Children’s Fundamental Rights in Human-Robot Interaction Research: A Systematic Review

Daniella DiPaola†
dipaola@media.mit.edu
MIT Media Lab
Cambridge, Massachusetts, USA

Cynthia Breazeal
cynthiab@media.mit.edu
MIT Media Lab
Cambridge, Massachusetts, USA

Vicky Charisi†
vasiliki.charisi@ec.europa.eu
European Commission, Joint Research Centre
Seville, Spain

Selma Šabanović
selmas@indiana.edu
Indiana University Bloomington
Bloomington, Indiana, USA

ABSTRACT
Citizens and policy institutions increasingly express their concerns regarding the emerging challenges in the context of Artificial Intelligence (AI) and have concrete demands for the protection of human rights. In parallel, studies in the field of AI and Human-Robot Interaction (HRI) indicate the impact of social robots on children’s development. We conducted a systematic review based on UNICEF’s AI Policy Guidance to map the landscape of research on social robots and children’s rights. We used the PRISMA method and identified N = 37 papers that address one of the rights, which we then annotated to indicate tendencies and areas of alignment and misalignment with the UNICEF guidance. Our findings reveal that although the field of HRI is looking at specific rights, with a focus on inclusion, some of the rights have been under-researched. Furthermore, we observed a misalignment between HRI and UNICEF regarding the terminology. With this paper, we hope to bring awareness to the field of HRI regarding children’s rights and to highlight directions for alignment among research, societal needs, and policy.

CCS CONCEPTS
• Social and professional topics → Computing / technology policy; • General and reference → Surveys and overviews; • Computing methodologies → Artificial intelligence.

KEYWORDS
human-robot interaction, human rights, technology policy

ACM Reference Format:

†Both authors contributed equally to this research.

1 INTRODUCTION
The current momentum of advancements in Artificial Intelligence (AI) and the ways AI systems impact human behaviour and our societies have triggered policy institutions to propose mechanisms to systematically facilitate the emerging opportunities associated with AI systems and at the same time to safeguard all citizens from possible harms. One of the frameworks that has been used as a foundation for the evaluation of AI systems is the Universal Declaration of Human Rights by the United Nations [47]. For example, the Blueprint for an AI Bill of Rights [31] proposes a set of principles and associated practices to help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of AI. Towards a similar direction but with certain differences, the AI Act of the European Commission [17] proposes a risk-based legal framework for the regulation of AI systems in Europe and suggests that as risk increases, stricter rules should apply to protect all citizens’ rights.

In this context, international organizations that specialize in children’s protection have proposed that special attention is needed in regulating AI systems for children. Children and young people have the same general human rights as adults and also specific rights that recognize their particular characteristics and needs. In 1989, governments across the world promised all children the same rights as adults in the United Nations Convention on the Rights of the Child (UNCRC) [46], the international human rights treaty which sets out the civil, political, social and cultural rights of children.

While for the analog world, there are well-established mechanisms that support the implementation of the UNCRC, the global community has several concerns about the lack of mechanisms for the protection of children’s rights in the digital world and even more so in the context of children’s activities that are algorithmically mediated. Given the evolving and expanding digital environment and the opportunities and the risks that the digital world brings for children, in 2021, the United Nations (UN) adopted the Comment 25 of the rights of the child in the digital environment and children’s possible interactions within digital networks, content, services and applications including robotics. Considering the potential risk that AI systems might bring, the United Nations International Children’s Emergency Fund (UNICEF) has published a set of policy guidelines on AI and children [15]. In both documents, Comment 25 and UNICEF’s guidance, there is a reference to robotics as a
distinct application with embedded AI components, with specific characteristics that relate to the embodied nature of robots.

However, to our knowledge, none of the existing initiatives elaborate on the emerging challenges and opportunities that are specific to social robots. A recent review of papers that examine age-appropriate AI systems in relation to children’s rights reports that out of the 188 papers included in the review, only 13 papers refer to social robots [50]. The inclusion criteria of this review were not specific to existing work on child-robot interaction, which means that more targeted investigation might be needed.

Indeed, the field of child-robot interaction is still in its early stages; however, an increasing number of studies indicate that certain features of social robots might affect children’s social behaviour [9, 32], their learning in various settings [3], and their problem-solving skills [7], among others. The majority of the studies in child-robot interaction consider features related to the physical nature of social robots such as body motion, gestures, facial expressions, and head orientation that might stand alone or accompany vocalizations or speech. In addition, in terms of perception, physical robotic agents have the potential to be embedded in children’s physical and social environment which generate robot-specific opportunities and challenges that should be considered for informed decisions. For well-informed policy-making and for the field of HRI to identify possible gaps and research need, a systematic overview of the existing work is needed.

This paper aims to map the current state-of-the-art in the field of child-robot interaction in relation to children’s rights with a systematic literature review on existing works that directly or indirectly address at least one of the rights as introduced by the UNICEF Policy Guidance on AI and children and to articulate current tendencies.

2 METHODS

We conducted a systematic literature review by collecting peer-reviewed papers from three databases: ACM Digital Libraries, IEEE Explore, and Web of Science. The search terms regarding children’s rights were based on the policy requirements proposed by UNICEF as appearing in Table 1. It should be noted that in the search, we did not include three of the recommendations from the UNICEF guidance for the following reasons: (i) “Empower governments and businesses with knowledge of AI and children’s rights” since this requirement is not research specific; (ii) “Support children’s development and well-being” because this requirement is too generic for the purposes of the current review; and (iii) “Prepare children for present and future developments in AI” because this requirement is beyond the scope of the current review as it focuses on AI literacy and will be considered separately in our future work.

We followed the PRISMA procedure by identifying keywords, inclusion and exclusion criteria. Figure 1 indicates the selection process of eligible papers. Abstracts were searched for the words “social AND robot*” AND “child* OR kid* OR teen*” AND the presence of a rights word as appears in Table 1. The decision to include “social AND robot*” was based on our goal to exclude papers that use robots as tools for children’s programming skills [35]. Preliminary searches yielded a total of 158 papers: accountability (3), explainability (10), fairness (13), inclusion (88), privacy (10), safety (28), and transparency (5). In the case that a paper elaborates on more than one right, this paper is categorized multiple times.

Two researchers screened the 158 papers against the inclusion and exclusion criteria, whether the paper was peer-reviewed, focused on children, and social robots. Papers were then divided into user studies, position papers, and literature reviews, and annotated accordingly. A subset of the papers were screened and annotated by two researchers and a percent agreement was reached.
3 RESULTS

3.1 Overview of Papers

Thirty-seven out of 158 papers were screened to contain the main criteria. This included 36 distinct papers as one was included in both accountability and privacy. Most papers were found for inclusion (14) and the least amount in accountability (1). Additionally, a user study paper was the most common type of paper (27) as compared to position papers (7) and literature reviews (4). For more information on the papers, see Figure 2. We further evaluated the largest group of papers, user studies. Each study was annotated for the participants’ ages, sample size, methodology, setting, and geographic location. The majority of studies included children between 8-9 (17), 6-7 (15), and 10-11 (12). The most common sample size was less than 10 people (10), followed by 20-50 people (6). 20 of the studies took place in Europe, 2 in North America, 2 in South America, 1 in Australia, 1 in Asia, and none in Africa. They most commonly took place in a school (11) or lab (10), and the average length was a single session (15). Nine studies lasted longer than 2 days, with three of them longer than 2 months. The most common robot used across all studies was the Nao robot (12) followed by CASTOR (2). All other robots (16) were found in only one study. Seven of the studies used a tele-operated robot. The robot’s apparent functionality, regardless if it was controlled, was mapped by and the most common functionalities were Verbal (19), Physical (16), Vision (9), and Navigation (8). Finally, we mapped the robot’s functionalities to the “most used techniques found in common AI applications”, as defined by UNICEF. Natural language processing and rule-based models (15) were the most commonly used, followed by computer vision (9), planning techniques (3), learning from examples (1), and reinforcement learning (1). No papers used predictive analytics and four papers did not include any of UNICEF’s listed techniques.

3.2 Accountability

The search revealed one paper that directly addresses the notion of accountability in the context of social robots and children. The paper [38] refers to accountability as one of the robot characteristics that connects with the robot’s explainability and transparency. While this aspect seems catalytic for child-robot interaction, the UNICEF policy guidance refers to accountability as a policy action rather than as a system’s characteristic; this indicates that the term can be used in different ways in different contexts. While the term accountability can be used in reference to the robot behaviour, the UNICEF guidance proposes ways in which the national governments and the involved parties should exercise their accountability for children’s rights including the establishment of oversight bodies compliant with principles and regulations as well as the auditing and the monitoring of the systems against children’s rights.

3.3 Explainability

The search revealed three papers which were all user studies that directly address the issue of explainability but from three different angles. In the first case [51], the robot intervenes in a way to trigger explanatory behaviour by children; in the second paper [36], the robot uses task-specific explanations in order to make the content more accessible by children which results in a positive impact, especially for slower learners; in the third paper [22], the robot was used for a behaviour change setting with children with diabetes. We identified more papers that provide insights on explainability in child-robot interaction from a more technical point of view such as in [45] but did not meet the inclusion criteria of our review. Considering the current work on explainable robots for children in relation to the UNICEF’s guidelines on explainability, we can discern that scholars focus on the examination of the impact of age-appropriate robot explanations on children’s behaviour as well as on technical advancements on autonomous explanation generation. Our review did not reveal papers that consider the explanations towards parents or caregivers, but probably this is related to the inclusion and exclusion criteria of this review that might have excluded papers that address other stakeholders than children.

3.4 Fairness

Our search for the term fairness resulted in two papers, both of them including a user study that directly aimed to research the concept of fairness, albeit from a different perspective [21, 37]. The first paper, explored children’s perceptions about fairness towards the robot while the second one took the angle of fairness from a robot-tutor in comparison with a human-tutor. While the definition of fairness and non-discrimination given by UNICEF focuses on the user diversity in terms of age, ethnicity, gender identities, socio-economic status of children etc, which is required for the development of fair and non-discriminatory intelligent systems for children, the studies that were revealed with our search take a slightly different perspective and contribute to our understanding about children’s perceptions of fairness in their interaction with the robot. While this research direction provides insights on how to design robots that can be perceived as fair by children, at the same time, in order for the HRI field to address the concept of fairness, as proposed by UNICEF, the field needs to consider more marginalized populations.

3.5 Inclusion

The search on inclusion yielded 14 papers, the largest amount for any right. Thirteen of these papers included a user study [11–14, 19, 20, 24–26, 28, 30, 33, 40], four of which used co-design practices, which were outlined by UNICEF as a preferred methodology for inclusion. The other paper was a review paper about telepresence robots fostering inclusion [10]. Many of the studies were conducted in a school setting and focused on inclusive teaching and learning practices. For example, De Carolis et al. [12] used a robot to foster inclusion between migrants by teaching them culture-related gestures. Seven of the studies focused on inclusion for children with autism spectrum disorder. The work shows that there are opportunities for the UNICEF guidelines to encourage using AI to promote inclusion for children with mixed abilities. There are also opportunities for the HRI community to incorporate the specific groups that UNICEF mentions (i.e. gender identity, culture, and geographic diversity).

3.6 Privacy

Six papers were found to include privacy in the abstract, all of which were written to address privacy indirectly. Three of the papers were user studies where privacy was explicitly mentioned.
or brought up by participants [4, 41, 43]. For example, in Cagiltay et al [4], parents who were interviewed expressed privacy concerns between household members with a shared robot [4]. These papers all include results about privacy because they were brought up by participants or are an ethical concern for a particular domain of robotics. Future work should center on privacy-by-design features and/or privacy perceptions.

### 3.7 Safety

The search for safety led to eight papers; three user studies [5, 39, 44], three position papers [1, 18, 23], and 2 review papers [34, 52]. All eight focused on physical safety. An example of physical safety is a user study by Taufatodua et al., which concludes that robots must be designed for safe contact with children [44]. The literature review and UNICEF mainly talk about physical safety with robots. While the focus on physical safety might be expected since robots are embodied agents, they can still have implications for the mental safety of the child, for example, deception. In the guidelines, there are issues of psychological safety that need to be researched (i.e. advertising deception [16]). There are opportunities for the HRI community to publish work on psychological and emotional safety.

### 3.8 Transparency

Our search on transparency revealed three papers, two of which reported user studies [42, 48] and one as a position paper [27]. All three papers focused on the importance of informing the user about the capabilities of the robot as a means to mitigate their tendency to attribute anthropomorphic characteristics. The user studies reported that transparency affected neither children’s perceptions of the robot’s social presence nor their closeness to and trust in the robot. As such, informing the user about the capabilities of the robot can increase the user’s perception of the trustworthiness of the system while not affecting the interaction. Although in some occasions robot designs are moving in an opposite direction to increase children’s engagement with the robot, the included studies appear in accordance with the provided definitions by UNICEF.

### 4 CONCLUSION

We synthesized existing research in the field of HRI regarding social robots and children’s rights. We took as a starting point the UNICEF’s policy guidance of AI and children to map the existing landscape of research in the field of child-robot interaction and identified current tendencies.

Our analysis revealed that the HRI community identifies the non-verbal interaction related to the physical nature of robots as equally important to the verbal interaction and that the embodiment renders robots distinct from other AI-based devices, which is in accordance with the existing literature [49]. However, often the physical nature of a robot monopolizes the interest of research; the majority of the papers of our review, for example, express concerns about children’s physical safety, and only a small number of papers refer to the psychological or mental safety of children.

This review illuminated a misalignment between the use of terms by UNICEF and the ways scholars in the field of HRI use specific terms, such as transparency, explainability, and accountability. With our current and future work, we hope to contribute to the creation of a common ground between policymakers and researchers in the field of social robots and children’s rights (e.g [6]), including the specification of a common terminology.

Lastly, while the identified rights can be addressed from various perspectives, we observed that in the field of social robots, there are homogeneous directions. For example, while inclusion seems one of the most popular topics for research, the included papers focus on one area of inclusion (autistic children) while research with children from under-represented geographical areas, such as with children in the global south (e.g. [8]) are rare.

With this systematic review, we provide an initial mapping of the current research on social robotics and children’s rights. Although researching children’s interaction with social robots is among the popular areas of research in the field of HRI [3], our work reveals a relatively limited amount of papers in relation to children’s rights. Interestingly, we observed that many of these studies use robots to promote children’s rights. We hope this review brings awareness to the HRI community about children’s rights in order for our systems and robotic agents to have the best possible societal impact.

### REFERENCES


