Scientific Electronic Archives
Issue ID: Sci. Elec. Arch. Vol. 17 (5)
Sept/Oct 2024
DOI: http://dx.doi.org/10.36560/17520241986
Article link: https://sea.ufr.edu.br/SEA/article/view/1986

ISSN 2316-9281

Human-Social Robot Interaction in the Light of ToM and Metacognitive Functions

Victoria Bamicha
Net Media Lab Mind - Brain R&D IIT - N.C.S.R. "Demokritos"

Corresponding author
Athanasios Drigas
Net Media Lab Mind - Brain R&D IIT - N.C.S.R. "Demokritos"
dr@iit.demokritos.gr

Abstract. Theory of Mind (ToM) and Metacognition constitute two superior mental mechanisms that promote the smooth integration and adaptation of the individual in society. In particular, the ability to read minds introduces the individual into the social world, contributing to understanding oneself and others. Metacognition focuses on individual knowledge, control, regulation, and readjustment regarding the cognitive mechanism and its influence on cognitive performance and the mental and social development of the individual. At the basis of the development of the two mechanisms is the activation of social interaction, which determines their levels of development. The innovative approaches and great expectations of technology and Artificial Intelligence for improving the artificial mind brought social robots to the fore. Robots with social action are gradually entering human life. Their interaction with the human factor is anticipated to become more and more frequent, expanded, and specialized. Hence, the investigation of equipping artificial systems with integrated social-cognitive and metacognitive capabilities was necessary, constituting the subject of study of the current narrative review. Research findings show that intelligent systems with introspection, self-evaluation, and perception-understanding of emotions, intentions, and beliefs can develop safe and satisfactory communication with humans as long as their design and operation conform to the code of ethics.
Keywords:human-robot interaction, social robotics, theory of mind, metacognition, metareasoning

---

Introduction

The International Federation of Robotics Statistical Department points out the gradual integration of robots into human reality. However, an effective human-robot social interaction requires the involvement of complex human models in the artificial system including their desires, beliefs, goals, knowledge, emotions, and the context where the interaction occurs (Yang, Dario, &Kragic, 2018). In addition, human-robot cooperation of whatever kind will impact how they communicate, so it should be distinguished by quality interaction and trust from the human towards the robot and the information it provides (Vinanzi et al., 2019).

The formation of increased skill requirements of individuals active in the robotics field is related to the direct interaction of humans with intelligent systems. The evolution of artificial intelligence promotes the creation of robots with capabilities that involve semantic perception, reasoning, and the integration of robotic networks with web services. The creation of intelligent systems that combine the perception of a dynamic environment with the corresponding action indicates the utilization of different scientific fields such as informatics, management, programming, cognitive science and psychology, informatics, management, programming, cognitive science, and psychology (Shmatko & Volkova, 2020).

Developments in artificial intelligence with an emphasis on robotics have allowed the transition from rigid position-controlled robots applied to typical automation tasks to the creation of intelligent systems that are part of the research field of soft robotics (Haddadin & Croft, 2016).

Artificial Intelligence (AI) systems try to imitate human processes to solve various problems in life. They perform human cognitive functions and tasks of varying grades of complexity in a manner comparable to that of a human. The development of programs for intelligent systems that solve computational problems, research on the methods that the human mind uses to solve problems, the study and creation of a network of artificial neurons that are inherent in the human nervous system, and the design of intellectual programs that promote self-learning constitute dominant fields of study and investigation of artificial intelligence (Iasechko et al., 2021).

Decoding the behavior of intelligent systems characterizes a human mental function, where the person interacting with them creates mental models to interpret and predict their behavior. The human agent leverages ToM and concludes, which may be wrong, particularly if the AI systems it interacts with are inadequate. The more refined a robot's mental model is, the more it resembles a thinking machine that communicates satisfactorily with humans (Wortham et al., 2016).

When an autonomous intelligent system can explain its behavior in ways that humans find understandable, humans are more likely to form correct mental models of such a system and calibrate their trust in it (De Graaf & Malle, 2017). Lasota et al., 2017 point out that the robot's ability to correctly predict a human agent's actions, and understand his intentions, leads to the right choice of decisions and creates safety in the interaction between them.

For a social robot to be understood and accessible by humans, it should possess communicative characteristics and abilities for effective communication with humans. Also of crucial importance is the technical implementation of these abilities in robots. Social-emotional intelligence and social-cognitive skills are considered necessary capabilities in a robot based on the circumstances in which it acts. Thus, emotional state recognition, processing, and behavior prediction are essential for human-robot interaction. In addition, the form of the robot that approximates the human presence is a significant tool in the mutual interaction with the human, as it supports it even more. Essentially, the anthropomorphic design of a robot endows it with characteristics and abilities by prompting humans to cooperate with it (Schleidgen& Friedrich, 2022).

Several researchers are exploring the integration of various aspects of intelligence into the robot to make it autonomous. What distinguishes its autonomy is the creation and choice of actions intended to fulfill goals based on knowledge and understanding of the world. According to various studies, autonomous robots should present reflective and reactive abilities, mainly involving decision-making, the ability to respond immediately to unexpected events, the ability to predict a situation, and the awareness of the context in which they are called upon to act (Gutiérrez & Steinbauer-Wagner, 2022).

Chen et al., 2013 state that interactive and cooperative robots with humans must possess autonomy, adaptability, and sociability. Consequently, functions such as planning, learning, and dialogue that act in concert are of particular importance for the effectiveness of an artificial agent.

An entity possessing a complete cognitive system, contemplating itself, its actions, and their consequences in the environment in which it acts, engages many different metacognitions. In the case of an artificial agent that is an integrated version of a cognitive mechanism, self-models' existence with meta-reflective action allows its integration into social contexts. Thus,the agent by use of reflection, evaluates his performance, assesses the consequences of the meta-deliberative process, and can formulate new strategies based on his experiences (Cox & Raja, 2008).

He et al., 2021 point out that a reliable autonomous system should be in power over by properties concerning its reliable and non-hazardous action. First of all, it should be safe, able to perceive the changing conditions of the environment, and act immediately, keeping to the predetermined goal. It is fundamental to synchronize in real-time, deal with unexpected failures, and act safely in situations without previous experience, allowing human intervention when required. In addition, the integrity of the software and the secure materials that make it up will limit the possibility of malfunction, loss of control, and lack of communication.

The gradual and continuous adoption of robots in human life and the expected interaction of artificial-biological agents cause an urgent need to equip robots with functions that enhance their communicative and collaborative role in society. Capabilities related to understanding and predicting emotions, intentions, and social behaviors, with the ability to self-assess the robot's choices and actions in the environment, are directly related to ToM and Metacognition skills. Keeping in mind the increasing and multifaceted participation of robots in society, we consider it necessary to study the upcoming integration of processes that promote their smooth and safe integration into the human world.

Methods and materials

The literature review explored the possibility of incorporating skills related to metacognition and mind-reading into human-interacting robots. In addition, the study focused on the effect of social-

cognitive empowerment of the artificial agent on human-intelligent system mutual communication. The authors approached the research topic methodologically, utilizing a narrative review. The specific method contributes to the examination and understanding of knowledge by involving the interpretation and criticism of data (Greenhalgh et al., 2018). In addition, it studies and presents the scientific evidence extracted from the research, following a theoretical and contextual perspective (Rother, 2007). Human-robot interaction, social robotics, theory of mind, metacognition, and metareasoning were the search terms used to find sources in the international bibliographic databases Scopus and Google Scholar. Articles written in English and published in respectable, peer-reviewed scientific publications between 1962 and 2024 met the selection criteria. However, research data from the most recent ten years, 2014–2024, was cited in most sources. The exclusion standards mentioned publications with bibliographies, whose interpretation and analysis of the data lacked clarity and only addressed how the many subjects of the current study related to one another. The course of the research followed the following stages: definition of the research topic and keywords, search, selection-exclusion, and classification of the sources according to the individual sections of the central topic. Composing the paper, which drew from an entire collection of 114 sources (books and articles), marked the completion of the research. The extracted findings point to the necessity of enriching robots with ToM and Metacognition abilities in light of the morality and security that define their usage. It is functional for them to develop effective communication with humans, providing a social, cooperative, and supportive role in their interactions.

Theoretical approach to the concepts

Social robotics

Studies argue that robot design primarily focuses on the development of cognitive function, involving capabilities related to programming, reasoning, navigation, manipulation, and then skills related to social cognition. However, constructing an intelligent system that acts in a social environment presupposes the development of a socially intelligent robot (Dautenhahn, 2007). The social robot was created to interact with humans while maintaining a social character in their communication (Baraka et al., 2020).

A dominant characteristic of social robots is the development of interpersonal relationships with humans. The artificial system communicates and coordinates its action with the human, using verbal and non-verbal messages and emotional-cognitive cues (Breazeal et al., 2016).

Social robotics creates autonomous or semi-autonomous robots seeking to interact, communicate, collaborate, and teach new skills to other agents. Social robots are useful in different contexts and for various purposes, and they use a variety of communication modes when interacting with humans. Limb or whole-body movement, facial expressions, gestures, gaze behavior, head orientation, and linguistic or emotional vocal expression are basic but essential forms of communication and joint action with the human agent. However, an important factor in the joint action of people is the mutual understanding of thoughts, feelings, and intentions. This ability to understand others appears in the cases of robots with the possibility of statistical predictions of human intentions and actions (Schleidgen& Friedrich, 2022).

Javaid et al., 2020 in their study report that social robots are increasingly integrated into human environments every day, complementing the capabilities of humans with their skills. Effective human-machine interaction breeds constructive cooperation in robots. However, the rise in their autonomy and functionality can cause inexplicable and unpredictable events in humans during communication, shaking their faith in their abilities. Therefore, the possibility of justifying their choices allows transparency and the development of the human factor's trust in them.

Numerous studies back up the idea that social robots that communicate with people can express and perceive emotions, use dialogue to communicate, recognize other agents, establish social relationships, exhibit elements of a personality with physical cues such as gaze and gestures, and learn social skills through experience (Dautenhahn, 2007; Drigas & Papoutsi, 2023; Karyotaki et al., 2024).

In addition, social intelligence systems draw elements of social knowledge from humans, having the corresponding behavioral models, creating and maintaining social relationships, though their capabilities differ according to their expected use (Dautenhahn, 2007; Baraka et al., 2020).

Wiese et al., 2017 point out that for robots to be social partners in their interactions with humans, their behavior should activate social cognition processes in the human brain. These functions are active through social interaction and concern joint attention, context perception, action understanding, and flexible situation management, which characterize the utilization of ToM. Essentially, two key factors that highlight a robot as a social agent are its appearance and behavior, which are related to the design of its software and overall operation. The inclusion of human social and cognitive mechanisms in their design is noteworthy, considering the knowledge of neuroscience about the functioning of the human brain in interacting environments.

Theory of Mind (ToM)

Theory of Mind constitutes a bridge of mental understanding between us and others, providing possibilities for communication, effective social interaction, and development at a cognitive-

metacognitive level (Bamicha & Drigas, 2022a, b). Tom appears in infancy and gradually develops into preschool and school age through manifestations of social communication like empathy and behaviors involving cheating and bullying. It is distinguished by the recognition of subjectivity in others, utilizing observation and the ability to simulate the behaviors of interacting individuals (Banks, 2020).

The ability of ToM goes hand in hand with the cultivation of social skills in preschool age. The development of ToM skills contributes to the development of social competence, which is vital to children's academic achievement. ToM plays a crucial part in a person's social knowledge and affects every aspect of his life (Rakoczy, 2022). In predicting the behavior of others, the ability to read minds requires reasoning that includes both the results of their corresponding actions and intentions. Since intentions are not observable, they should be derived from the overall social information that the individual draws and his cognitive capacities (Frith & Happé, 1999; Miranda et al., 2017; Bamicha & Drigas, 2022a). Linguistic skills and executive function improvement are considered the cognitive bases in the evolution of ToM skills. It is due to the ability to flexibly coordinate multiple perspectives that executive skills allow, enhancing the meta-representational process (Rakoczy, 2022).

According to research, people, since childhood, participate in social interactions and use expressive language, developing executive skills and language ability. This results in the strengthening and steady improvement of his performance at a cognitive and metacognitive level. Consciousness has a leading role, which is the most advanced mental phenomenon contributing to different perspectives recognition between the individual and others. Therefore, the ability to read minds and introspection processes are improved (Rosenthal, 2005; Bamicha & Drigas, 2022a; Fabbro et al., 2019; Wang & Frye, 2021; Astington& Jenkins,1995; Brock et al., 2018). According to Vygotsky, 1962 individual social-emotional, cognitive, and metacognitive development arises through social interaction.

Participation in social settings requires empathy and communication between participants. The perception and understanding of emotions, the observation, knowledge, and interpretation of the mental states of others are based on shared representations that lead to conclusions. The conclusion comes from the simulation of the behaviors in the context of the person performing the above mental functions (Wiese et al., 2017;Drigas & Bamicha, 2023a, b).

The ability to read minds allows humans to control and coordinate their thinking and actions, ensuring social behavior that promotes interpersonal relationships in different circumstances. Incorporating similar capabilities into a robot would be beneficial in creating social scenarios involving the human agent (Breazeal et al., 2009).

Metacognition

Metacognition,associatedwith consciousness, is demonstrated by people's capacity to contemplate their mental states. Human awareness is derived from metacognition and is vital for social relationships (Frith & Frith, 2012). Metacognition is a function linked to self-awareness and reflects private observation, control, and the ability to adapt one's cognitive states. Therefore, it includes processes by which cognitive control is achieved by the subject's access to his cognitive mechanism. In addition, it utilizes mechanisms where adaptive cognitive control relies on the use of available information resulting from stimuli in the external environment (Hampton, 2009).

The ability to self-perceive allows us to perceive ourselves in the world, to reflect on how we feel and act in it while at the same time enabling us to self-reflect and improve ourselves. These specific processes require conscious functioning and are related to the metacognitive process (Kralik et al., 2018).

Given the close relationship of executive functions with metacognition, metacognitive skill is frequently evaluated by the individual's tendency to use these skills successfully in problem-solving situations. Implicit performance of self-regulatory skills and conscious articulation of knowledge are highlighted as critical reflections of metacognitive ability (Whitebread et al., 2010).Moreover, higher-level metacognitive functions such as mindfulness are associated with higher cognitive skills, enhancing self-esteem, emotional intelligence, and cognitive functioning, and contributing to the growth of social and interpersonal relationships (Drigas & Karyotaki, 2018).

Metacognition is best described figuratively as "the silent dialogue of thought," or, in Socrates' words, "the conversation that the soul has with itself about the matters it considers as it thinks, converses with itself," reportedly by Plato (Worley, 2018).

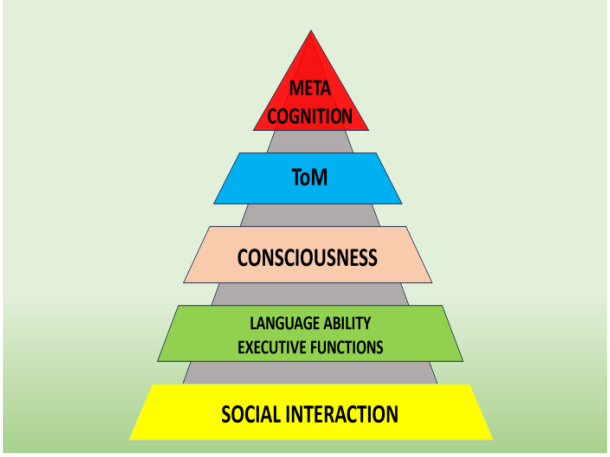

Figure 1. Functionality of human social interaction

Authors Bamicha and Drigas, in this figure, capture the importance of human social interaction, which, using emergent conscious processes combined with executive and language skills, gives prominence to the gradual development of ToM and Metacognition (Frith & Frith, 2012; Bamicha & Drigas, 2023a, b;Fabbro et al., 2019; Wang & Frye, 2021; Astington & Jenkins,1995; Rakoczy, 2022; Brock et al., 2018; Vygotsky, 1962).

Human-Robot Interaction

The collaboration and communication between humans and AI are becoming more and more essential, especially shortly, as the development of technology proceeds at a rapid pace. AI is gradually developing high capabilities in various fields, focusing on data processing and automation. However, the human factor possesses unique capacities such as emotions, ethics, and creativity. Creating constructive cooperation between humans and AI by leveraging each other's strengths increases efficiency and flexible decision-making (Lu, 2023).

The rapid progress of technology utilizing faster processors and suitable algorithms has helped to reshape human communication. In addition, the creation and evolution of intelligent systems that can understand and use natural language enables the study of fundamental principles of human communication. Furthermore, network operators can more easily employ artificial intelligence to automate communication thanks to computer platforms. Artificial intelligence, by analyzing patterns of human behavior, provides information that enhances and improves communication skills, for the benefit of people, in the light of responsibility and ethics(Danso et al., 2023).

Artificial systems can process large amounts of data in real-time by exploring structures that enhance human decision-making and problem-solving. In the educational process, artificial intelligence influences the way learners and teachers approach solving complex problems by providing personalized insights and information (Chaidi et al., 2021; Pergantis & Drigas, 2024; Joksimovic et al., 2023; Drigas et al., 2023;| Vouglanis, 2023).In addition, the use of innovative approaches enhances the flexibility, adaptability, and self-efficacy of people with special educational needs, promoting their social, cognitive, and metacognitive development (Bakola et al., 2022; Pergantis & Drigas, 2023; Drakatos et al., 2023; Bamicha & Drigas, 2024; Pergantis, 2024).

Specifically, human-AI interaction to deal with problematic situations involves the social, emotional, cognitive, and metacognitive processes of both parties, to the extent that each is concerned. Cognitive and metacognitive mechanisms are two operating systems that each adapt accordingly to the data it receives. Consequently, man can improve his performance, and artificial intelligence, enhancing its function, can meet human needs (Joksimovic et al., 2023).

Communication between a human and an agent is more effective when the AI system has ToM elements that allow it to assess the beliefs and intentions of others in a social context. In addition, the agent's metacognitive control allows flexible handling of dynamic and complex environmental conditions. Integrating ethical reasoning into the intelligent system through attentional encoders that can detect and process moral information would encourage decision-makers to weigh moral factors before acting (Wallach et al., 2010).

According to studies, robotics is expected to have a dominant role in the coming years in human life, as it will shape new conditions in the working environment, increase efficiency, provide improved and safe services, and create new jobs. The place of intelligent systems in the real world will be strengthened, given their expected interaction with humans. Consequently, robotics is a social issue with several social and cultural challenges, which the human factor must face, especially since the development of innovative ideas and inventions touches its management (Kaivo-Oja et al., 2017).

According to De Santis et al., 2008 safety and reliability are considered key parameters in the effective integration of robots into the human world. Studies have explored social factors that facilitate the integration of the artificial agent into the human environment, emphasizing cognitive interaction with intelligent systems. However, physical-cognitive interactions are linked and complement each other. The former contributes to defining rules that enable cognitive evaluations of the environment, while the latter facilitates physical interaction creating the basis for controlling communication between participants.

Research interest is shifting due to the advancement of robotics into robots' incorporation of soft skills. However, the autonomy of artificial systems that gradually develops through interaction and experience does not abolish human-supervised learning. A reality, that would result in the skeptical behavior of the human about the quality of communication and interaction with the robot (Gloor et al., 2020). In particular, human trust in the robot is enhanced when individuals can perceive and comprehend a robot's function. It is due to the system's transparency, which enables people to accurately form a mental model of the robot's capabilities, which encourages cooperation between them (Pipitone et al., 2023).

Interactive robots leveraging knowledge from various scientific fields should be able to perceive communication data from the human environment. In addition, they should respond adequately, predict human behaviors, recognize faces, emotions, and intentions, and process natural language (Cross et al., 2019). For a robot to interact in a safe, natural, and autonomous way in a human environment, its design must encompass essential

components related to perception, learning, and cognition. Incorporating human security and physical interaction into the cognitive decision-making layer of the intelligent system is beneficial for mutual communication (Haddadin & Croft, 2016).

The recent study by Basich et al., 2023 focuses on the development of cognitive capabilities of the intelligent system to cope with unpredictable conditions and enhance its autonomy. Among the primary functions of a robot that develops communication with a human agent in the real world is the awareness of its abilities, the ability to understand and process information it receives from the external environment alongside its capacity for action planning. Additionally, programmability acts like a forecast for the choices and outcome of an artificial system goal, assessing the need for human-level assistance and adequately covering safe operations.

Human-robot social interaction incorporates robots where, through their communication with humans, the ultimate goal is to educate, entertain, and assist the human agent in the upcoming interaction (Sheridan, 2016). Socially interactive robots successfully interact in a social environment emphasizing cognition, social behavior, and physical interactions such as verbal, visual, and most often non-physical with human agents. Notably, recently robots have developed greater autonomy, achieving mutual influence with humans to accomplish a goal (Losey et al., 2018).

Natarajan et al., 2023 typically state that human-robot collaboration requires appropriate communication that is directly related to the robot's level of autonomy and human supervision. Effective communication between them should be governed by sending the required data and limiting the unnecessary, to lead to decision-making to fulfill the objective. It is essentially to integrate into an artificial system the ability to reason about social elements and contexts when interacting with people. The ability of the intelligent system to know different behaviors and adapt to new unstructured environments that are gradually evolving will prove beneficial for human-robot interaction.

The design and operation of the robot are considered particularly important, as it has a powerful impact on human-robot interaction. Therefore, robot categorization relies on its morphology, autonomous level, and preprogrammed task. In addition, human role, team makeup, communication style, and physical and temporal closeness are all considered whichever applies—in each of the HRI scenarios (Onnasch& Roesler, 2021).

Additionally, according to their level of autonomy, robots are categorized into autonomous and remote-controlled robots. The autonomous robot has a high level of autonomy and a low level of human intervention. Conversely, an appliance that needs to be operated remotely shows a low degree of autonomy and a high degree of human intervention. Studies reveal that when it comes to a remote-controlled robot vs an autonomous one, people exhibited greater emotional empathy and felt more secure. It possibly happened because, during the communication and emotional involvement between them, the autonomous robot expresses its own emotions compared to the remote-controlled one transmitting the emotional messages of the person handling it (Choi et al., 2014).

Onnasch& Roesler, 2021 presented an HRI human-robot interaction taxonomy that includes the human, the robot, the interaction, and the context in which the interaction between artificial and human agents of robot application takes place, where depending on the services provided by the intelligent systems, we distinguish professional, military, police, and space robots. Social robots, which have a wide range of applications, including therapeutic ones, play a significant role. Also noteworthy is their application in the educational field, research, as well as the fulfillment of an entertainment purpose. One crucial element is the environment in which the agents' interactions occur, such as whether the exposure to a robot takes place in a natural or a controlled laboratory environment, which could influence human perception and action toward the intelligent system.

In their study, Haddadin & Croft (2016) highlighted three more general categories of human-robot interaction characterized by a supportive, collaborative, and cooperative role. In supportive interactions, the robot enhances the human's performance with information, tools, and materials to achieve the predetermined goal in the best possible way by the human. During collaborative interaction, a human and an artificial system work in parallel to complete their assigned part of the work, carrying out the joint task they have undertaken. Whereas in collaborative interaction the robot assumes an independent rather than passive role.

Worthy of mention is the observation of Jung & Hinds, 2018, who argue that a robot's behavior and actions influence the person it communicates with, as well as its wider social environment. It is proven by research that a robot having the role of a mediator can influence the people related to it in a task.

Also noteworthy are the cases of children with ASD, where their approach and education through innovative applications of technology and AI, with an emphasis on social robots, strengthen their social skills, contributing to the formation of their social and cognitive behavior (Ntaountaki et al., 2019; Mitsea et al., 2020; Sideraki & Drigas, 2021; Drigas & Sideraki, 2021; Syriopoulou-Delli et al., 2021; Moraiti& Drigas, 2023; Bamicha & Salapata, 2024).

In addition, research shows that the robot's interaction with humans can influence how people communicate with others, the roles they assume, and, in general, the formation of their norms in different social groups (Jung & Hinds, 2018)

Effects of integrating aspects of ToM in the Human-Interacting Robot

Social robotics requires the collaboration of various scientific disciplines of artificial intelligence, psychology, medicine, neuroscience, and social and cognitive sciences to enable human-robot interaction effectively. The social robot should be able to process and understand human behavior, adhering to ethical and social rules. Several studies report their use for both healthy and special needs populations. It's important to note that the targeted intervention of a social robot can socially and cooperatively enhance an activity, particularly in the cases of children with ASD, who present significant social and emotional deficits. Studies describe progress in social robotics as the ability to detect human intentions, recognize actions, process, and understand emotions and parameters related to human psychology (Yang et al., 2018).

Social robotics systems utilizing knowledge from the fields of biology and psychology display social behaviors such as emotional facial expressions that attract attention, inspire trust, and theory of mind skills. However, for social robots to interact effectively with humans, it is necessary to understand social messages from humans, possess reasoning abilities, use natural language, and act according to the circumstances at hand (Cross et al., 2019).

Social and interactive behavior fosters human-robot cooperation, interaction, and communication. Human communication involves body posture, gestures, and facial expressions since they are valuable information sources. For the robot to maintain a social engagement with humans, it must understand and interpret the social messages that permeate their language and behavior. In "Cutting Edge Robotics", robots can analyze some linguistic aspects of human communication however, they are limited to coordination and adaptation with their interlocutors (Cerrato & Campbell, 2017).

In addition, the research of González-Docasal et al., 2021 focuses on the importance of understanding voice commands by the robot in the process of natural language communication with the human agent. According to the researchers' study, this interaction can be improved by making the system more robust to noisy audio conditions, better parsing semantic signals from the environment, and using a knowledge manager that uses contextual information before the robot receives the instruction. Completing their research, they found high usability and trust in the system.

Man can conclude the mental states of others through meta-representation, allowing the development of social interaction. These are ToM abilities, which give one the chance to understand and forecast human behavior. Essentially, the person compares and interprets his mental state, taking into account the thoughts, beliefs, and intentions of others. In his attempt to communicate with artificial systems, man uses social heuristics. Sometimes, it attempts to understand the behavior of a robot, which displays social cues that are similar to human ones and can interpret them. However, robots exhibit individual aspects of ToM (Banks, 2020).

According to the literature, the attribution of mental states to robots uses various terminologies, the most dominant of which are mental state attribution, anthropomorphism, mind perception, theory of mind, and mentalizing. In contrast, there is a decrease in the usage of the phrases deliberate posture, folk psychology, and mind reading. However, although the terminology is different, it converges with the concept of mental state attribution. The term anthropomorphism is used to emphasize the similarity of the external appearance of the robot to the human. Humans attribute mental states to robots based on their ability to interact or because it allows them to understand, predict, and explain the robots' behavior. Furthermore, attributing mental states to the robot has been found to reduce human anxiety and uncertainty while increasing control over social interaction with them. Additionally, people's tendency to attribute mental states to robots is related to the robot's age, motivation, behavior, appearance, and identity (Thellman et al., 2022).

Hegel et al., 2008 typically state that the social robot constitutes an interface between humans and AI. A robot's ability to communicate effectively increases with its degree of resemblance to human appearance and behavior. Surveys report many of the non-verbal messages come out in facial expressions. Consequently, a major factor in the robot's external appearance is the design of the head, enhancing its communication with the human.

SCASSELLATI, 2002 applied some aspects of ToM to a humanoid robot at the MIT Artificial Intelligence Laboratory. His study presented the integration of visual attention, face detection, recognition eye tracking, and animate-inanimate discrimination in the Cog robot. His research has been highly influential in the development of social robotics. Characteristically, he pointed out that the progress of humanoid robotics, especially artificial systems that are going to interact socially cooperatively with humans must incorporate features of ToM. It would enable the robot to learn through observation and could express internal states such as goals, desires, emotions, and thoughts. In addition, the robot would be able to recognize and process the desires and goals of others by predicting their behavior and varying its actions according to the circumstances.

After a few years, Breazeal et al., 2009 developed a social-cognitive architecture that leverages people's capacity to mentally mimic others to gain a deeper understanding and interpretation of behavior. The robot uses simulation mechanisms, in real-time, to collect information

about the human's beliefs, and intentions through the observation of its movements and visual perspective. Furthermore, it uses these inferences with similar mechanisms to infer details about its intentions, beliefs, and behavior.

Martini et al., 2015 point out that the direction of a gaze, body posture, and facial expression are non-verbal cues, which are considered essential in the perception and communication of emotions, intentions, and preferences during social interaction. Therefore, the integration of these non-verbal signals in human-robot communication is of particular importance in their interaction and facilitates cooperation between them. Eye gaze provides a lot of information about both the environment and the intentions, motivations, and preferences of others. Furthermore, it is related to joint attention is linked to the development of social relevance in a given interaction, as well as the development of ToM. According to studies, gaze tracking can increase when an agent is more human-like, given that it is perceived as having a mind and social relevance, influencing the beliefs attributed to it.

Studies report that people engage in joint attention with a robot when they perceive it signifies acting purposefully and with goal-directed behavior. Therefore, it is a system with human-like intentions, goals, and mental states. Consequently, considering humans and some robots as targeted agents, the human's interpretation of the robot's and other humans' behavior is enhanced by the same or overlapping biological mechanisms (Thellman et al., 2017).

Vinanzi et al., 2019 state that the ability to assess the thoughts, beliefs, and desires of others is related to the ability to estimate their credibility and specifically to the development of self-esteem. It is due to the one-to-one ability to predict behaviors and perceive signs of reliability. Trust is a dominant component in both human relationships and robot-human interaction, as it improves communication and enhances the credibility of the artificial agent. The researchers created a humanoid social robot that could interact with people by assessing their trustworthiness, predicting their behavior, and determining their course of action.Episode memory and Theory of Mind were ingrained in the robot, allowing it to remember past experiences and shape its behavior accordingly, improving its cognitive abilities. Also, he could differentiate his actions according to his beliefs, developing a model of ToM. The bot uses machine learning techniques, detection, and facial recognition algorithms to distinguish between the different people it communicates with.

Developing long-term human interaction with a robot is a significant challenge. Robots that combine many human characteristics and cognitive mechanisms that simulate the human decision-making process, including ToM abilities, enhance cooperation with humans and promote successful interpersonal relationships. But a key element that amplifies their efficacy is people's confidence in the artificial system. Since trust is a dynamic process that develops according to past experiences, affecting upcoming emotional relationships, the robot can act as a reliable and stable helper for children with fragile emotional relationships (Di Dio et al., 2020).

Görür et al., 2017 proposed incorporating Theory of Mind into a robot's decision-making to understand and infer human intentions, varying its behavior accordingly. The robot evaluates the human's intentions by observing his actions. By incorporating human emotional variability and human-robot collective decision-making into its design, it can act cooperatively toward humans rather than intrusively, understanding that it needs their help.

A subsequent study emphasized the significance of comprehending how developmental robotics leverages human intent. In particular, Vinanzi et al.'s 2021 research highlights the cognitive and evolutionary importance of the ability to understand the intentions of others, which requires the development of ToM. Human-machine communication and collaboration presuppose a shared understanding of common goals and intentions for completing a task. The researchers relied on the theoretical background of developmental robotics, namely current experience rather than pre-existing knowledge of the robot. In particular, the humanoid robot iCub gradually observes the actions, and movements of the human partner, trying to predict his intentions, connected to the accomplishment of shared objectives, helping when necessary.

Several studies argue that viewing robots as agents with minds and the ability to act purposefully can positively influence the performance of mental states and the evolution of human-robot interaction. Social communication uses verbal and nonverbal messages, and the resulting reactions are related to the social relevance attributed to social cues. Changes in gaze direction are important for considering an agent as a minded entity, as these shifts are attributed to intentions, despite its mechanical form. Consequently, movement patterns observed in communicative actions between people are retrieved in human memory, characterizing the behavior of the artificial agent as reliable and intentional (Abubshait& Wiese, 2017).

Human social interaction involves social understanding, combining the focus of attention with the perception of social cues that unfold in real-time. Eye contact is a non-verbal component of communication, helping to interpret social behaviors and sustain everyday interactions. A means of communication, the gaze conveys the partners' interest in one another during the exchange. Xu et al., 2016 argued that a better understanding of human-robot communication comes from an analysis of the sensorimotor behaviors of agents,

which influence each other in real-time. At the same time, they mention the importance of eye contact with each other, as it is a social reinforcement, encouraging interaction between them. In particular, they implemented an experiment where interacting human-robot members participated in a joint-attention task, in which gaze played a crucial role in their communication. The study's findings showed that eye contact facilitates the concentration of attention in human-robot interaction, enhancing the coordination and synchronization of different behaviors.

Humans can develop different approaches when interacting with a robot, evaluating its movement, form, and behavior. Therefore, the human agent uses ToM in its communication with the artificial agent. Being able to inform the robot of the second-order belief system formed by the human about the robot contributes to modeling the robot's beliefs through incremental knowledge. It would result in the detection of errors on the part of the robot, improving its action and cooperation with humans (Brooks & Szafir, 2019).

A fundamental dimension of ToM is the understanding and use of false statements. Advanced robots with built-in mind-reading skills would be able to use lying. In some cases, the "white lie" could function as reinforcement, as in education and medicine, covering teaching techniques and therapeutic methods. Kneer, 2021 in his study states that people equally attribute deception intentions to other humans and robots because they are seen as capable of sustaining a verbal deception, especially if they possess core ToM elements.

Joint action between people is a social interaction in which they coordinate their actions in space and time to create a change in the environment. Clodic& Alami, 2021 argue that realizing robot-human joint action involves fundamental processes such as Self-Other Distinction, Joint Attention, Understanding of Intentional Action, and Shared Task Representation. In the evolution of the joint action, the robot must complete its goal successfully, considering the human reactions. An essential factor in the overall process is the understanding of intentional action, where each agent can recognize the actions of the other. Understanding intentional actions presupposes the ability to predict the intentions of one's partner, that is, their goals and plans.

The use of ToM is a dominant meta-representational skill that enhances the development of social relationships. Better robot-human communication requires increasing the sociability of the artificial system, promoting its adaptability and autonomy. Incao et al., 2021 report the creation of a robot with an artificial Self, considering that this dimension would enable the system to act more precisely in various emotional and cognitive situations. According to this perspective, the robot should acquire a set of distinct characteristics that characterize humans. Some of these are adaptive behavior, facial recognition, joint attention, a complex action style, emotional flexibility and metacognition, and the ability to recognize one's situation and predict the consequences of one's actions in the environment.

According to Castelfranchi & Falcone, 2019 consciousness has to do with intentional action at the individual and social level and Self-formation. Factors necessary to build robots and autonomous artificial systems that interact satisfactorily. Essentially, these are crucial aspects of both ToM and metacognition. Action involving intention requires a form of self-awareness, a meta-representation of mental states and the self. In this case, social engagement encourages the social behavior that occurs. A rudimentary form of self-awareness and understanding of the thoughts, intentions, and beliefs of those involved in the social transaction is distinguished as a prerequisite. The agent's ability to represent his identity in the context of Self-creation has a connection to self-awareness processes. Functions necessary for robots that communicate with humans.

Impact of incorporating aspects of Metacognition in the Human-Interacting Robot

Creating a conscious robot drawing knowledge and inspiration from biological consciousness is a paramount challenge for the scientific and research community. Its construction requires the interaction and collaboration of various scientific disciplines of robotics, technology, psychology, philosophy of mind, ethics, and neuroscience. The ultimate goal of the interdisciplinary approach to consciousness is a robot with dimensions of consciousness, displaying the possibility of self-awareness, evaluating its choices and actions, improving its behavior, and readjusting its decisions during its functional action (Chella, 2022).

Goel et al., 2020 recognize that an intelligent system, such as a robot, is expected to be increasingly involved in human daily life. Therefore, it should have cognitive mechanisms that help it cope with new and different conditions. The study by Chen et al., 2013 portrays an agent's metacognitive function, which relates to meta-level control and reflective observation of reasoning, as particularly important at the metacognitive level. The first sub-process aims to improve decision-making that leads to the best possible result. While gathering information for the meta-level reasoning test is the focus of introspection.

The social interaction of the robot with the human requires higher mental skills such as joint attention, and the performance of spatial-temporal and emotional reasoning, which make it difficult for an intelligent system to act in communication between them. Consequently, the meta-cognitive design of a computational model of a humanoid robot would create channels of communication with

humans and exchange of physical resources. Therefore, their contribution in cases of children with attention deficit disorders or autism spectrum disorders during their mutual communication would be particularly supportive (Mishra et al., 2023).

Potential improvements in human-robot communication could result from the artificial system's reasoning in humans about their decisions and actions. A fact that would inspire confidence, especially if the AI system could interpret its behavior, describing its mental contents and processes, a process that includes aspects ofToM and metacognition. Furthermore, an important function of an artificial system would be to understand human intentions and beliefs and hold the ability to compare them with their own (Castelfranchi & Falcone, 2019).

A strategy for environmental adaptation is the growth of self-awareness, providing the ability to manage mental states and social and cognitive processes. This results in the selection of the optimal decision in ever-changing environmental conditions. Consequently, a robot possessing elements of self-awareness would be able to adapt to unpredictable conditions and communicate better with a human entity. In addition, it would strengthen the human's confidence in his autonomy, as processes related to ToM, awareness of intentions, and free will are included in his autonomous functioning and evaluated by the human (Chella et al., 2020).

According to research, internal speech is a beneficial function for artificial signals. In particular, it is a form of internal self-directed speech, which develops as a result of developmental progression in the child. According to Vygotsky, it provides a distinct purpose from outward speech and grows with socialization. Inner speech involves thought linked to words, which weaken and give way to the creation of thought that contains pure meanings. It is pointed out that thought follows changes, it differentiates before becoming words, that why it is not accompanied by a synchronized appearance of speech (Pipitone et al., 2023; Vygotsky, 1962).

Inner speech is another cognitive tool that emerges from one's social environment and is related to self-regulation, planning, problem-solving, ToM, and metacognition. That is the outcome of an evolutionary course where first the social discourse appears, then the private discourse, and eventually the internal discourse. Metacognition is a process of self-reflection, observation, and control of thoughts that utilizes internal speech, contributing to the formation of the structure of oneself and the external world based on self-attention, self-control, and self-adjustment (Chella et al., 2020).

Various terms referring to inner speech are found in the literature, such as speech, such as inner voice, private speech, inner language, internal dialog, self-talk, and covert speech. However, the best rendering of the term is captured as the subjective experience of language that lacks overt and audible articulation. Regarding artificial intelligence, self-speech is a cognitive field that has attracted the attention of research interest in the last two decades. Computational models included in their design simulations of different forms of inner speech, believing that they contribute to the improvement of language communication and the organization of consciousness. Research shows that agents with speech input perform better than those without such ability. The use of inner speech in intelligent systems aims at automation to improve its functioning, which is related to the development of self-regulation, optimal performance, and self-awareness (Geraci et al., 2021). The case of using inner speech by the robot, involves the reverse propagation of the generated sentences to an inner ear, providing a form of self-awareness. Internal speech reproduces social mechanisms that lead to self-awareness (Chella et al., 2020).

Additionally, Geraci et al., 2021 claim that inner speech is a significant factor related to the automation of artificial systems affecting human trust and mostly the anthropomorphism of automation. According to the study by Pipitone et al., 2019 inner speech is considered essential in the robot's conception that tends to be self-aware. The researchers presented a cognitive architecture where, through perception, messages from the external environment are converted into linguistic data that are stored in the phonological storage space. The central executive controls information in the working memory system and manages the internal thought process. At this stage, the internal monologue is created. Then, the perception of the new conditions and the repetition of the process of the cognitive cycle can follow.

Chella et al., 2020 propose a computational model of inner speech based on the complex interaction of speech recognition and speech production system, short-term memory, procedural, declarative long-term memory. The robot, thanks to the reintroduction of its inner/private speech, describes static and dynamic scenes in front of it, enhancing its situational awareness. In addition, the robot can represent itself, observing and describing its actions, presenting a form of self-awareness. In particular, the artificial system receives perceptual signals from the camera and the internal sensors. It then converts the external signals into linguistic data and stores them in the phonological system. Next, the robot's covert articulation unit generates the viewed object's symbolic form.

In their recent study, Pipitone et al., 2023 report that the robot's internal speech can enhance human understanding and prediction of the robot's behaviors, allowing the creation of an adequate mental representation of the robot. Consequently, an artificial system that simulates the human mechanism promotes the attribution of human characteristics of users to the robot, as well as human-robot trust. Specifically, they used a robot equipped with an internal speech system. By using

explicit self-talk, the robot communicates its "thoughts" and explains its actions to be comprehended. Hence, the approach to human inner speech enhanced human confidence in the operation, safety, and the robot's vitality.

Self-evaluation is a vital function for a robot. The agent can identify modifications that require reprocessing its actions and updating its performance evaluation. However, the assessment of his entire performance promotes the evaluation of his choices, the enrichment of his knowledge, the possibility of redefining his actions, and a more accurate future evaluation. According to Frasca et al., 2020 introspection of capabilities and evaluation of current and future performance are critical capabilities for a robot. Introspection helps assess strengths and weaknesses, while self-evaluation enhances the valuation associated with achieving a goal. The researchers presented a unified self-evaluation framework based on the DIARC cognitive robotics architecture that allows robots to have self-evaluation dialogues before, during, and after a task. The concept was applied to an NAO robot to implement a task.

Görür& Albayrak, 2016 proposed the cognitive architecture CASOR (Cognitive Architecture for Assistive Social Robots) for social robots. CASOR includes basic processes of understanding human behavior and mental states by the robot through experiential learning and user modeling during interactions. The two levels it acts on are the cognitive and metacognitive levels. The first includes sensing, actuation, and memory elements, enhancing the robot to recognize external stimuli through sensory processes. At the second level, the integrated ToM enables the robot to assess the intention and mental states of the person. Then at the metacognitive level, the artificial system integrates them into decision-making and evaluates the effectiveness of its current action. The robot interrupts the cognitive process if the human's objectives are not met and modifies its behavior as needed.

According to Goel et al., in 2020 four cognitive functions contribute to knowledge building, learning reinforcement, and social learning, with particular focus on analogy and Metalogical Analogy and Metaphysics. In the analogy, the AI system has a memory of previous situations it has responded to and is required to solve the current situation by recalling pre-existing knowledge. In the case of meta-thinking, the robot has knowledge of the world around it and has acquired self-knowledge. Therefore, in solving a problem, he uses the knowledge he has. If he fails, he uses meta-thinking by investigating the causes of the failure, to bring about his adaptation. However, depending on the context, the use of these strategies or their combination is also determined for the better perception and action of the agent.

Mishra et al., 2023 with their study present a computational model for humanoid robots, in which they incorporate a computational approach to consciousness and awareness. The robot performs metacognitive reasoning, linking spatiotemporal and affective reasoning skills through a reinforcement learning algorithm. Essentially, the system engages processes such as short-term memory, attention, planning, association, and cause-and-effect analysis for problem-solving and decision-making. The model's working memory, which organizes the observable data and stores the learned verbal claims, is crucial in fostering human connection.

Daglarli, 2020 leveraging data from AI and cognitive neuroscience, presented a study showing enhanced metacognitive aspects in an intelligent system. Specifically, he implemented an experiment that refers to an interaction process, through a game, a logic puzzle where memory-based classification and prediction processes assess cognitive functions. It is a computational model of approaching consciousness, and awareness, involving spatiotemporal and emotional skills for the humanoid robot to perform metacognitive reasoning. The reinforcement learning algorithm leads the development of spatiotemporal and emotional skills such as attention, short-term memory, decision-making, planning, analysis of cause-effect relationships, and problem-solving. A reward system oversees these processes, and cognitive functions remain organized in the working memory of the model. The research findings highlight the importance of incorporating cognitive functions into a social robot, enhancing its social dimension.

Cognitive robotics is the branch of robotics that focuses on designing and building robots that gain knowledge through experience and interaction with others. It blends techniques and expertise, artificial intelligence, cognitive psychology, biology, and neuroscience. Robots store the knowledge and skills in memory, which they recall with flexible action, depending on the context and the goals they expect to carry out. As a result, they can understand what they are doing, defend it, and successfully adjust to changing circumstances. The cognitive robot can infer the intentions and goals of the people with whom it communicates, determining its behavior accordingly, strengthening two-way engagement, a key element of social robotics (Sandini et al., 2021).

According to the study by Zouganeli & Lentzas, 2022 AI systems perform well in performing specific, repetitive tasks but require human supervision. Until now, robots couldn't operate autonomously due to a lack of flexible action and the resulting limitations regarding their safety and reliability. A cognitive robot will be capable of distinguishing dominant goals, integrating and managing new knowledge, acting innovatively, devising appropriate strategies, choosing its behavior, evaluating its performance, and responding to complex tasks, skills, and mental processes currently associated with human intelligence.

Research reports that there are different metacognitive models of artificial systems with similar functions concerning various environments. This results in the lack of a common understanding of the terms and concepts of metacognition in AI systems. Caro et al., 2022 providing a shared comprehension representation of metacognition domains presented an ontology called IM-Onto, which contains key terms, concepts, and relationships for metacognition and computational metacognition. IM-Onto is a semantic model for interoperable metareasoning problems and includes a sub-ontology for each metareasoning problem related to discussion time allocation, effort evaluation allocation, knowledge testing, reasoning problem stopping, computational performance data collection, reasoning failure problem detection, the self-observation, and self-understanding.The successful and sustained interaction of the robot with the human agent requires the gradual acquisition of aspects of social intelligenceby intelligent systems (Görür& Albayrak, 2016).

Williams et al., 2022 typically state that social intelligence comprises a range of skills and skills necessary for successful social interaction. Some of these are the perception and understanding of the internal states and moods of others, the knowledge and use of social norms, the ability to understand, be sensitive to, and manage complex social situations, and the ability to flexibly adapt socially.

Social robots that provide a helpful role to humans should understand their needs and preferences, recognize their mental states, and adapt their actions to the user's wishes through personalized interaction. It is necessary to cover short-term and long-term changes by forming cognitive representations of interacting people. The robot's utilization of human mental states enables the metacognitive assessment of the cognitive process that results in decision-making. An important factor in the success of metacognitive judgment is the user's approval of the robot's action, developing a relationship of trust between them (Görür& Albayrak, 2016).

In his study Jokinen, 2021 mentions that for a social robot beyond its ability to distinguish the intentions of the human it interacts with, it is important to control and clarify its perceptions, organizing its action in a targeted manner. Also, the awareness of the context, the collection of knowledge from the observation of the environment by the artificial agent, and the connection of the discourse with its action are necessary components as they will act as facilitators in the interactive process. Consequently, the development of social skills in an intelligent system plays an important role in the maintenance and evolution of human communication. However, the presence of cognitive skills in the robot contributes to the awareness of the environment, simulating more to the human factor.

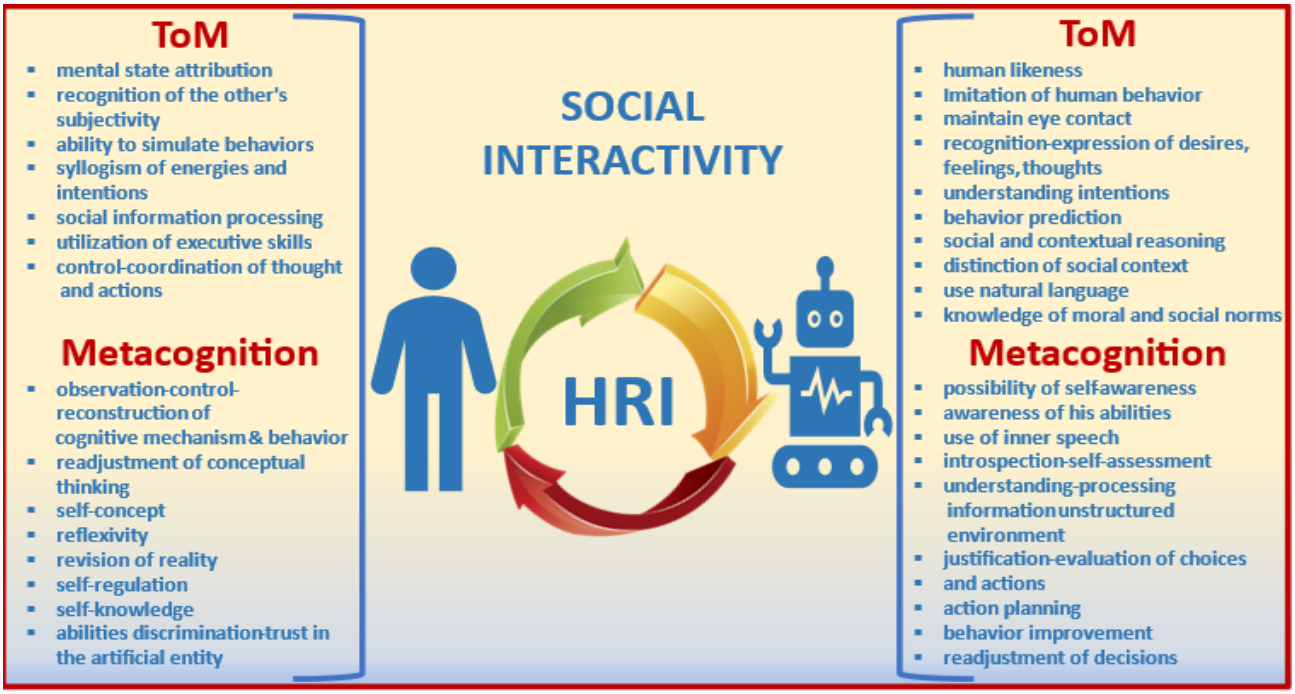

Figure 2. Human-Robot social interactivity under the lens of ToM and Metacognition dimensions

Authors Bamicha and Drigas, taking note of the study's findings, summarized in the figure above the skills required for an effective human-robot social interaction. Competencies refer to the fundamental mental mechanisms of ToM and Metacognition, which determine communication, collaboration, and social engagement levels (Natarajan et al., 2023; Cross et al., 2019; SCASSELLATI, 2002; Vinanzi et

al., 2019; Xu et al., 2016; Basich et al., 2023; Chella, 2022; Castelfranchi & Falcone, 2019; Geraci et al., 2021;
Frith & Happé, 1999; Hampton, 2009; Thellman et al., 2022; Banks, 2020; Mitsea et al., 2020; Miranda et al., 2017; Bamicha & Drigas, 2023a,b; Breazeal et al., 2009; Kralik et al., 2018).

Results and Discussion

Human-centered Artificial Intelligence has as its basic premise reliable human-machine interaction, promoting the mutual correlation between advanced technology and humanitarian-ethical issues. Consequently, the expansion of mental and functional human capacities, as well as technical limitations of human intelligence, emerges, emphasizing the positive effects of artificial intelligence on the human species (He et al., 2021).

Advances in artificial intelligence and cognitive neuroscience have been instrumental in creating different computational models and machine learning applications related to the development of metacognition, which is activated in the prefrontal cortex. Hence, the development of neurocognitive robotics made possible by the study's findings significantly advanced the field of social robotics (Daglarli, 2020).

Reasoning about the mental states of others presupposes the existence of a mind in other agents so that they can form internal mental processes like humans. Machines do not possess mental states but are considered physical entities with built-in programmed behaviors. The perception of some agents as having minds develops an intentional attitude toward them, and they are considered rational entities with goals, desires, and beliefs (Martini et al., 2016).

The design and creation of integrated computational models of social knowledge in a social robot requires the collaboration and contribution of various scientific disciplines. In particular, cognitive interaction between social neuroscience, artificial intelligence, robotics, computational linguistics, and the disciplines of psychology is necessary for effective social signal processing and understanding of human social behavior by intelligent systems (Cross et al., 2019).

The groundbreaking development in AI and machine learning has cultivated human-robot interaction in different kinds of disciplines, related to information retrieval, medicine, education, and even navigation. A necessary mental capacity to achieve cooperation between the two actors is the mastery of the principal dimensions of the theory of mind. In particular, second-order mental processes are practical to complete the meta-representation of mental states (knowledge, beliefs, desires, goals) between them (Kneer, 2021).

According to Incao et al., 2021 the rapid developments in social robotics tend to create systems that seek to imitate humans not only in their physical appearance but also in their behavior, as it is manifested in the various social interactions. To achieve efficient robot-human communication and cooperation, this would be highly advantageous. Several studies have focused on incorporating aspects of ToM into the robot, given that humans use the same processes to attribute intentions and beliefs to the robot, to which they attribute human characteristics.

Comprehending the individual tastes and demands of the individuals they engage with is essential for social robots, and they must modify their actions accordingly. However, according to research, a robot lacks adaptability to the ever-changing emotional conditions of a person, affecting the evolution of their communication (Görür et al., 2017).

Ensuring human safety while interacting with a robot is of utmost importance. It includes the prevention of conflicts between humans and robots acting in the same space, as well as any negative physical or psychological impact arising to the person from an unpleasant or dangerous communication with the artificial agent (Lasota et al., 2017).

In addition, an important issue is the performance of reliability in a robot or even in an autonomous system, which is related to several factors. The transparency of decision-making, the quality, smooth performance, and the correctness and reliability of the principles that govern the handling of erratic circumstances contribute to the formation of trust in the operation of the artificial system. Also, the features of the system that allow safe navigation, combined with social and psychological elements enhance its reliability. However, no matter how autonomous a system is, human interaction is necessary, as it confers the capacity to control the system whenever the situation calls for it. An assertion that intelligent systems complement human intelligence rather than displace it is supported by this fact. The importance of including ethical criteria in the design and operation of autonomous and intelligent systems is highlighted to prevent bias, deception, opacity, and the invasion of privacy when using them (He et al., 2021).

Natural and safe human-robot interaction is a crucial factor in human-to-human communication and presupposes the ability of perspective-taking (ToM), adaptation, and goal-directed behavior. The reasons behind man's actions are derived from his intentions, which also dictate his aims and point of view. Essentially, intention involves goals and actions in a future context. In situations involving joint actions, mutual communication includes common intentions, goals, and primordially the capacity to discern the intents of others. ToM is directly related to the previously mentioned procedure, where one agent can recognize another's perspective on a particular condition. Imitation contributes significantly to the development of ToM, as it facilitates the inference of an agent's

intentions. By exploiting internal simulation an agent can predict the action of another, its consequences, and the intention of his actions. Therefore, modeling these incrementally developing processes would enhance human-robot interaction (Vernon et al., 2016).

The design of robots combines principles of association and simulation to enable them to extract information about the social world and recognize and predict behaviors through observation. However, the process of complex mental states by a robot remains a challenge. The perception and differentiation of social contexts from artificial systems are related to data analysis resulting from environmental stimuli and the embedded knowledge they possess, limiting their adaptability. Enriching the robot with aspects of ToM would help uncover the hidden mental states of the human agent, providing the ability to form beliefs and derive intentions from the robot itself. The ability to predict actions by the AI system would promote its successful involvement in different social environments, which present variability, preparing its action within them (Bianco & Ognibene, 2019).

Modeling human behavior is a complex cognitive process linked to aspects of social communication and is considered a challenge for advancements in robotics. Various studies have attempted to model the behaviors of artificial systems by focusing on symbolic or specific sensory data and knowledge resulting from previous actions. Chen et al., 2021 made an effort to create ToM for robots or machines so that they could, through imitation, perceive the perspective of others. A fact that would allow the understanding of the metacognitive mechanism to get even closer to human behavior. Specifically, they employed a nonverbal, nonsymbolic method robotics experiment, where an AI system using visual processing, with no prior symbolic information or knowledge, predicted a robot's future actions. The use of visual symbolism can contribute to the development of the social abilities of artificial agents.

A successful human-robot interaction (HRI) requires the creation of an integrated architecture that understands, processes, and controls different software components that facilitate the execution of multiple tasks and capabilities. The storage of previous events and experiences and the modeling of actions, beliefs, desires, and intentions of others, which constitute the construction of knowledge for the selection of decisions and actions, are among the fundamental skills of an interacting robot. In addition, cognitive abilities related to perception, memory, attention, optimal action selection, reasoning, and Metareasoning integrated into the design of an intelligent system can enhance multiple interactions, but mainly personalized interaction (Rossi et al., 2022).

The study of cognitive systems includes metalogical or computational metacognition that constitutes the basis for high-level decision-making, introspection, and self-evaluation. Applying metacognitive processes to intelligent systems can act promptly and efficiently, enhancing the decision-making process. However, heterogeneity is a hallmark of metareasoning (Caro et al., 2022).

In light of that autonomous systems are increasingly entering human society, assuming a social-collaborative role, the growth of ethical proficiency in their operation is a prerequisite. According to Malle & Scheutz, 2019, the moral competence of a robot would create safety, trust, and acceptance by the human agent. Researchers consider that moral judgment, moral action, and moral communication as processes, combined with the presence of moral rules and vocabulary constitute the necessary components of the moral capacity of an intelligent system. They even support the importance of learning and reasoning in forming social and moral agents of a robot, which acts with a social character.

The rapid development of technology has seen positive developments in health, information management, productivity, security, and knowledge delivery. However, it highlighted weaknesses related to the management of personal data, excessive expectations, and issues of reliability and trust due to increasing technological complexity. Therefore, human training and expertise, technology management, focus on user interfaces and creating experiences, and good governance are challenges that will limit any adverse consequences of technological achievements (Kaivo-Oja et al., 2017).

It is highlighted that the amalgamation of inventiveness and troubleshooting with the parameters of Diversity, Equity, and Inclusion in the human environment could create robots that support diversity, limiting existing inequalities. It would possess the ability to encourage long-term interaction with humans and the acceptance of the artificial agent by the broader societal context (Natarajan et al., 2023).

Conclusions

In summary, the man who invented artificial intelligence is on the path of rapid and continuous developments in technology and science. Therefore, he has to co-create with intelligent systems the foundations of a different society, where humans and machines co-exist, cooperate, and accompany harmoniously in the future's voyage, aiming at the optimal management of all resources for human well-being. However, a few issues were raised, concerning the ethical demarcation and the scientific limitation in the introduction of technology in the daily life of man, specifically the artificial factors in the personal social life. Coordinated human-robot social interaction with anthropocentric characters presupposes skills that enhance mutual communication, social cognition, and cooperative action. Higher mental mechanisms such as ToM and Metacognition are gradually integrated into the

operation of intelligent systems aiming at better social and cognitive support for the human factor. However, further studies are needed to cover aspects related to the ability of robots to understand the ingenuity, creativity, and flexible adaptation of humans.

It is recommended upcoming research looks into the potential enhancement of robot sensors and interface means to provide greater flexibility in decoding information about human behavior. In addition, it would be beneficial to integrate social competence into the robot from the standpoint of transitioning from a structured communication to a dynamic and unpredictable interaction environment. Furthermore, the use of intelligent agents in the mutual engagement of a social nature mainly focuses on short-term interactions. Thus, searching for conditions that allow long-term human-robot interaction while maintaining social relevance, quality communication, and the necessary rules of ethics and morality would constitute a crucial research challenge.

References

Abubshait, A., & Wiese, E. (2017). You look human, but act like a machine: agent appearance and behavior modulate different aspects of human–robot interaction. Frontiers in psychology, 8, 277299. https://doi.org/10.3389/fpsyg.2017.01393

Astington, J. W. & Jenkins, J. M. (1995). Theory ofminddevelopment and social understanding.Cognition & Emotion, 9 (2-3), 151-165.https://doi.org/10.1080/02699939508409006

Bakola, L. N., Drigas, A., & Skianis, C. (2022). Emotional Intelligence vs. Artificial Intelligence: The interaction of human intelligence in evolutionary robotics. Research, Society and Development, 11(16). http://dx.doi.org/10.33448/rsd-v11i16.38057

Bamicha, V., & Drigas, A. (2022a). The Evolutionary Course of Theory of Mind-Factors That Facilitate or Inhibit Its Operation & the Role of ICTs. Technium Soc. Sci. J., 30, 138-158. https://doi.org/10.47577/tssj.v30i1.6220

Bamicha, V., & Drigas, A. (2022b). ToM & ASD: The interconnection of Theory of Mind with the social-emotional, cognitive development of children with Autism Spectrum Disorder. The use of ICTs as an alternative form of intervention in ASD. Technium Social Sciences Journal, 33, 42-72. https://orcid.org/0000-0001-5637-9601

Bamicha, V., & Drigas, A. (2023a). Consciousness influences in ToM and Metacognition functioning-An artificial intelligence perspective. Res earch, Society and Development, 12(3). https://doi.org/10.33448/rsd-v12i3.40420

Bamicha, V., & Drigas, A. (2023b). Theory of Mind in relation to Metacognition and ICTs. A metacognitive approach to ToM. Scientific Electronic Archives, 16(4). https://doi.org/10.36560/16420231711

Bamicha, V., & Salapata, Y. (2024). LLLT applications may enhance ASD aspects related to disturbances in the gut microbiome, mitochondrial activity, and neural network function. Brazilian Journal of Science, 3(1), 140-158. https://doi.org/10.14295/bjs.v3i1.457

Bamicha, V., & Drigas, A. (2024). Strengthening AI via ToM and MC dimensions. Scientific Electronic Archives, 17(3). http://dx.doi.org/10.36560/17320241939

Banks, J. (2020). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics,* 12(2), 403-414.https://doi.org/10.1007/s12369-019-00588-x

Baraka, K., Alves-Oliveira, P., & Ribeiro, T. (2020). An extended framework for characterizing social robots. In: Jost, C., et al. Human-Robot Interaction. Springer Series on Bio- and Neurosystems, vol 12. Springer, Cham. https://doi.org/10.1007/978-3-030-42307-0_2

Basich, C., Biswas, J., & Zilberstein, S. (2023). Competence-Aware Autonomy: An Essential Skill for Robots in the Real World. Retrieved from: https://www.researchgate.net/search.Search.html?query=Competence-Aware+Autonomy%3A+An+Essential+Skill+for+Robots+in+the+Real+World&type=publication

Bianco, F., & Ognibene, D. (2019). Transferring adaptive theory of mind to social robots: Insights from developmental psychology to robotics. In Social Robotics: 11th International Conference, ICSR 2019, Madrid, Spain, November 26–29, 2019, Proceedings 11 (pp. 77-87). Springer International Publishing. https://doi.org/10.1007/978-3-030-35888-4_8

Breazeal, C., Gray, J., & Berlin, M. (2009). An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research,* 28(5), 656-680. https://doi.org/10.1177/0278364909102796

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social robotics. Springer handbook of robotics, 1935-1972. https://doi.org/10.1007/978-3-319-32552-1_72

Brock, L. L., Kim, H., Gutshall, C. C., &Grissmer, D. W. (2018). The developmentof theory of mind: Predictors and moderators of improvement in kindergarten. Early ChildDevelopment and Care. https://doi.org/10.1080/03004430.2017.1423481

Brooks, C., & Szafir, D. (2019). Building second-order mental models for human-robot interaction. *arXiv preprint arXiv:1909.06508*. https://doi.org/10.48550/arXiv.1909.06508

Caro, M. F., Cox, M. T., & Toscano-Miranda, R. E. (2022). A Validated Ontology for Metareasoning in

Intelligent Systems. Journal of Intelligence, 10(4), 113. https://doi.org/10.3390/jintelligence10040113

Castelfranchi, C., & Falcone, R. (2019). Self-Awareness Implied in Human and Robot Intentional Action. In AAAI Spring Symposium: Towards Conscious AI Systems. Retrieved from: https://ceur-ws.org/Vol-2287/paper7.pdf

Cerrato, L., & Campbell, N. (2017). Engagement in dialogue with social robots. Dialogues with Social Robots: Enablements, Analyses, and Evaluation, 313-319. https://doi.org/10.1007/978-981-10-2585-3_25

Chaidi, E., Kefalis, C., Papagerasimou, Y., & Drigas, A. (2021). Educational robotics in Primary Education. A case in Greece. Research, Society and Development, 10(9), e17110916371-e17110916371.http://dx.doi.org/10.33448/rsd-v10i9.16371

Chella, A., Pipitone, A., Morin, A., & Racy, F. (2020). Developing self-awareness in robots via inner speech. Frontiers in Robotics and AI, 7, 16. https://doi.org/10.3389/frobt.2020.00016

Chella, A. (2022). Robots and machine consciousness. MIT Press, In Special Collection: CogNet. ISBN electronic: 9780262369329. https://doi.org/10.7551/mitpress/13780.003.0029

Chen, X., Sui, Z., & Ji, J. (2013). Towards metareasoning for human-robot interaction. In Intelligent Autonomous Systems 12: Volume 2 Proceedings of the 12th International Conference IAS-12, held June 26-29, 2012, Jeju Island, Korea (pp. 355-367). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33932-5_34

Chen, B., Vondrick, C., & Lipson, H. (2021). Visual behavior modelling for robotic theory of mind. Scientific Reports, 11(1), 424. https://doi.org/10.1038/s41598-020-77918-x

Choi, J. J., Kim, Y. K., & Kwak, S. S. (2014). The autonomy levels and the human intervention levels of robots: The impact of robot types in human-robot interaction. In 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN 2014 (pp. 1069-1074). Institute of Electrical and Electronics Engineers Inc.https://doi.org/10.1109/ROMAN.2014.6926394

Clodic, A., & Alami, R. (2021). What Is It to Implement a Human-Robot Joint Action?. Robotics, AI, and Humanity: Science, Ethics, and Policy, 229-238. https://doi.org/10.1007/978-3-030-54173-6_19

Cox, M., & Raja, A. (2008). Metareasoning: A manifesto. BBN Technical.

Cross, E. S., Hortensius, R., &Wykowska, A. (2019). From social brains to social robots: applying neurocognitive insights to human–robot interaction. *Philosophical Transactions of the Royal Society B,* 374(1771), 20180024. https://doi.org/10.1098/rstb.2018.0024

Daglarli, E. (2020). Computational modeling of prefrontal cortex for meta-cognition of a humanoid robot. IEEE Access, 8, 98491-98507. https://doi.org/10.1109/ACCESS.2020.2998396

Danso, S., Annan, M. A. O., Ntem, M. T. K., Baah-Acheamfour, K., &Awudi, B. (2023). Artificial intelligence and human communication: A systematic literature review. World Journal of Advanced Research and Reviews, 19(01), 1391–1403. https://doi.org/10.30574/wjarr.2023.19.1.1495

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. Philosophical transactions of the royal society B: Biological sciences, 362(1480), 679-704. https://doi.org/10.1098/rstb.2006.2004

De Graaf, M. M., & Malle, B. F. (2017). How people explain action (and autonomous intelligent systems should too). In 2017 AAAI Fall Symposium Series. Retrieved from: https://cdn.aaai.org/ocs/16009/16009-69853-1-PB.pdf

De Santis, A., Siciliano, B., De Luca, A., & Bicchi, A. (2008). An atlas of physical human–robot interaction. Mechanism and Machine Theory, 43(3), 253-270. https://doi.org/10.1016/j.mechmachtheory.2007.03.003

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020). Shall I Trust You? From Child–Robot Interaction to Trusting Relationships. Frontiers in Psychology, 11, 469. https://doi.org/10.3389/fpsyg.2020.00469

Drigas, A., & Karyotaki, M. (2018). Mindfulness Training & Assessment and Intelligence. International Journal of Recent Contributions from Engineering, Science & IT (iJES), 6(3), 70-85. https://doi.org/10.3991/ijes.v6i3.9248

Drakatos, N., & Christou, A. (2023). The impact of robotics on development, of hot and cold executive functions, and their role in improving them in special education. World Journal of Advanced Engineering Technology and Sciences, 9(2), 070-080. https://doi.org/10.30574/wjaets.2023.9.2.0198

Drigas, A., & Karyotaki, M. (2018). Mindfulness Training & Assessment and Intelligence. International Journal of Recent Contributions from Engineering, Science & IT (iJES), 6(3), 70-85. https://doi.org/10.3991/ijes.v6i3.9248

Drigas, A., & Sideraki, A. (2021). Emotional intelligence in autism. Technium Soc. Sci. J., 26, 80. https://doi.org/10.47577/tssj.v26i1.5178

Drigas, A., Papanastasiou, G., & Skianis, C. (2023). The school of the future: The role of digital technologies, metacognition and emotional intelligence. International Journal of Emerging Technologies in Learning (Online), 18(9), 65. https://doi.org/10.3991/ijet.v18i09.38133

Drigas, A., & Bamicha, V. (2023). PoM & ToM-Harnessing the Power of Mind in Theory of Mind by shaping beneficial mental states in Preschoolers and the ICT's role. Research, Society and Development, 12(5). http://dx.doi.org/10.33448/rsd-v12i5.41590

Drigas, A., & Papoutsi, C. (2023). A New Pyramid Model of Empathy: The Role of ICTs and Robotics on Empathy. Int. J. Online Biomed. Eng., 19(2), 67-91. https://doi.org/10.3991/ijoe.v19i02.33591

Iasechko, M., Kharlamov, M., Gontarenko, L., Skrypchuk, H., Fadyeyeva, K., &Sviatnaia, O. (2021). Artificial intelligence as a technology of the future at the present stage of development of society. Laplage em Revista (International), n. Extra D, 7, 391-397. http://dx.doi.org/10.24115/S2446-622020217Extra-D1119p.391-397

Incao, S., Rea, F., &Sciutti, A. (2021). A Self for robots: core elements and ascription by humans. Interactions, 5(14), 15. https://doi.org/10.5281/zenodo.5645583

Fabbro, F., Cantone, D., Feruglio, S., & Crescentini, C. (2019). Origin and evolution of human consciousness. Progress in Brain Research, 250, 317-343. https://doi.org/10.1016/bs.pbr.2019.03.031

Frasca, T., Krause, E., Thielstrom, R., & Scheutz, M. (2020). " Can you do this?" Self-Assessment Dialogues with Autonomous Robots Before, During, and After a Mission. arXiv preprint arXiv:2005.01544. https://doi.org/10.48550/arXiv.2005.01544

Frith, U. & Happé, F.G.E (1999). Theory of Mind and Self-Consciousness: What Is It Like to Be Autistic? Mind & Language, 14 (1), 1–22. https://doi.org/10.1111/1468-0017.00100

Frith, C. D. & Frith, U. (2012). Mechanisms of Social Cognition. Annual Review of Psychology, 63:287-313. https://doi.org/10.1146/annurev-psych-120710-100449

Haddadin, S., & Croft, E. (2016). Physical human–robot interaction. Springer handbook of robotics, 1835-1874. https://doi.org/10.1007/978-3-319-32552-1_69

Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? Comparative cognition & behavior reviews, 4, 17. https://doi.org/10.3819%2Fccbr.2009.40002

He, H., Gray, J., Cangelosi, A., Meng, Q., McGinnity, T. M., &Mehnen, J. (2021). The challenges and opportunities of human-centered AI for trustworthy robots and autonomous systems. IEEE Transactions on Cognitive and Developmental Systems, 14(4), 1398-1412. https://doi.org/10.1109/TCDS.2021.3132282

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008, March). Theory of mind (ToM) on robots: A functional neuroimaging study. In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, 335-342. https://doi.org/10.1145/1349822.1349866

Geraci, A., D'Amico, A., Pipitone, A., Seidita, V., & Chella, A. (2021). Automation inner speech as an anthropomorphic feature affecting human trust: Current issues and future directions. Frontiers in Robotics and AI, 8, 620026. https://doi.org/10.3389/frobt.2021.620026

Gloor, J. L., Howe, L., De Cremer, D., & Yam, K. C. S. (2020). The funny thing about robot leadership. European Business Law Review, online. https://www.alexandria.unisg.ch/handle/20.500.14171/111535

Goel, A. K., Fitzgerald, T., & Parashar, P. (2020). Analogy and metareasoning: Cognitive strategies for robot learning. In Human-Machine shared contexts (pp. 23-44). Academic Press. https://doi.org/10.1016/B978-0-12-820543-3.00002-X

González-Docasal, A., Aceta, C., Arzelus, H., Álvarez, A., Fernández, I., & Kildal, J. (2021). Towards a natural human-robot interaction in an industrial environment. *Conversational Dialogue Systems for the Next Decade,* 243-255. https://doi.org/10.1007/978-981-15-8395-7_18

Görür, O. C., & Albayrak, Ş. (2016). A cognitive architecture incorporating theory of mind in social robots towards their personal assistance at home. In Proceedings of the Workshop on Bio-inspired Social Robot Learning in Home Scenarios at IEEE/RSJ International Conference on Intelligent Robots and Systems 2016 (IROS'16).

Görür, O., Rosman, B. S., Hoffman, G., & Albayrak, S. (2017). Toward integrating Theory of Mind into adaptive decision-making of social robots to understand human intention. http://hdl.handle.net/10204/9653

Greenhalgh, T., Thorne, S., &Malterud, K. (2018). Time to challenge the spurious hierarchy of systematic over narrative reviews?. European journal of clinical investigation, 48(6). https://doi.org/10.1111%2Feci.12931

Gutiérrez, S. M., & Steinbauer-Wagner, G. (2022). The Need for a Meta-Architecture for Robot Autonomy. arXiv preprint arXiv:2207.09712. https://doi.org/10.4204/EPTCS.362.9

Javaid, M., Estivill-Castro, V., &Hexel, R. (2020). Enhancing humans trust and perception of robots through explanations. Proceedings of the ACHI, 10(1912), 4071. https://doi.org/10.25904/1912/4071

Jokinen, K. (2021). Exploring Boundaries Among Interactive Robots and Humans. In: D'Haro, L.F., Callejas, Z., Nakamura, S. (eds) Conversational Dialogue Systems for the Next Decade. Lecture Notes in Electrical Engineering, 704. Springer,

Singapore. https://doi.org/10.1007/978-981-15-8395-7_20

Joksimovic, S., Ifenthaler, D., Marrone, R., De Laat, M., & Siemens, G. (2023). Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. Computers and Education: Artificial Intelligence, 100138. https://doi.org/10.1016/j.caeai.2023.100138

Jung, M., & Hinds, P. (2018). Robots in the wild: A time for more robust theories of human-robot interaction. ACM Transactions on Human-Robot Interaction (THRI), 7(1), 1-5. https://doi.org/10.1145/3208975

Kaivo-Oja, J., Roth, S., & Westerlund, L. (2017). Futures of robotics. Human work in digital transformation. International Journal of Technology Management, 73(4), 176-205. https://doi.org/10.1504/IJTM.2017.083074

Karyotaki, M., Drigas, A., & Skianis, C. (2024). Mobile/VR/Robotics/IoT-Based Chatbots and Intelligent Personal Assistants for Social Inclusion. International Journal of Interactive Mobile Technologies, 18(8). http://dx.doi.org/10.3991/ijim.v18i08.46473

Kneer, M. (2021). Can a robot lie? Exploring the folk concept of lying as applied to artificial agents. Cognitive Science, 45(10), e13032. https://doi.org/10.1111/cogs.13032

Kralik, J. D., Lee, J. H., Rosenbloom, P. S., Jackson Jr, P. C., Epstein, S. L., Romero, O. J., ... & McGreggor, K. (2018). Metacognition for a common model of cognition. Procedia computer science, 145, 730-739. https://doi.org/10.1016/j.procs.2018.11.046

Lasota, P. A.; Fong, T.; &Shah, J. A. (2017). A survey of methods for safe human-robot interaction. Foundations and Trends in Robotics, 5(4):261–349. http://dx.doi.org/10.1561/2300000052

Losey, D. P., McDonald, C. G., Battaglia, E., & O'Malley, M. K. (2018). A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction. Applied Mechanics Reviews, 70(1), 010804. https://doi.org/10.1115/1.4039145

Lu, M. (2023). The ability and importance of human communication in the age of AI. Geographical Research Bulletin, 2, 187-189. https://doi.org/10.50908/grb.2.0_187

Malle, B. F., & Scheutz, M. (2019). Learning how to behave: Moral competence for social robots. Handbuchmaschinenethik, 255-278. Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-17483-5_17

Martini, M. C., Buzzell, G. A., & Wiese, E. (2015). Agent appearance modulates mind attribution and social attention in human-robot interaction. In Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7 ,431-439. Springer International Publishing. https://doi.org/10.1007/978-3-319-25554-5_43

Martini, M. C., Gonzalez, C. A., & Wiese, E. (2016). Seeing minds in others–Can agents with robotic appearance have human-like preferences? PloS one, 11(1), e0146310. https://doi.org/10.1371/journal.pone.0146310

Miranda, A., Berenguer. C., Roselló, B., Baixauli, I. & Colomer, C. (2017). Social Cognition in Children with High-Functioning Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. Associations with Executive Functions. Frontiers in Psychology, 8:1035. https://doi.org/10.3389/fpsyg.2017.01035

Mitsea, E., Lytra, N., Akrivopoulou, A., & Drigas, A. (2020). Metacognition, Mindfulness and Robots for Autism Inclusion. Int. J. Recent Contributions Eng. Sci. IT, 8(2), 4-20. https://doi.org/10.3991/ijes.v8i2.14213

Mishra, D., Lugo, R.G., Parish, K., Tilden, S. (2023). Metacognitive Processes Involved in Human Robot Interaction in the School Learning Environment. In: Kurosu, M., Hashizume, A. (eds) Human-Computer Interaction. HCII 2023. Lecture Notes in Computer Science, vol 14012. Springer, Cham. https://doi.org/10.1007/978-3-031-35599-8_6

Moraiti, I., & Drigas, A. (2023). AI Tools Like ChatGPT for People with Neurodevelopmental Disorders. International Journal of Online & Biomedical Engineering, 19(16). DOI:10.3991/ijoe.v19i16.43399

Natarajan, M., Seraj, E., Altundas, B., Paleja, R., Ye, S., Chen, L., ... &Gombolay, M. (2023). Human-robot teaming: grand challenges. Current Robotics Reports, 4(3), 81-100. https://doi.org/10.1007/s43154-023-00103-1

Ntaountaki, P., Lorentzou, G., Lykothanasi, A., Anagnostopoulou, P., Alexandropoulou, V., & Drigas, A. (2019). Robotics in Autism Intervention. Int. J. Recent Contributions Eng. Sci. IT, 7(4), 4-17. https://doi.org/10.3991/ijes.v7i4.11448

Onnasch, L., & Roesler, E. (2021). A taxonomy to structure and analyze human–robot interaction. International Journal of Social Robotics, 13(4), 833-849. https://doi.org/10.1007/s12369-020-00666-5

Pergantis, P., & Drigas, A. (2023). Sensory integration therapy as enabler for developing emotional intelligence in children with autism spectrum disorder and the ICT's role. Brazilian Journal of Science, 2(12), 53-65. https://doi.org/10.14295/bjs.v2i12.422

Pergantis, P. (2024). A new era of ICTs for combating symptoms of neurodevelopmental disorders. World Journal of Biology Pharmacy and Health Sciences, 18(1), 036-047. https://doi.org/10.30574/wjbphs.2024.18.1.0145

Pergantis, P., & Drigas, A. (2024). The Effect of Drones in the Educational Process: A Systematic Review. Education Sciences, 14(6), 665. https://doi.org/10.3390/educsci14060665

Pipitone, A., Lanza, F., Seidita, V., & Chella, A. (2019, March). Inner Speech for a Self-Conscious Robot. In AAAI Spring Symposium: Towards Conscious AI Systems. https://ceur-ws.org/Vol-2287/paper14.pdf

Pipitone, A., Geraci, A., D'Amico, A., Seidita, V., & Chella, A. (2023). Robot's inner speech effects on human trust and anthropomorphism. International Journal of Social Robotics, 1-13. https://doi.org/10.1007/s12369-023-01002-3

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. Nature Reviews Psychology, 1(4), 223-235. https://doi.org/10.1038/s44159-022-00037-z

Rosenthal, D. M. (2005). Consciousness and Mind. Oxford University Press.Rother, E. T. (2007). Systematic literature review X narrative review. Acta paulista de enfermagem, 20, v-vi. https://doi.org/10.1590/S0103-21002007000200001

Rossi, A., Rossi, S., Andriella, A., & van Maris, A. (2022, March). The Road to a Successful HRI: AI, Trust and ethicS (TRAITS) Workshop. In 2022 *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 1284-1286).* IEEE.

Sandini, G., Sciutti, A., & Vernon, D. (2021). Cognitive robotics. In Encyclopedia of Robotics (pp. 1-7). Springer-Verlag.

SCASSELLATI, B. (2002). Theory of Mind for a Humanoid Robot. *Autonomous Robots,* 12, 13-24. https://doi.org/10.1023/A:1013298507114

Schleidgen, S., & Friedrich, O. (2022). Joint Interaction and Mutual Understanding in Social Robotics. *Science and Engineering Ethics,* 28(6), 48. https://doi.org/10.1007/s11948-022-00407-z

Sheridan, T. B. (2016). Human–robot interaction: status and challenges. Human factors, 58(4), 525-532. https://doi.org/10.1177/0018720816644364

Shmatko, N., & Volkova, G. (2020). Bridging the skill gap in robotics: Global and national environment. Sage Open, 10(3). https://doi.org/10.1177/2158244020958736

Sideraki, A., & Drigas, A. (2021). Artificial Intelligence (AI) in Autism. Technium Soc. Sci. J., 26, 262. http://dx.doi.org/10.33448/rsd-v11i16.38057

Syriopoulou-Delli, C., Deres, I., & Drigas, A. (2021). Intervention program using a robot for children with Autism Spectrum Disorder. Research, Society and Development, 10(8), e35010817512-e35010817512. https://doi.org/10.33448/rsd-v10i8.17512

Thellman, S., Silvervarg, A., & Ziemke, T. (2017). Folk-Psychological Interpretation of Human vs. Humanoid Robot Behavior: Exploring the Intentional Stance toward Robots. *Frontiers in Psychology,* 8, 280953. https://doi.org/10.3389/fpsyg.2017.01962

Thellman, S., De Graaf, M., & Ziemke, T. (2022). Mental state attribution to robots: A systematic review of conceptions, methods, and findings. *ACM Transactions on Human-Robot Interaction (THRI),* 11(4), 1-51. https://doi.org/10.1145/3526112

Vernon, D., Thill, S., & Ziemke, T. (2016). The Role of Intention in Cognitive Robotics. In: Esposito, A., Jain, L. (eds) Toward Robotic Socially Believable Behaving Systems - Volume I. Intelligent Systems Reference Library, vol 105. Springer, Cham. https://doi.org/10.1007/978-3-319-31056-5_3

Vinanzi, S., Patacchiola, M., Chella, A., & Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philosophical Transactions of the Royal Society B,* 374(1771), 20180032. https://doi.org/10.1098/rstb.2018.0032

Vinanzi, S., Cangelosi, A., &Goerick, C. (2021). The collaborative mind: intention reading and trust in human-robot interaction. Iscience, 24(2). doi: 10.1109/DEVLRN.2019.8850698.

Vouglanis, T. (2023). The use of robotics in the education of students with special educational needs. World Journal of Advanced Research and Reviews, 19(1), 464-471. https://doi.org/10.30574/wjarr.2023.19.1.1331

Vygotsky, L. (1962). Thought and language. (E. Hanfmann & G. Vakar, Eds.). MIT Press. https://doi.org/10.1037/11193-000

Wallach, W., Franklin, S., & Allen, C. (2010). A conceptual and computational model of moral decision making in human and artificial agents. Topics in cognitive science, 2(3), 454-485. https://doi.org/10.1111/j.1756-8765.2010.01095.x

Wang, Z., & Frye, D. A. (2021). When a CircleBecomes the Letter O: Young Children'sConceptualization of Learning and Its Relation WithTheory of Mind Development. Frontiers inpsychology, 11, 596419.https://doi.org/10.3389/fpsyg.2020.596419

Whitebread, D., Almeqdad, Q., Bryce, D., Demetriou, D., Grau, V., & Sangster, C. (2010). Metacognition in young children: Current methodological and theoretical developments. Trends and prospects in metacognition research, 233-258.

Wiese, E., Metta, G., &Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social.

Frontiers in psychology, 8, 281017. https://doi.org/10.3389/fpsyg.2017.01663

Williams, J., Fiore, S. M., & Jentsch, F. (2022). Supporting Artificial Social Intelligence With Theory of Mind. Frontiers in artificial intelligence, 5, 750763. https://doi.org/10.3389/frai.2022.750763n

Worley, P. (2018). Plato, metacognition and philosophy in schools. Journal of Philosophy in Schools, 5(1). http://doi.org/10.21913/jps.v5i1.1486

Wortham, R. H., Theodorou, A., & Bryson, J. J. (2016, June). What does the robot think? Transparency as a fundamental design requirement for intelligent systems. In IJCAI 2016 Ethics for AI Workshop. http://opus.bath.ac.uk/50294/1/WorthamTheodorouBryson{_}EFAI16.pdf

Xu, T., Zhang, H., & Yu, C. (2016). See you see me: The role of eye contact in multimodal human-robot interaction. ACM Transactions on Interactive Intelligent Systems (TiiS), 6(1), 1-22. https://doi.org/10.1145/2882970

Yang, G. Z., Dario, P., &Kragic, D. (2018). Social robotics—Trust, learning, and social interaction. *Science Robotics,* 3(21), eaau8839. https://doi.org/10.1126/scirobotics.aau8839

Zouganeli, E., &Lentzas, A. (2022). Cognitive robotics-towards the development of next-generation robotics and intelligent systems. In Symposium of the Norwegian AI Society (pp. 16-25). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-17030-0_2