The human in the loop Perspectives and challenges for RoboCup 2050

Alessandra Rossi[1,2] · Maike Paetzel-Prüsmann[3] · Merel Keijsers[4] · Michael Anderson[5] · Susan Leigh Anderson[6] · Daniel Barry[2] · Jan Gutsche[7] · Justin Hart[8] · Luca Iocchi[9] · Ainse Kokkelmans[10] · Wouter Kuijpers[10] · Yun Liu[11] · Daniel Polani[2] · Caleb Roscon[12] · Marcus Scheunemann[2] · Peter Stone[8,13] · Florian Vahl[7] · René van de Molengraft[10] · Oskar von Stryk[14]

Abstract

Robotics researchers have been focusing on developing autonomous and human-like intelligent robots that are able to plan, navigate, manipulate objects, and interact with humans in both static and dynamic environments. These capabilities, however, are usually developed for direct interactions with people in controlled environments, and evaluated primarily in terms of human safety. Consequently, human-robot interaction (HRI) in scenarios with no intervention of technical personnel is under-explored. However, in the future, robots will be deployed in unstructured and unsupervised environments where they will be expected to work unsupervised on tasks which require direct interaction with humans and may not necessarily be collaborative. Developing such robots requires comparing the effectiveness and efficiency of similar design approaches and techniques. Yet, issues regarding the reproducibility of results, comparing different approaches between research groups, and creating challenging milestones to measure performance and development over time make this difficult. Here we discuss the international robotics competition called RoboCup as a benchmark for the progress and open challenges in AI and robotics development. The long term goal of RoboCup is developing a robot soccer team that can win against the world's best human soccer team by 2050. We selected RoboCup because it requires robots to be able to play with and against humans in unstructured environments, such as uneven fields and natural lighting conditions, and it challenges the known accepted dynamics in HRI. Considering the current state of robotics technology, RoboCup's goal opens up several open research questions to be addressed by roboticists. In this paper, we (a) summarise the current challenges in robotics by using RoboCup development as an evaluation metric, (b) discuss the state-of-the-art approaches to these challenges and how they currently apply to RoboCup, and (c) present a path for future development in the given areas to meet RoboCup's goal of having robots play soccer against and with humans by 2050.

Alessandra Rossi, Maike Paetzel-Prüsmann and Merel Keijsers have equally contribution to this work. The other authors are in alphabetical order.

✉ Alessandra Rossi

1   University of Naples Federico II, Naples, Italy

2   University of Hertfordshire, Hertfordshire, UK

3   Independent Researcher, Zürich, Switzerland

4   John Cabot University, Rome, Italy

5   University of Hartford, Hartford, USA

6   University of Connecticut, Storrs, USA

7   University of Hamburg, Hamburg, Germany

8   University of Texas at Austin, Austin, USA

9   Sapienza University of Rome, Rome, Italy

10   Eindhoven University of Technology, Eindhoven, The Netherlands

11   Zhejiang Lab, Hangzhou, China

12   Universidad Nacional Autonoma de Mexico, Mexico City, Mexico

13   Sony AI, Tokyo, Japan

14   Technische Universität Darmstadt, Darmstadt, Germany

1 Introduction

Robots are complex systems that require hardware and software components working together, supporting and at times compensating for each other.[1] From a scientific perspective, these requirements make reviewing progress in robotics difficult: How does a robot that folds clothes compare to one that finds human victims in a disaster scenario? How can we measure a fleet of robots organizing a warehouse against a single robot watering a plant? One could measure the stability of the shell material, the accuracy of the computer vision components, or the precision of the actuators. This, however, only provides us with a partial picture of the robots' performance. What is missing is the evaluation of the system as a whole, and how the components work together in solving a specific task.

RoboCup (RC), one of the largest annual robotics competitions, is aimed at providing a benchmark for such evaluations: Robots from all over the world compete in several leagues offering unique challenges with well-defined sets of rules. The major leagues range from the *@Home league*, in which robots are tasked with household chores and interact with humans in social environments, to the *Rescue league*, in which robots need to find victims in realistic disaster scenarios, to the *@Work and Logistics leagues*, in which robots assemble objects or optimize a production chain. The *Soccer leagues* are the most well-known, as it is their goal to beat the best human soccer team by 2050. In the soccer leagues, teams of robots of different sizes and hardware configurations play soccer against each other. While these leagues address different scientific aspects, they are united in their aim to foster scientific development, by presenting increasing yearly challenges and favoring scientific collaborations between the different leagues.

Given the goal of creating robots that can beat humans at soccer, one might reasonably ask, "how will we know if they can?" On the surface, this seems like an easy question to answer – organise a soccer game: robots against the current FIFA World Cup Champion and if the robots win the RoboCup challenge has been met. However, it may not be quite so simple. Even *if* the World Cup Champions were to agree to play such a match, what would the rules be? If we built a "robot" the size and shape of a goal and place it in the goal (an invincible goalie), or if we built a robot that could place the ball in a cannon, and then shoot it towards the corner of the goal at high speed (an invincible attacker), nobody would be particularly impressed.

While the current FIFA rules do not place any restrictions on the size, shape, or "actuators" of the players, these are examples of issues that would need to be considered prior to assessing whether robots are better than people at soccer. A

few similar issues arose in prior contests of humans versus machines, such as DeepBlue vs. Gary Kasparov at Chess, and AlphaGo vs. Lee Sedol at Go. The rules of these purely cognitive challenges, however, were relatively straightforward to define – the computers could use any means to decide what next move to make, and if they won, they were better than their opponent at the game in question. Soccer, instead, has cognitive and physical challenges. It is much less straightforward to define rules such that if the robots were to win, people would generally agree that robots are better than people at "soccer." Thus, this question gets at a somewhat philosophical issue: what is the *essence* of soccer? Is it still soccer if one player can run twice as fast as all the other players, or if they can score without passing, or if the players are all controlled by a single program? These questions need to be answered so that we can ensure that the robots are really playing *soccer*.

Stone, Quinlan and Hester considered this question more than a decade ago in the Chapter "Can Robots Play Soccer?" from a popular philosophy book called "Soccer and Philosophy: Beautiful Thoughts on the Beautiful Game" Stone et al. (2010). They laid out a set of restrictions on the form and capabilities of individual robots to ensure that they will not be too fast, too strong, or too precise to be considered "human-like". They also considered restrictions on team composition and communication, such as ensuring that the teammates have at least somewhat differing capabilities from one another, and that they can only communicate via human-perceptible sounds. And finally, they considered restrictions on coaching to place the robot coach on a similar footing as human coaches.

When looking at the abilities of the robots competing in the RC soccer leagues today, these considerations seem rather futuristic, given that the bipedal human-sized robots are so unstable and fragile that they need a human robot-handler walking behind them to catch them when they fall. In a recent survey (Paetzel-Prüsmann et al. 2023) which we distributed to students, researchers, and professors engaged in RC activities, locomotion was identified both as the most important and the most difficult research area when preparing to play against humans in 2050. Other areas that were considered of great importance and difficulty were awareness of the environment, robustness, and decision-making. While scientific progress in these areas can be seen as a pre-requisite to the more future-looking considerations made by Stone et al. (2010), these responses also indicate that many researchers are currently overlooking the importance of the human in the loop as they are designing robots that can play against humans. Safety ranked fifth in perceived importance and difficulty to achieve, while HRI was considered quite challenging (ranked 6/12), but less important (ranked 9/12), and natural-language understanding, a key aspect in creating fair communication, was ranked last in importance.

---

[1] Writing led by Peter Stone and Maike Paetzel-Prüsmann.

This article can be seen as a natural revision and extension of the work by Stone et al. (2010), fleshing out the desiderata they laid out in more detail. We aim to give an overview of the state-of-the-art in robot hardware, cognition, behavior, and human-robot relational dynamics, as well as point out current challenges that robotics researchers are facing. The article however goes beyond these contemporary issues by identifying future challenges for the goal of 2050, and aims to prepare the research needed to create the robots that will eventually play with and against humans.

The remainder of this article is structured to provide state-of-art and current open challenges in the following areas: Sect. 2 discusses hardware and motion design; Sect. 3 presents cognitive capabilities and robot behaviors, including perception; Sect. 4 deliberates the complex dynamics in humans-robots soccer games; and Sect. 5 summarises the identified future research directions in unstructured HRI.

2 Hardware requirements

Robots that play soccer come in very different shapes and sizes. In the MiddleSize League (MSL), robots use wheels to get around the field and Lidars to create a three-dimensional map of the environment. In the Humanoid League (HL), robots are constrained to human-like locomotion and sensing. Scaling the robots to human size (which is likely necessary to match the running and kicking speed of humans) comes with unique challenges in the robots' hardware design and motion control, many of which are unsolved to date. In this section, we give an overview of the current state of the art in hardware design and motion control for human-like soccer robots, and discuss a road to a more stable and safe robot design in the future.

2.1 Human-sized robot design

In order to meet the RC challenge and more generally unlock the potential of humanoid robots,[2] numerous research groups have been working on the hardware required for locomotion. For example, Honda Corporation developed the humanoid robot ASIMO (Sakagami et al. 2002), which has 34 DoF, is 120 cm tall, weighs 43 kg, and can kick a ball and shoot a goal. Boston Dynamics developed Atlas, a 150 cm tall research platform designed to push the limits of whole-body mobility. It has 20 DoF and weighs 80 kg. Atlas' advanced control system and state-of-the-art hardware give the robot the power and balance to demonstrate human-level agility.[3] Georgia Institute of Technology developed humanoid robot DURUS which is 180 cm tall and weighs 79.5 kg, and which

is one of the most efficient robots when it comes to energy consumption for walking (Reher et al. 2016). Finally, the Technical University of Munich developed the humanoid robot LOLA with 25 DoF, which is 180 cm tall and weighs 60 kg (Buschmann et al. 2012).

2.1.1 Open challenges

Although a range of different humanoid robots have been developed, the design of a more powerful robot body remains a prerequisite for the RC 2050 goal. Robot configuration has always been one of the biggest challenges in robot design, with the main decisions revolving around the selection of DoF and the arrangement of the drive mechanism. Robot soccer requires a very flexible robot body that has the ability to walk, run, throw the ball, stand up, as well as a variety of other humanoid movements. First, this requires the robot to have sufficient DoF. For the humanoid robots currently participating in RC, there is a minimum of 6 DoF per leg, 3 per arm, and 2 in the neck joint, amounting to at least 20 DoF for a full robot. However, 20 is far from sufficient for more complex movements, which will be needed for competition with humans. Unfortunately, increasing the DoF leads to a dramatic increase in robot design complexity, control difficulty, and cost.

In terms of drive mechanism arrangement, the robot leg mechanisms of LOLA, ASIMO, and DURUS are designed in a very inspiring way. For example, the motor position of the knee and ankle joint of LOLA are improved by adding tandem and parallel drive mechanisms (see Fig. 1). In this way, the inertia of the robot's legs is significantly reduced, it is more humanoid, and easier to control.

Finally considering motor power, existing motors are still far from being comparable to human muscles in terms of energy, efficiency, and torque output density. Among the existing motor-driven robots, the fastest humanoid robot known to be able to run is ASIMO, which can reach a maximum speed of 9 km/h (Sakagami et al. 2002). The fastest known human running speed lies almost 5 times higher at $44.72 km/h$, which was achieved by Usain Bolt. At the same time, the power density of the ASIMO motor solution cannot support the completion of some highly explosive movements, such as parkour and backflips shown by ATLAS2.[4] As a result, a number of research institutions are now turning their attention to hydraulic solutions, such as IHMC, which is developing the full-size humanoid robot Nadia.[5] The difficulty with hydraulic drive solutions, however, is the lack of marketable integrated hydraulic drive units

---

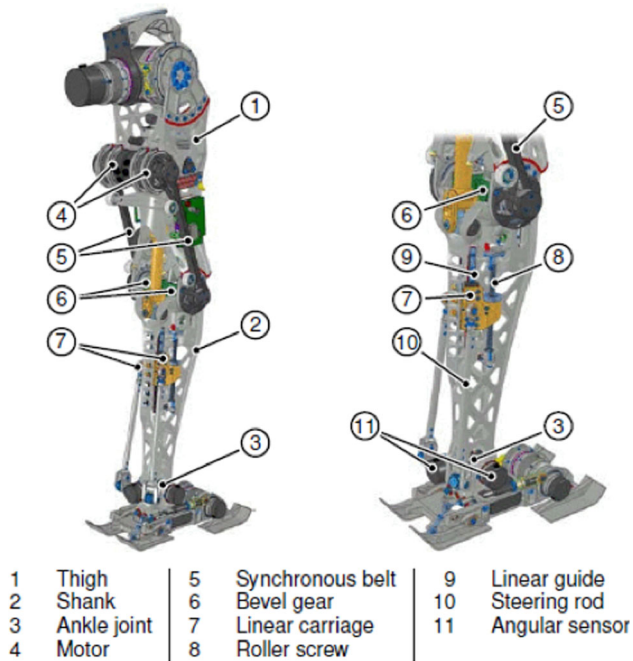|					
1	Thigh	5	Synchronous belt	9	Linear guide
2	Shank	6	Bevel gear	10	Steering rod
3	Ankle joint	7	Linear carriage	11	Angular sensor
4	Motor	8	Roller screw		

Fig. 1 Structural of robot LOLA's leg [10]

and the R&D costs which may be prohibitive for general research institutes and universities. Therefore, most research institutions and universities are still considering the use of electric motors to design relatively lightweight bipedal robots through weight reduction and non-full-size arrangements. Currently, many bipedal research institutions are studying Electro-active Polymer artificial muscles (Kim and Kim 2023) in the hope of obtaining drive units that are comparable to human muscle capabilities. This research direction could prove to be very interesting.

With the development of new drive units, such as carbon nanotube yarns, robot joints can now produce up to 85 times more force than human muscle (Lima et al. 2012). Furthermore, the capacitive dependence of artificial muscle drive performance has been solved which helps designing high-performance drivers with non-toxic, low drive voltages (Chu et al. 2021). The physical performance of future robots is thus expected to break through rapidly, and more and more robust robots will emerge to achieve the goals of RC 2050.

2.2 Motion engine

The HL and Standard Platform[6] League (SPL)[7] both require humanoids that use bipedal locomotion to compete in the

RC competition. In both leagues, there have been successful approaches to enable robust and dynamic walking on mostly flat artificial grass. Herein, we consider bipedal locomotion to be a subset of all robot motion, including actions such as standing-up or kicking. RC has proved as a useful test bed for the current applied state of robot motion in a challenging environment, where humanoid robots have been able to successfully walk on artificial grass with little or no falling. Most approaches within the HL and SPL utilize zero moment point (ZMP) based step planning or computing walk trajectory. Although robust, the humanoids are yet unable to run, jump, stand-up or operate on non-flat terrain using ZMP-based motion and it does not appear to be a suitable candidate for a generalised motion engine (Vukobratovic and Borovac 2004).

Realizing a dynamic bipedal walk for robots is very difficult, and this is why most approaches have "typically been achieved by considering all aspects of the problem, often with explicit consideration of the interplay between modeling and feedback control" (Reher and Ames 2020). This is also true for RC where prominent candidates explicitly compute the center of mass using the ZMP (Czarnetzki et al. 2009), or use a central-pattern generator (Behnke 2006) to compute a suitable walking trajectory for the robot. Teams then use the robot's sensory input to satisfy the computed trajectory. These methods need extensive parameter tweaking and rely on a growing number of assumptions about the environment. A popular assumption to render the methods computable is an approach that assumes a mostly flat and even terrain. The environment complexity will further increase the parameter space with moving towards a real-world soccer pitch, and with humans entering the competition as players.

2.2.1 Open challenges

When considering open challenges, we first propose agent self-modelling, where agents should be able to model their own non-linear control with meaningful abstraction from the environment. We expect this process to somewhat resemble a baby learning to walk, a process that is often linked to curiosity and intrinsic motivation (Scheunemann et al. 2022). This would include complex control variables, such as actuator behaviour under load, temperature, voltage and wear for example, where the behaviour is expected to change over time. We propose the challenge for agent local-world modelling, where the agent builds a model of the local environment abstracted from its self-model, to allow future planning of movements. This would include other robots and humans in the near vicinity, nearby terrain the robot is likely to interact with, and other useful observations.

Using mechanisms that allow robots to self-model their environment and adapt to unknown situations opens new issues. Teams in RC typically use algorithms that are compu-

tationally inexpensive due to the full autonomy constraints, whilst research labs want to use motion generations with a high level of control. There is evidence that an agent's ability to create intelligent behaviours depends on the sensory motor loop, where an agent tends to benefit from embodiment due to environment complexity (Kubisch et al. 2011). Intrinsic motivation (IM) has been used to feed reinforcement learning for motion acquisition in simulation (Peng et al. 2018), but it also shows the learning of motion skills on real robots (Schillaci et al. 2016). IM has been shown to produce perceivably social motion behaviours, suggesting that it is suitable for complex human-robot interactions, such as a soccer game (Scheunemann et al. 2019).

2.3 Discussion and conclusion

Designing robots that are comparable to humans in their speed of locomotion, stability and robustness remains a major open challenge. Whilst the current approaches to motion in humanoid robotics have proven to be successful in more controlled scenarios, it remains to be seen how these solutions will scale to more complex real-world environments, where there is a larger number of unknown complex variables. These problems are not unique to humanoid soccer players. There is a considerable effort to get robots into dynamic environments, where most successful implementations have reduced motor capabilities, such as smart vacuums or toys. One reason we may see a reduced DoF for commonly deployed platforms is due to the cost and performance of actuators. We suggest that even with low-cost, high-performance actuators, robots are more generally still missing the motion framework for the required control in dynamic environments.

Another challenge to the design of robots that goes beyond the application of robot soccer is the development of bodies that are robust enough to survive and recover from a fall with minimal damage to the hardware system. Especially in social environments, human-like bodies are both desirable from an interaction point of view as from a locomotion perspective (as human environments are often designed to suit their bodily abilities well). However, even robots smaller than human size still risk permanent damage when falling down. Moreover, the potential threats to a robot's bodily integrity don't stop at the damage from a fall: They can also break small parts like a finger when getting tangled into another robot, human, or an obstacle in the environment. Apart from the motors and the outer shell of the robot, its inner parts can face failures like short circuits and cable breaks. While shielding these parts from extraneous interference can help to prevent some of the failures, it also makes it difficult to repair them on the spot.

One potential solution in making robots more robust could be cover materials that are harder to physically break. Especially when combining these with powerful motion engines,

however, serious safety concerns arise for human players. One potential solution to this problem is the implementation of advanced safety procedures in the motion control loop, as is already standard in industrial robots. These robots recognize and stop a collision between their hardware and an obstacle within milliseconds, which minimizes their physical impact on a potential human getting in their way. While this works well for robots that interact with humans within a constrained space, robots that could potentially fall onto a human or find themselves in an otherwise unstable position need to find different strategies to minimize damage. Another potential solution to ensure human safety independent of the current physical state of the robot could be the application of materials and joints currently researched in the area of soft robotics. These materials require further advancement for being robust enough to work in an environment with as much physical contact as in robot soccer. As was pointed out by many researchers participating in our survey, hardware and motion control is still one of the main factors that needs to advance in order to play with or against human soccer players. However, as we will see in the next section, there are still many open research questions that can be tackled independently of the improvements in the robots' hardware.

3 Cognitive capabilities & robot behavior

During a soccer game, robots need to proactively plan, manage and execute their playing goals – both collaborative/cooperative and for personal gain – while modeling their surroundings including human players. Therefore, robots need to be able to formulate purposeful conscious observations, build their knowledge of the context and the agents (human or machine) in the environment, and both plan and act accordingly (Rossi et al. 2020a). Humans are able to naturally communicate among each other using verbal and non-verbal signals. However, robots' ability to generate verbal and non-verbal expressive behaviors (such as natural spoken language, gestures, affective responses) does still not match their capability of understanding the situational context. This is particularly relevant if we want to simulate cognitive capabilities based on human-like senses, as is the case in the HL. This section presents an overview of existing techniques based on basic human-like abilities such as vision and audio sensors to build a robot's awareness, and subsequently provides future scientific challenges to be addressed.

3.1 Audio in human-multi-robot systems

There is a growing interest in the use of auditory perception in robotic systems (Rascon and Meza 2017) which has been shown to be an important part of the interaction scene

between a robot and a human (Meza et al. 2016).[8] In fact, it has been a part of other service robotics competitions (such as *RC@Home*) for several years (RoboCup@Home Technical Committe 2024). In terms of a human-robot soccer match, there is an important amount of relevant information that can be extracted from the auditory scene, such as the location and intentions of the human adversaries, as well as the robot's teammates; even the audience noise during the match can be integrated in the robots' decision making process (Antonioni et al. 2021). Since audio can be perceived in an omnidirectional way, it is well suited to complement information that is extracted by other means (e.g., vision) which can benefit strategy planning and safety.

Pragmatically, auditory perception in robots (or robot audition) entails three main tasks: (1) localizing the sound sources in the environment given a frame of reference (usually, with the robot at its origin), (2) separating the audio data of each sound source from others such that each sound source has its own audio channel, and (3) classifying the sound source from each sound source channel. These three tasks are typically carried out in a serial manner, since the location of a sound source can be used to separate it from the captured audio mixture into its own channel. Once separated, a mono-source classifier can be used, instead of relying on far more complex techniques that carry out multi-source classification.

In terms of localization, the ODAS library (Grondin & Michaud, 2019) provides good localization performance, while requiring a relatively small software footprint. A deep-learning approach (Nakadai et al., 2020) outperforms it, but requires more computational power. It is also worth mentioning a few-microphone approach that can outperform them in certain scenarios (Gato, 2020). A beamforming-based approach (Grondin et al., 2020) requires knowing the location of the sound sources but can run in relatively light hardware. A deep-learning approach (Liu et al., 2020a) provides an important jump in separation performance in real environments, although it requires an important amount of computational resources. A hybrid approach (Maldonado et al., 2020) provides a middle-ground between the acceptable performance and low computational requirements. In terms of classification, and particularly speaker identification, a deep-learning-based approach (Xie et al., 2019) can carry out this task "out in the wild", but requires more computational power. A "lighter" approach (Vélez et al., 2020) provides lower-but-still-acceptable performance. It is also worth mentioning the HARK library (Nakadai et al., 2010), since it has been a tried-and-true audition workhorse for more

than a decade, and carries out all three auditory tasks in conjunction.

3.1.1 Open challenges

The challenges proposed here will go through several iterations, with rising difficulty as time goes on. The initial version is to estimate and track the relative direction of human adversaries in the near vicinity of the robot. Recordings of human adversaries can be used, or actual human volunteers, vocalizing specific utterances that can be expected to be heard during a soccer match, such as "I'm open", "pass me the ball", etc. The difficulty can be later increased by: a) using shorter utterances, such as "hey", or non-linguistic vocalizations (grunts or mono-vowel yelling); and b) activating multiple human sound sources at the same time. The location of each human sound source can be used to quantify the precision of the robot's localization performance.

In a subsequent version, the location information of the human sound source is to be integrated with the audio estimations of other robots, as well as their available visual data, to provide a shared robust localization of the human adversaries. This is proposed to eliminate the need for external sensing, which is typically used in indoor robot-robot matches, but is impractical to use in an outside environment. A robot will not be able to sense (either acoustically or visually) a human adversary on the other side of the field, but a nearby robot teammate should be able to. Thus, the robots themselves should aim to create an ad-hoc network through acoustic means to share the information perceived from their immediate surroundings to the rest of the robots. The acoustic parameters of the robot vocalization should be in the human-hearing range, so that it falls within the restrictions set by Stone et al. (2010). To evaluate the efforts of creating an ad-hoc acoustic network, a version of the challenge can be carried out using mobile human sound sources which will no doubt introduce localization errors in the estimation carried out by one robot. Thus, redundancy between the estimations of several robots should surmount these issues, and will be evaluated as such. To transition between using a common wireless network (e.g. WiFi or Bluetooth) and the acoustic network proposed here, a version of the test can simulate a situation where the wireless network "fails" by manually disabling one or more of the wireless sensors/antennae that the robots use to communicate with each other, and forcing them to use audio as a backup to continue such communication while a time-out is called. It is important to mention that such type of communication should not be required to be speech, and should be accepted in any form as long as the robots are able to communicate the relevant information to each other acoustically, without requiring wireless sensors, and without causing hearing discomfort to the human adversaries. However, it is also important to consider that not

---

[8] Caleb Rascon; RC@Home Technical Committee 2012–2017, Executive Committee 2017–2019, Universidad Nacional Autonoma de Mexico, Mexico.

using speech will make the robots' behaviours and intentions entirely non-transparent and impossible to infer for humans. As a consequence, human players will be less inclined to accept and trust to play with robots Nesset et al. (2021), and their interaction will be negatively affected and induce people to toss robots away de Graaf et al. (2017).

Other types of audio-based human-robot interactions can also be evaluated, such as making the robot verbally announce to the human referee if a human adversary made an illegal move (such as a foul or violating the offside rule).

In the final version of this challenge, the robot assesses the humans' intentions and strategies via the analysis of the paralinguistic characteristics of the vocal utterances emitted to each other during the match, such as prosody, pitch, volume, and intonation, as well as the sound of stepping patterns. Professional players are well aware that yelling out a phrase such as "pass the ball" announces to their adversaries their intent. However, human adversaries may not be aware that they emit some vocalizations in critical moments even when not meant to be (a deep breath before a sprint, a small sigh when a play didn't go as planned, a slight wail when they are free to receive the ball, etc.), which can be used to the robotic team's advantage. This can also be used for the human team's safety. For example, if a human would yelp right before crashing into the ground or another agent, or if they would scream when they are hurt. In addition, other auditory cues can be used that are not specific to speech, specifically that of the sound of human feet running or walking in the grass. Recordings of human volunteers during human-human matches can be used to evaluate the robot's ability to recognize such activities, and communicate them to the rest of the team to be used for strategy planning and safety precautions.

The final outcome of a robot team that is able to solve all the proposed challenges is the localization and intention estimation of each human adversary through auditory perception without the use of external sensing.

3.2 Robot vision

Computer vision techniques have been used in many domains such as medical image processing (Ronneberger et al. 2015),[9] autonomous driving (Janai et al. 2020), and robotics (Jamzad et al. 2001) for several years. Computer vision enables autonomous robots to visually perceive their environment and offers a challenging testing ground for applied computer vision in complex and dynamic real world scenarios.

Currently computer vision used in humanoid robotics (and especially in the RC context) is transitioning from hand-crafted model-based algorithms (Fiedler et al. 2019) to more robust and powerful data-driven ones (Vahl et al. 2021). The

model-based approaches include conventional methods like the usage of color lookup tables or color clustering for simple segmentation tasks (Freitag et al. 2016), Hough lines for line fitting (Szeliski 2010), or filtering in the frequency domain to generate regions of interest for later classification. Currently available data-based approaches include simple CNN classifiers which classify candidates generated by a model-based approach. More complex data-based methods include the YOLO architecture (Redmon et al. 2016) which directly detects objects in an image, or architectures like SegNet (Badrinarayanan et al. 2017) or UNet (Ronneberger et al. 2015) which generate pixel precise segmentation maps. Data-driven approaches such as convolutional neural networks (CNNs) are very powerful in terms of accuracy, robustness to noisy data, and the overall generalization. But they are computationally expensive and hard to modify or debug after the training. The data-driven approaches need large amounts of training data. This is an issue for many domains, but in the RC domain large quantities of annotated data for supervised learning are available as part of open data projects (Bestmann et al. 2022). While very powerful data-driven approaches exist, real-time constraints are still a limiting factor on embedded platforms like the autonomous robots used in the RC domain. Due to this limitation only subscale versions of models like YOLO and nearly no Region based Convolutional Neural Networks (RCNNs) (Girshick 2015) or Vision Transformers (Liu et al. 2021) are used. While being also computationally expensive, frameworks like OpenPose (Cao et al. 2019) enable 2D and 3D human pose estimation which is a growing field of interest in the humanoid RC domain.

3.2.1 Open challenges

One major challenge of the computer vision system is to perceive the state of the whole environment in a short amount of time. This requires the fast and reliable detection of various small objects in a large image space. For humanoid robots in the soccer context, this means that the comparably small soccer ball is one of the most important items that must be localized from a maximum distance of over 100 ms. On the other hand, a wide field of view is required to minimize the head movement needed for the observation of multiple targets. Head movements take time and limit the ability of tracking (e.g. the position of the opponents) which is a safety concern when playing against humans.

Adaptive resolution, which is dynamically changing the resolution of parts of images, could result in an efficient way of handling very high resolution images (Mnih et al. 2014). Various fast region of interest proposal methods or attention-based mechanisms could be used for such a task and need to be evaluated in the RC context.

As discussed before, there is a large amount of environment information that is critical for both strategy building

---

[9] Florian Vahl and Jan Gutsche; HL Team Hamburg Bit-Bots, University of Hamburg, Germany.

and humans' safety, and which can be transferred with and gathered from audio-based data. However, it is unreliable for long range communication, since the energy of acoustic signals drops faster compared to vision-based signals. Thus, gestures are essential for intuitive non-verbal long range communication and are therefore used by humans in everyday situations as well as in many different sports. As the soccer field size in the RC competition grows and the wireless communication gets more restricted, it is a feasible way of communicating with other players, referees or the trainer. Understanding gestures of the opposing team brings also tactical advantages. A more general version of the gesture recognition is the pose estimation. The robot's behavior could use the pose of opponent's legs and torso to outplay them or more importantly avoid injuries among the opponents when playing in proximity to them. There are state-of-the-art pose estimation frameworks, but further research regarding the integration into a dynamic gameplay and the reliability and safety impact of such approaches should be done. A classification of facial features expressing emotions, exhaustion, or the intentions of an opponent could also be used by the robot's behavior when playing against humans. There are approaches, such as FER (Goodfellow et al., 2013), which could be adapted to this specific domain.

We expect that the robotic soccer games will be played more dynamically in the future. Such a play style includes faster movements, higher passes and less predictable surroundings. This implies that visual processing needs to be faster while remaining reliable. Currently, most of the RC robots do not feature any depth sensing, because LIDAR sensors are not allowed in the HL as there is no equivalent human sense (HL Laws of the Game 2019/2020). Instead, object's relative positions are estimated based on the assumption that it is located on the same ground plane that the robot is standing on. This approach will no longer work when objects (e.g., the soccer ball) leave the ground. We therefore assume that a combination of both stereoscopic imaging for accurate short distance depth estimation and a quasi-monocular method for long range measurements as well as featureless regions is needed (Smolyanskiy et al., 2018). This is based on the fact that the distance between the cameras is small and the angular differences get too small for far away objects.

On the way out of the laboratory and onto the field, we also encounter environmental effects such as natural light, which can drastically change in brightness, cast shadows, or glare the robot's vision system. Other effects include disturbances due to rain, snow or dirt both in the air and on the ground. As long as these disturbances are included in the datasets, data-driven approaches appear to be robust against them to a certain degree. See figure 2 for an example.

Fig. 2 Learning based approach [full-size YOLOv4 (Bochkovskiy et al., 2020)] in natural light conditions. Source (Bestmann et al., 2022)

3.3 Discussion and conclusion

The cognitive capabilities implemented in the state-of-art robots allow them to elaborate static and dynamic scenarios that do not take into account people's fast and complex reactions. Perceiving as much information as possible about the state of the other players is crucial to avoid injuries and damage.

Multiple senses, such as hearing and vision perception, could be fused to improve robots' perception and decision making process of the environmental context.

While the field of computer vision made large steps in the past years, there are still open challenges. For example, robots will need to be able to adapt to natural conditions of the weather and illuminations as well as expand the amount of observed information to include detailed information regarding enemy poses which are crucial for a dynamic and safe behavior. Learning based approaches are promising for these purposes, as they perform well in many domains and are distantly related to the way humans solve these challenges.

Moreover, in a such dynamic and near vicinity context, we can expect bidirectional communications. We want to optimize the ability of robots to communicate with each other, as well as their ability to infer the humans' intentions, through sounds, natural languages and non-verbal modes. However, it is important that robots still perform transparent motions and behaviors that can be clearly recognized by the humans (Holthaus & Wachsmuth, 2021).

4 Human-robot relational dynamics

While anyone who watched the most recent RC matches will agree that playing against human teams is still a far way off, this is the eventual goal. Playing against and with humans

opens up new challenges and dilemmas related to the HRI, for which no simple solution may exist. For example, a delicate balance will need to be struck between ensuring the robots are safe for humans to play with (and are *perceived* to be so by the human players so that they will actually agree to play) on the one hand, while on the other hand ensuring that the robots have enough opportunity to win so that they will give the human team a run for their money. The following section aims to highlight some of the most pressing issues and illustrate how they create a paradox that may prove to be unsolvable.

4.1 Playing against humans

Every year after RC's MSL final,[10] the fresh world champion demonstrates its soccer skills against a team of human players. This annual match is an exposure of the worldwide state-of-the-art in human versus robot soccer playing (Soetens et al., 2015). The first goal against the human team was scored in 2014 and multiple goals have followed since. The human team, consisting of RC Trustees, continues its winning streak ever since. In RC, the MSL is well suited to the 'robot versus human' soccer play due to both its focus on robot teamwork, and its accessibility for humans by using the standard size FIFA ball and by its playing field dimensions.

The regular matches during the tournament are however without direct human interaction. The human referee team interfaces with the robots through a league-specific Refbox application (Dias et al., 2020) on a computer that is connected with both teams.

4.1.1 Open challenges

Various challenges have been identified towards a more sophisticated human interaction in the league. As a first step, robots have to be safe, not harming humans or themselves. As a second step, anticipating human behavior and, thirdly, cooperation can be aimed for. These three steps will be treated in the remainder of this section.

a) Safety The first challenge in a more sophisticated human interaction in the MSL is the safety of the human players. Ensuring human safety can be achieved both by considering the robot's design and by considering its behavior. Currently, the robots in the MSL must not exceed the regulatory dimensions of $50 \times 50 \times 80$ cm and weigh no more than 40 kg (MSL Technical Committee (2020)). The robots can achieve speeds of up to 4 m/s without controlling the ball (Soetens et al., 2015). Even though collisions are to be avoided at any given moment, a collision with a human with the aforementioned

weight and speed should not result in an injury. In a collision of 0.01 s, the kinetic energy of the robot, 160 $kg.m.s^{-1}$, would result in an interaction force of 16 kN.

The weight of the robot is mostly constituted by the weight of the electronic solenoid used to shoot the ball (4.5 kg (Meessen et al., 2010)), the frame of the robot and the motors used. Reducing weight is one of the possible solutions to improve the safety of human players. Within the league, however, most robots weigh close to the maximum. With the state-of-the-art in sensors, actuators and materials, it is difficult to have competitive robot specifications (e.g., driving speed, kicker force) with reduced weight. Adding soft material on the outside of the robot, i.e. a bumper, and thus extending the duration of the collision, will result in smaller interaction forces and will enable safe feedback control actions. A robot should detect a collision via its compliant skin and react accordingly. Passive compliance should prevent initial damage, while further damage should be mitigated through active compliance. Even though the rulebook states a bumper has to be included in the design of the robot (MSL Technical Committee, (2020)), the specifications are based on robot-robot collisions, which will result in too high interaction forces for humans.

Another approach to increasing the safety of human players is behavioral; i.e. to prevent high-speed collisions. For this to work, the robot has to detect the human. The current obstacle detection of most MSL teams uses a combination of a camera and a parabolic mirror, often referred to as Omnivision. This catadioptric vision system enables a 360° view with a range of up to 11 metres (Dias et al., 2020), see Fig. 3. The camera is pointing upwards and looking into a downwards-mounted parabolic mirror, hence it is impossible to detect objects above the height of the robot (80 cm). This not only hampers the detection of the ball once it is airborne and above the height of the robot, but also the detection of humans. Thanks to the increase in available computing power, many teams equip robotic players or goalkeepers with forward-facing cameras such as Kinect cameras (Dias et al., 2020) and use those as either main camera systems (Schreuder et al., 2019) or complementary systems.

b) Anticipation The second challenge is to play against human players and to be able to anticipate their actions. The latter will require the detection and tracking of the human's position on the soccer field. To detect opponent robots, most teams use the aforementioned color segmentation and vision system. For tracking, most teams filter the detections from the catadioptric vision system using extended Kalman filters or particle filters to be able to handle false positive detections, occlusions and to estimate the velocity of the opponent robots (Dias et al., 2017). These filters typically employ constant velocity models for the opponent robots.

The view from a catadioptric vision system will not be optimal, if sufficient at all, to provide accurate detections of

---

Fig. 3 Image captured from the catadioptric vision system. A human is observed in the top left corner of the image and a MSL robot in the top right corner

a human and estimate its velocity. However, once a qualitative detection has been established for humans, for example using forward-facing cameras, similar filters can be employed to track opponent players and estimate their velocity. The possibilities for qualitative detections of humans have considerably changed over the last few years due to the use of the Kinect camera, state-of-the-art image-based human pose detection software (Cao et al., 2021), and other classifiers. These detections could be enhanced by using human motion patterns or gait patterns to provide better detections and/or estimate their velocity (Cao et al., 2021). In Dolatabadi et al. (2020), for example, the output of OpenPose is combined with a model for the position and velocity of the hip, knee, and ankle in typical human motion patterns, resulting in better tracking of humans.

Aside from technical questions, this also raises the question of to what extent collision should be anticipated in a human versus robot match. In human soccer, collisions frequently occur when opponents try to gain control over the ball. An open question thus remains, to what extent should a robot prevent collisions while maintaining a strong competitive intercept action?

c) Cooperation The third challenge is to eventually cooperate with human players. The teams of robots currently communicate information, such as detections, planned actions, and strategies, over a WiFi connection. A team of humans communicates by means of speech, gestures (Lim et al., 2017) and other subtle non-verbal cues. Even though communication through gestures was introduced in the MSL as a means to coach the robots in-between plays, this has yet to be attempted in dynamic play. Coaching, allowed by

the rules, up to now included the use of QR codes shown by humans, voice coaching, and gesture coaching.

With the high-paced developments in the MSL, the league is likely to prepare itself for the first competitive or collaborative matches with humans. The increasing attractiveness of the league combined with this grand challenge steers developments into this direction.

4.2 Value-Driven Players

When considering the scenario where a robot team takes on a human team in a soccer match,[11] it is important to realize that the rules of the soccer game itself comprise only a subset of obligations that the robot has towards its human opponents. When circumstances warrant, say when an injury or some other incident not specifically covered by the rules occurs, other duties are likely to be added to or even take priority over the rules of the game. For instance, in case of an injury to a human player, the robot may be required to stop playing and prioritize providing whatever assistance it is capable of. Given this, a number of research questions arise, such as: 1) In which circumstances do the rules of the game no longer apply and how might these be discerned by an AI system? 2) What other obligations does a robot player have towards its human opponents and, when they conflict, how might the strongest obligation be determined? 3) How might a robot system be designed to meet these obligations in such circumstances? These and other such questions comprise an ethical dimension of the game, and provide an opportunity for research in this domain to contribute to the greater concerns regarding the ethical behavior of artificially intelligent agents operating autonomously in the world.

Although literature pertaining directly to the goal at hand is scarce, there have been efforts in related areas such as the ethics of sport (e.g., Boxill 2002) and machine and robot ethics (e.g., Anderson and Anderson 2011).

4.2.1 Open challenges

Central to ethical behavior in every domain are ethically-relevant features, duties to minimize or maximize these features, and a set of principles that prescribe which duties will prevail if they are in conflict. Ethically-relevant features may have a positive value, like sportsmanship; or a negative value, like harm. It is incumbent upon agents acting in any domain to not only minimize ethically-relevant features that have a negative value, but also to maximize features with a positive value. These considerations comprise the agent's duties in that domain. Duties are likely to

---

[11] Michael Anderson and Susan Leigh Anderson; MachineEthics.com, Dept. of Computer Science,, US; Dept. of Philosophy, University of Connecticut, US.

be context dependent. That is, which duties pertain will be contingent upon the current circumstances and the actions available to the agent within those circumstances. Furthermore, these circumstances will also determine which actions satisfy and/or violate these duties, as well as by how much. Thus, determining the correct action in any given set of circumstances is dependent upon how strongly each action satisfies and/or violates the applicable duties. This decision may be straightforward, as in the case where only one available action satisfies any duty. However, it is more likely that more than one action will satisfy and/or violate one or more duties. In such cases, a means (or set of principles) must be provided to choose between conflicting duties. Principles are the crux of ethical decision making and, in general, can be contentious. That said, even though many ethical dilemmas may still be unresolved, it seems more likely that a consensus may be reached in constrained domains as this one. In particular, we might find agreement on how we would like robots to behave towards us, the crux of the matter in this domain. An example of the approach we are advocating can be found in Anderson et al. (2019). Within the domain of healthcare robots, ethically relevant features and corollary duties are discovered though a dialogue with ethicists regarding straightforward cases of ethical dilemmas that such robots are likely to encounter. From determining in these example cases which actions are correct and why, machine learning is used to abstract an overarching principle that balances duties when they conflict. In a robot's daily routine, sensors provide raw data from which a representation of the current situation may be abstracted. The robot can apply the learned principle to this representation in order to determine which of its possible actions is most ethically correct in the current situation. As any interaction a robot has with a human being will have ethical ramifications, this principle is used to determine all behavior of the robot (Berenz and Schaal 2018).

It is our hope that the investigation of such domain-specific value-driven agents will help illuminate the path to a better understanding of the ethical behavior of artificially intelligent agents in general.

4.3 Trust

The HL aims to have robots with humanlike appearance,[12] ability to sense, and functionality by 2050. This robotic design may have both positive and negative consequences for the trust that people place in the robot. While social robots are perceived more positively and have both higher quality and more effective interactions with humans than non-social robots (Holler and Levinson 2019), the same significant fac-

tors that improve perceived human likeness can negatively affect people's acceptance of, and trust in, a robot.

Trust is considered to be a critical construct for establishing successful and lasting human-agent (i.e., human, computer or robot) interaction (Ross 2008). In the psychological literature (Szczesniak 2012), trust is a multidimensional reality that includes cognitive, emotional and behavioral components. It allows people to take decisions that will impact their everyday lives based on rational judgements (i.e., cognitive trust), affective interpersonal relationships (i.e., emotional trust), and their own or others' actions (i.e., behavioral trust). For example, people decide to take a leap of trust while investing in a portfolio, buying a house, picking out an outfit or holiday destination, sharing working responsibilities with a team of other people, or passing a ball to their teammates hoping they will catch it and not score in their own goal.

Researchers in HRI (Rossi et al. 2017; Hancock et al. 2011; Cameron et al. 2015) highlighted several principles and factors that affect someone's (i.e., the trustor or trusting) trust in a robot (the trustee or trusted). These factors can be related to the person, such as demographics, personality, prior experiences, self-confidence; to the robot, such as the robot's reliability, transparency; and to the context of the interaction, e.g. communication modes and shared mental models. We believe that there is a correspondence between the multifaceted nature of human-human trust and the factors affecting people's trust in robots. Firstly, cognitive trust is based on the trustees reliability, dependability, and competence (Szczesniak 2012). In the context of HRI, it is thus built on and affected by a robot's performances and faults. People's expectations of the capabilities of a robot depend on its appearance (Bernotat et al. 2021), its characteristics (Hancock et al. 2011), and the magnitude and timing of the errors it makes (Rossi et al. 2017). Secondly, emotional trust is based on the interpersonal relationships built between trustor and trustee (Szczesniak 2012). Similarly, human-robot trust is stronger when people are more familiar with robots HT et al. (2011), especially with their capabilities and limitations (Rossi et al. 2019). Thirdly, behavioral trust is affected by the trustee's behavior and risk taking in untried and uncertain situations (Szczesniak 2012). Trust also depends on the trustor's belief in the trustee's positive attitude and credibility towards the trustor and a common goal (Simpson 2007). An example of how risk-taking behaviors affect the credibility of an opponent can be found in the popular game of poker where it is important that players gain a good reputation (Billings 1995). Similarly, a robot that builds a good "reputation" is trusted more by its human opponent in human-robot games (Correia et al. 2016).

---

[12] Alessandra Rossi; HL Executive Committee, University of Naples "Federico II", Italy.

4.3.1 Open challenges

Two interrelated challenges in the current state of the art are the baseline level of trust that people may put in robots, and how to manage people's expectations of a robot to ensure those are realistic. Due to the particularly dynamic and unpredictable actions that a robot can perform during a soccer game, human supervised intervention (i.e., using the robot's safety button) will be impossible. Thus, notwithstanding the state of development of the technological and cognitive abilities of robots, the question remains whether people will be willing to engage in a soccer match where there is no option for human supervised intervention. Here, trust drives individuals' choice to rely on others (opponents and teammates) if they are in a vulnerable and uncertain situation. This trust depends on others' choices, including behaviours, actions and motivations (Lee and See 2004). It is important that those observed choices can be interpreted along realistic expectations. For example, encountering a robot that looks very humanlike can lead people to believe that this robot has the ability to sense and respond to their actions and intentions. When these expectations are not met, people lose trust in the robot (Rossi et al. 2020b). People lose trust when the robot makes errors or has non-transparent behaviours that are perceived as errors (Rossi et al. 2017). It is fundamental to understand how to balance robots' appearances to enhance people's trust without setting too high expectations. Robots with human-like appearances might be perceived as more aggressive and less friendly than a machine-like robot (Woods et al. 2006), which might lead to them being perceived as a threat. While people's physical safety is well-investigated in the literature, particularly concerning industrial settings, their *perceived* safety is still overlooked (Akalin et al. 2021). People's discomfort or stress during their interactions with robots can be prevented by manipulating the robot's motions, social behaviors, or attitude (e.g., speech, gaze, posture) (Lasota et al. 2017). Perceived safety is also enhanced by producing higher predictability with legible robot behaviors (Rossi et al. 2020a). Even assuming that transparent behaviors can be implemented by improving a robot's modes of communications (verbal and non-verbal), soccer players act instinctively and use implicit communication signals that are difficult to identify and reproduce with robots.

4.4 Taking advantage of the robot

While it is important that human players feel safe enough to engage in a game of soccer with a robotic team,[13] perceived safety and predictability may have the unintended side

---

[13] Merel Keijsers; HL Team Electric Sheep, John Cabot University, Italy.

effect of humans trying to take advantage of the robot and the restrictions on its behaviors. In non-soccer settings humans have been observed abusing robots that were deployed in public spaces, such as shopping malls, museums, and restaurants (Brscić et al. 2015), even when the robot is supposed to assist the human (Mutlu and Forlizzi 2008) or when it could result in dangerous situations for all parties involved, including any bystanders (Liu et al. 2020b). In one way or another, these behaviors concern humans taking advantage of the robot – an entity that either by programming or sheer lack of comprehension will not retaliate against exploitation or misconduct. While unprovoked aggression purely for the sake of damaging the robot seems unlikely during a soccer match, it is easy to imagine humans searching for the loopholes in the robot's programming that can be used to their advantage. For example, it would be naive to assume that human players will not try to capitalize on a robot's built-in tendency to avoid conflict; this behavior has already been observed in interactions between human drivers and self-driving cars (Liu et al. 2020b). Human drivers become more reckless around autonomous cars as they expect the autonomous car to prioritize safety over traffic rules.

4.4.1 Open challenges

In previously described scenarios, moreover, opportunistic behavior could emerge unintentionally. Social exchanges require a constant interpretation of others' behavior and intentions in order to update evaluations of what the other parties might do next. This interpretation is often done automatically and without much thought, and is not only shaped by societal rules and norms but also on experience related to what others will (not) do or allow. For example, when two opposing human players are running towards the ball, each has to monitor on one hand their belief that the other player will avoid a collision and on the other hand whether this risk of colliding (and potential injury) is worth the potential reward. If one party knows that the other will avoid collision at all costs (including tackles or other risky methods of obtaining the ball), that gives them leverage. Thus, if robotic players avoid any and all situations where a human could get harmed, negotiations like these will be heavily skewed in favor of the humans.

Value-Driven players discussed the ethical implications of this conflict between "keeping human players safe" and "be a successful soccer player", and *Trust* approached it from the perspective of human players' perceived safety. However, the tension between these two values and how it is resolved will have further implications still. On one hand, robots need to place the bodily integrity of the human players above winning or no sensible human player would ever agree to play a game of soccer against a robotic team. At the same time, the robot players cannot afford to be too cautious

as that would be a great disadvantage. A possible solution could be to impose harsher punishments and more meticulous monitoring of players' behavior. However, this would probably only have limited effects: players could claim that their tackle was unintentional (which may result in unjust sanctions), and the potential advantages could be large enough to entice players to try their luck anyway. Alternatively, one could design a feedback loop within the robot decision making process that balances the risk and severity of possible negative consequences of any behavior against the odds and positive outcomes of it. In a sense, humans do this continuously (although our estimates may be biased by heuristics, mood, attention span, energy levels, and so on) and scientists "merely" need to find a way to formalize this constant updating of a cost-vs-benefits model of behavior. This way of decision making could introduce enough assertiveness in the robot team that human players cannot take full advantage of their programmed caution. Moreover, such a loop would imply that the robotic team will adapt their behavior during the match in order to counter their opponents' playing style. If this is rather aggressive, the costs of a defensive play style would become higher, inducing robots to adopt a more assertive playing style themselves too. This leaves the question of how much harm inferred by a robot we are willing to theoretically suffer. In autonomous vehicles, humans are unforgiving of the slightest margin of error. We hold robots to different ethical standards as other humans (Malle et al. 2015) and view reactive aggressive behavior as a lot more maleficent and unacceptable when it comes from a robot than when it comes from a human (Bartneck and Keijsers 2020). However, we will need to come to terms with a certain degree of risk, if only to prevent humans causing far more risky scenarios while attempting to play the robot's programming.

4.5 Mixed teams

In human-robot (HR) soccer teams,[14] the goal is to perform joint soccer tasks in order to achieve common shared objectives, such as scoring in the opposite goal, defending the own goal, and eventually winning a match or a tournament. HR teams have been studied for several application domains, including search and rescue (Nourbakhsh et al. 2005), and surveillance (Srivastava et al. 2013). HR mixed soccer teams (Argall et al. 2006) are very relevant examples of HR collaboration, as a soccer environment provides for interesting and challenging features, such as real time perception and action, dealing with naive users, competitive scenario (possibly two HR mixed teams playing against each other), and an attractive, engaging and easy to understand problem. Solutions validated in HR mixed soccer teams can be transferred,

---

[14] Luca Iocchi; Vice President of the RC Federation; Università di Roma "La Sapienza", Italy.

adapted and extended in many other industrial applications, bringing several advantages and contributions to improve human-robot collaboration in such domains. There are several relevant properties of HR mixed teams. Firstly, the presence of humans and robots in the same team implies a high degree of heterogeneity. Indeed, the interaction mechanisms in HR settings are very different from those used in robot-robot teams, since in many cases HR teams are forced to use natural human-like communications. Moreover, if we consider mixed HR teams with robots developed by different researchers, a suitable common language must be defined to account for the diversity of the agents in the team. A major consequence of such heterogeneity is that most of the elements that are relevant to define a joint behavior (such as communication, players' actions, intentions, etc.) cannot be standardized and limited to a known predefined set of elements. Moreover, each agent has specific skills and abilities that should be exploited to optimize the overall team performance. Although heterogeneous, team members can interchange their roles among each other when this is beneficial to increase the performance. Secondly, the team goals are common and shared. Common goals refer to the notion of having the same goals for all the agents in the team. When a goal is achieved, all the team members will get the same benefit from it. If the goal is not achieved, all the team members will get the same disadvantage. Shared goals refer to explicit knowledge: all the agents know about the common goal, they know that all the agents know about the common goal, etc. Notice that in some cases of human soccer, individual goals are also present: e.g., a player wants to score to gain some personal benefit not completely shared with the team. We will not consider individual goals in this section. We also assume that team members trust each other. In particular, any agent expects that all the other agents in the team will act to achieve the common goal. Thirdly, when executing the task, the agents have to deal with limited resources (such as time space, energy, etc.) not only among themselves, but also with respect to the agents of the opponent team. We cannot assume the presence of a central processing unit, so strategic and tactical decisions must be distributively taken. Finally, as humans and robots share the same physical space, safety must be guaranteed with the maximum priority.

4.5.1 Open challenges

The properties described above make HR soccer teams very challenging to design, develop and deploy. Several research topics must be addressed, which are briefly summarized in the following.

(a) Design of HR teams HR team design should mainly focus on collaboration and interaction (Ma et al. 2017), possibly exploiting existing models of human-human interaction or defining new specific models. Dimensions for a taxon-

omy of existing methods (e.g., Jiang and Arkin 2015) can be helpful to identify specific design elements. Some general architectures for HR teams have been proposed (e.g., Lallée et al. 2010) to identify the main components needed for the development of such systems. The current achievements are still far from providing a concrete methodology or guidelines to design effective HR teams.

(b) Cooperative perception HR teams need sophisticated distributed perception abilities that allow all the team members to have a clear understanding of the situation. Moreover, simple assessment of the current situation is often insufficient, and predicting intentions of other agents in the environment is necessary. Typical solutions rely on sensor analysis and sensor fusion and are suitable in many practical applications, such as industrial environments (e.g., Bonci et al. (2021)). Cooperative perception in HR soccer teams is even more challenging, due to the possibly high speed of operations and to the safety risks for humans involved in the task.

(c) Knowledge alignment A main use of cooperative perception is to align the knowledge states of all the agents in a team, which is necessary for a fully comprehensive situation assessment. For example, a complete shared understanding of the soccer play state (position and dynamics of all the players and of the ball) can enable the team members to distribute themselves in the field in a convenient formation. Designing proper models that allow humans and robots to efficiently share their knowledge (obviously individually represented in a very different way) is one of the most challenging research objectives in HR teams.

(d) Coordinated actions HR soccer teams need to properly coordinate their physical actions to affect the environment. Although some basic actions (e.g., kicking the ball) are executed by each single team member independently of the others, joint actions (e.g., passing) are very relevant in this domain. In addition to reactivity, which requires the team members to directly perform actions based on sensor stimuli, anticipating behaviors and pro-activity, based on prediction of future states of the environment, are extremely important. For example, predicting the intention of an opponent provides advantage in the choice and timing of executing suitable actions. Balancing reactivity, pro-activity, and anticipating behaviors in a heterogeneous HR team is a completely open problem.

(e) Interactions Interactions in HR soccer teams must be multi-modal (speech, non-verbal vocalizations, gestures, body postures, etc.) as many different situations may occur that make some modalities more appropriate than others. These interactions are often used to provide or exchange information, affecting the knowledge (or mental) state of the agents. For example, gestures can be used to indicate where or to whom to pass the ball. Developing effective interactions

in the soccer domain is thus another interesting research challenge.

(f) Decision making Distributed decision-making and coordination are necessary abilities for soccer agents who need to balance decisions considering both short- and long-term goals. The soccer domain is inherently dynamic and dynamic forms of distributed coordination (Dias et al. 2006) are needed. The autonomy in decision-making by each team member must be considered as a dynamic aspect (Dias et al. 2008) in order to adapt to different situations that may occur during a game. For example, an agent may have a better view of the situation and can suggest another agent what to do. Individual decision-making must take into account teamwork elements, such as negotiation, commitment, and anticipation. If an agreement is taken (e.g., a pass), the decisions should be finalized to fulfil it.

(g) Learning and adaptation Team learning and adaptation is also of crucial importance for effective HR collaboration, due to the presence of an opponent team for which a precise model is not available beforehand and thus optimal behaviors cannot be precisely planned before the game. Techniques like Multi-Agent Reinforcement Learning (MARL) have been successfully used in robot soccer teams. However, the application to HR teams, i.e., the development of Human-Robot Reinforcement Learning is a very interesting novel research challenge for HR soccer.

(h) Benchmarking Benchmarking HR teams has been considered both in general cases (e.g., Groom and Nass (2007)) and for specific tasks (e.g., Xin and Sharlin (2007)). HR soccer games can provide for a very interesting and challenging benchmarking scenario for HR collaboration, due to the features of the problems and the many open research areas that have been illustrated in this section.

4.6 Discussion and conclusion

At the moment, RC features almost exclusively matches between robotic teams. Since the long term goal is to have matches against human teams, human-robot relational dynamics will have to be considered at some point in the near future. One step towards this direction has been taken by MSL which introduced a rule for the 2022 competition. This rule allows a human player to take the place of a robot player.[15]

The MSL new rule highlights a few interesting dilemmas with mixed teams, and there are connected principally to the human players' safety, game's dynamics and communications, and liability of any possible injuries to the human players.

---

[15] MSL rule number 7 of the Competition Rule Section https://msl.RC.org/wp-content/uploads/2022/01/Rulebook_MSL2022_v23.0.pdf

This brings to the attention that there is most likely going to be a considerable tension between two conflicting goals: on one hand, the need for the robotic players to keep the humans safe, and on the other the need for the robotic team to not be (perceived as) pushovers. This is a complex dilemma to solve, as it involves the robot's ability to dynamically evaluate many different and opposing goals (e.g., "pass the defenders of the opposing team while they are trying to take the ball from me" vs "avoid injuring the defenders of the opposing team"); the humans' perception of the robot's ability to evaluate opposing goals and make the right (moral) call; and finally ensuring that this trust of the human player in the robot's morality doesn't result in the human taking advantage of the robot (e.g., "the robots are programmed to avoid harming me, so if I go for a tackle they'll abandon the ball to avoid the possibility of harming me"). This is a non-trivial issue as it depends not only on the robot's ability to juggle a complex interplay of values, but also on the human's perception of the robot's ability to do so, and on balancing out those values in such a way that humans will still be willing to play against the robotic team without taking advantage of it. This may be a paradox that cannot be solely solved through robot design, but will require humans as well to adapt, e.g. through accepting a risk of being injured by a robot player.

A second issue that most sections touched upon but may not have discussed as in-depth as the trust dilemma, is the relevance of communication (both verbal and non-verbal). Successful communication of intentions and current states, both between members of the same team and also (maybe especially) between members of opposing teams, will be of tremendous importance if we are to see human-robot soccer matches in the future. Communication is key to all open questions discussed above. Without it, ethical behaviour cannot be designed, nor can trust be gained or boundaries set, and collaboration will be impossible.

Finally, a third issue resides in the identification of the legal and moral responsible actors in case of injuries to human players or broken property of robotic players. Several RC Leagues, such as the HL, have rules in place since a long time to prevent the damage to robots or the game fields, and the MSL stated in their new rule that the liability of injuries to human players falls on the team of the human player. The liability does not necessarily rely on one part, and the robot may be partially or fully responsible for an incident (e.g., if it applies a more forceful contact game with the human player). Legal responsibilities also do not only depend on the RC Federation's regulations, but it could vary according to the country where the RC is played. For this reason, it is important to firstly define a global legal regulation for the whole RC, and then define a complaint mechanism with respect to the regulations of the host countries of the competition.

5 Conclusions

RoboCup provides one of the best benchmarks for autonomous robotics in unstructured environments due to the multitude of its open challenges. For example, to effectively play soccer, the robots need to perceive and interpret data from the external environment, collecting information about themselves, their teammates, and their opponents (e.g., position in the field); they need to be able to understand and communicate using verbal and non-verbal cues, and so on. However, not only do robots need to be designed using appropriate materials, but roboticists need to model their behaviors and mechanisms to allow human players to trust that robots are able to play in a safe and secure way. To explore such research directions, here, we contextualized RoboCup within the state-of-art of in the fields of Robotics, Engineering, Material Science, Ethics, and HRI, and presented the requirements that researchers in such areas need to address and develop in order to bring solutions/systems together in a safe, coherent and testable way for both human and robot players. We invite and encourage researchers to use the RC 2050 challenge to inspire, evaluate, and promote their work, ideally in collaboration with one another throughout the world.

Author Contributions AR contributed to the manuscript conception, and AR, MPP and MK contributed to the design and preparations. The first and final draft of the manuscript was written by AR, MPP and MK, and all authors commented on previous versions of the manuscript. This work had contributions from several authors attributed in their sections. In particular, PS and MPP wrote the Introduction section; YL wrote the Human-sized robot design subsection; DB and MS wrote Motion Engine subsection; CR wrote the Audio in Human-Multi-robot Systems subsection; FV and JG wrote the Robot Vision subsection; WK, AK and RM wrote the Playing against humans subsection; MA and SLA wrote the Value-Driven Players subsection; AR wrote the Trust subsection; MK wrote the Taking advantage of the robot subsection; LI wrote the Mixed teams subsection. All authors read and approved the final manuscript.

Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

References

Akalin, N., Kristoffersson, A., & Loutfi, A. (2021). Investigating the multidisciplinary perspective of perceived safety in human-robot interaction.

Anderson, M., & Anderson, S. L. (2011). *Machine Ethics*. Cambridge University Press.

Anderson, M., Anderson, S. L., & Berenz, V. (2019). A value-driven eldercare robot: Virtual and physical instantiations of a case-supported principle-based behavior paradigm. *Proceedings of the IEEE, 107*(3), 526–540.

Antonioni, E., Suriani, V., Solimando, F., Bloisi, D., & Nardi, D. (2021). Learning from the crowd: Improving the decision-making process in robot soccer using the audience noise. In *Proceeding on RoboCup Symposium*.

Argall, B., Gu, Y., Browning, B., & Veloso, M. (2006). The first segway soccer experience: Towards peer-to-peer human-robot teams. in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, ser. HRI '06. New York, NY, USA: Association for Computing Machinery, 321–322.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481–2495.

Bartneck, C., & Keijsers, M. (2020). The morality of abusing a robot. *Paladyn, Journal of Behavioral Robotics, 11*(1), 271–283.

Behnke, S. (2006). Online trajectory generation for omnidirectional biped walking, In: *Proceedings 2006 IEEE International conference on robotics and automation, 2006. ICRA 2006.*, pp. 1597–1603.

Berenz, V., & Schaal, S. (2018). The playful software platform: Reactive programming for orchestrating robotic behavior. *IEEE Robotics Automation Magazine, 25*(3), 49–60.

Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (fe)male robot: How robot body shape impacts first impressions and trust towards robots. *International Journal of Social Robotics, 13*, 06.

Bestmann, M., Engelke, T., Fiedler, N., Güldenstein, J., Gutsche, J., Hagge, J., & Vahl, F. (2022). Torso-21 dataset: Typical objects in robocup soccer 2021. In *Robot World Cup*. Springer, 65–77.

Billings, D. (1995). Computer poker. *University of Alberta M.Sc. thesis*.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv:2004.10934.

Bonci, A., Cen Cheng, P. D., Indri, M., Nabissi, G., & Sibona, F. (2021). Human-robot perception in industrial environments: A survey. *Sensors, 21*(5), 1571.

Boxill, J. (2002). *Sports ethics: An anthology*. Wiley.

Brscić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015). Escaping from children's abuse of social robots. In *Proceedings of the 10th ACM/IEEE international conference on Human-robot interaction (HRI)*, ACM. Portland, USA: ACM/IEEE, 59–66.

Buschmann, T., Schwienbacher, M., Favot, V., Ewald, A., & Ulbrich, H. (2012). The biped walking robot lola-hardware design and walking control-. *Journal of the Robotics Society of Japan, 30*(4), 363–366.

Cameron, D., Aitken, J. M., Collins, E. C., Boorman, L., Chua, A., Fernando, S., McAree, O., Martinez-Hernandez, U., & J. Law (2015). Framing factors: The importance of context and the individual in understanding trust in human-robot interaction. In *International Conference on Intelligent Robots and Systems*.

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., & Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on pattern analysis and machine intelligence*.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172–186.

Chu, H., Hu, X., Wang, Z., Mu, J., Li, N., Zhou, et al. (2021). Unipolar stroke, electroosmotic pump carbon nanotube yarn muscles, *Science (New York, N.Y.)*, pp. 494–498.

Correia, F., Alves-Oliveira, P., Maia, N., Ribeiro, T., Petisca, S., Melo, F. S., & Paiva, A. (2016). Just follow the suit! trust in human-robot interactions during card game playing. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 507–512.

Czarnetzki, S., Kerner, S., & Urbann, O. (2009). Observer-based dynamic walking control for biped robots. *Robotics and Autonomous Systems, 57*(8), 839–845.

de Graaf, M., Ben Allouch, S., & van Dijk, J. (2017). Why do they refuse to use my robot? reasons for non-use derived from a long-term home study, In *Proceedings of the 2017 ACM/IEEE International conference on human-robot interaction*, ser. HRI '17. New York, NY, USA: Association for Computing Machinery, 224–233.

Dias, R., Amaral, F., Angelico, I., Azevedo, J., Cunha, J., Dias, P., & Silva, J. (2020) CAMBADA'2020: Team description paper.

Dias, R., Cunha, B., Sousa, E., Azevedo, J. L., Silva, J., Amaral, F., & Lau, N. (2017). Real-time multi-object tracking on highly dynamic environments. In *2017 IEEE International conference on autonomous robot systems and competitions (ICARSC)*. Coimbra, Portugal: IEEE, 178–183.

Dias, M., Harris, T., Browning, B., Jones, E., Argall, B., Veloso, M., Stentz, A., & Rudnicky, A. I. (2006). Dynamically formed human-robot teams performing coordinated tasks. In *AAAI Spring Symposium: To boldly go where no human-robot team has gone before*.

Dias, M., Kannan, B., Browning, B., Jones, E., Argall, B., Zinck, M., Veloso, M., & Stentz, A. (2008). Sliding autonomy for peer-to-peer human-robot teams. In *Proceedings of the international conference on intelligent autonomous systems*.

Dolatabadi, M., Elfring, J., & van de Molengraft, R. (2020). Multiple-joint pedestrian tracking using periodic models. *Sensors, 20*(23), 6917.

Fiedler, N., Brandt, H., Gutsche, J., Vahl, F., Hagge, J., & Bestmann, M. (2019). An open source vision pipeline approach for robocup humanoid soccer, In *Robot World Cup*. Springer, 376–386.

Freitag, L., Draegert, J., Gottlieb, S. G., Barth, G., Pluhatsch, M., Schmidt, A. et al. (2016). Berlin united-fumanoids team description paper for robocup.

Gato, L. (2020). Localización y rastreo de múltiples hablantes para robots de servicio usando un arreglo triangular de micrófonos," Master's thesis, Universidad Nacional Autonoma de Mexico, Mexico, [Online]. Available: https://github.com/lmiguelgato/DAP_project

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448.

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., et al. (2013). Challenges in representation learning: A report on three machine learning contests. In *International conference on neural information processing*. Springer, 117–124.

Grondin, F., Lauzon, J.-S., Vincent, J., & Michaud, F. (2020). Gev beamforming supported by doa-based masks generated on pairs of microphones. *Proceeding Interspeech, 2020*, 3341–3345.

Grondin, F., & Michaud, F. (2019). Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations. *Robotics and Autonomous Systems, 113*, 63–80.

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human-robot teams. *Interaction Studies, 8*(3), 483–500.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of Human Factors and Ergonomics Society, 53*(5), 517–527.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences, 23*(8), 639–652.

Holthaus, P., & Wachsmuth, S. (2021). It was a pleasure meeting you - towards a holistic model of human-robot encounters. *International Journal of Social Robotics, 13*(7), 1729–1745.

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology, 101*(3), 557–570. https://doi.org/10.1037/a0022885

Jamzad, M., Sadjad, B., Mirrokni, V. S., Kazemi, M., Chitsaz, H., Heydarnoori, A., Hajiaghai, M., & Chiniforooshan, E. (2001). A fast vision system for middle size robots in robocup, In *Robot Soccer World Cup*. Springer, 71–80.

Janai, J., Güney, F., Behl, A., & Geiger, A. (2020). Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends in Computer Graphics and Vision, 12*(1–3), 1–308.

Jiang, S., & Arkin, R. C. (2015). Mixed-initiative human-robot interaction: Definition, taxonomy, and survey. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*.

Kim, J.-H., & Kim, J. (2023). High performance electro-active artificial muscles based on 3d dendritic nickel cobalt sulfide nanorods-graphene electrodes. *Sensors and Actuators, B: Chemical, 389*, 133915.

Kubisch, M., Werner, B., & Hild, M. (2011). Using co-existing attractors of a sensorimotor loop for the motion control of a humanoid robot. In *International conference on neural computation theory and applications (NCTA 2011)*.

Lallée, S., Lemaignan, S., Lenz, A., Melhuish, C., Natale, L., Skachek, S., van Der Zant, T., Warneken, F., & Dominey, P. F. (2010). Towards a platform-independent cooperative human-robot interaction system: I. perception. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4444–4451.

Lasota, P. A., Song, T., & Shah, J. A. (2017). *A Survey of Methods for Safe Human-Robot Interaction*. Now Foundations and Trends, 104.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(1), 50–80.

Lim, G. H., Pedrosa, E., Amaral, F., Lau, N., Pereira, A., Dias, P., & Reis, L. P. (2017). Rich and robust human-robot interaction on gesture recognition for assembly tasks. In *2017 IEEE International conference on autonomous robot systems and competitions (ICARSC)*. Coimbra, Portugal: IEEE, 159–164.

Lima, D.M., Li, N., Andrade, J.D.M., Fang, S., Oh, J., Spinks, M.G., et al. (2012). Electrically, chemically, and photonically powered torsional and tensile actuation of hybrid carbon nanotube yarn muscles, *Science (New York, N.Y.)*, pp. 928–932.

Liu, Y., Delfarah, M., & Wang, D. (2020). Deep casa for talker-independent monaural speech separation, In *ICASSP 2020-2020 IEEE International conference on acoustics, speech and signal processing (ICASSP)*. IEEE, pp. 6354–6358.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Liu, P., Du, Y., Wang, L., & Da Young, J. (2020). Ready to bully automated vehicles on public roads? *Accident Analysis & Prevention, 137*, 105457.

Ma, L. M., Fong, T., Micire, M. J., Kim, Y., & Feigh, K. (2017). Human-robot teaming: Concepts and components for design. In *FSR*.

Maldonado, A., Rascon, C., & Vélez, I. (2020). Lightweight online separation of the sound source of interest through blstm-based binary masking, *Computación y Sistemas*, 24.

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015). Sacrifice one for the good of many?: People apply different moral norms to human and robot agents. In *Proceedings of the 10th ACM/IEEE international conference on human-robot interaction (HRI)*. ACM, 117–124.

Meessen, K. J., Paulides, J. J. H., & Lomonova, E. A. (2010). A football kicking high speed actuator for a mobile robotic application. In *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*. Glendale, AZ, USA: IEEE, 1659–1664.

Meza, I., Rascon, C., Fuentes, G., & Pineda, L. A. (2016). On indexicality, direction of arrival of sound sources, and human-robot interaction. *Journal of Robotics*. https://doi.org/10.1155/2016/3081048

Mnih, V., Heess, N., Graves A. *et al.* (2014). Recurrent models of visual attention. In *Advances in neural information processing systems*, vol. 27.

MSL Technical Committee 1997–2021, *Middle Size Robot League Rules and Regulations for 2021*, 2020.

Mutlu, B., & Forlizzi, J. (2008). Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction (HRI)*. ACM, 287–294.

Nakadai, K., Masaki, S., Kojima, R., Sugiyama, O., Itoyama, K., Nishida, K. (2020). Sound source localization based on vonmises-bernoulli deep neural network. In *IEEE/SICE International symposium on system integration (SII).IEEE, 2020*, 658–663.

Nakadai, K., Takahashi, T., Okuno, H. G., Nakajima, H., Hasegawa, Y., & Tsujino, H. (2010). Design and implementation of robot audition system'hark'-open source software for listening to three simultaneous speakers. *Advanced Robotics, 24*(5–6), 739–761.

Nesset, B., Robb, D. A., Lopes, J., & Hastie, H. (2021). *Transparency in HRI: Trust and decision making in the face of robot errors* (pp. 313–317). New York, NY, USA: Association for Computing Machinery.

Nourbakhsh, I., Sycara, K., Koes, M., Yong, M., Lewis, M., & Burion, S. (2005). Human-robot teaming for search and rescue. *IEEE Pervasive Computing, 4*(1), 72–79.

Paetzel-Prüsmann, M., Rossi, A., & Keijsers, M. (2023). Current and future challenges in humanoid robotics —An empirical investigation.

Peng, X. B., Abbeel, P., Levine, S., & van de Panne, M. (2018). Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transcation on Graphics, 37*(4), 1–4.

Rascon, C., & Meza, I. (2017). Localization of sound sources in robotics: A review. *Robotics and Autonomous Systems, 96*, 184–210.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788.

Reher, J., & Ames, A. D. (2020). Dynamic walking: Toward agile and efficient bipedal robots.

Reher, J., Cousineau, A. E., Hereid, A., Hubicki, M. C., & Ames, D. A. (2016). Realizing dynamic and efficient bipedal locomotion on the humanoid robot durus, *ICRA*, pp. 1794–1801.

RoboCup Soccer Humanoid League Laws of the Game 2019/2020. https://humanoid.robocup.org/wp-content/uploads/RCHL-2020-Rules-Dec23.pdf, (accessed November 21, 2020).

RoboCup@Home Technical Committee. (2024). *Rulebooks*. Available: https://athome.robocup.org/rules/

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

Ross, J. M. (2008). Moderators of trust and reliance across multiple decision aids (doctoral dissertation), University of Central Florida, Orlando.

Rossi, A., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2017). How the timing and magnitude of robot errors influence peoples' trust of robots in an emergency scenario. In A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, & H. He (Eds.), *Social Robotics* (pp. 42–52). Cham: Springer International Publishing.

Rossi, A., Dautenhahn, K., Lee Koay, K., & Walters, M. L. (2020). How social robots influence people's trust in critical situations. In *2020 29th IEEE International conference on robot and human interactive communication (RO-MAN)*, 1020–1025.

Rossi, A., Moros, S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2019). Getting to know kaspar : Effects of people's awareness of a robot's capabilities on their trust in the robot. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–6.

Rossi, S., Rossi, A., & Dautenhahn, K. (2020). The secret life of robots: Perspectives and challenges for robot's behaviours during non-interactive tasks. *International Journal of Social Robotics, 12*(6), 1265–1278.

Sakagami, Y., Watanabe, R., Aoyama, C., & Matsunaga, S. (2002). "The intelligent asimo: system overview and integration," Intelligent Robots and Systems, 2002. *IEEE/RSJ International Conference,3*, 2478–2483.

Scheunemann, M., Salge, C., & Dautenhahn, K. (2019). Intrinsically motivated autonomy in human-robot interaction: Human Perception of Predictive Information in Robots. Springer, *6*, 325–337.

Scheunemann, M. M., Salge, C., Polani, D., & Dautenhahn, K. (2022). Human perception of intrinsically motivated autonomy in human-robot interaction. *Adaptive Behavior, 30*(5), 451–472.

Schillaci, G., Hafner, V. V., & Lara, B. (2016). Exploration behaviors, body representations, and simulation processes for the development of cognition in artificial agents. *Frontiers in Robotics and AI, 3*, 39.

Schreuder, E., Feitsma, J., Kouters, E., & Vos, J. (2019). Falcons team description paper 2019.

Simpson, J. A. (2007). Foundations of interpersonal trust. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 587–607). Guilford Press.

Smolyanskiy, N., Kamenev, A., & Birchfield, S. (2018). On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops,* 1007–1015.

Soetens, R., van de Molengraft, R., & Cunha, B. (2015). RoboCup MSL - History, Accomplishments, Current Status and Challenges Ahead. in RoboCup 2014: Robot World Cup XVIII, R. A. C. Bianchi, H. L. Akin, S. Ramamoorthy, and K. Sugiura. *Eds. Cham: Springer International Publishing,8992*, 624–635.

Srivastava, V., Surana, A., Eckstein, M. P., & Bullo, F. (2013). Mixed human-robot team surveillance.

Stone, P., Quinlan, M., & Hester, T. (2010). *"Can Robots Play Soccer?" In "Soccer and Philosophy: Beautiful thoughts on the beautiful game" by Richards, T.* Open Court Publishing, 51.

Szczesniak, M. (2012). Development of interpersonal trust among children and adolescents. *Polish Psychological Bulletin, 43*, 50–58.

Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Springer

Vahl, F., Gutsche, J., Bestmann, M., & Zhang, J. (2021). Yoeo–you only encode once: A cnn for embedded object detection and semantic segmentation, In *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 619–624.

Vélez, I., Rascon, C., & Fuentes-Pineda, G. (2020). Lightweight speaker verification for online identification of new speakers with short segments, *Applied Soft Computing*, vol. 95.

Vukobratovic, M., & Borovac, B. (2004). Zero-moment point - thirty five years of its life. *International Journal Humanoid Robotics, 1*, 157–173.

Woods, S., Walters, M., Kheng Lee Koay, & Dautenhahn, K. (2006). Comparing human robot interaction scenarios using live and video based methods: Towards a novel methodological approach. In *9th IEEE International workshop on advanced motion control, 2006.*, 750–755.

Xie, W., Nagrani, A., Chung, J. S., & Zisserman, A. (2019). Utterance-level aggregation for speaker recognition in the wild. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5791–5795.

Xin, M., & Sharlin, E., (2007). Playing games with robots – a method for evaluating human-robot interaction. In *Human Robot Interaction*. IntechOpen.
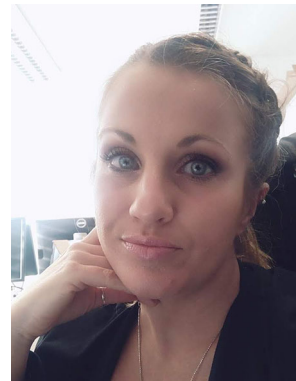
Alessandra Rossi is Assistant Professor working as part of Italian PON R&I 2014-2020 - REACT-EU (CUP E65F21002920003). She was previously a postdoctoral researcher on the BRILLO project (PON I&C 2014-2020 MISE) at the University of Naples "Federico II". Her PhD thesis was part of the Marie Sklodowska-Curie Research ETN SECURE project (https://secure-robots.eu/) at the University of Hertfordshire (UK). Alessandra is co-PI of the AFOSR ERROR project, Project Manager of Marie Sklodowska-Curie Research ETN PERSEO, and involved in the scientific coordination of several national and international projects. Alessandra is currently Exec member of the RoboCup Humanoid League since 2022. Her research interests include Human-(Multi) Robot Interaction, social robotics, trust, XAI, multi-agent systems and user profiling.
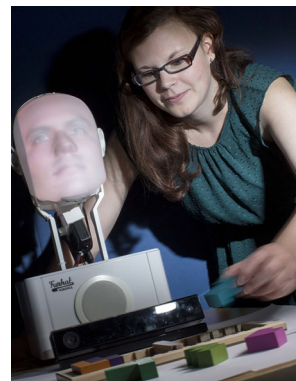
Maike Paetzel-Prüsmann received her PhD from Uppsala University, Sweden, in Computer Science with a Specialization in Human-Computer Interaction in 2020. Her research is focused on how robots and AI agents interact with humans and how we can design better interactive systems to meet people's expectations. In addition to Human-Robot Interaction, she is involved with the RoboCup community since 2011. After first co-founding a student-led RoboCup team in Hamburg, Germany, were

she studied for her Bachelor in Computer Science, she later got actively engaged in the RoboCup organization, serving in the Technical and Executive Committee for the Humanoid League and on the Board of Trustees.

Merel Keijsers has a double masters degree in social psychology and statistics, and researched the psychological processes underlying robot bullying as part of her PhD. She currently is an assistant professor in psychology at John Cabot University in Rome.

Michael Anderson professor of computer science at the University of Hartford, earned his Ph.D. in computer science and engineering at the University of Connecticut. He has been instrumental in establishing machine ethics as a bona fide field of study, co-chairing/authoring with Susan Leigh Anderson the AAAI Fall 2005 Symposium on Machine Ethics, a IEEE Intelligent Systems special issue on machine ethics, and an invited article for Artificial Intelligence Magazine on the topic. Further, this research in machine ethics was selected for Innovative Applications of Artificial Intelligence as an emerging application in 2006. Scientific American (Oct. 2010) features an invited article on this research in which the first robot whose behavior is guided by an ethical principle is debuted. They have published "Machine Ethics" with Cambridge University Press (2011). The most complete statement of this research was published in the March 2019 issue of the Proceedings of the IEEE.

Susan Leigh Anderson Professor Emerita at the University of Connecticut, received her Ph.D. in Philosophy from UCLA. A member of the American Philosophical Association, she has received fellowships and grants from NEH, Yale University, NASA and NSF. She has published three books in the Wadsworth Philosophers Series and numerous articles in applied ethics, particularly bio-medical ethics.

Daniel Barry got a MEng acc (hons) at the University of Hertfordshire. He is full-time Robotics Engineer at Theia Robotics, developing navigation aids for the visually impaired. Daniel Barry is also a Visiting Lecturer at the University of Hertfordshire. He is a RoboCup team member for 10+ years and team leader of the Bold Hearts Humanoid League RoboCup team at the University of Hertfordshire. His core research areas include real-time autonomy, humanoid motion, embedded electronics and computer vision.
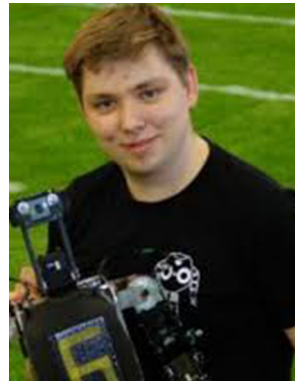
Jan Gutsche is associated with the Hamburg Bit-Bots, a RoboCup Humanoid League team. His contributions focus on vision and active vision systems.

Justin Hart is an assistant professor of practice teaching the Autonomous Robots stream of the Freshman Research Initiative at UT Austin. He works under the supervision of Professor Peter Stone in the Learning Agents Research Group in the Department of Computer Science. Justin Hart co-leads the Building-Wide Intelligence Project, and supervise the UT Austin Villa@Home RoboCup@Home team. He completed my doctoral thesis under the supervision of Professor Brian Scassellati in the Department of Computer Science at Yale University in 2014. He researches artificial intelligence and human-robot interaction. Significant to his work at UT is the development of comprehensive systems and enabling technologies for general purpose service robots. Topics include the construction of architectures for long-term autonomy, knowledge representation, semantic mapping as it relates to both planning and scene understanding, and autonomous human-robot interaction.

Luca Iocchi (www.diag.uniroma1.it/iocchi) is Full Professor at Sapienza University of Rome, Italy, mainly teaching in the Master in Artificial Intelligence and Robotics. His main research interests include cognitive robotics, task planning, multi-robot coordination, robot perception, robot learning, human-robot interaction, and social robotics. He is the author of over 200 referred papers in journals and conferences in artificial intelligence and robotics. He is currently Associate Editor of Artificial Intelligence Journal and Vice-President of RoboCup Federation. He has been principal investigator of several international, EU, national and industrial projects in artificial intelligence and robotics. He has organized several scientific events and contributed to benchmarking domestic service robots through scientific competitions within RoboCup@Home and the European Robotics League Service Robots (ERL-SR).

Ainse Kokkelmans is a a mechatronics engineer working at Thermo Fisher in July 2022. She is CST (Mechanical Engineering) of the Tech United Middle Size League, Eindhoven University of Technology.

Wouter Kuijpers was born on June 1, 1992 in Geldrop, the Netherlands. He studied Mechatronics Engineering at the Fontys Hogeschool in Venlo, the Netherlands, where he obtained a Bachelor of Engineering in 2013 (cum laude). He received the Bachelor of Science and Master of Science degree (with great appreciation) in Mechanical Engineering from the Eindhoven University of Technology (TU/e), Eindhoven, the Netherlands, in 2014 and 2016, respectively. He defended his thesis on Model Selection and Optimal Control Design for Automatic Greenhouse Climate Control on Monday 8th of March 2021. He is currently a Program Officer at the Dynamical Systems Design division at TU/e's Mechanical Engineering department.

Yun Liu is Team leader of ZJLabers. He is part of the ZHEJIANG LAB, Intelligent Robot Reasearch Center, Intelligent Robot Research Center, and works on hardware design and behavior algorithm.

Daniel Polani is Professor of Artificial Intelligence at the Department of Computer Science, Director of the Centre for Artificial Intelligence and Robotics (CAIR) and Head of the Adaptive Systems Research Group at the University of Hertfordshire. His research interests concentrate on the understanding and modeling of collective complex systems and intelligent decision-making, especially in the context of cognition in artificial and biological agents. His research ranges from fundamental questions, such as the role of embodiment, intrinsic motivations, taskless utilities, self-organization and Artificial Life, to questions from cognitive science, psychology, social science, and biology. His interest in bridging the gap between fundamental questions with practical implementation drove his interest and participation in the RoboCup competition series; amongst other, he has been President of the RoboCup Federation during the period July 2017-July 2019.

Caleb Roscon is a Principal Researcher at the Universidad Nacional Autonoma de Mexico (UNAM). His work is centered in Auditory Scene Analysis and Robot Audition, as well as in general audio signal processing in real environments, the results of which have been in constant publication since 2010. He won, together with the Golem Group, the Innovation Award at the RoboCup@Home Netherlands 2013 competition, and was a member of their Executive Council. He received the Under 35 Innovator Award from the MIT Technology Review, Mexico Edition 2014. Currently he's the president of the Digital Signal Processing Group of the Electrical Engineering Postgraduate program of the UNAM.

Marcus Scheunemann is the Head of Autonomy at Dexory, a Scientific Advisor for Theia Robotics and a Visiting Research Fellow at the University of Hertfordshire. His main work currently involves the development of fully autonomous mobile robots for the logistic industry. His science interests are driven by the ambition to understand the emergence of behaviour and social interactions, he is interested in the principles of learning and decision-making, as well as implementational details of how to let robots employ said principles to create meaningful behaviour autonomously. In his PhD titled "Autonomous and Intrinsically Motivated Robots for Sustained Human-Robot Interaction" he investigated the human perception of autonomously generated robot behaviour in human-robot interaction experiments.

Peter Stone is founder and director of the Learning Agents Research Group (LARG) within the Artificial Intelligence Laboratory in the Department of Computer Science at The University of Texas at Austin, as well as associate department chair and Director of Texas Robotics. I was a co-founder of Cogitai, Inc. and am now Executive Director of Sony AI America. My main research interest in AI is understanding how we can best create complete intelligent agents. I consider adaptation, interaction, and embodiment to be essential capabilities of such agents. Thus, my research focuses mainly on machine learning, multiagent systems, and robotics. To me, the most exciting research topics are those inspired by challenging real-world problems. I believe that complete successful research includes both precise, novel algorithms and fully implemented and rigorously evaluated applications. My application domains have included robot soccer, autonomous bidding agents, autonomous vehicles, and human-interactive agents.
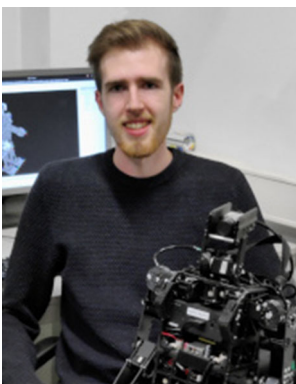
Florian Vahl has been associated with the Hamburg Bit-Bots, a RoboCup Humanoid League team, since 2018. His contributions focus on vision and active vision systems, the ROS ecosystem and overall robot design. Through his work, Vahl has published several papers, aiming to advance the capabilities of robots, especially humanoid ones.

René van de Molengraft is an Associate Professor and Chair Robotics at the Mechanical Engineering Department at Eindhoven University of Technology (TU/e). His main research interest is in autonomous robotic systems that can perform tasks in the living and working environments of humans, encompassing all the variations that occur therein. René van de Molengraft received his MSc degree (with honors) in Mechanical Engineering from TU/e in 1986. From 1986 until 1990 he was a research assistant at Eindhoven University of Technology. In 1990 he received his PhD from TU/e for this thesis on the subject of Identification of Mechanical Systems for Control. He was Assistant Professor between 1990 and 2008 and Associate Professor since 2008. In 2005, he founded the Tech United Robocup team (world champion MSL robot soccer in 2012, 2014, 2016, 2018, 2019, 2022 and 2023 and world champion @Home competition for domestic service robots in 2019 and 2022).

Oskar von Stryk is Full Professor at the Department of Computer Science of the Technical University Darmstadt. He is acting Deputy Chairman of the German Rescue Robotics Center DRZ, which he co-founded. He also has been founding trustee of the Association for the Understanding of Artificial Intelligence as well as long-standing trustee and vice-president of the International RoboCup Federation. He has published more than 200 international peer-reviewed publications and co-founded the robotics startups Bionic Robotics, Energy Robotics and Freemotion Systems. He was the key driving force in establishing the innovative interdisciplinary study programs on Autonomous Systems and Robotics as well as Computational Engineering. His research on cooperating autonomous mobile robots, biologically inspired robots and numerical optimization and control methods has been funded in more than 30 projects by national, European and US research agencies and companies. The research results have received several awards, e.g., a first prize of the European Robotics Technology Transfer Award, the Hessian Cooperation Award, the Louis Vuitton Best Humanoid Award and numerous international competitions for autonomous mobile robots (e.g, ARGOS Challenge, ENRICH, RoboCup, World Robot Summit). Academic Influence ranks him among the 25 most influential computer scientists worldwide in 2010 to 2020.