thought, in the case of automata, is nothing but the thought of their human inventors.

At some moment, machines incorporated control in the form of self-regulation. The most illustrative device devised to this end is the centrifugal regulator (or governor), invented by Christiaan Huygens and later (1788) adapted by James Watt to control his steam engine. The point is that the engine’s output (the rotational velocity of a shaft) is used via a closed

---

[1] Simulations have, of course, their own laws, which may be realistic or follow the whims of the designer.

[2] Wiktionary, “automatic”, https://en.wiktionary.org/wiki/automatic (last edited on 29 January 2025, at 01:48).

feedback loop to maintain this very velocity within an established range around the desired value of this velocity. This feedback loop is embodied in a fly-ball mechanism, whose motion is connected to the engine's output shaft, and which regulates the steam admission to the cylinders of the steam engine. Modern automatic machines integrate some form of control, generally built in the electronic circuit governing the machine. Thus, in the modern sense, automatic means not only self-powered and self-moving, but also self-controlled.

Robots belong to a relatively recent category of automatic machines whose behavior is far richer, in the sense of variability and functionality, than conventional machines, thanks to a new feature they are furnished with, as mentioned previously: they are programmable. A program can be interpreted as a set of instructions that have to be followed by the machine, generally in sequential order (including conditional jumps and loops as well). The same machine can perform quite different tasks just by changing the program it is following. Modifying a program, or substituting one program by another one, is relatively easy. The first programs were written in assembly or machine language, requiring a high degree of expertise in the human programmers, but they evolved towards programming languages closer to human natural expressivity. Most operative robots nowadays, in industry and domestic environments, fit within this category. However, programming enables robots to perform a qualitative step towards a new condition: autonomy. In contrast to automatism, autonomy involves decision making at an abstract, higher level: it means self-determination, "the state or condition of self-governance, or leading one's life according to reasons, values, or desires that are authentically one's own" (Taylor, 2017) [3]. Etymology reveals the meaning of the second part of the word: νόμος (*nomos*) means rule, law. In a strict sense , then, no robot is "autonomous", and so the distinction should be always kept in mind between 'autonomy' referred to a human capacity, and that what we loosely term as "autonomy" to refer to a technological condition. As for robotics, programming is the entrance door to furnish robots with AI, that is, with cognitive faculties such as learning, decision making, reasoning or planning. These faculties are paramount, if not essential, to some sort of rationality, of deliberative processes with the own rules that are the premise for autonomy. What is the picture at this moment? It is clear that today there are robots with different cognitive capacities (if they have any at all). Can we in consequence talk about different degrees or kinds of autonomy? Is robot autonomy in conflict with the autonomy of their human users[4] or collaborators? And which ethical consequences may robot autonomy entail as for humans and environment? These questions are addressed in the following sections.

9.2. **A characterization and a categorization of robots.**

Robots are machines running through a perceive—process—act cycle:

- they perceive their surroundings, often (e.g. in industrial operations) just in a very focused way, to answer binary questions such as "is there a certain workpiece in a predefined location or not?", but increasingly in a more elaborated fashion, such as the determination of the mood of the human counterpart, required in some social robotics applications;

[3] Taylor (2017). For the sake of clarity, it is important noticing that, drawing from significant philosophical work on the notion (Frankfurt, 1971), desires here are not to be contemplated as blind actions triggers, but submitted to a deliberative evaluation process.

[4] By human user we will refer both to the formed human operator as well as to the unskilled end-user

- they have to process the incoming information from perception and elaborate the appropriate answer, and again, this information processing ranges from reactively executing a simple program, to using AI in complex deliberative processes; and
- they act in the physical world, by changing their configuration (their pose in the case of robot arms, their position and orientation in the case of mobile robots), by closing or opening their gripper, by performing an operation with a tool or a measurement with an instrument, by communicating with a user, etc.

The way perceptual information is processed in order to give a response in form of an action and the control share between human and robot provides the following categorization:

Telerobots: These are remotely guided robots by a human operator, and it is reasonable to wonder if they can be considered as robots at all. The answer is affirmative if perception takes place through the robotic system itself, and/or if human action commands are elaborated by the robot's software. In contrast, telemanipulators or remote manipulators, arm-like mechanisms directly actuated by a human through mechanical, hydraulic, or electronic linkages, to manipulate radioactive, chemical dangerous, or biological hazardous material inside a so-called hot cell (a sealed chamber), are not robots. These devices, developed in the second half of the 1940s, are just safe extensions of the human's hands, who observes the manipulation directly through a glass. But the Da Vinci surgical robot really qualifies as such, because it not only magnifies the view that the surgeon has from the ongoing operation from in-body cameras, but also because the surgeon's movements are scaled to the minute precise motions required for the instruments, and because these movements can be guided by software to avoid critical areas. As for Unmanned Aerial Vehicles (UAVs), aka drones, there may be different levels of autonomy, whether the operator has to take care of all the flight operations, or the system is able to perform take-off and landing, for example, on its own. If the human pilot is in charge of a swarm of UAVs, it is obvious that these drones will have a high degree of autonomy to coordinate their flight and avoid collisions.

Reactive robots: In their simplest version, these are robots following "bug algorithms" for sensor-based planning, which are simple obstacle avoidance strategies for mobile robots headed towards a goal location. There is a robotic control architecture based on atomic behaviors, which are direct connections between simple perceptions and basic actions. This so-called subsumption architecture, quite popular in the 1980s and 90s, advocates the emergence of complex behaviors from subsuming (i.e. integrating) simpler ones. These robots are as autonomous as an insect can be. The human designer may be certain about the robot's immediate behavior in front of specific stimuli, but may easily get lost on the long run, or about the higher-level response to a set of simultaneous perceptions.

Programmed robots: In one sense or another, all robots are programmed, what is meant with this category is the set of robots that have explicitly been programmed to follow specific instructions in a controlled environment. Robot programming allows for sensor-(i.e., perception of the environment-) dependent choice between alternative courses of action: a defect piece, for example, triggers it to being discarded. This can be seen as a rule of behavior for the robot. But this rule has been provided by the human programmer. The task the robot has to execute is well-known by the programmer, and all the possible contingencies can be foreseen. Conventional industrial robotic manipulators fit into this

class. Human involvement takes place prior to operation; once the task starts the robot is entirely on its own, unless some sort of malfunctioning requires to halt the operation.

Cognitive robots: In the previous categories, robots either follow directly human commands, on the fly, or they follow a set of owned (as they are in their software) but not own (they are not developed by themselves) instructions and rules. To be truly autonomous, to follow their own, self-acquired rules, robots have to be endowed with cognitive faculties: learning, to develop their own set of rules based on their experience, and decision-making (which of these rules are applied and how) via reasoning and planning (to fulfill not just immediate goals but also long-term objectives). Humans supervise and guide the learning process, but in some settings there is no real insight in what the robot is actually learning. This is particularly notorious in the most popular learning paradigms, namely those based on deep neural networks. Many efforts are devoted to make such systems explainable. Still, even if research succeeds in this endeavor, will humans always be able to control the data sources that feed the robot learning process? The problem is similar to the one parents are confronted with when their children begin to be submitted to external influences, outside the family sphere. The more the robot becomes autonomous, the more it may search information we don't know about, and the more a stranger it may become.

9.3. **Kinds of autonomy in robotics.**

Different views on the concept of autonomy have arisen in the literature on human-robot systems. One of the early insights is that autonomy isn't just a feature of a given agent (be it a human, a robot, a software program…), but is shaped by all the intervening elements. In the context of multiagent systems, Brainov and Hexmoor identify up to five such elements: the subject of autonomy (in our case, a robot), the influencer (e.g. the human user), the scope of autonomy or means used by the influencer to affect the subject's autonomy, the object or elements respect to which the subject is considered autonomous (e.g. a single action or a complete task), and the degree of autonomy (its measure of quantization)[5]. In their work, autonomy is not only considered with respect to the human user, but also with respect to the interaction with the environment (e.g., the degree of robustness of the agent's autonomy before environmental changes and uncertainties) and with respect to other agents that may be present. In Huang et al.'s Autonomous Levels for Unmanned Systems (ALFUS) framework, their Contextual Autonomous Capability (CAC) model characterizes the system's (e.g. a robot's) autonomy along three axes, the environmental difficulty, the mission's (task's) complexity, and the human/operator's involvement[6]. Beer et al propose a definition, based on a survey of definitions of autonomy in the robotic literature[7], that takes the sense – plan (or process) – act cycle into account: "The extent to which a robot can sense its environment, plan based on that environment, and act upon that environment with the intent of reaching some task-specific goal (either given to or created by the robot) without external control."[8] These three activities are considered as subcomponents of a given task, and the degree of autonomy of the robot may vary from one subcomponent to the other. Moreover, the authors also highlight the fluid nature of autonomy in a given robot, as it may change over time, depending on task and environmental circumstances. In his definition of autonomy, Antsaklis insists in the

[5] Brainov and Hexmoor (2001).
[6] Huang et al (2007).
[7] some of which statements they gather in a table.
[8] Beer et al (2015).

circumstantial nature of the concept: a system is autonomous with respect to a given set of goals and a specific set of environmental uncertainties[9]. According to this researcher, autonomy can be increased by augmenting the system's aptness to deal with uncertainties, via feedback control, adaptation, learning, planning, failure detection and reconfiguration capabilities.

Compared to the polyhedral vision on the concept of autonomy in other disciplines, the treatment in human robot systems, and in particular in human-robot interaction (HRI), is rather one-dimensional and ambiguous, claim Kim et al. 2024. To remedy this, they present a taxonomy of different types[10] of autonomy obtained from a thorough literature review from 2008 to 2022, focusing on full papers of the four most relevant HRI venues with the words autonomy or autonomous in the title, abstract or keywords. From analyzing the resulting 500 papers, a taxonomy of 6 categories emerged (all the definitions are literally theirs):

- Human and robot involvement at runtime
  - **Operational autonomy**: the degree of human operator disinvolvement[11] at runtime. Most works in the literature refer to this type of autonomy.
  - **Intentional autonomy**: the degree of robot goal-oriented involvement at runtime.
  - **Shared autonomy**: the degree to which both human operator and robot are intentionally involved at runtime.
- Human involvement before runtime
  - **Non-deterministic autonomy**: the degree to which the robot's behavior is not specified prior to runtime.
- Expressions of autonomy at runtime
  - **Cognitive autonomy**: the degree to which the robot takes cognitive action.
  - **Physical autonomy**: the degree to which the robot takes physical action.

It should be clear that these are not mutually exclusive classes, but evaluation criteria that allow to attribute to each robot setting a precise definition of the autonomy it enjoys. The authors provide some illustrative examples, which we leverage together with other settings that belong to the robot categories mentioned above:
Telerobots: While having low operational autonomy, we may distinguish low intentional autonomy robots such as plain telerobots, as used for example in marine exploration or in pipeline inspection, being the eyes and hands of the human operator in harsh environments, from higher intentional autonomy robots, such as the planetary rovers, that due to the signal travelling time delay have to be able to autonomously respond to situations such as imminent collisions, the surgical robots (e.g. Da Vinci), or swarms of UAVs. The latter examples illustrate as well high degree of shared autonomy: while the surgeon is in charge of practicing the required incision, the surgical robot system cares about the scalpel not affecting sensible tissues such as nerves or blood vessels. And the drone swarm pilot may focus on covering a specific area in a search-and-rescue task,

[9] Antsaklis (2020).
[10] types, and not levels!
[11] They use the term "disinvolvement" when referring to human involvement in the task for consistency, to denote higher robot autonomy the more disinvolved the human is in the task. However, the domain is considered to be "general-sum", that is, with human and robot involvement to be independent factors, as opposed to the "zero-sum" approach, where one is detrimental to the other.

while the UAVs process the individual images gathered by their cameras and coordinate their flight to avoid collisions.
Reactive robots: Robots with simple stimulus-response architectures will display a high operational autonomy, while their intentional autonomy will depend strongly on the task they have to perform. Kim et al put the examples of a wandering (without a specific destination) robot and a vacuum cleaner robot as low and high intentional autonomy, respectively. Intentional autonomy may be higher for behavior-based subsumption architectures, because of the emergence of higher-level behaviors and a wider task (and contingencies) spectrum they can cope with. Generally, such robots are not designed for HRI, and shared autonomy will be rather low.
Programmed robots: Once programmed and as long as no maintenance is needed nor malfunctioning appears, such robots display high levels of operational autonomy. Intentional autonomy will be high (if we assume the programmer's goals to become the robot's ones), and will increase with the robot's robustness to deal with uncertainties. Shared autonomy will be rather low, although some instances of collaborative robots (belonging in general to the next category) may work alternatively with the human operator on the same task/workpiece, in which case shared autonomy is high.
Cognitive robots: The latter category obviously scores high in all the autonomy types, the highest on intentional and cognitive autonomy, and variable on operational and physical autonomy, depending on the specific embodiment. As for shared autonomy, it depends of course on the task at hand, interaction-intensive tasks may demand a high degree of shared autonomy, which such robots are able to deliver, whereas other tasks will favor the autonomy share to lie entirely on the robot's shoulders.

As for the category of non-deterministic autonomy, it is quite suitable as it stresses the distinction of the domains before and during runtime. Embodiment and programmability are distinctive features of robotics, and non-deterministic autonomy accurately addresses the robustness of the hardware designers' and the programmers' work in coping with environmental and task uncertainties. A robot where all the low-level instructions have to be explicitly specified will have much lower non-deterministic autonomy than a robot that just needs a high-level goal specification and is left on its own capabilities to determine the intermediate steps and negotiate contingencies.

9.3.1. Quantifying autonomy.

As stated in Beer et al. robot autonomy is not a binary quality but "a continuum ranging from no autonomy to full autonomy"[12]. In robotics, considering the degree of autonomy of a robot is a practical requirement, if an appropriate performance in the application at hand has to be met. To this end, different criteria can be met. The first and most obvious one is to account for the (relative) time of human intervention and neglect (the latter meaning "not taking care of", in this case referred to the robot). As straightforward as it seems (just measuring the percentage of the time the robot is working on its own, without human intervention), it is not unproblematic, in the first term for the possible confusion of intervention with interaction. A social robot, for instance, is precisely addressed to have large amounts of interaction with people, while at the same time being autonomous. In HRI, two ways of conceptualizing autonomy as for interaction coexist, according to Beer et al[13]: one way considers the robot more autonomous the less frequent interaction with a

[12] Beer et al (2015).
[13] Beer et al (2015).

human it requires, while the other way considers higher autonomy the higher level or the more sophisticated is the interaction. Interaction is a quite generic term that ranges from interchanging information to physical collaboration, whereas intervention is a specific form of interaction that involves the human performing some part of the task. Moreover, not taking the complexity of the task into account may lead towards considering a robot that executes a simple and repetitive task with just a little previous programming as highly autonomous (as long as no unforeseen cause of malfunctioning emerges). However, efficiency in performing a task is neither necessarily correlated with autonomy. It is obvious that the less error-prone the robot is in addressing a given task, the more autonomous it is perceived by humans. But, as Brainov and Hexmoor point out[14], autonomous behavior can be inefficient, and, conversely, efficient behavior may not necessarily be met by autonomous agents. Instead, they propose measures of *relative performance*: the degree of individual autonomy (i.e., autonomy with respect to the user), is computed as the ratio of performance of the agent acting autonomously and its performance under the user's supervision[15]. Similar measures are defined for the agent's autonomy with respect to environmental resources and interference of other agents[16]. Antsaklis defines two metrics for computing the autonomy level L, one with respect to the environment's uncertainties, as the product of the measure of the set of goals MG and the measure of the set of uncertainties MU, L= MG x MU, and another one with respect to external (human) intervention, MI, as the quotient L=MG/MI. He proposes performance as MG, and robustness as MU. For the latter he considers entropy as well[17].

Assigning a numerical value to performance (and, indirectly, to autonomy) may be not straightforward. To determine which kind of robot is more suitable for a given task and environment, as function of its autonomy, possibly it is enough to find out to which category it belongs. Some attempts to define such a categorization can be found in the literature. Here we point at two of such classifications, Huang et al.'s LoA (levels of autonomy) of UMS (unmanned systems), focused on, but not exclusive of, military robots, and Beer et al.'s levels of robot autonomy (LoRA[18]), which is inspired in a similar categorization of automation[19], but adapted to robot specifics. We summarize their discretization of the continuum of robot autonomy into categories in the following table:

Huang et al (2007)					Beer et al (2015)			
LoA	S	P	A	C	LoRA	S	P	A
Level 1: Remote control	H	H	H/r	H/r	(Manual, robot as a tool at most)	H	H	H
					Tele-operation	H/R	H	H/R
Level 2: Tele-operation	H/r	H	H/r	h/R	Assisted tele-operation	H/R	H	H/R
					Batch processing	H/R	H	R
Level 3: Human directed	H/R	H/r	H/R	h/R	Decision support	H/R	H/R	R
					Shared control with human initiative	H/R	H/R	R
Level 4: Human aided	h/R	H/R	h/R	h/R	Shared control with robot initiative	H/R	H/R	R
Level 5: Autonomous	h/R	h/R	R	h/R	Executive control	R	H/R	R

[14] Brainov and Hexmoor (2001).
[15] Not to be confused with the compound performance agent-user.
[16] By the way, they demonstrate that determining the multiagent group with the highest autonomy is a hard computational problem (NP-complete).
[17] Antsaklis (2020)
[18] The acronym is originally written as LORA in Beer et al (2015), but we prefer this notation, employed in Kim et al (2024).
[19] The taxonomy of Endsley and Kaber (1999), which is organized around the funcions of monitoring, generating and selecting (options or strategies), and implementing.

					Supervisory control	H/R	R	R
					Full autonomy	R	R	R

Table 1. Comparison of the taxonomies presented in Huang et al (2007) and Beer et al (2015)[20]. Both works analyze autonomy in terms of sensing (S, subdivided in Huang et al. (2007) into *sensing*, *perception* and *analyzing*)), processing (P, subdivided in in Huang et al. (2007) into *planning* and *decision making*, and called *plan* in Beer et al (2015)), and acting (A). Huang et al. (2007) include communicating (C) as well. H stands for human intervention, R for the robot share (the UMS in Huang et al (2007)), uppercase and lowercase are used to indicate higher or lower dominance (i.e. share of the task). Human dominance in perceptual tasks in Huang et al (2007) refers of course to the higher function of analyzing.

Beer et al propose a framework, based on this categorization, to determine a given robotic setting's autonomy, as well as the HRI variables influenced by autonomy. This framework goes successively along five guidelines, being the first about task and environment variables (which task is the robot to perform?), the second about the task aspects assigned to the robot (i.e., decomposing the task in the sense – plan – act parts and the possible use of the robot), third measure the robot autonomy (or, conversely, the amount of human intervention) in each one of these parts, fourth categorize by level (align with Table 1), and fifth, evaluate the influence of robot autonomy on robot-related, human-related, and social variables of HRI. In the next section, we dedicate some space to the human-related aspects of robot autonomy.

9.4. **Human perception of robot autonomy and user satisfaction.**

We have already considered the robot's capabilities as for their influence on the robot's autonomy. Embodiment is crucial, of course, and has to be adapted to the task and the environment, ensuring appropriate mobility in the medium (e.g., immersion and propelling capabilities in underwater robots), sensing, and/or manipulating hardware and skills. Learning, planning, decision making capabilities, on the other hand, enhance their adaptability and robustness in changing environments or task specifications. But how does robot autonomy influence how the robot is perceived by the human user? Beer et al. stress the relevance of robot autonomy in HRI even to the point of stating that "understanding robot autonomy is essential to understand HRI". This is due to its incidence on three factors:

- the tasks the robot is able to perform
- the level and frequency of interaction with the human user
- the reliability of the robot's performance.[21]

These are observable factors that lead towards a certain perception of the robot autonomy and which have measurable consequences as for the human's goals. However, this perception is conditioned as well by more subjective human features, such as situational awareness, mental (and physical) workload, trust and acceptance.

Sousa Silva et al[22] developed a Robot Autonomy Perception Scale (RAPS) around five autonomy components identified on Kim et al's aforementioned categorization. For each one of these components, they pose three representative questions, extracted from

[20] The categories are not exactly coincident, in some cases LoAs blurr into neighboring LoRAs as well. The original works are quite more nuanced, consider this table as a rough approximation. Take also the slightly different robotic targets of the two references into account.
[21] Beer et al (2015).
[22] Sousa et al (2024).

robotics literature on autonomy, and present them on a 7-point Likert scale. The components, their correspondence to Kim et al's autonomy types, and one of the representative questions for each one are as follows: Sense (cognitive autonomy, "The robot is able to detect variations in its surroundings"), Plan (cognitive autonomy, "The robot is able to establish a course of action"), Act (physical autonomy, "The robot is able to take actions to affect its environment", Goal (intentional autonomy, "The robot would modify its behavior if this were necessary to pursue its goals"), and Control (operational autonomy, "The robot can operate without the direct intervention of humans"). In a preliminary study, using a resource allocation game with a virtual tele-operated robot with different performative autonomy levels (low - asking for arbitration, medium - asking for confirmation, and high – just reporting) the authors confirm the suitability of RAPS to assess the users' perception of changes in the robot's autonomy.
The users' background is also essential to their perception and acceptance of robot autonomy. This may be illustrated by two works performed on expert users and the general public, on compelling domains such as space exploration and security, respectively. In the first of these works[23], interviews with five space industry professionals were conducted. It should be stressed that in the professional domain, besides environmental complexity, two task-related key factors condition the appreciation of robot autonomy in semiautonomous systems: task criticality and accountability (i.e. who is to blame when things do not go right). The focus of the interviews was put on how robot autonomy and human decision-making were valued in space-mission contexts (including planetary exploration and on-orbit robotics). Technical factors such as latency in communications or on-board computing limitations are crucial for the actual needs of autonomy (e.g. the higher the latency, the more the robotic system has to be able to respond on its own, especially in time-critical tasks). Human factors do impact choices on levels of autonomy as well: human involvement in the mission may be prioritized even if human decisions may be detrimental to the mission's outcome (e.g. the capacity to abort a mission, which may be riskier than delegating this decision on an automated system, especially under pressure). As for future projections, higher latency and increase in mission complexity, involving robotic swarms for example, will foster the desirability for more autonomy, but at the same time higher-level decisions will continue to be delegated on humans due to their better holistic understanding of a given situation.

In the second work[24], an online survey with 236 participants was conducted, with a questionnaire (with a 7-point Likert scale) referring to a video and explanations about different security robot settings. These included 3 autonomy levels (from tele-operated to fully autonomous) and 2 risk levels (just observation and physical involvement). The research question was about the influence of risk on the acceptance of robot autonomy. Experimental results confirmed the generic aspects of the Technology Acceptance model (TAM) for the case of security robots, that is, that perceived usefulness, perceived ease of use, and trust have a positive incidence on acceptance. As for autonomy and risk, they confirmed as well the two formulated hypotheses, namely that increasing the degree of autonomy decreases the acceptance of security robot, and that risk modulates the impact of autonomy on acceptance, mainly lowering acceptance with increasing autonomy in the low-risk situation (with no significant influence in the high-risk situation, where acceptance was low in any case).

[23] Smith et al (2024).
[24] Ye et al (2024)

These findings reveal that human concerns about robot autonomy are mediated through the trade-off of benefits and risks, that in one case are about professional performance and in the other about sense of security in given places. But many other concerns related to potential threats linked to robot autonomy may arise in other areas of deployment of such technologies. These are briefly tackled in the next section.

9.5. **Robot autonomy in application domains and the rise of ethical concerns.**

Decades ago, professionals in robotics considered two large application domains: industrial and service robots, the latter being everything that wasn't in a factory. Nowadays, there is a general consensus on a more refined division, which can be summarized in the following image (with the category of experimental robots being present in all the others as an early stage of development):

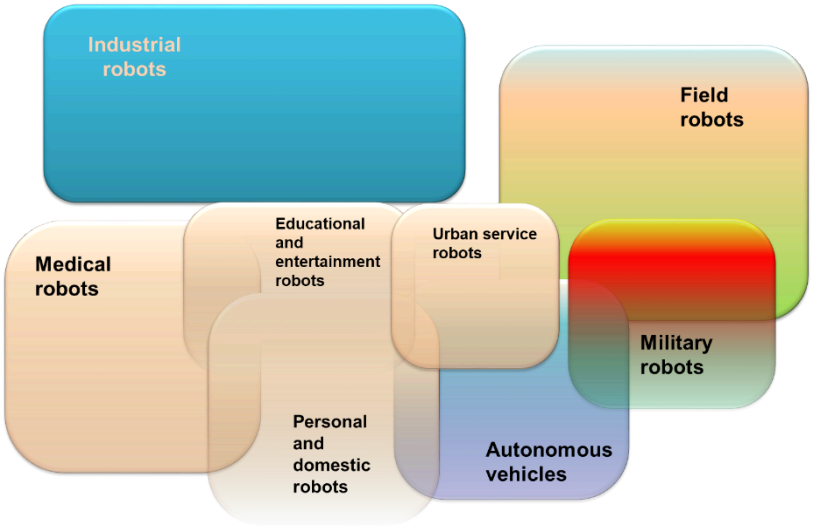


Fig. 1 Application domains of Robotics.

Many of these application domains appear to be overlapping: a domestic elderly care robot, for instance, fits in the medical (health), as well as in the personal and domestic (and occasionally also in the entertainment) categories.

Advances in technological autonomy[25] have substantially enlarged the range of tasks and roles that robots can perform, unprecedently extending the scope of robotics application domains. The new condition of "autonomy" grants these robotic "agents" the ability to assume not only automatable tasks, but also those that require to process information about the situation and to make choices on the basis of it, always in view of a predefined goal. This has placed robotics as a promising tool-provider for certain activities that were so far still mostly reserved to human agency, inasmuch their development demands interaction with and adaptability to the environment and to users, as it is the case of care-related activities (healthcare, assistance, education, etc.).

---

[25] Funk and Coeckelbergh (2020).

Along with the new applications it enables, there are potential risks related to technological autonomy that turn it into a matter of particular ethical concern, and which reinforce the need to take robotics as a field for ethics, that is, as a technoscientific field that raises issues of a normative kind –i.e., having to do with the question of how *should things be* (or what *should be* done), and which therefore differ from that other kind of questions relative to theoretical knowledge, of a descriptive nature and concerned with the knowledge of what it *is*–. In fact, it is greatly the rise of technological autonomy what triggers the contemporary articulation of ethics within robotics, as the very same birth of the so-called Roboethics[26] already shows.

Despite the historical engagement of philosophy with technology (Franssen, Lokhorst and van de Poel, 2023)[27] and thus its acknowledgeable role to the initiative, Roboethics is an interdisciplinary field of research that was officially established at the hands of engineers with the aim of addressing the challenge of developing robots in line with fundamental human rights and values, particularly regarding the intelligent/autonomous ones (Tzafestas, 2018). Specifically, taken in the original disciplinary sense of the term –i.e., as the applied ethics to robotics as historically formulated by engineers[28] (Veruggio & Abney, 2014)–, Roboethics encompasses three major lines of research:

(1) The first line of research is known as 'roboethics' in a strict sense. It focuses on the ethical issues that arise from human action in relation to the different stages of robotics development: design, research, implementation, and use of products. These issues relate to the implications and risks that this development may entail for the different levels of human life (individual, interpersonal, professional, sociostructural, etc.) in physical, psychological, sociopolitical and economic terms. It is thus a line of ethical research that addresses the grounding of decision-making in the several spheres of human activity that come into play in the development of robotics, and which thereby reflects on the actions of subjects belonging to very diverse fields of activity: engineers, professionals of the particular context in which robots are to be implemented, institutions that promote the adoption of such intelligent systems, individual or collective users of the same, etc.

(2) The second line of research refers to the so-called ‘robot ethics’ or ‘machine ethics’[29]. Here the focus is on the behavior of intelligent autonomous systems and so the ethical issues that are addressed have their root in the *agent* dimension of the robot –as an intelligent system that processes and selects courses of action, which then translates into a behavior within our world, thanks to its corporeality. It is thus a line of research aimed at ensuring that robot behavior is aligned with human values, by means of programming moral codes in robots or even endowing them with faculties for moral reasoning[30]–in a pursuit of the so-called “artificial moral agents (AMAs) (Allen et al., 2006).

(3) The third line of research relates to what has been named 'Robot’s Ethics'. It is a more speculative line of reflection, focused on ethical issues linked to a stage of

---

26 Operto and Veruggio (2008).
27 Franssen et al. (2023).
28 Veruggio and Abney (2014).
29 Veruggio et al. (2011).
30 Wallach and Allen (2009).

technological development that is still far from reality, such as the rights of robots[31] or the question of how to guarantee human interests and values in the face of the potential emergence of artificial intelligence in a strong or general sense[32]. Although this line also focuses on the agent dimension of intelligent robots, it does so from a futuristic perspective.

The more recent integration of ethics in the European Union (EU) AI and robotics deployment policy this last decade[33] is also an expression of the central place of technological autonomy in normative concerns on robotics. As part of the European commitment to responsible technology development[34], the turn to ethics of the EU tech policy[35] is highly responsive to the increasing technological autonomy of AI systems[36] –robots among them– and the risks it may pose in the different domains of application of these technologies.

Among them, we propose here to briefly review the impact of robot autonomy in the assistive domain, which is one of the application areas of robotics under greater transformation. We will focus on the case of social assistive robotics in particular, for it is notably representative of the novel ethical challenges of this technological condition of autonomy.

9.5.1. Technological autonomy in assistive robotics: The emergence of social assistive robots.

The kind of human-robot interaction (HRI) enabled by the advances in technological autonomy in robotics, very particularly in that of a cognitive type, have placed social assistive robotics (SAR) as a new promising branch of intelligent robotics, and as among the ones undergoing a greater development[37].

In general terms, SAR focuses on developing intelligent robots that socially interact with humans as a means to carry out tasks related to assistance, that is, having to do with helping people with special needs (elderly people with a certain degree of dependency, people with dementia, convalescent patients, people with mental health diseases, and people with functional diversity) in various activities of their daily lives. This unprecedented capacity of robots to assume tasks requiring interpersonal interaction with people is what turns them into a differential support when it comes to practices of a relational kind, such as care, and what substantively expands their potential as tools for the provision of assistance.

In the European context, social assistive robots (SARs) are designed to be used in professional assistive activities mostly related to healthcare (physical and cognitive therapy, as well as rehabilitation, and support for dependent individuals), and in different care ecosystems (daycare centers, long-term care homes, hospitals and domestic settings), where SARs can perform a variety of functions and roles such as coaching in

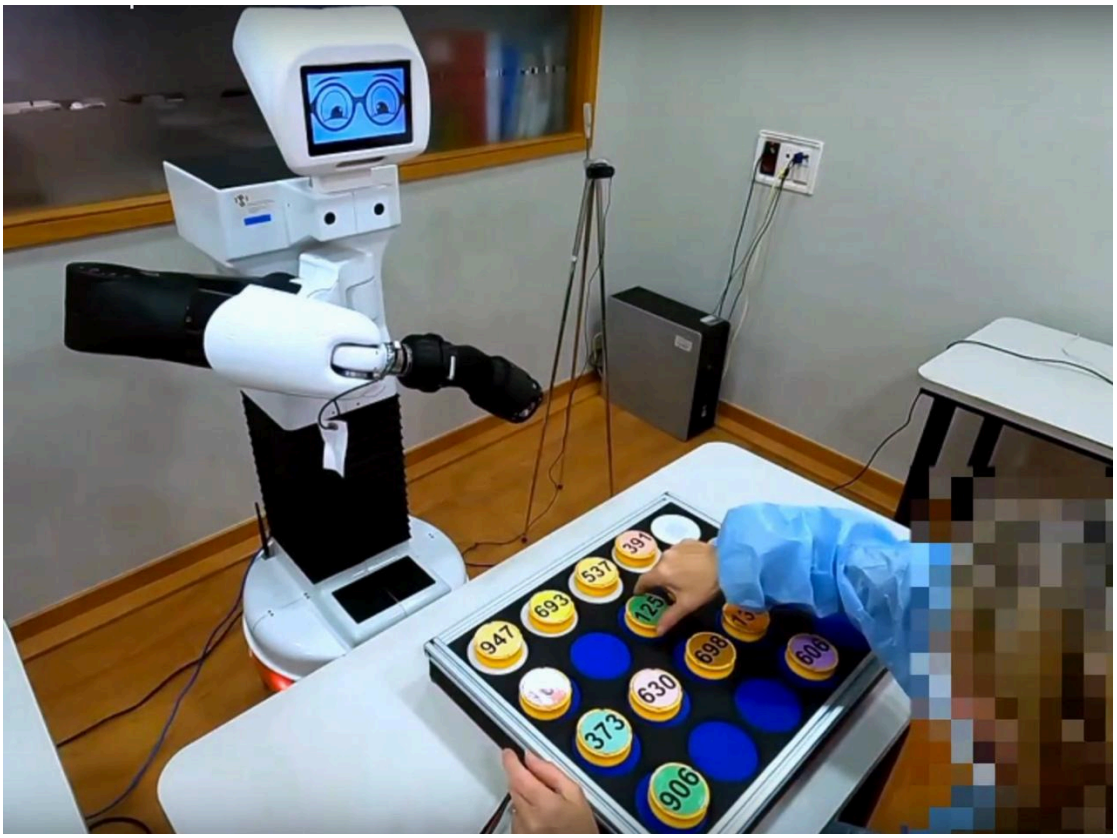
---

31 Gunkel (2018).
32 Bostrom and Yudkowsky (2011).
33 de Pagter (2023).
34 European Commission (2018).
35 Carlsson and Rönnblom (2022).
36 de Pagter (2023).
37 Mejia & Kajikawa (2017).

health restoration or maintenance exercises, assisting in tasks such as eating or dressing, or serving as a companion for mental health treatments[38]. Note that the focus of application in healthcare is in line with the interest in deploying AI, robotics and digitalisation primarily in that sector[39] as a strategy to tackle the "care crisis"[40] of public health systems in the face of an increasing population with special needs[41] –which is in turn greatly due to the ageing of global population[42] –, and that it aligns with the current healthcare logic of dehospitalization that converges with the European commitment to active aging and independent living[43].

To list just some illustrative examples of SARs, which moreover demonstrate the institutional expectations of introducing this technology into healthcare practices, especially in relation to the care of the elderly, it is useful to recall the following ones:

(1) The socially interactive AI robotic system developed by the Institut de Robòtica I Informàtica Industrial (CSIC-UPC) in collaboration with the ACE Foundation – Alzheimer Center Barcelona[44]. Conceived within the framework of the European project SOCRATES–Social Cognitive Robotics in the European Society, it is a social assistive robot that interacts with dementia patients to help them perform a cognitive training exercise that is part of the ACE Foundation's therapeutic activities.



38 Rabbitt et al. (2015).
39 European Commission (2020).
40 Liedo (2022).
41 Dolic et al. (2019).
42 United Nations Department of Economic and Social Affairs (2022).
43 Maibaum et al. (2021).
44 Andriella et al. (2020).

Fig. 2 - Prototype of a robotic system for providing cognitive training, co-designed by the Institut de Robòtica i Informàtica Industrial (CSIC-UPC) and the ACE Alzheimer Center.

(2) The social assistive robot ARI (previously Misty II) that helps dependent elderly people with their needs in domestic daily life, and which is part of a Barcelona City Council's pilot project to improve the quality of life of elderly people with some form of dependency who live alone[45] [46] –an initiative that is already being deployed throughout the Catalan territory[47].

(3) The robot 'Nyam,' developed by the Institut de Robòtica i Informàtica Industrial (CSIC-UPC) in collaboration with the Parc Sanitari Pere Virgili, as part of the AI-EAT project[48], designed to assist in the feeding process of people who cannot do it by themselves in hospital settings.

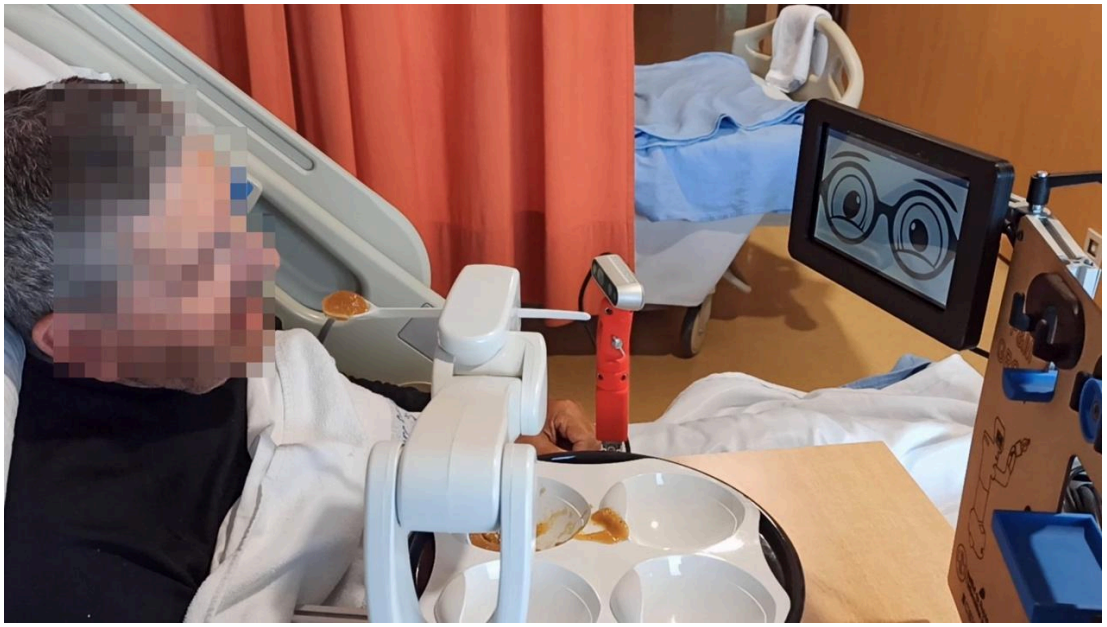


Fig. 3 Prototype of a robotic feeding system, co-designed by the Institut de Robòtica i Informàtica Industrial (CSIC-UPC) and the Pere Virgili Health Park in Barcelona.

9.5.2 The risks of technological autonomy in social assistive robots

Technological autonomy in social assistive robots, and the novel kind of technological mediation of care practices it enables –especially in virtue of the new (human-robot) alterity relation it affords–, is a specific source of ethical concerns. From the risk-logics underlying the European normative approach to technology[49], these concerns mostly

[45] Ajuntament de Barcelona (2020).
[46] Fundació iSocial (2020).
[47] Corporació Catalana de Mitjans Audiovisuals (2023).
[48] Barrué et al. (2024).
[49] European Commission (2021).

relate to the negative consequences that robot autonomy could entail from the perspective of fundamental human rights and values, both at the individual and societal level[50].

Think about the current ethical scholar discussion: concerns on privacy, deception and human autonomy, which occupy a central place within reflection[51], and other key correlated ones – such as dignity (along with objectification and emotional bonding), loss of human contact and responsibility[52]–, have to do with the potential threats of autonomous features of these AI-powered systems. Sticking to the categorization of Kim et al.[53], it is worth noticing that these features may link to different kinds of autonomy-related capacities –be them physical, cognitive, operational–, and their correlative degree of fulfillment in each case. Still, in the framework of ethical attention on SAR, concerns commonly rely on quite an imprecise or raw notion of technological autonomy, from which no fine-grained correlation of risks can be easily drawn.

At the normative level, and despite the risk-based approach of the EU regulations on AI systems[54], to which intelligent robots belong, SAR risks are not sufficiently estimated upon the specific particularities of robotic autonomy either. While in such regulations SARs are not addressed in a differentiated way, the social interactivity of AI systems –and thus one of the distinctive traits of SARs–, is indeed a (growing) focus of particular concern.

The very same reference EU document on robotics, namely the 2017 Civil Law Rules on Robotics[55], already includes what could be considered as a first allusion to the risks linked to the social interaction capacity of robots. These are posed in terms of a potential dehumanization of care practices, was human social interaction to be replaced by interaction with these devices. In the subsequent 2019 Ethics Guidelines for Trustworthy AI[56], the negative effects that interacting with social AI systems can have on people's physical and mental well-being, as well as on human skills, relationships and social bonds, are introduced more specifically, therefore urging to its evaluation and monitoring. More recently, the first legal regulation on AI[57] –the so-called *AI Act*– includes the social interaction capacity of conversational agents in the AI risk-cataloguing of the European Union[58]. Specifically, unless interaction is expressly intended for manipulation (in which case the system is classified as posing an unacceptable risk), the AI social interaction capacity is catalogued as of a limited risk. This is a risk level against which Europe prescribes transparency, in the sense of ensuring that end-users have knowledge of being interacting with an AI system (which responds to the need of avoiding deception and what are taken as other correlative threats to human dignity and autonomy).

Insofar as the interactive capacity of AI systems relates to high degrees of what has been categorized as cognitive and operational autonomy[59], the current EU normative focus on this feature represents at least a first step in an estimation of risks upon the particularities of technological autonomy. However, a risk analysis of SAR from a more nuanced

---

50 Torras (2024).
51 Pareto Boada et al. (2021).
52 Vallès-Peris and Domènech (2020).
53 Kim et al. (2024).
54 European Parliament and Council of the European Union (2024).
55 European Parliament (2017).
56 High-Level Expert Group on AI (2019).
57 European Parliament and Council of the European Union (2024).
58 European Parliament (2024).
59 Kim et al. (2024).

perspective on robot autonomy could help normatively respond to the ethical specificity of this branch of intelligent robotics. Therefore, we propose a first tentative application of the European AI risk-categorization to the specific case of SAR, considering the different types of autonomy that may come into play and with some use-cases serving as an example of SARs' level of risk:

Risk levels	**Description**	**SARs autonomy**	**SAR use-case**
Unacceptable risk	Clear threats to human safety or fundamental rights	Full operational, intentional, cognitive and physical autonomy	Adaptive robot companions for children with unbounded internet connexion and without human supervision
High risk	Potential serious threats to health, safety or fundamental rights	Shared autonomy and either physical or cognitive autonomy	Learning robots in therapeutic contexts without constant expert supervision
Limited risk	Potential threats to users' informed decision-making	High cognitive and operational autonomy	Robots nudging people to do certain actions without they being aware and having given previous consent
Minimal risk	Potential threat of malfunctioning	High operational and physical autonomy	Robots carrying out household tasks as handy appliances

Table 2. SARs' levels of risk related to autonomy.

9.6. **Conclusions.**

Robots display different degrees and types of autonomy, being cognitive robots the ones which potentially (if physically enabled) exhibit the highest level of intentional, operational, and non-deterministic autonomy. Quantifying robot autonomy is not straightforward, nor necessarily linked to performance, as it is heavily setting-dependent (both for the task to carry out and the environment). Nonetheless, useful categorizations can be found in literature taking into account the relative involvements of the human and robotic parts of the team in the different aspects of the development of the task, namely perception, decision-making and action-execution (and communication). As for the human perception of robot autonomy, which is related to their acceptance and trust on the robot partner, there is clearly a trade-off of what the human (and the task) may benefit from the robot's involvement, and the risks, linked to the sense of security (whether they are real or just perceived as such).

Although the lack of transparency may easily contribute to the human perception of the robot as pursuing its own goals, it should be kept in mind that all robots are designed by humans and thus, at the end, determined by human will, in one way or another[60]. That is, by not being living creatures, robots have no biological drives to self-preservation and

[60] Free will (or rather the lack of it) is one of the key differences between artificial and natural (human) autonomy, as pointed in Beghili et al. (2025), although the authors admit that this question is debatable in the human case as well.

reproduction as ultimate goals. However, as the setting becomes more complex and the cognitive skills of the robot escalate in sophistication, the possibilities of misinterpretation or robotic operative misalignments with human will, goals and values, may increase as well.

Within the risk-based logic of the European tech policy, this places technological autonomy as a focus matter of normative concern to be approached in its particularity. Despite being a central trigger for the so-far undertaken ethical and legal initiatives on AI systems, technological autonomy is still a quite underdeveloped attribute, from which no fine-grained correlation of risks can be easily estimated. In a context in which new fields of intelligent robotics are emerging (social assistive robotics among the main ones) and are to be normatively regulated in their different domains of application, this renders the EU risk-based approach not very effective as a navigating tool for responsible robotics deployment. Therefore, in this chapter we have provided an overview of the different kinds of autonomy-related features that may come into play in robotics, which we have then taken into account to advance a first tentative risk-level analysis of SAR from the perspective of robot autonomy.

Acknowledgements

This work was partially funded by the European project SoftEnable (HORIZON-CL4-2021-DIGITAL-EMERGING-01-101070600), and the CSIC project ArmTracEthic (Ref.202550E127).

References

Ajuntament de Barcelona. (2020). *Misty II the social robot becomes part of the lives of twenty senior citizens*. https://www.barcelona.cat/infobarcelona/en/tema/senior-citizens/misty-ii-the-social-robot-becomes-part-of-the-lives-of-twenty-senior-citizens_907645.html

Allen C, Wallach W and Smit I. (2006). Why machine ethics? *IEEE Intelligent Systems*, *21*(4), 12–17. https://doi.org/10.1109/MIS.2006.83

Andriella A, Torras C and Alenyà G (2020). Cognitive System Framework for Brain-Training Exercise Based on Human-Robot Interaction. *Cognitive Computation*. https://doi.org/10.1007/s12559-019-09696-2

Antsaklis P (2020). Autonomy and metrics of autonomy. Annual Reviews in Control 49, 15-26. https://doi.org/10.1016/j.arcontrol.2020.05.001

Barrué C, Suárez A, Inzitari M, Ribera A and Alenyà G (2024). NYAM: The Role of Configurable Engagement Strategies in Robotic-Assisted Feeding. *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*, 228–232.

Beer J, Fisk AD and Rogers WA (2014). Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction. Journal of Human-Robot Interaction, 3(2), 2014, 74-99, https://doi.org/10.5898/JHRI.3.2.Beer

Beghili M, Chetouani M and Barberousse A (2025). Ethical Perspectives on Natural Autonomy
and Artificial Autonomy. Social Robots with AI: Prospects, Risks, and Responsible Methods. Proceedings of Robophilosophy 2024. 77-84. https://doi.org/10.3233/FAIA241491

Bostrom, N., & Yudkowsky, E. (2011). The Ethics of Artificial Intelligence BT - Cambridge Handbook of Artificial Intelligence. *Cambridge Handbook of Artificial Intelligence*, 1–20. http://www.nickbostrom.com/ethics/artificial-intelligence.pdf%5Cnpapers2://publication/uuid/95B5D360-790E-4AB3-B572-7E337D11D892
Brainov S and Hexmoor H (2001). Quantifying Relative Autonomy in Multiagent Interaction. IJCAI-01 workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents.

Carlsson V, and Rönnblom M. (2022). From politics to ethics: Transformations in EU policies on digital technology. *Technology in Society*, *71*(October). https://doi.org/10.1016/j.techsoc.2022.102145

Corporació Catalana de Mitjans Audiovisuals. (2023, November 24). *ARI, el robot que facilita tasques a la gent gran i arribarà a un miler de llars catalanes*. https://www.ccma.cat/324/ari-el-robot-que-facilita-tasques-a-la-gent-gran-i-arribara-a-un-miler-de-llars-catalanes/noticia/3262661/

de Pagter J. (2023). From EU Robotics and AI Governance to HRI Research: Implementing the Ethics Narrative. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-023-00982-6

Dolic Z, Castro R, and Moarcas R. (2019). Robots in healthcare: a solution or a problem?, Study for the Committee on Environment, Public Health, and Food Safety. In *European Parliament* (Issue April).

Endsley MR and Kaber DB (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42(3), 462-492. https://10.1080/001401399185595

European Commission. (2018). Artificial Intelligence for Europe. *COM(2018) 237 Final*, 20.

European Commission. (2020). *White Paper on Artificial Intelligence - A European approach to excellence and trust*. https://doi.org/10.1017/CBO9781107415324.004

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council. Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*.

European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. http://data.europa.eu/eli/reg/2024/1689/oj

European Parliament. (2017). Civil Law Rules on Robotics. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2013(INL)). In *Official Journal of the European Union*. http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html#title10

European Parliament. (2024). Briefing “EU Legislation in Progress”. Artificial intelligence act. In *Official Journal on* (Vol. 12).

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68*, 5-20.

Franssen, M., Lokhorst, G.-J. and van de Poel, I. (2023) ‘Philosophy of Technology’, *The Stanford Encyclopedia of Philosophy (Spring 2023 Edition)*. Edward N. Zalta & Uri Nodelman (eds.). Available at: https://plato.stanford.edu/archives/spr2023/entries/technology/.

Funk M, and Coeckelbergh M. (2020). (Technical) Autonomy as Concept in Robot Ethics. In *Biosystems and Biorobotics* (Vol. 25). Springer. https://doi.org/10.1007/978-3-030-24074-5_12

Gunkel DJ (2018). The other question: can and should robots have rights? *Ethics and Information Technology*, *20*, 87–99. https://doi.org/10.1007/s10676-017-9442-4

High-Level Expert Group on AI. (2019). *Ethics Guidelines for Trustworthy AI* (pp. 1–41). European Commission.

Huang H-M, Pavek K, Ragon M, Jones J, Messina E and Albus J (2007). Characterizing Unmanned System Autonomy: Contextual Autonomous Capability and Level of Autonomy Analyses. Proceedings of SPIE Defense and Security Symposium, 2007, Orlando (FL). https://doi.org/10.1117/12.719894

Kim S, Anthis JR and Sebo S (2024). A Taxonomy of Robot Autonomy for Human-Robot Interaction. HRI '24: Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. Pages 381 – 393 https://doi.org/10.1145/3610977.3634993

Liedo, B. (2022). Cuidar en común. *Isegoría*, *66*. https://doi.org/10.3989/isegoria.2022.66.15

Maibaum, A., Bischof, A., Hergesell, J., & Lipp, B. (2021). A critique of robotics in health care. *AI and Society*, *0123456789*. https://doi.org/10.1007/s00146-021-01206-z

Mejia C, and Kajikawa Y. (2017). Bibliometric Analysis of Social Robotics Research: Identifying Research Trends and Knowledgebase. *Applied Sciences (Switzerland)*, *7*(12). https://doi.org/10.3390/app7121316

Operto F, and Veruggio G. (2008). Roboethics: Social and Ethical Implications of Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1499–1524). Springer. https://doi.org/10.1007/978-3-540-30301-5

Pareto Boada J, Román Maestre B and Torras C. (2021). The ethical issues of social assistive robotics: A critical literature review. *Technology in Society*, *67*. https://doi.org/10.1016/j.techsoc.2021.101726

Pareto J, and Coeckelbergh M. (2024). Social Assistive Robotics: An Ethical and Political Inquiry Through the Lens of Freedom. *International Journal of Social Robotics*, *16*(8), 1797–1809. https://doi.org/10.1007/s12369-024-01161-x

Rabbitt SM, Kazdin AE, and Scassellati B. (2015). Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical Psychology Review*, *35*, 35–46. https://doi.org/10.1016/j.cpr.2014.07.001

Smith C, Mott T, and Williams T (2024). Robot, Take the Joystick: Understanding Space Robotics Experts' Views on Autonomy. 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN), Pasadena, CA, USA, 2024, pp. 2182-2188, https://doi.org/10.1109/RO-MAN60169.2024.10731167

Sousa Silva R, Smith C, Bezerra L, and Williams T (2024). Toward RAPS: the Robot Autonomy Perception Scale. arXiv preprint arXiv:2407.11236

Torras C. (2024). Ethics of Social Robotics: Individual and Societal Concerns and Opportunities. *Annual Review of Control, Robotics, and Autonomous Systems*, *7*(1), 1–18. https://doi.org/10.1146/annurev-control-062023-082238

Taylor J S (2017). "autonomy". *Encyclopedia Britannica*, 20 Jun. 2017, https://www.britannica.com/topic/autonomy. Accessed 31 March 2025.

Tzafestas SG. (2018). Roboethics: Fundamental concepts and future prospects. *Information (Switzerland)*, *9*(6). https://doi.org/10.3390/INFO9060148

United Nations Department of Economic and Social Affairs. (2022). *World Population Prospects 2022. Summary of Results*. www.un.org/development/ desa/pd/.

Vallès-Peris N, and Domènech M. (2020). Robots para los Cuidados. La Ética de la Acción Mesurada frente a la Incertidumbre. *Cuadernos de Bioética : Revista Oficial de La Asociación Espanola de Bioética y Etica Médica*, *31*(101), 87–100. https://doi.org/10.30444/CB.54

Veruggio G and Abney K. (2014). Roboethics: The Applied Ethics for a New Science. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot Ethics. The Ethical and Social Implications of Robotics*. MIT Press.

Veruggio G, Solis J, and Van Der Loos M. (2011). Roboethics: Ethics applied to robotics. *IEEE Robotics and Automation Magazine*, *18*(1), 21–22. https://doi.org/10.1109/MRA.2010.940149

Wallach W, and Allen C. (2009). *Moral Machines. Teaching Robots Right from Wrong*. Oxford University Press.

Ye X, Jo W, Ali A, Bhatti SC, Esterwood C, Kassie HA, and Robert LP (2024). Autonomy Acceptance Model (AAM): The Role of Autonomy and Risk in Security Robot Acceptance. Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24). Association for Computing Machinery, New York, NY, USA, 840–849. https://doi.org/10.1145/3610977.3635005