Doncieux *et al. AI Perspectives* (2022) 4:1
https://doi.org/10.1186/s42467-021-00014-x

AI Perspectives

POSITION PAPER Open Access

Human-centered AI and robotics

Stephane Doncieux[1*], Raja Chatila[1], Sirko Straube[2] and Frank Kirchner[2,3]

Abstract

Robotics has a special place in AI as robots are connected to the real world and robots increasingly appear in humans everyday environment, from home to industry. Apart from cases were robots are expected to completely replace them, humans will largely benefit from real interactions with such robots. This is not only true for complex interaction scenarios like robots serving as guides, companions or members in a team, but also for more predefined functions like autonomous transport of people or goods. More and more, robots need suitable interfaces to interact with humans in a way that humans feel comfortable and that takes into account the need for a certain transparency about actions taken. The paper describes the requirements and state-of-the-art for a human-centered robotics research and development, including verbal and non-verbal interaction, understanding and learning from each other, as well as ethical questions that have to be dealt with if robots will be included in our everyday environment, influencing human life and societies.

Keywords: AI, Human-centered, Robotics, Human robot interaction

Introduction

Already 30 years ago, people have learned in school that automation of facilities is replacing human workers, but over time people recognized in parallel that working profiles are changing and that also new type of work is created through this development, so that the effect was rather a change in industry and not a mere replacement of work. Now, we see that AI systems are getting increasingly powerful in many domains that were initially solvable only using human intelligence and cognition, thus starting this debate anew. Examples for AI beating human experts in Chess [1] or Go [2], for instance, cause significant enthusiasm and concerns at the same time about where societies are going when widely using robotics and AI. However, we see at the same time with a closer look, that although the performance of AI in such selected domains may outrun that of humans, the mechanisms and algorithms applied do not necessarily resemble human intelligence and methodology, and may even not involve any kind of cognition. In addition, AI algorithms are application specific and their transfer to other domains is not straightforward [3].

*Correspondence: stephane.doncieux@sorbonne-universite.fr
[1]Institute of Intelligent Systems and Robotics (ISIR), Sorbonne Université, CNRS, Paris, France
Full list of author information is available at the end of the article

Robots using AI means an advancement from pure automation systems to intelligent agents in the environment that can not only work in isolated factory areas, but also in an unstructured or natural environment as well as in direct interaction with humans. Then, the application areas of robots are highly diverse, such that robots might influence our everyday life in the future in many ways. Already without direct contact to a human being required, robots are sought to support human ambitions, e.g. for surface exploration or installment, inspection or maintenance of infrastructure in our oceans [4, 5] or in space [6–8]. Everywhere, the field of robotics is an integrator for AI technology, since complex robots need to be capable in many ways, because they have the ability to act and thus have a physical impact on their environment. Robots therefore create opportunities for collaboration and empowerment that are more diverse than what a computer-only AI system can offer. A robot can speak or show pictures through an embedded screen, but it can also make gestures or physically interact with humans [9], opening many possible interactions for a wide variety of applications. Interactions that can benefit to children with autism [10, 11] or elderly [12] have been shown with robots that are called social[13, 14] as they put a strong emphasis on robot social skills. Mechanical skills are also important for empowering humans, for instance through

a collaborative work in teams involving both robots and humans [15, 16]. Such robots are called cobots: collaborative robots that share the physical space of a human operator and can help to achieve a task by handling tools or parts to assemble. Thus cobots can help the operator to achieve a task with a greater precision while limiting the trauma associated to repetitive motions, excessive loads or awkward postures [17]. Similar robots can be used in other contexts, for instance in rehabilitation [18, 19].

If humans and robots work together in such a close way, then it is required that humans have a certain trust in the technology and also an impression of understanding what the robot is doing and why. Providing robots with the ability to communicate and naturally interact with humans, would minimize the required adaptation from the human side. Making this a requirement such that humans can actually work and interact with robots in the same environment, complements the view of Human-Centered AI as a technology designed for collaboration and empowerment of humans [20].

After examining the specificity of robotics from an AI point of view in the next section, we discuss the requirements of human-centered robotics and, in the light of the current research on these topics, we examine the following questions: How can a robot interact with humans? How can it understand and learn from a human? How can the human understand the robot? And finally what ethical issues does it raise?

AI and robotics

A robot is a physical agent that is connected to the real world through its sensors and effectors [21]. It perceives the environment and uses this information to decide what action to apply at a particular moment (Fig. 1). These interactions of an autonomous robot with its environment are not mediated by humans: sensor data flows shape perceptions which are directed to the decision or planning system after some processing, but without any human intervention. Likewise, when an autonomous robot selects an action to apply, it sends the corresponding orders directly to its motors without going through any human mediated process. Its actions have an impact on the environment and influence future perceptions. This direct relation of the robot with the real world thus raises many challenges for AI and takes robotics away from the fields in which AI has known its major recent successes.

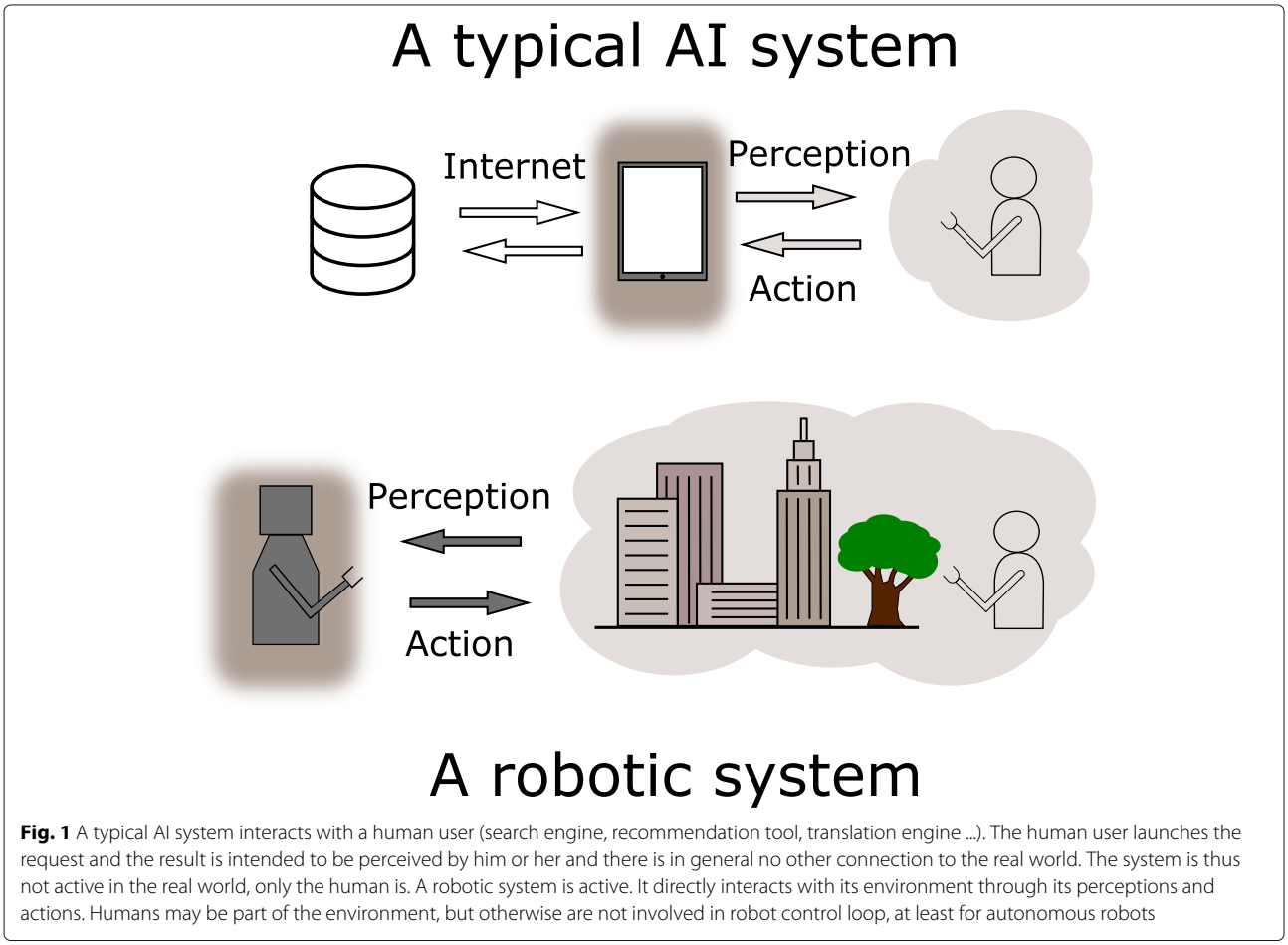

Fig. 1 A typical AI system interacts with a human user (search engine, recommendation tool, translation engine ...). The human user launches the request and the result is intended to be perceived by him or her and there is in general no other connection to the real world. The system is thus not active in the real world, only the human is. A robotic system is active. It directly interacts with its environment through its perceptions and actions. Humans may be part of the environment, but otherwise are not involved in robot control loop, at least for autonomous robots

When it was first coined in 1956 at the Dartmouth College workshop, AI was defined as the problem of *"making a machine behave in ways that would be called intelligent if a human were so behaving"* [22]. This definition has evolved over time, with a traditional definition now that states that *"AI refers to machines or agents that are capable of observing their environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions"*[23]. This view of AI includes many of the impressive applications that have appeared since Watson's victory at the Jeopardy! quiz show in 2011, from recommendation tools or image recognition to machine translation software. These major successes of AI actually rely on learning algorithms and in particular on deep learning algorithms. Their results heavily depend on the data they are fed with. The fact that the design of the dataset is critical for the returned results has been clearly demonstrated by Tay, the learning chatbot launched in 2016 by Microsoft that twitted racist, sexist and anti-Semitic messages after less than 24 h of interactions with users [24]. Likewise, despite impressive results in natural language processing, as demonstrated by Watson success at the Jeopardy! show, this system has had troubles to be useful for applications in oncology, where medical records are frequently ambiguous and contain subtle indications that are clear for a doctor, but not straightforward to extract for Watson's algorithm [25]. The "intelligence" of these algorithms thus again depends heavily on the datasets used for learning, that should be complete, unambiguous and fair. They are external to the system and need to be carefully prepared.

Typically, AI systems receive data in forms of images or texts generated or selected by humans and send their result directly to the human user. Contrary to robots, such AI systems are not directly connected to the real world and critically depend on humans at different levels. Building autonomous robots is thus part of a more restrictive definition of AI based on the *whole intelligent agent* design problem: *"an intelligent agent is a system that acts intelligently: What it does is appropriate for its circumstances and its goal, it is flexible to changing environments and changing goals, it learns from experience, and it makes appropriate choices given perceptual limitations and finite computation"* [26].

The need to face the whole agent problem makes robotics challenging for AI, but robotics also raises other challenges. A robot is in a closed-loop interaction with its environment: any error at some point may be amplified over time or create oscillations, calling for methods that ensure stability, at least asymptotically. A robot moves in a continuous environment, most of the time with either less degrees-of-freedom than required – underactuated system, like cars – or more degrees-of-freedom than required – redundant systems, like humanoid robots. Both conditions imply the development of special strategies to make the system act in an appropriate way. Likewise, the robot relies on its own sensors to make a decision, potentially leading to partial observability. Sensors and actuators may also be a source of errors because of noise or failures. These issues can be abstracted away for AI to focus on high level decision, but doing so limits the capabilities that are reachable for the robot, as building the low-level control part of the robot requires to make decisions in advance about what the robot can do and how it can achieve it: does it need position control, velocity control, force control or impedance control (controlling both force and position)? Does it need a slow but accurate control or a fast and rough one? For a multi-purpose robot like a humanoid robot, deciding it a priori limits what the robot can achieve and considering control and planning or decision in a unified framework opens the possibility to better coordinate the tasks the robot has to achieve [27, 28].

In the meantime, robotics also creates unique opportunities for AI. A robot has a body and this embodiment produces alternative possibilities to solve the problems it is facing. Morphological computation is the ability of materials to take over some of the processes normally attributed to control and computation [29]. It may drastically simplify complex tasks. Grasping with rigid grippers requires, for instance, to determine where to put the fingers and what effort to exert on the object. The same task with granular jamming grippers or any other gripper made with soft and compliant materials is much simpler as there is basically just to activate grasping without any particular computation [30]. Embodiment may also help to deal with one of the most important problems in AI: symbol grounding [31]. Approaches like Watson rely on a huge text dataset in which the relevant relations between symbols are expected to be explicitly described. An alternative is to let the robot experience such relations through interactions with the environment and the observation of their consequences. Pushing an object and observing what has moved clearly shows object boundaries without the need to have a large database of similar objects, this is called interactive perception [32]. Many concepts are easier to understand when interaction can be taken into account: a chair can be characterised by the sitting ability, so if the system can experience what sitting means, it can guess whether an object is a chair or not without the need to have a dataset of labelled images containing similar chairs. This is the notion of affordance that associates perception, action and effect [33]: a chair is sittable, a button pushable, an object graspable, etc.

Robots are a challenge for AI, but also an opportunity to build an artificial intelligence that is embodied in the real world and thus close to the conditions that allowed the emergence of human intelligence. Robots have another specificity: humans are explicitly out of the interaction

loop between the robot and its environment. The gap between robots and humans is thus larger than for other AI systems. Current robots on the market are designed for simple tasks with limited or even no interactions (e.g. vacuum cleaning). This situation can be overcome only if the goal of a human-centered robotic assistant is properly addressed, because the robot has to reach a certain level of universality to be perceived as an interaction partner. One component alone, like, e.g., speech recognition, is not enough to satisfy the needs for proper interaction.

Requirements of human-centered AI and robotics

All humans are different. If they share some common behaviours, each human has their specificities that may further change along time. A human-centered robot should deal with this to properly collaborate with humans and empower them. It should then be robust and adaptive to unknown and changing conditions. Each robot is engaged in an interaction with its environment that can be perturbed in different ways. A walking robot may slip on the ground, a flying one may experience wind gusts. Adaptation is thus a core objective of robotics since its advent and in all fields of robotics, from control to mechanics or planning. All fields of robotics aim thus at reaching the goal of a robot that can ultimately deal with the changes it is confronted with, but these changes are, in general, known to the robot designer that has anticipated the strategies to deal with them. With these strategies one tries to build methods that can, to some extent, deal with perturbations and changes.

Crafting the robot environment and simplifying its task is a straight-forward way to control the variability the robot can be subject to. The application of this principle to industry has lead to the large deployment of robots integrated in production lines built explicitly to make their work as simple as possible. New applications of robotics have known a rapid development since the years 2000: autonomous vacuum cleaners. These robots are not locked up into cages as they move around in uncontrolled environments, but despite the efforts deployed by engineers, they may still have some troubles in certain situations [34]. When a trouble happens, the user has to discover where the problem comes from and make whatever change to its own home or to the way the robot is used so that the situation will not occur again. Adaptation is thus on the human user side. Human-centered robotics aims at building robots that can collaborate with humans and empower them. They should then first not be a burden for their human collaborators and exhibit a high level of autonomy [35].

The more variable the tasks and the environments to fulfil them, the more difficult it is to anticipate all the situations that may occur. Human-centered robots are supposed to be in contact with humans and thus experience their everyday environment, that is extremely diverse. Current robots clearly have trouble to appropriately react to situations that have not been taken into account by their designer. When an unexpected situation occurs and results in a robot failure, a human-centered robot is expected to, at least, avoid to infinitely repeat this failure. It implies an ability to exploit its experience to improve its behaviour: *a human-centered robot needs to possess a learning ability*. Learning is the ability to exploit experience to improve the behaviour of a machine [36]. Robotics represents a challenge for all learning algorithms, including deep learning [37]. Reinforcement learning algorithms aim at discovering the behaviour of an agent from a reward that tells whether it behaves well or not. From an indication of what to do, it searches how to do it. It is thus a powerful tool to make robots more versatile and less dependant on their initial skills, but reinforcement learning is notoriously difficult in robotics [38]. One of the main reasons is that a robot is in a continuous environment, with continuous actions in a context that is, in general, partially observable and subject to noise and uncertainty. A robot that successfully learns to achieve a task owes a significant part of its success to the appropriate design of the state and action spaces that learning relies on. Different kinds of algorithms do exist to explore the possible behaviours and keep the ones that maximise the reward [39], but for all of them holds, the larger the state and action spaces, the more difficult the discovery of appropriate behaviours. In the meantime, a small state and action space limits robot abilities. A human-centered robot is expected to be versatile, it is thus important to avoid too strong limitations of their capabilities. A solution is to build robots with an open-ended learning ability [40, 41], that is with the ability to build their own state and action spaces on-the-fly [42]. The perception of their environment can be structured by their interaction capability (Fig. 2). The skills they need can be built on the basis of an exploration of possible behaviours. In a process inspired from child development [43], this search process can be guided by intrinsic motivations, that can replace the task oriented reward used in reinforcement learning, for the robot to bootstrap the acquisition process of world models and motor skills [44]. This adaptation capability is important to make robots able to deal with the variability of human behaviours and environments and to put adaptation on the robot side instead of the human side, but it is not enough to make robots human-centered.

The main reason is that humans play a marginal role in this process, if any. A human-centered robot needs to have or develop human-specific skills. To do so, they first need to be able to *interact with humans*. It can be done in different ways that are introduced, with the challenges it raises, in "Humans in the loop" section. They also need to *understand humans*. "Understanding humans

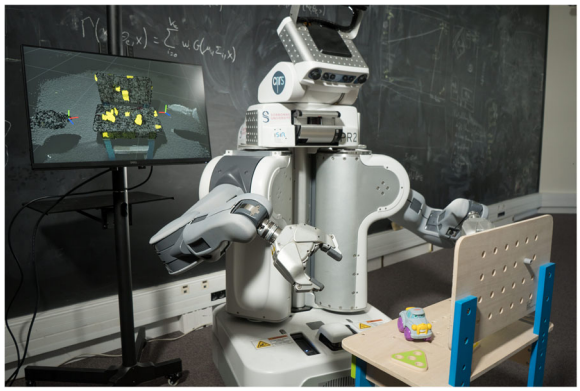

Fig. 2 A PR2 robot engaged in an interactive perception experiment to learn a segmentation of its visual scene [93, 94]. The interaction of the robot with its surrounding environment provides data to learn to discriminate objects that can be moved by the robot from the background (Copyright: Sorbonne Université)

and human intentions" section discusses this topic. Based on this understanding, robots may have to adapt their behaviour. Humans are used to transmit their knowledge and skills to other humans. They can teach, explain or show the knowledge they want to convey. Providing a robot with a particular knowledge is done through programming, a process that requires a strong expertise. A human-centered robot needs to provide other means of knowledge transmission. It needs to be able to *learn from humans*, see "Learning from humans" section for a discussion on this topic. Last but not least, humans need to understand what robots know, what they can and what they cannot do. It is not straightforward, in particular in the context of the current trend of AI that mostly relies on black-box machine learning algorithms [45]. "Making robots understandable for humans" section examines this topic in a robotics context.

Humans in the loop

The body of literature about the interaction of humans with computers and robots is huge and contains metrics [46, 47], taxonomies [48] and other kinds of descriptions and classifications trying to establish criteria for the possible scenarios. Often, a certain aspect is in the focus, like e.g. safety [49]. Still, a structured and coherent view is not established, such that it remains difficult to directly compare approaches in a universal concept [50]. Despite this ongoing discussion, we take a more fundamental view in the following to describe what is actually possible. A human has three possibilities to interact with robots: physical interaction, verbal interaction and non-verbal interaction. Each of these interaction modalities has its own features, complexities and creates its own requirements.

Physical interaction

As a robot has a physical body, any of its movements is likely to create a physical interaction with a human. It may not be voluntary, for instance if the robot hits a human that it has not perceived, but physical interaction is also used on purpose, when gestures are the main target. Physical interaction between humans and robots has gained much attention over the past years since some significant advancements have been made in two main areas of robotics. On the one hand, new mechanical designs of robotic systems integrate compliant materials as well as compliant elements like springs. On the other hand, on the control side, it became possible to effectively control compliant structures because of increased computational power of embedded micro-controllers. Another reason is also the availability of new, smaller and yet very powerful sensor elements to measure forces applied to the mechanical structures. It has lead to the implementation of control algorithms that can react extremely rapidly to external forces applied to the mechanical structure. A good overview of the full range of applications and the several advancements that have been made in recent years can be found in [51].

These advancements were mandatory for a safe use of robotic systems in direct contact with human beings in highly integrated interaction scenarios like rehabilitation. Rehabilitation opens up enormous possibilities for the immediate restoration of mobility and thus quality of life (see, e.g. the scene with an exoskeleton and a wheel chair depicted in Fig. 3), while at the same time promoting the human neuronal structures through sensory influx. Furthermore, the above-mentioned methods of machine learning, especially in their deep (Deep-Learning) form, are suitable methods to observe and even predict accompanying neural processes in the human brain [52]. By observing the human electro-encephalogram, it becomes possible to predict the so-called lateral readiness potential (LRP) -that reflects the process of certain brain regions to prepare deliberate extremity movements- up to 200ms before the actual movement occurs. This potential still occurs in people even after lesions or strokes and can be predicted by AI-methods. In experimental studies, the prediction of an LRP was used to actually perform the intended human movement via an exoskeleton. By predicting the intended movement at an early stage and controlling the exoskeleton mechanics in time, the human being experiences the intended movement as being consciously performed by him or herself.

As appealing and promising such scenarios sound, it is necessary to consider the implications of having an 'intelligent' robot acting in direct contact with humans. There are several aspects that need to be considered and that do pose challenges in several ways [53]. To start with, we do need to consider the mechanical design and the kine-

Fig. 3 An upper-body exoskeleton integrated into a wheel chair can support patients in doing everyday tasks as well as the overall rehabilitation process. (Copyright: DFKI GmbH)

matic structure in much deeper way as we would have to in other domains. First of all, there is the issue of safety of the human. In no way can it be allowed for the robot to harm the human interaction partner. Therefore safety is usually considered on three different levels:

1. On the level of mechanical design we must ensure that compliant mechanisms are used that absorb the energy of potential impacts with an object or a human. This can be done in several ways by integrating spring like elements in the actuators that work in series with a motor/gear setting. This usually allows the spring to absorb any impact energy but on the other hand it decreases the stiffness of the system which is a problem if it comes to very precise control with repeatable motions even under load.
2. on the second level the control loops can be used to basically implement an electronic spring. This is done by measuring the forces and torques on the motor and by controlling the actuators based on these values instead of position signal only. The control based on position ensures a very stiff and extremely precise and repeatable system performance while torque control is somewhat less precise. It further requires a nested control approach which combines position and torque control in order to achieve the desired position of the joint while at the same time respecting torque limits set by the extra control loop. Overall the effect is similar to that of a mechanical spring as the robot will immediately retract (or stop to advance) as soon as external forces are measured, and torque limits are violated. Even though this sounds like it is a pure control problem and AI-Technologies are not required. The problem quickly becomes NP Hard if the robot actually consists of many degrees of freedom like e.g. a humanoid robot. In these cases, deep neural network strategies are used to find approximations to the optimal control scheme [54]. Yet there are cases when even higher levels of cognitive AI approaches are required, and this is in cases where the limitations of torques to the joints contradict the stability of the robot standing or walking behavior, for instance, or when it comes to deliberately surpass the torque limits if e.g. the robot needs to drill a hole in the wall. In this case some joints need to be extremely stiff in order to provide enough resistance to penetrate the wall with the drill. These cases require higher levels of spatio-temporal planning and reasoning approaches to correctly predict context and to adjust the low-level control parameters accordingly and temporarily.
3. on the level of environmental observation there are several techniques that use external sensors like cameras, laser range finders and other kinds of sensors to monitor the environment of the robot and to intervene with the control scheme of the robot as soon as a person enters the work cell of the robotic system. Several AI technologies are used to predict the intentions of the person entering the robots environment and can be used to modify the robots behavior in an adequate way: instead of just a full stop if anything enters the area, it is a progressive approach with a decrease of robot movement speed if the person comes closer. In most well-defined scenarios these approaches can be implemented with static rule-based reasoning approaches, however, imagine a scenario where a robot and a human being are working together to build cars. In this situation there will always be close encounters between the robot and the human and most of them are wanted and required. There might even be cases where the human and the robot actually get into physical contact, for instance when handing over a tool. Classical reasoning and planning approaches have huge difficulties in adequately representing such situations [55]. What is needed instead is an even deeper approach to actually make the robot understand intentions of the human partner [56].

Verbal interaction

"Go forward", "turn left", "go to the break room", it is very convenient to give orders to robots using natural language, in particular when robot users are not experts or physically impaired [57]. Besides sending orders to the robot (human-to-robot interaction), a robot could answer questions or ask for help (robot-to-human interaction) or engage in a conversation (two-way communication) [58]. Verbal interaction has thus many different applications in robotics and contrary to physical interactions, it does

not create strong safety requirements. A human cannot be physically harmed through verbal interaction, except if it makes the robot act in a way that is dangerous for the human, but in this case the danger still comes from the physical interaction, not from the verbal interaction that has initiated it.

Although a lot of progress has been made on natural language processing, robotics creates specific challenges. A robot has a body. Robots are thus expected to understand spatial (and eventually temporal) relations and to connect the symbols they are manipulating to their sensorimotor flow [59]. This is a *situated* interaction. Giving a robot an order as "go through the door" is expected to make the robot move to the particular door that is in the vicinity of the robot. There is a need to connect words to the robots own sensorimotor flow: each robot has specific sensors and effectors and it needs to be taken into account. If the robot needs to understand a limited number of known words, it can be hand-crafted [57]. It can also rely on deep learning methods [60], but language is not static, it dynamically evolves through social interaction, as illustrated by the appearance of new words: in 2019, 2700 words have been added to the Oxford English Dictionary[1]. Furthermore the same language may be used in a different way in distant places of the world. French as talked in Quebec, for instance, has some specificities that distinguishes it from the French talked in France. A human-centered robot needs to be able to adapt the language it uses to its interlocutor. It raises many different challenges [61], including symbol grounding, that is one of the main long-standing AI challenges [31]. Using words requires to know their meaning. This meaning can be guessed from a semantic network, but as the interaction is situated, at least some of the words will need to be associated with raw data from the sensorimotor flow, for instance the door in the "go through the door" order needs to be identified and found in the robot environment. This is the grounding problem.

The seminal work of Steels on language games [62, 63] shows how robots could actually engage in a process that converges to a shared vocabulary of grounded words. When the set of symbols is closed and known beforehand, symbol grounding is not a challenge anymore, but it still is if the robot has to build it autonomously [64]. To differentiate it from the grounding of a fixed set of symbols, it has been named *symbol emergence* [65, 66]. A symbol has different definitions. In symbolic AI, symbols are basically a pointer to a name, a value and possibly other properties, like a function definition, for instance. A symbol carries a semantic which is different for the human and for the robot, but enables them to partially share the same grounds. In the context of language study, the definition of a symbol is different. Semiotics, the study of signs that mediate communication, defines it as a triadic relationship between an object, a sign and an interpretant. This is not a static relationship, but a process. The interpretant is the effect of a sign on its receiver, it is thus a process relating the sign with the object. The dynamic of this process can be seen in our ability to dynamically give names to objects (may they be known or not). Although many progresses have been made recently on these topic [58, 66], building a robot with this capability remains a challenge.

[1] https://public.oed.com/updates/

Non-verbal interaction

The embodiment of robots creates opportunities to communicate with humans by other means than language. It is an important issue as multiple nonverbal communication modalities do exist between humans and they are estimated to represent a significant part of communicated meaning between humans. Non verbal cues revealed for instance to help children to learn new words from robots [67]. Adding nonverbal interaction abilities to robots thus opens the perspective of building robots that can better engage with humans [68], i.e. social robots [13]. Nonverbal interaction may support verbal communication, as lip-syncing or other intertwined motor actions as head nods [69], and may have a significant impact on humans [70], as observed through their behaviour response, task performance, emotion recognition and response as well as cognitive framing, that is the perspective humans adopt, in particular on the robot they interact with.

Different kinds of nonverbal communications do exist. The ones that incorporate robots movements are kinesics, proxemics, haptics and chronemics. Kinesics relies on body movements, positioning, facial expressions and gestures and most robotics related research on the topic focus on arm gestures, body and head movements, eye gaze and facial expressions. Proxemics is about the perception and use of space in the context of communication, including the notions of social distance or personal space. Haptics is about the sense of touch and chronemics with time-experiencing. Sanuderson and Nejat have reviewed robotics research work on these different topics [70].

Besides explicit non-verbal communication means, the appearance of a robot has revealed to impact the way humans perceive a robot and engage in a human-robot interaction [71, 72]. It has been shown for instance that a humanlike-shape influences non-verbal behaviors towards a robot like delay of response, distance [73] or embarrassment [74]. Anthropomorphic robots significantly draw the attention of the public and thus creates high expectations in different service robotics applications, but the way they are perceived and their acceptance is a complex function involving multiple factors, including user culture, context and quality of the interaction or even degree of human likeness [75]. The impact of

this last point, in particular, is not trivial. Mori proposed the uncanny valley theory to model this relation [76, 77]. In this model, the emotional response improves when robot appearance gets more humanlike, but a sudden drop appears beyond a certain level: robots that look like humans but still with noticeable differences, can thus create a feeling of eeriness resulting in discomfort and rejection. This effect disappears when the robot appearance gets close enough to humans. The empirical validation of this model is difficult. Some experiments seem to validate it [78], while others lead to contradicting results [79]. For more details, see the reviews by Fink [80] or Złotowski et al. [81].

Understanding humans and human intentions

There are situations in which robots operate in isolation, such as in manufacturing lines for welding or painting, or in deep sea or planetary exploration. Such situations are dangerous for humans and the robot task is provided to it through pre-programming (e.g. welding) or teleprogramming (e.g., a location to reach on a remote planet). However, in many robotic application areas, be it in manufacturing or in service, robots and humans are starting to more and more interact with each other in different ways. The key characteristics making these interactions so challenging are the following:

- Sharing space, for navigation or for reaching to objects for manipulation
- Deciding for joint actions that are going to be executed by both the robot and the human
- Coordination of actions over time and space
- Achieving joint actions physically

These characteristics lead to many different scientific questions and issues. For example sharing space requires geometric reasoning, motion planning and control capabilities [82]. Deciding for joint actions [83] requires a mutual representation of human capabilities by the robot and vice-versa, e.g., is the human (resp. robot) capable of holding a given object? It also requires a Theory of Mind on the part of the robot and of the human: what are the robot's representations and what are the human's representations of a given situation? What is the human (resp. robot) expected to do in this situation?

The third mentioned characteristic, coordination of action, requires in addition to what has been mentioned above signal exchanges between human and robot to ensure that each is indeed engaged and committed to the task being executed. For example gaze detection through eye trackers enables to formulate hypotheses about human visual focus. The robot in turn has to provide equivalent information to the human, since the human usually cannot determine the robot's visual focus from only observing its sensors. In this case, it becomes therefore necessary that the robot signals explicitly what is its focus or what are its intentions (see "Making robots understandable for humans" section).

Now, when it comes to physical interaction, robot and human are not only in close proximity, but they also exchange physical signals such as force. Consider for example a robot and a human moving a table together. Force feedback enables to distribute the load correctly between them, and enables to coordinate the actions. In the case of physical interaction, another important aspect is to ensure human safety, which puts constraints on robot design and control. Compliance and haptic feedback become key (see "Physical interaction" section).

In all these interaction scenarios, the robot must already have all the autonomous capacities for decision-making and task supervision. Indeed the robot must be able to plan its own actions to achieve a common goal with the human, taking into account the human model and intentions.

Take the simple example of a human handing an object to the robot. The common goal is that, in the final state, the robot is holding the object, whereas in the initial state the human is holding it. The goal must be shared right from the beginning of the interaction, for example through an explicit order given by the human. Alternatively the robot might be able to determine the common goal by observing the human's behavior, which requires the robot to have the ability to deduce human intentions from their actions, posture, gestures (e.g., deictic gestures) or facial expressions. This cannot be but a probabilistic reasoning capacity, given the uncertainties of observation and of prior hypotheses. Then the robot must plan its actions according to its human model, and this cannot be but a probabilistic planning process, e.g., using markovian processes, because of the inherent uncertainties of the observations – and therefore the robot's beliefs – and of action execution. Robot task supervision must also ensure that the human is acting in accordance to the plan, by observing actions and posture.

Another essential dimension for complex interactions is communication using dialogue. The robot can start such a dialogue for example when it detects that some information is needed to complete its model, or to reduce its uncertainties. Formulating the correct questions requires the robot to have a self assessment capacity of its own belief state.

Learning from humans

Using the human as a teacher to train robotic systems has been around for some time [84]. Many cases and scenarios, like the hybrid team scenario (see example depicted in Fig. 4) where humans and robots are building cars

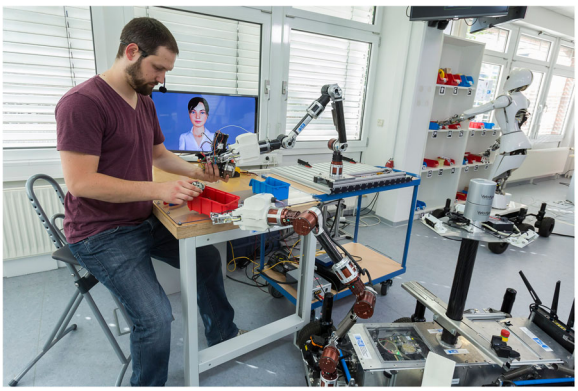

Fig. 4 Examples for humans, robots and other AI agents working in hybrid teams. Due to the possible applications and scenarios robots can be configured here as stationary or mobile systems up to even complex systems with humanoid appearance. (Copyright: Uwe Völkner/Fotoagentur FOX)

together acting as a team, are too complex to be completely modelled. Consequently, it is difficult or impossible to devise exact procedures and rule-based action execution schemes in advance. One example here could be to formulate the task to have a robot pack a pair of shoes in a shoebox [85]. Even a task that sounds as simple as this proved to be impossible to be completely modeled. Therefore, a learning by demonstration method has been applied to teach the robot the task by a human demonstrator. In such cases learning, or said differently a step-wise approximation and improvement of the optimal control strategy, is the most straightforward option available. In situations where enough a priori data is available, this can be done offline and the robotic system can be trained to achieve a certain task. However, in many cases, data is not available and therefore online strategies are needed to acquire the desired skill. The learning by demonstration approach can already be implemented quite successfully by e.g. recording data from human demonstrators that are instrumented with reflectors for image capturing devices and then feeding skeleton representations of the human movements as sample trajectories into the learning system which in turn uses e.g. Reinforcement Learning techniques to generate appropriate trajectories. This approach usually leads to quite usable policies on the side of the robotic system, yet in many cases when applied in a realistic task scenario it turns out that "quite good" is not good enough and online optimization has to be performed. Here it turns out to be advantageous to include approaches like discussed in the previous section on understanding human intentions or state of mind.

Using this general idea, it was possible to online improve the performance of an already trained robot by applying a signal generated by the human brain on a subconscious level providing it as a reinforcement signal back to the robot [56]. The signal is the so-called Error potential. This is an event related potential (ERP) generated by brain areas when a mismatch between expected input and actual input occurs. In many real-world situations such a signal is produced e.g., when a human observes another human to perform a movement in an obviously wrong way in the correct context or the correct movement is performed but in the wrong context. The beauty about this signal is that it is generated on subconscious levels, so before the human actively is aware of it. This is important for two reasons:

1. When the human becomes aware of the signal that means that it was already analyzed and modulated by other brain regions. This means that a cognitive classification of the subconscious signal has taken place which will disassociate the original signal.
2. The second reason why it is important that the signal occurs before evaluation by other brain areas is that it does not have to be externalized e.g. by verbalization. Imagine a hybrid team scenario where the human in the team has to explicitly verbalize each error that he or she observes in the performance of the robot. First, the above mentioned disassociation process will lead to a blurriness or haziness of the verbalized feedback to the robot but more importantly as a second result the human would probably not verbalize each and every error due to fatigue and information valuable for interaction is lost.

To summarize, the learning could either happen using external information available, like getting commands or watching humans demonstrating a task, or implicit signals during interaction like evaluation of facial expressions or by using brain signals like certain ERPs to provide feedback. The latter is of course using information from the human interaction partner that is not directly controlled by the human and also not per se voluntarily given. This raises ethical and legal questions that have to be addressed when using this as a standard procedure for interaction (see also "Ethical questions" section), underlining the fact that Human-centered AI and robotics ultimately include the involvement of disciplines from social sciences. At the same time, we have outlined that making use of such information can be highly beneficial for fluent and intuitive interaction and learning.

Making robots understandable for humans

In "Understanding humans and human intentions" section, it was discussed how the robot can better understand humans and how this can be achieved to some point.

It is rather straightforward to equip the robot with the necessary sensors and software to detect humans and to interpret gestures, postures and movements, as well as to detect their gaze and infer some intentions. Even if it is not the whole complexity of human behavior, these capacities can capture enough of human intentions and actions to enable task sharing and cooperation. Equally important however in an interaction is the opposite case, that is how can the human better understand the robot's intentions and actions.

In most scenarios, we can safely assume that the human does have some a priori knowledge about the framework of action that the robot is equipped with. That is to say that the human can infer some of the physical capabilities and limitations of the system from its appearance (e.g., a legged robot vs. a wheeled robot) but not of its power e.g., can the robot jump or climb a given slope? Even if the human could have some general ideas of the spectrum of robot sensing possibilities, it is not clear whether the robot perceptive capabilities and their limits can be completely and precisely understood. This is e.g., a result of the fact that it is difficult for humans to understand the capabilities and limitations of sensors that they don't have e.g., infrared sensors or laser-rangefinders providing point-clouds. It is fundamentally impossible for a human being to understand the information processing going on in robot systems with multi-level hierarchies, from low-level control of single joints to higher levels of control involving deep neural networks and finally to top level planning and reasoning processes that all interact with each other and influence each other's output. This is even extremely difficult for trained computer science experts and robot designers. It represents a complete field of research that deals with the problems of how to manage the algorithmic complexity that occurs in structurally complex robotic systems that act in dynamic environments. Actually the design of robot control or cognitive architectures is an open research area and still a big challenge for AI-Based-Robotics [86].

Attempts to approach the problem of understanding robots by humans have been made in several directions. One attempt is the robot verbally explaining its actions [16]. This is to say that the robot actually tells (or writes on a screen) the human what it is doing and why a specific action is carried out. At the same time, it is possible for the human to ask the robot for an explanation of its action(s) and the robot gives the explanation verbally, in computer animated graphics or in iconized form on a screen installed on the robot. The hope behind such approaches is that the need for explanations deliberately uttered by the robot as well as the quest for answers from the side of the human will decrease over time as learning and understanding occurs on the side of the human. Of course this is difficult to assess as long term studies so far have not been carried out or could not be carried out because of the unavailability of appropriate robots. But one assumption that we can safely make is that the explicit answering or required listening to explanations by the human will not be highly appreciated when it comes to practical situations, and the repetitive explanatory utterances of the robot will quickly bother humans.

Therefore it is necessary to think about more subtle strategies to communicate robot internal states and intentions to the human counterpart e.g., its current goals, its knowledge about the world, its intended motions, its acknowledgement of a command, or its requests for an action by the human. Examples of such approaches are to use mimics and gestures. Robots equipped with faces - either just as computer screens where the face is generated or by actually actuated motors forming faces under artificial skin covered robotic heads (if such devices are deemed acceptable - see "Ethical questions" section - in order to produce facial expressions which gives some information about the internal state of the robot. These approaches could successfully be applied in e.g. home and elderly care scenarios. However, the internal states being externalized here are rather simple ones that are meant to stimulate actions on the human side like in the pet robot Paro.

However, we can assume that it should be possible in well known scenarios, such as in manufacturing settings, to define fixed signals for interaction made from a set of gestures, including deictic gestures, facial expressions or simply graphical patterns that can be used to externalize internal robot states to human partners. Such a model of communication can be described as the first steps towards achieving a more general common alphabet [87] as the basis for a language between humans and robots. It is likely that such a common language will be developed or more likely emerge, from more and more robot human interaction scenarios in real world applications as a result of best practice experiences.

It is certain that the corresponding challenges on the robotic side go beyond what was described earlier like the soft and compliant joints that are used for safety reasons. It will be necessary to develop soft and intelligent skin as a cover of the mechanical robot structures that can be used not just as an interface for expressions -in the case of facial skin- but also as a great and powerful sensor on other parts of the robot body for improving and extending the range of physical interactions with humans [88]. Just a simple example that we all know is that in a task performed by two humans it is often observed that one of two partners slightly pushes or touches the other on the shoulder or the arm in order to communicate e.g. that a stable grip has been achieved or to say: 'okay I got it, you can let it go...' This kind of interaction could also be verbally transmitted to the interaction partner, but humans have

the ability to visualize the internal states of their human counterparts, because we share the same kinematic structure and disposition. It is thus in this case not necessary to speak. Just a simple touch suffices to transmit a complex state of affairs. Yet, the interaction of humans with robots that are equipped with such kind of advanced skin technologies can be expected to be a starting point for a common language. The physical interaction will therefore enable new ways of non-physical interaction and very likely the increased possibilities for nonphysical interaction will in turn stimulate other physical interaction possibilities. In summary, it will be an interesting voyage to undertake if in fact intelligent and structurally competent robotic systems will become available as human partners in various everyday life situation. Like in all other technologies, the human designer will shape the technology, but at the same time the technology will shape the human, both as a user of the technology but also as the designer of this technology.

Ethical questions

There are several issues which raise questions of ethics of robotic technologies considered as interaction partners for humans [89]. To list but a few:

- Transformation of work in situations where humans and robots interact together. Depending on how it is designed, the interaction might impose constraints on the human instead of making the robot adapt to the human and carry the burden of the interaction. For example the human is given more dexterous tasks such as grasping, which end up being repetitive and wearing when robot speed doing simpler tasks imposes the pace.
- Mass surveillance and privacy issues when personal or domestic robots collect information about their users and households, or self-driving cars which are permanently collecting data on their users and their environments.
- Affective bonds and attachment to personal robots, especially those made to detect and express emotions.
- Human transformation and augmentation through exoskeletons or prosthetic devices.
- Human identity, status of robots in society (e.g;, legal personality), especially for android robots mimicking humans in appearance, language and behavior.
- Sexbots designed to be sexual devices that can be made to degrade the image of women, or to look like children
- Autonomous weapon systems - which are not so to speak "interacting" with humans, but which are endowed with recognition capacities to target humans.

If we speak about ethics in the context of robots and AI technologies, what we fundamentally mean is that we want to make sure that this technology is designed and used for the good of mankind and not for the bad. The first problem is obviously how do we define good and bad? There are the obvious answers implying that a robot should not harm a person. No question, but what about a surgical robot that needs to inject a vaccine into the arm of a person with a syringe, thus physically injuring her at the moment, but for her benefit? How can we make the distinction between these cases in a formal way? This is the core of the problem.

If we speak about ethics and how to design ethical deliberation into technical systems so that the robot decision-making or control system behaves for "the good", we are fundamentally required to come up with a formalization of ethics. In some form or the other we will be required to put down in expressions of logic and numerical values what is ethical and what is not. In our understanding this will not be possible in a general form, because human ethical judgment and moral thinking is not amenable to algorithmic processing and computations. For example, how would we define algorithmically a principle of respect for human dignity? The concept of dignity itself is complex and has several moral and legal interpretations.

Ethical deliberation cannot be reduced to computing and comparing utilities, as we often see in publications on ethical dilemmas for self driving cars for example. The car could only make computations based on data acquired by its sensors, but the ethical choices would have actually been already made by the designers. Even deciding that the passengers can customise ethical choices, or to let the system learn [90], for example in simulations, to determine values to be optimized is a negation of what ethical deliberation is. Indeed this would entail an a priori decision on a situation to come, or to decide that ethical deliberation is based on statistics of past actions.

We will of course be able to formalize ethical guidelines (to the designers) for robot design and control if concrete well specified domains are regarded. We could e.g. solve the syringe problem easily if we built a surgical robot that is used and operated only in hospitals and that has a clearly defined set of tasks to fulfill in e.g. the vaccination department of the hospital. And then this becomes a matter of safety design, similar to any other technical device. But what about a household service robot that is designed to clean the floors and wash the dishes... Wouldn't we want this robot also to be able to perform first aid services e.g. if the person in the household suffers diabetics and need insulin injections from time to time... Cases can be constructed where we come to the problem that a complete and full formalization of ethics is impossible.

Carrying a responsible approach or a value-based design procedure [91] can help to conceive robots and AI systems for which ethical issues are actually solved by the human designers and manufacturers beforehand, during specification, development and manufacturing. The robot itself will not be endowed with moral judgment. But we will have to make sure that the humans will abstain from misusing the technology.

But more profound questions arise when it comes to the last three issues listed above. For example, building android human-like robots can be considered a scientific research topic, or a practical solution to facilitate human-robot interaction. However, the confusion this identification of humans with machines provokes requires a reflection on the nature of human identity as compared to machines, that needs to address all aspects and consequences of such technical achievements.

A reflection grounded on philosophical, societal and legal considerations is necessary, beyond sole scholarly studies, to address the impact of these technologies on society. Indeed, there are numerous initiatives and expert groups who have actually already issued ethics recommendations on the development and use of AI and Robotics systems, including the European High-Level Expert Group on AI (HLEG-AI), the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, the UNESCO COMEST, and the OECD (see [92] for a comprehensive overview). As an example of commonly accepted ethics recommendations are the seven "requirements for trustworthy AI[2]" issued by the HLEG-AI in 2019:

1. "Human agency and oversight": AI systems should be subject to human oversight and they should support humans in their autonomy and decision-making
2. "Technical Robustness and Safety" should be provided. Systems should be reliable and stable also in situations with uncertainty, they should be resilient against manipulations from outside
3. "Privacy and Data Governance" should be guaranteed during the lifecycle with data access controlled and managed, and data quality provided.
4. "Transparency": Data and processes should be well documented to trace the cause of errors. Systems should become explainable to the user on the level appropriate to understand certain decisions the system is making.
5. "Diversity, Non-Discrimination and Fairness" should be ensured by controlling for biases that could lead to discriminatory results. Access to AI should be granted to all people.
6. "Societal and Environmental Well-Being": The use of AI should be for the benefit of society and the natural environment. Violation of democratic processes should be prevented.
7. "Accountability" should be provided such that AI systems can be assessed and audited. Negative impacts should be minimised or erased.

However there are still open issues, mostly related to how to translate principles into practice, or topics subject to hard debates such as robot legal personality, advocated by some to address liability issues. Furthermore, when considering specific use-cases, tensions between several requirements could arise, that will have to be specifically addressed.

Conclusion

Most AI systems are tools for which humans play a critical role, either at the input of the system, to analyse their behavior, or at the output, to give them an information they need. Robotics is different as it develops physical systems that can perceive and act in the real world without the mediation of any humans, at least for autonomous robots. Building human-centered robots requires to put humans back into the loop and to provide the system with the ability to interact with humans, to understand them and learn from them while ensuring that humans will also understand what they can and cannot do. It also raises many ethical questions that have been listed and discussed. Human centered AI and Robotics thus create many different challenges and require the integration of a wide spectrum of technologies. It also highlights that robots assisting humans are not only a technological challenge in many aspects, but rather a socio-technological transformation in our societies. In particular, the use of this technology and how it is accessible, are important topics involving actors in dealing with social processes, public awareness and political and legal decisions.

Authors' contributions
All authors have contributed to the text and approved the final manuscript.

Funding
The project has received funding from the European Union's Horizon 2020 research and innovation programme Project HumanE-AI-Net under grant agreement No 952026.

Availability of data and materials
Not applicable.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Institute of Intelligent Systems and Robotics (ISIR), Sorbonne Université, CNRS, Paris, France. [2] Robotics Innovation Center, DFKI GmbH (German Research Center for Artificial Intelligence), Bremen, DE, Germany. [3] Faculty of

[2] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Mathematics and Computer Science, Robotics Group, University of Bremen, Bremen, DE, Germany.

Received: 2 June 2021 Accepted: 27 October 2021
Published online: 28 January 2022

References

1. Campbell M, Hoane Jr AJ, Hsu F.-h. (2002) Deep blue. Artificial intelligence 134(1-2):57–83
2. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, et al. (2016) Mastering the game of go with deep neural networks and tree search. Nature 529(7587):484–489
3. Torrey L, Shavlik J (2010) Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI global, Hershey. pp 242–264
4. Yuh J, West M (2001) Underwater robotics. Adv Robot 15(5):609–639. https://doi.org/10.1163/156855301317033595
5. Kirchner F, Straube S, Kühn D, Hoyer N (2020) AI Technology for Underwater Robots. Springer, Cham
6. Yoshida K (2009) Achievements in space robotics. IEEE Robot Autom Mag 16(4):20–28. https://doi.org/10.1109/MRA.2009.934818
7. Yoshida K, Wilcox B (2008) Space robots. In: Springer handbook of robotics. Springer, Berlin. pp 1031–1063
8. Yangsheng X, Kanade T (1993) Space Robotics: Dynamics and Control. Springer
9. Goodrich MA, Schultz AC (2008) Human-robot Interaction: a Survey. Now Publishers Inc
10. Ricks DJ, Colton MB (2010) Trends and considerations in robot-assisted autism therapy. In: 2010 IEEE International Conference on Robotics and Automation, Anchorage. pp 4354–4359
11. Boucenna S, Narzisi A, Tilmont E, Muratori F, Pioggia G, Cohen D, Chetouani M (2014) Interactive technologies for autistic children: A review. Cogn Comput 6(4):722–740
12. Shishehgar M, Kerr D, Blake J (2018) A systematic review of research into how robotic technology can help older people. Smart Health 7:1–18
13. Breazeal C, Dautenhahn K, Kanda T (2016) Social robotics. In: Springer Handbook of Robotics. Springer, Berlin. pp 1935–1972
14. Sheridan TB (2020) A review of recent research in social robotics. Curr Opin Psychol 36:7–12
15. Schwartz T, Feld M, Bürckert C, Dimitrov S, Folz J, Hutter D, Hevesi P, Kiefer B, Krieger H, Lüth C, Mronga D, Pirkl G, Röfer T, Spieldenner T, Wirkus M, Zinnikus I, Straube S (2016) Hybrid teams of humans, robots, and virtual agents in a production setting. In: 2016 12th International Conference on Intelligent Environments (IE). IOS Press, Amsterdam. pp 234–237
16. Schwartz T, Zinnikus I, Krieger H-U, Bürckert C, Folz J, Kiefer B, Hevesi P, Lüth C, Pirkl G, Spieldenner T, Schmitz N, Wirkus M, Straube S (2016) Hybrid teams: Flexible collaboration between humans, robots and virtual agents. In: Klusch M, Unland R, Shehory O, Pokahr A, Ahrndt S (eds). Multiagent System Technologies. Springer, Cham. pp 131–146
17. Peshkin M, Colgate JE (1999) Cobots. Ind Robot Int J 26(5):335–341
18. Maciejasz P, Eschweiler J, Gerlach-Hahn K, Jansen-Troy A, Leonhardt S (2014) A survey on robotic devices for upper limb rehabilitation. J Neuroeng Rehabil 11(1):3
19. Kumar S, Wöhrle H, Trampler M, Simnofske M, Peters H, Mallwitz M, Kirchner EA, Kirchner F (2019) Modular design and decentralized control of the recupera exoskeleton for stroke rehabilitation. Appl Sci 9(4). https://doi.org/10.3390/app9040626
20. Nowak A, Lukowicz P, Horodecki P (2018) Assessing artificial intelligence for humanity: Will ai be the our biggest ever advance? or the biggest threat [opinion]. IEEE Technol Soc Mag 37(4):26–34
21. Siciliano B, Khatib O (2016) Springer Handbook of Robotics. Springer, Berlin
22. McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI Mag 27(4):12–12
23. Annoni A, Benczur P, Bertoldi P, Delipetrev B, De Prato G, Feijoo C, Macias EF, Gutierrez EG, Portela MI, Junklewitz H, et al. (2018) Artificial intelligence: A european perspective. Technical report, Joint Research Centre (Seville site)
24. Wolf MJ, Miller KW, Grodzinsky FS (2017) Why we should have seen that coming: comments on microsoft's tay "experiment," and wider implications. ORBIT J 1(2):1–12
25. Strickland E (2019) Ibm watson, heal thyself: How ibm overpromised and underdelivered on ai health care. IEEE Spectr 56(4):24–31
26. Poole D, Mackworth A, Goebel R (1998) Computational intelligence
27. Salini J, Padois V, Bidaud P (2011) Synthesis of complex humanoid whole-body behavior: A focus on sequencing and tasks transitions. In: 2011 IEEE International Conference on Robotics and Automation, Changaï. pp 1283–1290
28. Hayet J-B, Esteves C, Arechavaleta G, Stasse O, Yoshida E (2012) Humanoid locomotion planning for visually guided tasks. Int J Humanoid Robotics 9(02):1250009
29. Pfeifer R, Gómez G (2009) Morphological computation–connecting brain, body, and environment. In: Creating Brain-like Intelligence. Springer, Berlin. pp 66–83
30. Shintake J, Cacucciolo V, Floreano D, Shea H (2018) Soft robotic grippers. Adv Mater 30(29):1707035
31. Harnad S (1990) The symbol grounding problem. Physica D Nonlinear Phenom 42(1-3):335–346
32. Bohg J, Hausman K, Sankaran B, Brock O, Kragic D, Schaal S, Sukhatme GS (2017) Interactive perception: Leveraging action in perception and perception in action. IEEE Trans Robot 33(6):1273–1291
33. Jamone L, Ugur E, Cangelosi A, Fadiga L, Bernardino A, Piater J, Santos-Victor J (2016) Affordances in psychology, neuroscience, and robotics: A survey. IEEE Trans Cogn Dev Syst 10(1):4–25
34. Vaussard F, Fink J, Bauwens V, Rétornaz P, Hamel D, Dillenbourg P, Mondada F (2014) Lessons learned from robotic vacuum cleaners entering the home ecosystem. Robot Auton Syst 62(3):376–391
35. Kaufman K, Ziakas E, Catanzariti M, Stoppa G, Burkhard R, Schulze H, Tanner A (2020) Social robots: Development and evaluation of a human-centered application scenario. In: Human Interaction and Emerging Technologies: Proceedings of the 1st International Conference on Human Interaction and Emerging Technologies (IHIET 2019), August 22-24, 2019, Nice, France, vol. 1018. Springer Nature, Berlin. pp 3–9
36. Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. Science 349(6245):255–260
37. Sünderhauf N, Brock O, Scheirer W, Hadsell R, Fox D, Leitner J, Upcroft B, Abbeel P, Burgard W, Milford M, et al. (2018) The limits and potentials of deep learning for robotics. Int J Robot Res 37(4-5):405–420
38. Kober J, Bagnell JA, Peters J (2013) Reinforcement learning in robotics: A survey. Int J Robot Res 32(11):1238–1274
39. Sigaud O, Stulp F (2019) Policy search in continuous action domains: an overview. Neural Netw 113:28–40
40. Doncieux S, Filliat D, Díaz-Rodríguez N, Hospedales T, Duro R, Coninx A, Roijers DM, Girard B, Perrin N, Sigaud O (2018) Open-ended learning: a conceptual framework based on representational redescription. Front Neurorobotics 12:59
41. Doncieux S, Bredeche N, Goff LL, Girard B, Coninx A, Sigaud O, Khamassi M, Díaz-Rodríguez N, Filliat D, Hospedales T, et al. (2020) Dream architecture: a developmental approach to open-ended learning in robotics. arXiv preprint arXiv:2005.06223
42. Lesort T, Díaz-Rodríguez N, Goudou J-F, Filliat D (2018) State representation learning for control: An overview. Neural Netw 108:379–392
43. Cangelosi A, Schlesinger M (2015) Developmental Robotics: From Babies to Robots. MIT press
44. Santucci VG, Oudeyer P-Y, Barto A, Baldassarre G (2020) Intrinsically motivated open-ended learning in autonomous robots. Front Neurorobotics 13:115
45. Hagras H (2018) Toward human-understandable, explainable ai. Computer 51(9):28–36
46. Steinfeld A, Fong T, Kaber D, Lewis M, Scholtz J, Schultz A, Goodrich M (2006) Common metrics for human-robot interaction. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI '06. Association for Computing Machinery, New York. pp 33–40. https://doi.org/10.1145/1121241.1121249
47. Murphy R, Schreckenghost D (2013) Survey of metrics for human-robot interaction. In: Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction, HRI '13. IEEE Press. pp 197–198

48. Yanco HA, Drury J (2004) Classifying human-robot interaction: an updated taxonomy. In: 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583) Vol. 3. pp 2841–28463. https://doi.org/10.1109/ICSMC.2004.1400763
49. Pervez A, Ryu J (2008) Safe physical human robot interaction–past, present and future. J Mech Sci Technol 22:469–483
50. Onnasch L, Roesler E (2021) A taxonomy to structure and analyze human–robot interaction. Int J Soc Robot 13(4):833–849
51. Haddadin S, Croft E (2016) Physical Human–Robot Interaction. In: Siciliano B, Khatib O (eds). Springer Handbook of Robotics. Springer, Cham. pp 1835–1874. https://doi.org/10.1007/978-3-319-32552-169
52. Gutzeit L, Otto M, Kirchner EA (2016) Simple and robust automatic detection and recognition of human movement patterns in tasks of different complexity. In: Physiological Computing Systems. Springer, Berlin. pp 39–57
53. Kirchner EA, Fairclough SH, Kirchner F (2019) Embedded multimodal interfaces in robotics: applications, future trends, and societal implications. In: The Handbook of Multimodal-Multisensor Interfaces: Language Processing, Software, Commercialization, and Emerging Directions-Volume 3. pp 523–576
54. Haarnoja T, Ha S, Zhou A, Tan J, Tucker G, Levine S (2018) Learning to walk via deep reinforcement learning. arXiv preprint arXiv:1812.11103:1–10
55. Tsarouchi P, Makris S, Chryssolouris G (2016) Human–robot interaction review and challenges on task planning and programming. Int J Comput Integr Manuf 29(8):916–931. https://doi.org/10.1080/0951192X.2015.1130251
56. Kim S, Kirchner E, Stefes A, Kirchner F (2017) Intrinsic interactive reinforcement learning–using error-related potentials for real world human-robot interaction. Sci Rep 7
57. Williams T, Scheutz M (2017) The state-of-the-art in autonomous wheelchairs controlled through natural language: A survey. Robot Auton Syst 96:171–183
58. Tellex S, Gopalan N, Kress-Gazit H, Matuszek C (2020) Robots that use language. Annu Rev Control Robot Auton Syst 3:25–55
59. Landsiedel C, Rieser V, Walter M, Wollherr D (2017) A review of spatial reasoning and interaction for real-world robotics. Adv Robot 31(5):222–242
60. Mei H, Bansal M, Walter MR (2016) Listen, attend, and walk: neural mapping of navigational instructions to action sequences. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. pp 2772–2778
61. Taniguchi T, Mochihashi D, Nagai T, Uchida S, Inoue N, Kobayashi I, Nakamura T, Hagiwara Y, Iwahashi N, Inamura T (2019) Survey on frontiers of language and robotics. Adv Robot 33(15-16):700–730
62. Steels L (2001) Language games for autonomous robots. IEEE Intell Syst 16(5):16–22
63. Steels L (2015) The Talking Heads Experiment: Origins of Words and Meanings, vol. 1. Language Science Press
64. Steels L (2008) The symbol grounding problem has been solved. so what's next. Symbols Embodiment Debates Meaning Cogn:223–244
65. Taniguchi T, Nagai T, Nakamura T, Iwahashi N, Ogata T, Asoh H (2016) Symbol emergence in robotics: a survey. Adv Robot 30(11-12):706–728
66. Taniguchi T, Ugur E, Hoffmann M, Jamone L, Nagai T, Rosman B, Matsuka T, Iwahashi N, Oztop E, Piater J, et al. (2018) Symbol emergence in cognitive developmental systems: a survey. IEEE Trans Cogn Dev Syst 11(4):494–516
67. Westlund JMK, Dickens L, Jeong S, Harris PL, DeSteno D, Breazeal CL (2017) Children use non-verbal cues to learn new words from robots as well as people. Int J Child-Computer Interact 13:1–9
68. Anzalone SM, Boucenna S, Ivaldi S, Chetouani M (2015) Evaluating the engagement with social robots. Int J Soc Robot 7(4):465–478
69. Mavridis N (2015) A review of verbal and non-verbal human–robot interactive communication. Robot Auton Syst 63:22–35
70. Saunderson S, Nejat G (2019) How robots influence humans: A survey of nonverbal communication in social human–robot interaction. Int J Soci Robot 11(4):575–608
71. Mathur MB, Reichling DB (2009) An uncanny game of trust: social trustworthiness of robots inferred from subtle anthropomorphic facial cues. In: 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE. pp 313–314
72. Natarajan M, Gombolay M (2020) Effects of anthropomorphism and accountability on trust in human robot interaction. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. pp 33–42
73. Kanda T, Miyashita T, Osada T, Haikawa Y, Ishiguro H (2008) Analysis of humanoid appearances in human–robot interaction. IEEE Trans Robot 24(3):725–735
74. Bartneck C, Bleeker T, Bun J, Fens P, Riet L (2010) The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. Paladyn 1(2):109–115
75. Murphy J, Gretzel U, Pesonen J (2019) Marketing robot services in hospitality and tourism: the role of anthropomorphism. J Travel Tourism Mark 36(7):784–795
76. MORI M (1970) Bukimi no tani [the uncanny valley]. Energy 7:33–35
77. Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. IEEE Robot Autom Mag 19(2):98–100
78. De Visser EJ, Monfort SS, McKendrick R, Smith MA, McKnight PE, Krueger F, Parasuraman R (2016) Almost human: Anthropomorphism increases trust resilience in cognitive agents. J Exp Psychol Appl 22(3):331
79. Bartneck C, Kanda T, Ishiguro H, Hagita N (2009) My robotic doppelgänger-a critical look at the uncanny valley. In: RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication. IEEE. pp 269–276
80. Fink J (2012) Anthropomorphism and human likeness in the design of robots and human-robot interaction. In: International Conference on Social Robotics. Springer. pp 199–208
81. Złotowski J, Proudfoot D, Yogeeswaran K, Bartneck C (2015) Anthropomorphism: opportunities and challenges in human–robot interaction. Int J Soc Robot 7(3):347–360
82. Khambhaita H, Alami R (2020) Viewing robot navigation in human environment as a cooperative activity. In: Amato NM, Hager G, Thomas S, Torres-Torriti M (eds). Robotics Research. Springer, Cham. pp 285–300
83. Khamassi M, Girard B, Clodic A, Sandra D, Renaudo E, Pacherie E, Alami R, Chatila R (2016) Integration of action, joint action and learning in robot cognitive architectures. Intellectica-La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo) 2016(65):169–203
84. Billard AG, Calinon S, Dillmann R (2016) Learning from Humans(Siciliano B, Khatib O, eds.). Springer, Cham
85. Gracia L, Pérez-Vidal C, Mronga D, Paco J, Azorin J-M, Gea J (2017) Robotic manipulation for the shoe-packaging process. Int J Adv Manuf. Technol. 92:1053–1067
86. Chatila R, Renaudo E, Andries M, Chavez-Garcia R-O, Luce-Vayrac P, Gottstein R, Alami R, Clodic A, Devin S, Girard B, Khamassi M (2018) Toward self-aware robots. Front Robot AI 5:88. https://doi.org/10.3389/frobt.2018.00088
87. de Gea Fernández J, Mronga D, Günther M, Knobloch T, Wirkus M, Schröer M, Trampler M, Stiene S, Kirchner E, Bargsten V, Bänziger T, Teiwes J, Krüger T, Kirchner F (2017) Multimodal sensor-based whole-body control for human–robot collaboration in industrial settings. Robot Auton Syst 94:102–119. https://doi.org/10.1016/j.robot.2017.04.007
88. Aggarwal A, Kampmann P (2012) Tactile sensors based object recognition and 6d pose estimation. In: ICIRA. Springer, Berlin
89. Veruggio G, Operto F, Bekey G (2016) Roboethics: Social and Ethical Implications(Siciliano B, Khatib O, eds.). Springer, Cham
90. Iacca G, Lagioia F, Loreggia A, Sartor G (2020) A genetic approach to the ethical knob. In: Legal Knowledge and Information Systems. JURIX 2020: The Thirty-third Annual Conference, Brno, Czech Republic, December 9–11, 2020. IOS Press BV, 2020, 334. pp 103–112
91. Dignum V (2019) Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer, Berlin
92. Jobin A, Ienca M, Vayena E (2019) The global landscape of ai ethics guidelines. Nat Mach Intell 1(9):389–399. https://doi.org/10.1038/s42256-019-0088-2
93. Goff LKL, Mukhtar G, Coninx A, Doncieux S (2019) Bootstrapping robotic ecological perception from a limited set of hypotheses through interactive perception. arXiv preprint arXiv:1901.10968
94. Goff LKL, Yaakoubi O, Coninx A, Doncieux S (2019) Building an affordances map with interactive perception. arXiv preprint arXiv:1903.04413

Publisher's Note