Weaving a Theory of Artificial Minds

A synthesis of multidisciplinary approaches describing children's understanding of artificial social entities, such as computers and AI systems

Nandini Asavari Bharadwaj†
Department of Educational and Counselling Psychology
McGill University
Montreal, Canada
nandini.bharadwaj@mail.mcgill.ca

Adam K. Dubé†
Department of Educational and Counselling Psychology
McGill University
Montreal, Canada
adam.dube@mcgill.ca

ABSTRACT

Children in the developed world can encounter sophisticated AI technology in their homes, right from birth. These intelligent, interactive, and embodied technologies may leave lasting impacts on children's understanding of themselves and the world around them. Understanding how children at critical periods for developing social cognition skills (4-8 years) conceive of AI systems, such as voice assistants or robots, is crucial. These form the interdisciplinary foundation of a Theory of Artificial Minds (ToAM) that concerns fields of psychology, human-computer interaction, artificial intelligence, computer science, and communication studies.

A scoping review (5 databases, 2540 articles screened, 70 reviewed) relating to theories describing children's interactions with computers and AI systems was conducted to map out the breadth of human-computer and child development literature. Results: a) The most popular theories used to describe children's understanding of artificial social entities originate from established theories of understanding of other human entities; b) There are disciplinary inconsistencies in how theoretical ideas are being used to describe similar phenomena and c) There is a research gap with opportunities to build a dedicated theoretical model of humans and AI system understanding. Hence, this review provides first steps to guide the creation of a fulsome and inclusive ToAM.

CCS CONCEPTS

• Applied Computing → Law, Social and Behavioural Sciences → Psychology; • Human Centred Computing → Human-Computer interaction → HCI theory, concepts, and models

KEYWORDS

Theory of Artificial Minds, Artificial Intelligence, Theory, Human-AI understanding

ToMinHAI at CHI 2024, May 12th, Honolulu, Hawaii
2024, ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

†Both authors: conceptualization. First author: data curation, formal analysis, original draft. Second author: supervision, review, and editing.

ACM Reference Format:

Nandini Asavari Bharadwaj and Adam K. Dubé. 2024. Weaving a Theory of Artificial Minds: A synthesis of multidisciplinary approaches describing children's understanding of artificial social entities such as computers and AI systems. In *Proceedings of Workshop on Theory of Mind in Human-AI Interaction at CHI 2024 (ToMinHAI at CHI 2024).* ACM, New York, NY, USA. 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1. INTRODUCTION

Artificial intelligence (AI) technology is affecting all spheres of human life today, including education, work, and home. In the western home, AI products such as digital assistants (DA) in smart speakers have become commonplace since their introduction in 2010 [1]. A third of American parents in a Pew Research Center study said their children under 12 interact with voice assistants and use them for information searches [2]. Moreover, commercial DAs such as Amazon Alexa, Google Assistant, and Samsung Bixby all feature child-focused functionalities such as parental controls, reading support, and games. In addition to more general-purpose AI systems like the DA, children are also being exposed to sophisticated smart toys [3, 4], educational robots [5, 6, 7] and other intelligent learning companions in the home [8]. Consequently, today more children than ever before may experience access to sophisticated AI technology from birth.

Previous studies on children's interactions with AI systems have focused on children's use, perceptions, and learning. Young children have been observed to enjoy their interactions with DAs and readily interact with them [9, 10, 11, 12, 13] and children can ascribe personality and social characteristics to DAs [11, 10, 14, 15]. Similarly, children enjoy their interactions with robots and interact with them socially [16, 17, 18]. A review of child-robot relationship studies found that the responsiveness, role of robots, and the type of interaction all increased the closeness between child and robot [19]. Studies with robotic animals have shown that children perceive them uniquely and challenge conventional ideas of categorisation and perception [20, 21].

While studies have found impactful preliminary results, the impact of AI use on child development is still under-researched given the novelty of the technology and its use cases (e.g., smart toys). As Danovitch [2019] highlights, children's interactions with sophisticated internet-based technologies can facilitate a reciprocal relationship with children's cognitive development, whereby children's cognition influences their understanding of technology and technology use can influence cognition. Hence, there is a need for fresh frameworks that can accommodate and explain new developments in child and even adult cognition regarding AI systems. For example, the Theory of Artificial Minds (ToAM) is a novel theoretical framework proposed by Bharadwaj et al. [2023] that extends the notion of theory of mind to AI systems, wherein ToAM is the reciprocal ability to infer the internal states of AI systems from a human perspective. However, ToAM is but one approach of many that exist across multiple fields studying child-AI interaction, such as psychology and human-computer interaction. To move this enquiry forward, it is essential to bring together empirical research and proposed frameworks to identify and compare approaches across disciplines. As such, a scoping review is suited for this effort.

2. SCOPING REVIEW

This review identifies theoretical frameworks that guide how children understand artificial social entities around them. These theories examine children's understanding of various technological systems capable of social interactions, such as personal computers, tablets, digital assistants, robots, chatbots, and intelligent tutoring systems. Given the broad interdisciplinary nature of this topic, a scoping review is preferred over other types of reviews (e.g., systematic review). Scoping reviews are ideal for providing an overview of an existent and complex area of research, highlighting underlying key concepts, theories, and practices [24]. The guiding question of this review is, "what are the prominent theories that describe children's understanding of computers and AI systems, as social agents?"

3. METHOD

Five databases were searched to identify literature from psychology, human-computer interaction (HCI), AI, computer science, and communication studies. These databases were PsycInfo (1806-Ovid), Scopus, ACM Digital Library, Web of Science, and Engineering Village. A keyword search strategy with similar inclusion and exclusion criteria (Appendix, Figure 1) was deployed across all databases; however, due to database idiosyncrasies, database-specific search methodologies were employed.

3. 1 Selection Criteria

The publication date (after 2010) was selected to coincide with the commercial release of digital voice assistants [1]. While children in the theory of mind development stage (ages 4-8 years) were the population of interest, studies with all participants under 18-years-of-age were included. Lastly, participants with developmental disorders, such as autism-spectrum disorders (ASD), were excluded from this review due to known issues with social cognition development and theory of mind tasks [25].

Additionally on content, children's understanding of computers and AI systems as social agents, is a key focus. Understanding computers and AI systems as social agents ensures that child-computer or child-AI system interactions being examined are complex and rich. It excludes one-way interactions or passive information consumption, such as a child merely looking at a computer. Further, "social agency" is defined as interactions involving bi-directional independent communication and mutual understanding [26]. This entails dynamic interactions where a child independently interfaces with the computer or AI system, without the help of parents or teachers, and where the child and computer or AI system develop some basic understanding of the other to accomplish a task or activity. For example, a child individually interacts and communicates with a DA to accomplish an information-search task. Hence, only articles that involve children's conceptions of AI and computers as social agents were included. A total of 70 articles met the criteria for inclusion and were reviewed for analysis (see Figure 1 in Appendix).

3. 2 Data Coding

For a synthesis of the final 70 articles, a coding scheme was adopted that captured content aspects of the articles. These are organised under broad categories as follows:

3.2.1 Article Features. The theme, discipline, and publication type for each article were identified and categorised using a mix of directly identifiable details and some generated frameworks (Appendix, Table 1). Themes were generated from article content (e.g., Child-Robot interaction or Child-AI understanding). Disciplines (e.g., Psychology or Interdisciplinary) were based on researcher affiliations and article focus. Studies on human-robot interactions were included under the larger umbrella of 'human-computer interaction'; articles with AI researchers and researchers from other fields were categorised as 'Interdisciplinary'. Publication types (e.g., conference proceedings or peer-reviewed journals) were also identified. Additionally, studies were categorised based on directly identifiable features. These included the population, location (e.g., in-person or virtual), and sample size (organised by sample size intervals of 20) (Appendix, Table 1)

3.2.2 Theoretical Approaches. Given the primary focus of this review, underlying theories and theoretical approaches adopted or discussed in the articles were categorised based on directly identifiable details in the articles (see below for process).

4. RESULTS

The scoping review results are presented here with high-level summaries of the article's features and theories. Given the review's focus, only the theoretical approach section will be presented in detail.

4. 1 Article Features

Most articles (61% of articles) focused on child-robot interactions and understanding. Many articles also focused on child-digital assistant interactions (23% of articles) (see Appendix, Table 1). Articles that focused on more general interactions between children and AI systems (e.g., programs with AI capabilities or more general discussions on AI technology and children) were categorised as child-AI interactions (7% of articles), and those that conceptualised child-technology interactions (e.g., tablets or general discussions on technology and children) were categorised as child-technology interactions (3% of articles). Overall, this reflects trends in the field that child-robot interactions, specifically, are a popular topic for academic investigations. Given that work with robots was observed as the most popular, it follows that most articles (41% of articles) were from the HCI discipline. After HCI, interdisciplinary articles (31% of articles) were common, which seems appropriate as the subject matter relates to multiple disciplines (e.g., psychology, robotics etc.). Across the review, most children being studied as the target population were under the age of 10 (over 46% of articles). This suggests that most researchers are interested in critical developmental periods for children (e.g., linguistic development periods or social cognition development).

4. 2 Theoretical Approaches

Given the multidisciplinary nature of this review, a broad conception of "theory" was employed that emphasises the role of description and explanation of phenomena (i.e., children's understanding of computers and AI systems as social agents) rather than any disciplinary-specific theoretical definitions. Two levels of categorisations were employed for the synthesis. Adopting a hierarchical structure in categorising theories, prominent theories and significant theoretical categories were identified at level 1, whereas less-significant theories were individually listed at level 2. Additionally, theoretical approaches were organised by citation, rather than by article due to a given article containing multiple theories and frameworks. Theoretical approaches with level 1 categorisations are presented (Appendix, Table 2), and detailed level 2 categorisations are available in supplementary materials (On request). The "Other theories" and "Social interaction frameworks" categories contained citations to diverse ideas and frameworks that were not attributable to a single unifying source or theory. Notably, 4 articles did not employ any discernible theoretical approaches and relied only on previous empirical work.

This section will discuss select theoretical approaches (Appendix, Table 2) from the scoping review. For sake of brevity, only prominent theories will be described in detail. Prominent theories were chosen by citation count and distinctiveness. For citation count, an individual theoretical approach with at least 4 citations (e.g., Theory of cognitive development) was chosen as a cut-off for inclusion. For distinctiveness, theoretical approaches that were novel, explanatory and useful were chosen for discussion (e.g., Theory of Artificial Minds).

4.2.1 Anthropomorphism. Anthropomorphism was one of the most prominent theoretical ideas present in the review (25 citations; 16% of all citations). Epley et al. [2007] discuss that anthropomorphism, as its name suggests, involves attributing human-like properties or mental states to real or imagined non-human objects or agents. Hence, it can involve attributing emotions, intentions, or even human-like expectations to non-human entities through induction, from known human qualities to those of unknown agents [27]. Across the review, there is a broad usage of "anthropomorphism" as a known theoretical concept without reference to a specific theoretical model, though some have been proposed (See [27]). Related concepts of personhood, personification, and human attribution are also used alongside the notion of anthropomorphism.

For this review, anthropomorphism is the tendency for children to conceive of computers and AI as social agents in human terms and make predictions about their behaviours, intentions and internal states based on what children know about other humans. Epley et al. [2007] highlight that for children, anthropomorphism likely develops after reasoning about self and others, given children's early exposure to only human entities. Children naturally over-extend human-like attributions to non-human entities (inanimate objects to non-human agents) as they concurrently develop their social cognitive skills regarding humans (see theory of mind) [27]

Several studies specifically observed the phenomena of anthropomorphism. Studies with DAs revealed that children were likely to anthropomorphise due to voice-based cues [14, 10, 28]. Strathmann et al. [2020] observe that the natural-sounding voice of the voice assistant is likely a potent trigger of anthropomorphism but that with more use, children were less likely to do so. Hoffman et al. [2021] suggest that the anthropomorphising of voice assistants likely leads to children developing parasocial relationships with them. Across the spectrum of robotics studies, children were observed to anthropomorphise all types of robots. A reading robot was treated as a valued social companion [29], a story-telling robot was viewed as more sociable if children anthropomorphised it [18], a teaching robot triggered anthropomorphism in children when children worked alone with a robot rather than in teams [7]. Similarly, Bjorling et al. [2020] observed that teens were readily willing to anthropomorphise the robot in their study and use gender pronouns to describe it, despite the researchers not referring to the robot in that way.

In addition to the presence of the phenomena, several studies also explored factors that affect the presence and development of anthropomorphism. These factors, such as age, agent qualities, type of tasks, and presence of other people, bring nuance to the discussion. Age mediated how readily children viewed social AI systems in several studies, with younger children more willing to anthropomorphise or attribute human-like qualities to AI systems than older children [20, 11, 15, 14, 28, 30, 31]. Qualities of the agent also affected how children were willing to anthropomorphise a robot. How transparent an AI system was about its capabilities decreased the tendency to anthropomorphise [32]. Curiously, framing an AI system as a machine instead of a social agent increased gaze time and behaviour [33]. Design choices like size also affected how children were willing to anthropomorphise a robot, with smaller child-size robots being more likeable than larger ones and humanoid robots perceived as more human-like than robots that resemble other entities like plants [34]. Lastly, the presence of others, such as peer groups, decreased the tendency to anthropomorphise [7], and parents, who heavily influenced children's perceptions of agent intelligence to be like theirs [36] also affected children's tendencies to anthropomorphise AI systems and computers.

4.2.2 Animism. Animism or animacy (15 citations; 10% of all citations) generally refers to attributing life or "aliveness" to inanimate objects [37]. It can refer to the perception of life in non-living objects such as tools or be used to explain the behaviour of objects [38]. Animism is related to the previous concept of anthropomorphism; however, it does not necessarily involve all types of human-like attributions (e.g., emotions, desires) and can merely refer to the perception of some lifelike quality in an object. Animism is widely observed in young children, who view the whole world as alive in various ways [39]. Given linguistic habits, adults can also regularly use animism in their language; however, this is considered to be more of a metaphorical usage [40].

For Piaget [1929], animism was a critical developmental phenomenon and explained why young children perceive objects as alive or having intentions. As children move through developmental stages, they work towards correctly distinguishing living and non-living entities [41]. Children initially may overgeneralise animism to different entities, such as objects and nature, as the attributions of intention are the only causal tools available to them [41]. In contrast, Carey [1985] offers a more biological account of animism, wherein children overattribute "aliveness" to inanimate objects as they do not yet have a strong grasp of biology. As children reach age 10, they have a firmer grasp of animism due to the restructuring of biological concepts they learn from birth [42]. While both models stand in contrast to each other, as Carey [1985] admits, understanding the notion of animism for young children is important regardless of which explanation is more accurate. Children use the available entities around them, from people to nature to objects, to form their theories of animism [39]. Bringing the conversation into the present day, as Beran et al. [2011] describe, understanding children's feelings of animism towards social AI entities, such as robots, has become increasingly pertinent in today's world, where children see AI systems increasingly in their daily life.

Several studies directly studied the concepts of animism [30, 34, 35]. Cameron et al. [2017] found age-related differences in their study, wherein younger children attributed personhood to humanoid robots more than older children and whose views on animism were influenced by how autonomous the robot was (whether directly controlled or not). Cameron et al. [2017] findings support Carey's [1985] model of animism in that children conceive of the animate based on their developing ideas about the biological world around them. In their study comparing perceptions of different robots, Søraa et al. [2021] found that not only were children's tendencies to anthropomorphise and attribute animism to the robots related but that key physical features can trigger more animism. The robot with more autonomy in movement and more idiosyncratic speech was judged to be more lifelike and alive [34]. In their study on factors that affect children's perceptions of animacy, anthropomorphism, trust and closeness in robots, van Straten et al. [2020] found that increasing transparency of a robot's machine capabilities led to a significant decrease in animacy ratings, compared to controls. Other studies explored whether children simply find social AI systems and computers to be "alive". Numerous studies found that children readily perceive AI systems, like DAs, to be alive [43, 44, 28]. Overall, animism is a useful construct to understand how children conceptually understand social AI entities. However, similar to anthropomorphism, animism is used broadly, and its usage can invoke ideas from life to human-like qualities to intentionality.

4.2.3 Media Equation Theory. Media equation theory (5 citations; 3% of all citations) is a communication paradigm initially proposed by Reeves and Nass [1996]. The media equation suggests that humans treat various types of media (e.g., television, computers) in natural social ways [45]. Due to natural language similarities in communication conventions, instructions and computer interactions invoke social responses in human users [45]. Through the straightforward proposal that "media experiences equal human experiences", the media equation theory explains that people's reactions to media are both automatic and simple [45, p. 251]. Importantly, people's perceptions are more important than the verifiable truth of what they believe about particular media, so if they perceive that a computer has a personality, then they will perceive it so and respond socially, even if they know it is not capable of having a personality [45]. The media equation is a vivid communication theory due to its intuitive premise and was ahead of its time in emphasising perceptions of media. Pashevich [2023], in recognising the legacy of the media equation paradigm, highlights that a vital principle in the HCI field today is that people prefer to interact with technologies naturally without specific training.

The media equation paradigm is increasingly relevant with advancements in social media and communication technologies [28]. In their study, Strathmann et al. [2020] found that voice-based DAs not only trigger social reactions in young children but that children assessed social reactions from DAs to be more appropriate over time. Garg and Sengupta [2020] discuss the affordances of DAs, such as their names (e.g., Alexa), gender, and personality, as key drivers of social cues and anthropomorphising for families and young children. In their study with an intelligent tutoring system, Ogan et al. [2012] found compelling evidence for effective social learning with AI systems through the appeal to social conventions and partnerships. Ogan et al. [2012] observed that students with less social rapport with the teachable agent were less likely to be direct about the teachable agent's shortcomings, which the researchers hypothesise to be about not wanting to appear impolite. This finding aligns with the media equation theory, which proposes politeness as a natural aspect of how humans interact with media [47]. While the voice-based nature of technologies such as DAs and robots lend themselves well to the media equation paradigm, Xu [2023] rightfully cautions its use with children as there are additional considerations of children's cognitive development. While interactive conversation-based technology may understandably trigger social cues in children, more research is needed to understand how conversational technologies uniquely affect children in context [48].

4.2.4 New Ontological Category Approaches. New ontological category (NOC) approaches (6 citations; 4% of all citations), as an umbrella term used here, encompass different approaches that propose that children may conceptualise social AI agents as unique and different from strict ontological categories such as living/non-living or human/machine. These approaches suggest that the specific affordances of social AI systems and computers, such as their interactivity, autonomous functioning, embodiment, and natural language capabilities, may result in children's perceptions being less strictly in any one ontological category and require a fresh approach [9, 17, 44].

Kahn et al. [2012] describe how over a third of children in their study were unwilling to choose between strict categories of living, non-living, or neither category when describing a humanoid robot. As humanoid robots cannot be easily mapped onto categories such as a person, animal or artefact, Kahn et al. [2012] suggest that they may require a new ontological category with its properties. Girouard-Hallam et al. [2021] findings also support an NOC approach, as children were more willing to attribute mental and social characteristics to voice assistants than any moral agency. Okita et al. [2015, p. 728] refer to AI devices and systems that can simultaneously have human-like and machine-like features as "technological boundary objects". These objects, such as robotic pets or friendly virtual pedagogical agents, challenge children and adults thinking about biological and social properties. In a study with young children and voice assistants, Festerling and Siraj [2020] found that while children held strong beliefs about what it meant to be a human or machine, they could also perceive both humanoid and non-humanoid interaction capabilities in voice assistants. Festerling and Siraj [2020] suggest that this sets up voice assistants, with their mixture of properties, as being on the edge of what children might consider living and non-living. Kory-Westlund and Breazeal [2019] use an innovative scale to uncover variations in how children conceive of various entities, from biological to machine. Children in the study classified one of the entities, the Tega robot, as somewhere between a biological entity and a machine, uncovering again a tendency to consider robots as an in-between entity [49]

NOC approaches are a promising and fecund area of research [50]. Focusing on children's developing ideas of social and personified AI systems, such as robots, can help clarify whether people actually believe social AI systems to be a different ontological entity or merely pretend to, as well as track generational differences in how AI systems are perceived [50].

4.2.5 Theory of Cognitive Development and Sociocultural Theory of Cognitive Development. Theoretical ideas from Piaget and Vygotsky that form the theory of cognitive development (4 citations; 3% of all citations) and sociocultural theory of cognitive development (5 citations; 3% of all citations), respectively, are some of the most influential theories of childhood learning and development in psychology and learning sciences. From a Piagetian [1964] perspective, learning is internally constructed through interactions with the environment and social entities. Further, according to Piaget [1964], children go through distinct stages of development as they develop a firmer understanding of the world and how to navigate it through their experiences. As discussed earlier, a vital aspect of this development is the role of animism. Animism is seen in children's pre-operational stage (ages 2-7) and describes how children see life in all objects around them [51]. Vygotsky's ideas more concretely centre social experiences as being responsible for children developing knowledge and conscience about the world [52]. For Vygotsky, social interactions with more capable parents, peers and cognitive tools allow children to build their knowledge of the world [52].

For several articles, the central assumptions of children's understanding of other entities and their development through observations, play, speech, and social interactions with AI systems are derived from the theory of cognitive development and sociocultural theory of cognitive development. In earlier parts of this section, several studies have already been mentioned in discussions on animism. Children's independent acquisition and manipulation of knowledge and their propensity to develop animism, ideas derived from Piaget [1964], form the theoretical basis of Druga et al. [2018] study with children and parents' attribution of intelligence to biological and artificial agents. The importance of social interactions and peer learning on children's cognitive development, per ideas from Piaget and Vygotsky, forms the foundation of the design of Charisi et al. [2021] study with child-robot team dynamics as well as Cagiltay et al. [2022] study with child-chatbot interactions. The influence of speech in children's learning and development derived from Vygotsky [1978] is used as a basis to explore the effect of a robot's vocalisation on children's speech and learning [55]. Lovato et al. [2019] rightly identify DAs as a more knowledgeable identity for the current age, previously only limited to human interaction partners (parents, teachers, peers, etc.) per Vygotsky [1978]. Overall, these mature theoretical ideas were seen across the review's articles as ways to provide context from children's development literature and guide understanding of children's conceptions of computers and AI systems.

4.2.6 Theory of Mind. Theory of mind (11 citations; 7% of all citations) refers to "the understanding that others have intentions, desires, beliefs, perceptions, and emotions different from one's own and that such intentions, desires, etc., affect people's actions and behaviours" [56]. This psychological capacity is known to develop from a young age (ages four onwards but sometimes even earlier) in typically developing children and can have wide-ranging cognitive impacts on human life [59]. Psychologists have used specialised tasks (e.g., false belief tasks) that assess different theory of mind competencies in cross-cultural studies to chart the developmental progression of theory of mind (ToM) in children, from infancy onwards [59]. With similar conceptual development seen worldwide, young children develop a ToM as they start to understand that people's behaviours are driven by their internal motivations and thoughts [59]. This similar development however is influenced by culture, especially the sequencing of ToM competencies, with variations seen in the order in which children in China, Iran and Turkey develop competencies compared to western children [59]. In this way, theory of mind and its extensions, theory of artificial minds, are helpful theoretical frameworks to help understand how children understand AI systems and computers as social agents.

Numerous studies in the review rely on theory of mind as a guiding theoretical construct to analyse children's interactions with various AI systems like DAs and robots [57, 58, 11, 36]. In Druga et al. [2017], children's interactions with different kinds of robots and DAs are observed through the lens of theory of mind age-related developmental progression. After interactions with AI systems, children's responses were found to be different on measures of perceived intelligence and identity attribution based on age, in line with developmental changes observed around ToM ages [11]. In their study, Yadollahi et al. [2022] explore a fundamental theory of mind concept, cognitive perspective taking between children and a robot. Based on study observations, Yadollahi et al. [2022] conclude that children created a mental model of the robot during their interactions and could adapt their perspectives based on how the robot behaved. Druga et al. [2018] investigated children's attributions of intelligence, socio-emotional capabilities, and strategies to different entities in an innovative study. While comparing the performances of a robot, mouse and human in a maze, Druga et al. [2018] found evidence for developmental differences among participants. Older children were more similar to adults in their understanding of the capabilities of robots; however, younger children were more open to evolving their understanding of robots through their experiences and observations [36]

4.2.7 Theory of Artificial Minds. While theory of mind has traditionally focused on competencies relating to understanding human minds, there is interest in children's understanding of "extraordinary minds" such as God, superheroes, and AI devices such as DAs and robots [59, p. 746]. To conceptualise what theory of mind can look like when extended to social AI systems, researchers have variously proposed a "Theory of Artificial Minds" [60, 23] or "Theory of Robot Mind" [25] or "Theory of AI Mind" [61]. Key shared ideas between these proposals are that humans and AI systems will need to understand the internal states of each other for successful social interactions (5 citations; 3% of all citations). Exploring a theory of artificial minds can also have the key benefit of increasing the transparency of AI systems and boosting community readiness to live with this technology [57].

Van Brummelen et al.'s [2021] study of children's perceptions of conversational AI is grounded in research on how interactions with agents can change people's theory of mind and considers theory of mind for AI systems, or theory of artificial minds. Approaching the topic from the view of AI literacy, Van Brummelen et al. [2021] found that children could change their perceptions of a DA and reported feeling closer to the DA after a programming and learning intervention. Dietz et al. [2023] found that children reason about DA beliefs similarly to how they reason about human agents' beliefs, even at ages 7-8. Based on study findings, it is suggested that young children may not possess an accurate theory of AI mind and rely on a theory of mind of human agents instead to understand AI systems [61]. Zhang et al. (2019) adapted false-belief tasks for robot settings, that were conducted with typically developing (TD) children and ASD children to explore a theory of mind for robots. In line with the focus of this review, only results for TD children are reported. Most children in the study, at theory of mind-sensitive ages (5-7 years old), could attribute false beliefs to a social robot and predict their subsequent actions [25]. Based on observed interactions between children in a robotic learning environment, Spektor-Precel and Mioduser [2015] propose a theory of artificial minds regarding behaving man-made objects or artefacts. It includes a first and second order understanding of artefacts, a continuum-based model of the artificial mind, and nuances to how their model differs based on robot programming and specifications [60]. This model of the artificial mind includes a continuum from an understanding of AI wholly based on a child's idea of a human mind to an entirely technological model of an artificial mind [60]. Bharadwaj et al.'s [2023] proposal for a theory of artificial minds rests on a descriptive claim and a prescriptive claim. The descriptive claim highlights the human tendency to anthropomorphise and interact naturally with social technology, whereas the prescriptive claim argues that there are tremendous opportunities available in the social world with a reciprocal theory of artificial minds [23]. Similar to ToM, the cross-cultural testing of ToAM will be important to understand similarities and variations in its development around the world.

5. LIMITATIONS

Limitations of this scoping review are as follows. First, the nature of the review necessitated the use of multiple databases across diverse fields. These databases each have peculiarities in how articles were catalogued and presented; hence, some articles were likely overlooked due to how the search was performed. Relatedly, article authors were categorised into disciplines based on article information however boundaries between disciplines may be more indistinct, in some cases. Second, the review limited its scope to theories. Articles were coded for their theoretical ideas and aspects of empirical studies such as sample sizes, statistical significance, or publication location prestige were not considered when selecting articles. Third, additional reviewers from different disciplines could be brought in to focus on the aforementioned themes and calibrate the single reviewer's selection and coding, to strengthen this scoping review.

6. CONCLUSION

Our review identifies prominent theories that describe children's understanding of computers and AI systems as social agents. Researchers are approaching this area from technical perspectives (e.g., how to build a social robot), psychological perspectives (e.g., children's perspectives), and educational perspectives (e.g., how to use robots and AI in the classroom). To conclude, three broad findings from the review will be discussed.

First, the most popular theoretical approaches to describe children's understanding of computers and AI systems as social agents derive from children's understanding of themselves, their environments, and other humans. Theoretical approaches such as anthropomorphism, animism, and theory of mind were some of the more popular ideas cited in this review. Broadly speaking, these ideas help to explain how children might understand AI systems but start from the lens of themselves and their immediate social experiences. Other social cognitive theories not discussed in detail in this paper but were otherwise present in the review included personification, social presence theory, empathy, social agency theory, and social cooperation. Hence, in the absence of a theory specific to understanding artificial agents, though some have been proposed (see Theory of artificial minds), the research community largely relies on theories that centre human beings, first and foremost.

Second, a diverse array of theoretical approaches is being used, but significant inconsistencies exist in how they are utilised and presented. Presentation of theory and theoretical approaches varies widely by discipline, wherein some theoretical approaches were more easily identifiable than others. For example, in studies from psychology or interdisciplinary studies (with psychology researchers participating), theoretical approaches are usually described under clear "Theoretical Background" headings. Whereas in studies from HCI or robotics, work is usually grounded in findings from previous empirical work with theoretical assumptions visible from brief introductory sections or references. Additionally, there is the issue of how specific theoretical ideas are being used. As has been discussed, some ideas like anthropomorphism and animism are sometimes used interchangeably to describe the understanding of artificial others or select traits as human-like. Further, some theories are attributable to a single source (see Piaget), and others are expressed as ideas (see NOC approaches) attributable to multiple sources. Hence, while findings highlight a rich diversity in approaches, there are opportunities to streamline and standardise use of theory for ease of interdisciplinary research.

Lastly, this review identifies a gap in theory regarding human understanding of AI. Instead of relying on a general theory of human understanding of others, a specific theory of understanding artificial entities is required for two reasons. First, artificial intelligence systems are unlike natural biological entities, that humans are already familiar with, and can confuse both adults and children when thinking about the boundaries between human and artificial [44]. Further, their complex affordances (e.g., information capacity) call for different frameworks of understanding. Second, specific theories allow for a more targeted approach to the enquiry of a phenomenon and create the ability to make precise predictions about human behaviour. For our purposes, it can provide insights into how humans might behave as they interact with different kinds of AI systems. With the advent of generative AI models in text, speech, and video, such as ChatGPT [62], AI systems will become increasingly adept at natural communication with humans and established in our social world. Hence, there is a strong need to build out a dedicated and robust theoretical model of how humans, from birth onwards, understand and interact with AI systems, such as ones that researchers have begun to propose [60, 23, 25, 61]. In this way, weaving a new theory of human understanding of AI involves consideration of established theories of human understanding of others as well as a focus on unique AI capabilities.

7. RECOMMENDATIONS

To further support the creation of a comprehensive theory of artificial minds (ToAM), we call on researchers to mobilize around three guiding questions. First, what are the salient competencies of a ToAM? This question is proposed to define and refine a common understanding of this term and its key underlying constructs (e.g., belief, intention etc.). Research in this vein would develop reliable measures of ToAM competencies and identify ones that best account for the developmental trajectory of and individual variability in ToAM. Second, how can we take this multidisciplinary patchwork of theories and approaches and weave them into a truly interdisciplinary ToAM? This question addresses the crucial need for disciplinary partnership on the creation and use of this framework and its measures. Third, what will ToAM mean cross-culturally? This question supports the need for empirical cross-cultural development and testing of ToAM similar to the theory of mind framework, given the global impact of AI. The consideration of these questions will ensure that we create and use a robust, inclusive, and safe model of human-AI understanding that continues to evolve with new developments in AI technologies.

ACKNOWLEDGMENTS

This work is supported by the Fonds de Recherche Société et Culture (FRQSC) Québec, bourses de doctorate en recherche (B2Z Doctoral Training Scholarship) awarded to first author in 2023.

REFERENCES

[1] Matthew B. Hoy. 2018. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly* 37, 1 (January 2018), 81–88. https://doi.org/10.1080/02763869.2018.1404391

[2] Brooke Auxier, Monica Anderson, Andrew Perrin, and Erica Turner. 2020. *Children's engagement with digital devices, screen time*. Pew Research Center. Retrieved December 14, 2021 from https://www.pewresearch.org/internet/2020/07/28/childrens-engagement-with-digital-devices-screen-time/

[3] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How smart are the smart toys? Children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 2018. 231–240.

[4] Randi Williams, Cynthia Breazeal, Christian Vazquez MacHado, Pattie Maes, and Stefania Druga. 2018. "My doll says it's ok": A study of children's conformity to a talking doll. (2018), 625–631. https://doi.org/10.1145/3202185.3210788

[5] Muneeb Imtiaz Ahmad, Omar Mubin, and Joanne Orlando. 2016. Children views' on social Robot's adaptations in education. *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (2016), 145–149. https://doi.org/10.1145/3010915.3010977

[6] Shruti Chandra, Raul Paradeda, Hang Yin, Pierre Dillenbourg, Rui Prada, and Ana Paiva. 2018. Do children perceive whether a robotic peer is learning or not? In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018. 41–49. .

[7] Alireza M. Kamelabad and Gabriel Skantze. 2023. I learn better alone! Collaborative and individualword learning with a child and adult robot. *ACM/IEEE International Conference on Human-Robot Interaction* (2023), 368–377. https://doi.org/10.1145/3568162.3577004

[8] Leona Chandra Kruse, Katrin Bergener, Kieran Conboy, Jenny Eriksson Lundström, Alexander Maedche, Suprateek Sarker, Isabella Seeber, Armin Stein, and Cathrine E. Tømte. 2023. Understanding the Digital Companions of Our Future Generation. *Communications of the Association for Information Systems* 52, 1 (2023), 22.

[9] Janik Festerling and Iram Siraj. 2020. Alexa, What Are You? Exploring Primary School Children's Ontological Perceptions of Digital Voice Assistants in Open Interactions. *Human Development* 64, (June 2020), 1–18. https://doi.org/10.1159/000508499

[10] Radhika Garg and Subhasree Sengupta. 2020. He is just like me: a study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–24.

[11] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is it OK if I eat you?": Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children* (*IDC '17*), June 27, 2017, New York, NY, USA. Association for Computing Machinery, New York, NY, USA, 595–600.. https://doi.org/10.1145/3078072.3084330

[12] Ying Xu and Mark Warschauer. 2020. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (*CHI '20*), April 21, 2020, New York, NY, USA. Association for Computing Machinery, New York, NY, USA, 1–13.. https://doi.org/10.1145/3313831.3376416

[13] Sara Aeschlimann, Marco Bleiker, Michael Wechner, and Anja Gampe. 2020. Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior* 112, (November 2020), 106466. https://doi.org/10.1016/j.chb.2020.106466

[14] Anna Hoffman, Diana Owen, and Sandra L. Calvert. 2021. Parent reports of children's parasocial relationships with conversational agents: Trusted voices in children's lives. *Human Behavior and Emerging Technologies* 3, 4 (2021), 606–617. https://doi.org/10.1002/hbe2.271

[15] Lauren N. Girouard-Hallam, Hailey M. Streble, and Judith H. Danovitch. 2021. Children's mental, social, and moral attributions toward a familiar digital voice assistant. *Human Behavior and Emerging Technologies* n/a, n/a (2021). https://doi.org/10.1002/hbe2.321

[16] Elin A Bjorling, Emma Rose, Andrew Davidson, Rachel Ren, and Dorothy Wong. 2020. Can we keep him forever? Teens' engagement and desire for emotional connection with a social robot. *International Journal of Social Robotics* 12, 1 (2020), 65–77. https://dx.doi.org/10.1007/s12369-019-00539-6

[17] Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen. 2012. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology* 48, 2 (March 2012), 303–314. https://doi.org/10.1037/a0027033

[18] Daan Robben, Eriko Fukuda, and Mirjam De Haas. 2023. The Effect of Gender on Perceived Anthropomorphism and Intentional Acceptance of a Storytelling Robot. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023. 495–499.

[19] Caroline L. van Straten, Jochen Peter, and Rinaldo Kühne. 2020. Child-Robot Relationship Formation: A Narrative Review of Empirical Research. *Int J Soc Robot* 12, 2 (2020), 325–344. https://doi.org/10.1007/s12369-019-00569-0

[20] Brendan Bartlett, Vladimir Estivill-Castro, and Stuart Seymon. 2004. Dogs or Robots: Why do Children see them as Robotic Pets rather than Canine Machines? 2004. .

[21] Gail F. Melson, Peter H. Kahn, Jr., Alan Beck, and Batya Friedman. 2009. Robotic Pets in Human Lives: Implications for the Human-Animal Bond and for Human Relationships with Personified Technologies. *Journal of Social Issues* 65, 3 (September 2009), 545–567. https://doi.org/10.1111/j.1540-4560.2009.01613.x

[22] Judith H. Danovitch. 2019. Growing up with Google: How children's understanding and use of internet-based devices relates to cognitive development. *Human Behavior and Emerging Technologies* 1, 2 (April 2019), 81–90. https://doi.org/10.1002/hbe2.142

[23] Nandini Asavari Bharadwaj, Adam K. Dubé, Victoria Talwar, and Elizabeth Patitsas. 2023. Developing a theory of artificial minds (ToAM) to facilitate meaningful human-AI communication. In *R. McEwen, A. L. Guzman, & S. Jones (eds.). Handbook of Human Machine-Communication*. Sage publishing.

[24] Micah D.J. Peters, Christina M. Godfrey, Hanan Khalil, Patricia McInerney, Deborah Parker, and Cassia Baldini Soares. 2015. Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare* 13, 3 (September 2015), 141–146. https://doi.org/10.1097/XEB.0000000000000050

[25] Yaoxin Zhang, Wenxu Song, Zhenlin Tan, Yuyin Wang, Cheuk Man Lam, Sio Pan Hoi, Qianhan Xiong, Jiajia Chen, and Li Yi. 2019. Theory of Robot Mind: False Belief Attribution to Social Robots in Children With and Without Autism. *Front Psychol* 10, (August 2019). https://doi.org/10.3389/fpsyg.2019.01732

[26] Rhonda McEwen and Adam Dubé. 2017. *Understanding Tablets from Early Childhood to Adulthood: Encounters with Touch Technology*. Routledge, New York. https://doi.org/10.4324/9781315389486

[27] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review* 114, 4 (2007), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

[28] Clara Strathmann, Jessica Szczuka, and Nicole Krämer. 2020. She talks to me as if she were alive: Assessing the social reactions and perceptions of children toward voice assistants and their appraisal of the appropriateness of these reactions. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 2020. 1–8.

[29] Joseph E. Michaelis and Bilge Mutlu. 2018. Reading socially: Transforming the in-home reading experience with a learning-companion robot. *Science Robotics* 3, 21 (August 2018). https://doi.org/10.1126/scirobotics.aat5999

[30] David Cameron, Samuel Fernando, Emily C. Collins, Abigail Millings, Michael Szollosy, Roger Moore, Amanda Sharkey, and Tony Prescott. 2017. You made

him be alive: Children's perceptions of animacy in a humanoid robot. 10384, (2017), 73–85. https://doi.org/10.1007/978-3-319-63537-8_7

[31] Suleman Shahid, Emiel Krahmer, Marc Swerts, and Omar Mubin. 2010. Child-robot interaction during collaborative game play: effects of age and gender on emotion and experience. *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (2010), 332–335. https://doi.org/10.1145/1952222.1952294

[32] Caroline L. van Straten, Jochen Peter, and Ronaldo Kühne. 2023. Transparent robots: How children perceive and relate to a social robot that acknowledges its lack of human psychological capacities and machine status. *International Journal of Human Computer Studies* 177, (2023). https://doi.org/10.1016/j.ijhcs.2023.103063

[33] Jacqueline M. Kory-Westlund, Marayna Martinez, Maryam Archie, Madhurima Das, and Cynthia Breazeal. 2016. Effects of framing a robot as a social agent or as a machine on children's social behavior. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, August 2016, New York, NY, USA. IEEE, New York, NY, USA, 688–693. . https://doi.org/10.1109/ROMAN.2016.7745193

[34] Roger Andre Søraa, Pernille Søderholm Nyvoll, Karoline Blix Grønvik, and J. Artur Serrano. 2021. Children's perceptions of social robots: a study of the robots Pepper, AV1 and Tessa at Norwegian research fairs. *AI & society* 36, (2021), 205–216.

[35] Caroline L. van Straten, Jochen Peter, Rinaldo Kuhne, and Alex Barco. 2020. Transparency about a Robot's Lack of Human Psychological Capacities: Effects on Child-Robot Perception and Relationship Formation. *ACM Transactions on Human-Robot Interactions* 9, 2 (February 2020). https://doi.org/10.1145/3365668

[36] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How smart are the smart toys? Children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 2018. 231–240.

[37] Tanya N. Beran, Alejandro Ramirez-Serrano, Roman Kuzyk, Meghann Fior, and Sarah Nugent. 2011. Understanding how children understand robots: Perceived animism in child–robot interaction. *International Journal of Human-Computer Studies* 69, 7–8 (2011), 539–550.

[38] Fritz Heider and Marianne Simmel. 1944. An Experimental Study of Apparent Behavior. *The American Journal of Psychology* 57, 2 (1944), 243–259. https://doi.org/10.2307/1416950

[39] Sherry Turkle. 2005. *The second self: computers and the human spirit* (20th anniversary ed., 1st MIT Press ed ed.). MIT Press, Cambridge, Mass.

[40] Kaveri Subrahmanyam, Rochel Gelman, and Alyssa Lafosse. 2002. Animates and other separably moveable objects. *Category specificity in brain and mind* (2002), 341–373.

[41] J. Piaget. 1929. The child's conception of the world Routledge and Kegan Paul. *Ltd: London* (1929).

[42] Susan Carey. 1985. *Conceptual Change in Childhood*. MIT Press. Retrieved August 8, 2023 from https://www.scribd.com/document/484114288/Carey-1985-Conceptual-Change-in-Childhood-pdf

[43] Silvia B. Lovato, Anne Marie Piper, and Ellen A. Wartella. 2019. Hey Google, Do Unicorns Exist? Conversational Agents as a Path to Answers to Children's Questions. *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (2019), 301–313. https://doi.org/10.1145/3311927.3323150

[44] Sandra Okita, Daniel Schwartz, Takanori Shibata, and Tokuda Hideyuki. 2015. Young children's preconceived notions about robots, and how beliefs may trigger children's thinking and response to robots. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* 2015, (2015), 728–733. https://doi.org/10.1109/ROMAN.2015.7333690

[45] Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996).

[46] Ekaterina Pashevich. 2023. Conceptualizing Empathic Child–Robot Communication. In *The Sage Handbook of Human–Machine Communication*. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road London EC1Y 1SP. https://doi.org/10.4135/9781529782783

[47] Amy Ogan, Samantha Finkelstein, Elijah Mayfield, Claudia D'adamo, Noboru Matsuda, and Justine Cassell. 2012. " Oh dear stacy!" social interaction, elaboration, and learning with teachable agents. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012. 39–48.

[48] Ying Xu. 2023. Talking with machines: Can conversational technologies serve as children's social partners? *Child Development Perspectives* 17, 1 (2023), 53–58. https://dx.doi.org/10.1111/cdep.12475

[49] Jacqueline M. Kory-Westlund and Cynthia Breazeal. 2019. Assessing children's perceptions and acceptance of a social robot. *Proceedings of the 18th ACM International Conference on Interaction Design and Children, IDC 2019* (June 2019), 38–50. https://doi.org/10.1145/3311927.3323143

[50] Rachel L. Severson and Stephanie M. Carlson. 2010. Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Networks* 23, 8 (2010), 1099–1103. https://doi.org/10.1016/j.neunet.2010.08.014

[51] Jean Piaget. 1964. Development and Learning. *Journal of Research in Science Teaching* 2, (1964), 176–186.

[52] Lev Semenovich Vygotsky. 1978. *Mind in society: Development of higher psychological processes*. Harvard university press.

[53] Vicky Charisi, Luis Merino, Marina Escobar, Fernando Caballero, Randy Gomez, and Emilia Gómez. 2021. The effects of robot cognitive reliability and social positioning on child-robot team dynamics. In *2021 IEEE international conference on robotics and automation (ICRA)*, 2021. IEEE, 9439–9445.

[54] Bengisu Cagiltay, Joseph E. Michaelis, Sarah Sebo, and Bilge Mutlu. 2022. Exploring Children's Preferences for Taking Care of a Social Robot. In *Interaction Design and Children*, 2022. 382–388.

[55] Lauren L. Wright, Aditi Kothiyal, Kai O. Arras, and Barbara Bruno. 2022. How a Social Robot's Vocalization Affects Children's Speech, Learning, and Interaction. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2022. IEEE, 279–286.

[56] APA. n.d. theory of mind – APA Dictionary of Psychology. Retrieved December 23, 2021 from https://dictionary.apa.org/theory-of-mind

[57] Jessica Van Brummelen, Tommy Heng, and Viktoriya Tabunshchyk. 2021. Teaching Tech to Talk: K-12 Conversational Artificial Intelligence Literacy Curriculum and Development Tools. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 17 (May 2021), 15655–15663.

[58] Elmira Yadollahi, Marta Couto, Pierre Dillenbourg, and Ana Paiva. 2022. Do Children Adapt Their Perspective to a Robot When They Fail to Complete a Task? (2022), 341–351. https://doi.org/10.1145/3501712.3529719

[59] Henry M. Wellman. 2018. Theory of mind: The state of the art. *European Journal of Developmental Psychology* 15, 6 (November 2018), 728–755. https://doi.org/10.1080/17405629.2018.1435413

[60] Karen Spektor-Precel and David Mioduser. 2015. The influence of constructing robot's behavior on the development of Theory of Mind (ToM) and Theory of Artificial Mind (ToAM) in young children. *Proceedings of IDC 2015: The 14th International Conference on Interaction Design and Children* (2015), 311–314. https://doi.org/10.1145/2771839.2771904

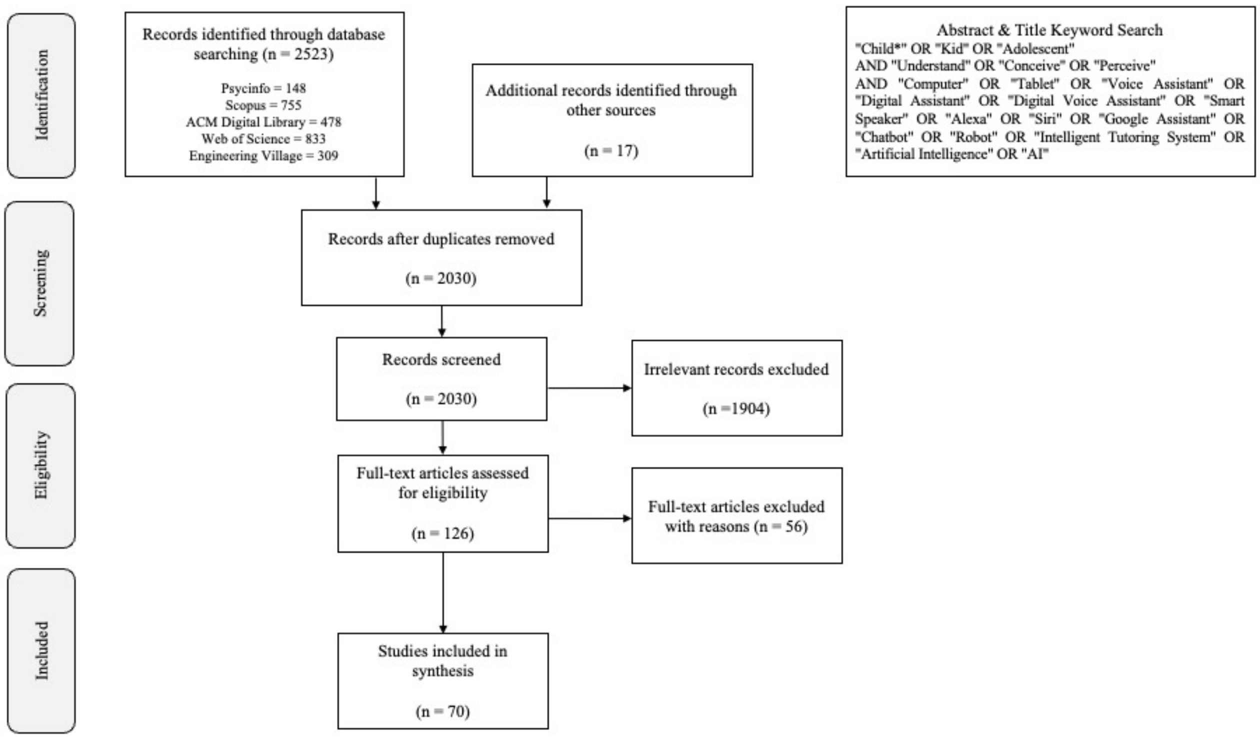
[61] Griffin Dietz, Joseph Outa, Lauren Lowe, James A Landay, and Hyowon Gweon. Theory of AI Mind: How adults and children reason about the "mental states" of conversational AI.

[62] Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, and Vanessa Parli. 2023. The AI index 2023 annual report. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA* (2023).

Works Included in Scoping Review

APPENDIX

Figure 1: Scoping review process using PRISMA Flow diagram.



[a] Flow diagram format based on Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-SCR) diagram (Tricco et al, 2018)

Table 1: Article Features

Category	No. of Articles	% of Articles
Article Features		
Article Themes		
Child-AI interactions and understanding	5	7%
Child-Chatbot interactions and understanding	1	1%
Child-Digital Assistant interactions and understanding	16	23%
Child-Intelligent Tutoring System interactions & understanding	1	1%
Child-Robot interactions and understanding	43	61%
Child-Smart Toy interactions and understanding	2	3%
Child-Technology interactions and understanding	2	3%
Article Disciplines		
Computer Science and Communication Studies	11	16%
Human-Computer Interaction	29	41%
Interdisciplinary	22	31%
Psychology	8	11%
Publisher Type		
Book	1	1%
Conference Proceeding	45	64%
Handbook	2	3%
Journal	22	31%
Target population		
Children (no age specified)	9	13%
Young Children (birth – 10 years of age)	32	46%
Children* (Includes children under 10 and over 10)	16	23%
Adolescent (11 years - 18 years of age)	10	14%
Parents and Children	3	4%
Study Location		
In person; at home	6	9%
In person; at home and in lab	1	1%
In person; at school	16	23%
In person; in lab	24	34%
In person; public engagement events**	7	10%
In person; multiple locations	1	1%
Virtual Study	7	10%
Not Applicable***	8	11%

Category	No. of Articles	% of Articles
Sample Sizes		
Under 20 participants	11	16%
20 – 40 participants	17	24%
40 – 60 participants	11	16%
60 – 80 participants	6	9%
80 – 100 participants	10	14%
Over 100 participants	7	10%
Not Applicable***	8	11%

Note. Total number of articles included in review is 70.
*Articles have a mixed group of ages under and over 10 years of age.
** Public engagement events include studies held at museums, science fairs and community centres
*** Not applicable indicates book chapters or articles not describing empirical work

Table 2: Theoretical Approaches

Theoretical Approach (Level 1)	Description		Computer Science & Communication Studies	Human-Computer Interaction	Interdisciplinary	Psychology
		# of citations (% total)	# of citations (% w/I discipline)	# of citations (% w/I discipline)	# of citations (% w/I discipline)	# of citations (% w/I discipline)
Animism	Theoretical idea that refers to the belief in life and "aliveness" for non-living objects	15 (10%)	1 (3%)	7 (12%)	5 (11%)	2 (8%)
Anthropomorphism	Theoretical idea that involves the attribution of human-like properties, mental states, and qualities to real or imagined non-human objects or entities	25 (16%)	5 (17%)	7 (12%)	11 (25%)	2 (8%)
Communication Theories*	Describes theories that refer to the practice of transmission and receipt of information as communication	7 (4%)	1 (3%)	2 (3%)	1 (2%)	3 (12%)
Media Equation Theory	Communication paradigm that posits that humans treat different media in natural social ways	5 (3%)	2 (7%)	2 (3%)	0 (0%)	1 (4%)
New Ontological Category Theories	Diverse set of approaches that suggest that understanding of AI may require the creation of a new ontological category to categorise entitles, different from existing ones	6 (4%)	0 (0%)	1 (2%)	1 (2%)	4 (15%)
Other Theories***	Diverse theoretical approaches and ideas that cannot be categorised together	39 (25%)	7 (24%)	14 (24%)	13 (30%)	5 (19%)
Social Interaction Frameworks**	Diverse frameworks that relate to social interactions and understanding of others in society.	33 (21%)	10 (34%)	14 (24%)	6 (14%)	3 (12%)
Sociocultural Theory of Cognitive Development	Vygotksy (1978) theory of the influence of social factors and others in children's development	5 (3%)	0 (0%)	3 (5%)	1 (2%)	1 (4%)
Theory of Artificial Minds	Proposed psychological reciprocal capacity to understand self and artificial intelligence others' beliefs, intentions, desires, and actions	5 (3%)	1 (3%)	1 (2%)	2 (5%)	1 (4%)
Theory of Cognitive Development	Piaget (1974) theory of how children develop cognitive abilities and learn through interactions with others	4 (3%)	0 (0%)	1 (2%)	1 (2%)	2 (8%)
Theory of Mind	Psychological capacity to understand self and human others' beliefs, intentions, desires, and actions	11 (7%)	1 (3%)	5 (9%)	3 (7%)	2 (8%)
Uncanny Valley	Theory that describes relation between appearance of objects, specifically their resemblance to humans and human reactions to them	2 (1%)	1 (3%)	1 (2%)	0 (0%)	0 (0%)
Grand Total		**157**	**29**	**58**	**44**	**26**

[a] Total number of articles included in review is 70.
* Category includes 6 different sub-categories
** Category includes 20 different sub-categories
*** Category includes 32 different sub-categories