

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,500

Open access books available

176,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Towards Children-Centred Trustworthy Conversational Agents

Marina Escobar-Planas, Vicky Charisi, Isabelle Hupont, Carlos-D Martínez-Hinarejos and Emilia Gómez

Abstract

Conversational agents (CAs) have been increasingly used in various domains, including education, health and entertainment. One of the growing areas of research is the use of CAs with children. However, the development and deployment of CAs for children come with many specific challenges and ethical and social responsibility concerns. This chapter aims to review the related work on CAs and children, point out the most popular topics and identify opportunities and risks. We also present our proposal for ethical guidelines on the development of trustworthy artificial intelligence (AI), which provide a framework for the ethical design and deployment of CAs with children. The chapter highlights, among other principles, the importance of transparency and inclusivity to safeguard user rights in AI technologies. Additionally, we present the adaptation of previous AI ethical guidelines to the specific case of CAs and children, highlighting the importance of data protection and human agency. Finally, the application of ethical guidelines to the design of a conversational agent is presented, serving as an example of how these guidelines can be integrated into the development process of these systems. Ethical principles should guide the research and development of CAs for children to enhance their learning and social development.

Keywords: chatbots, children, conversational agents, trustworthy AI, ethics

1. Introduction

Conversational agents (CAs) are computer programs designed to engage in a conversation with humans through voice or text-based interactions [1]. Nowadays, their availability and popularity are dramatically increasing. They start to be embedded in a wide range of devices used on a daily basis and in ubiquitous ways: mobile phones, smart cars [2], home devices and even social robots and toys [3].

Very recent improvements in this technology are driven by ground-breaking artificial intelligence (AI) techniques [4, 5]. They are allowing for a brand new generation of CAs, such as ChatGPT [6] and GPT-4 [7], which have demonstrated unprecedented levels of autonomy and natural language processing capabilities. These systems have the potential to transform the way humans interact with computers and machines,

offering simple and intuitive interfaces for a wide range of applications, from customer service [8] to entertainment [9], personal assistance [10], healthcare [11] or education [12]. However, this rapid progress also raises important ethical concerns, particularly when it comes to vulnerable populations such as children [13]. As such, there is a growing need for research on the ethical challenges to be tackled when it comes to the development and deployment of CAs that are conceived or that can potentially be used by children, to ensure that these systems are designed with the best interests of children in mind.

The goal of this chapter is to explore the unique considerations involved in the development and deployment of conversational agents for children. While there is a growing body of research on the ethical development of artificial intelligence in general, much of this work has yet to be fully applied to the specific challenges and opportunities of CAs for children. Our aim is to highlight the key ethical principles and best practices that should guide the design and implementation of these systems, with a focus on promoting safety, privacy and well-being for young users.

To achieve this goal, in Section 2, we begin by reviewing the related work on CAs and children, highlighting the most popular topics, such as educational and health applications. We also identify the opportunities and risks associated with the use of CAs for children. Next, we present ethical guidelines on the development of trustworthy AI, which provide a foundation for the ethical design and deployment of CAs with children. In Section 3, we present a case study that demonstrates the impact of CAs on children's learning and social development. Finally, in Section 4, we present an adaptation of the AI ethical guidelines to the specific case of CAs and children, highlighting the importance of data protection and human agency. We then demonstrate how these principles can be put into practice, offering a concrete example of how to design and deploy a CA that is both effective and ethical.

By the end of this chapter, readers will have a comprehensive understanding of the unique ethical challenges and opportunities presented by CAs for children, as well as a set of best practices for designing and implementing these systems in a responsible and ethical way.

2. Related work

2.1 Conversational agents

There are various terms used to describe conversational agents (CAs), such as dialog systems, virtual assistants and chatbots. These are all names for computer programs that allow people to interact with them through conversation [1] by using speech, text, or multimodal input/output. They have become popular in recent years and can be found on many devices such as smart speakers or cars. Traditional CAs typically consist of five main modules (**Figure 1**):

1. Automatic speech recognition (ASR): this module converts speech input into text, allowing the CA to understand and process what the user is saying.
2. Natural language understanding (NLU): this module takes the text generated by the ASR module and extracts the meaning behind it. The NLU module uses techniques such as semantic parsing to extract relevant information from the input.

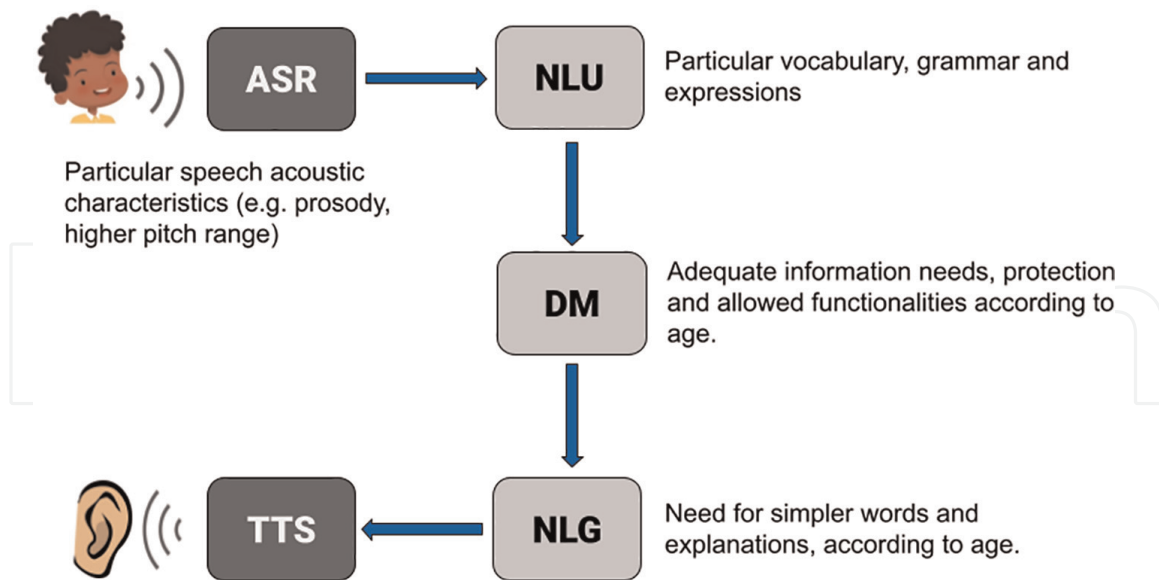


Figure 1.

Modules typically included in a conversational agent: ASR (automatic speech recognition), NLU (natural language understanding), DM (dialog management), NLG (natural language generation) and TTS (text-to-speech). The text close to each module summarizes the specific challenges that need to be tackled for their use with children.

3. Dialog management (DM): this module is responsible for managing the flow of the conversation and determining the appropriate response to the user's input based on the NLU module's understanding. It is usually based on identifying user intentions from this input to perform the needed actions and provide an answer to the user.
4. Natural language generation (NLG): this module takes the information obtained by the NLU and the answer type chosen by the DM module and generates a response in natural language for the user.
5. Text-to-speech (TTS): this module converts the text generated by the NLG module into speech output, allowing the CA to speak to the user in a natural way.

These modules work together to understand and respond to user input in a natural way. At present, many CAs are using advanced machine-learning methods to enhance their performance in one or more of these modules. For instance, BERT models are used to improve NLU [14], Reinforcement learning is used to improve DM [15] and ChatGPT utilizes deep neural networks to perform functions of NLU, DM and NLG in a unified manner.

CAs could be general, that is, without a specific purpose, such as the nowadays hugely popular ChatGPT [6] system. Alternatively, CAs could be intended to perform a quite specific mission, which is known as task-oriented CAs. This task-oriented nature makes CAs dependent on the corresponding use case they are developed for. For example, a CA that is devoted to home automation control expects interactions very different to a CA that is embedded in a social robot devoted to information tasks and guidance in commercial environments. The first system expects direct commands and it would ask for very few clarifications, whereas the other system will communicate in a more human-like form in order to establish a trusted linkage with human users.

Our focus in this chapter is on task-oriented CAs [16], which include tasks such as clothing selection [17], flight booking [18] and driving assistance [19]. CAs designed to assist users in completing specific tasks have become increasingly popular across various industries, such as e-commerce (e.g., purchasing), customer service (e.g., answering frequently asked questions) and healthcare (e.g., scheduling appointments). Some of these tasks are specifically designed for children, such as science learning [20], but CAs intended for adult-centred tasks are also utilized by children [21]. This may be due to the increasing accessibility and ease of use of these technologies for children.

2.2 Conversational agents and children

CAs are widely used devices, but it is important to consider their accessibility and popularity among children, as even young ones can interact with them through voice commands. In this section, we first present an overview of the literature in the field of conversational agents and children, and then an in-depth analysis of challenges, risks and opportunities identified in the related work.

2.2.1 Bibliometric study

We performed a bibliometric analysis to collect insights into how the research community has tackled the study of children and CAs. For that purpose, we used the *bibliometrix* tool [22] that allows extracting, among others, the most frequent keywords, clusters and co-occurrences of terms from a corpus of research papers. We collected the corpus of relevant papers from both Web of Science¹ and Scopus² as a result of the following search query over the papers' title, abstract and author keywords:

((“child” OR “children”) AND (“conversational agent” OR “conversational AI” OR “dialogue system” OR “dialogue systems” OR “chatbot” OR “chatbots” OR “virtual assistant” OR “home assistant” OR “voice assistant”))

We compiled a total of 440 papers, published from 2000 to 2022. Of the 440 papers, about 54% of them have been written in the period 2020–2022, with a clear increasing trend over the years, which is particularly remarkable since 2015 (from 7 papers per year in 2015 to 83 in 2022). This is probably a consequence of the recent popularity and market availability at a relatively low cost of conversational agents and assistants (e.g. Amazon's Alexa, Apple's Siri) and the emergence of large language models such as chatGPT [6].

Interestingly, when adding the terms “social robot” OR “robot interaction” to the search query, the number of papers increases from 440 to 2580, meaning that the research community has generally put a great effort into studying the effects of embodiment and the non-verbal side of interaction (e.g. through gestures, gaze or facial expressions).

For the verbal communication side, many studies on human-robot interaction with children actually rely on *Wizard-of-Oz* experimental settings [23, 24]. However, the full automation of voice interaction is already a reality and will be further enhanced by the aforementioned large language models in the near future. Nevertheless, careful attention has yet to be paid to their—still under-explored—maturity, moral capabilities [25] and trustworthiness when it comes to interacting with children.

¹ Web of Science: <https://www.webofscience.com/>

² Scopus: <https://www.scopus.com/>

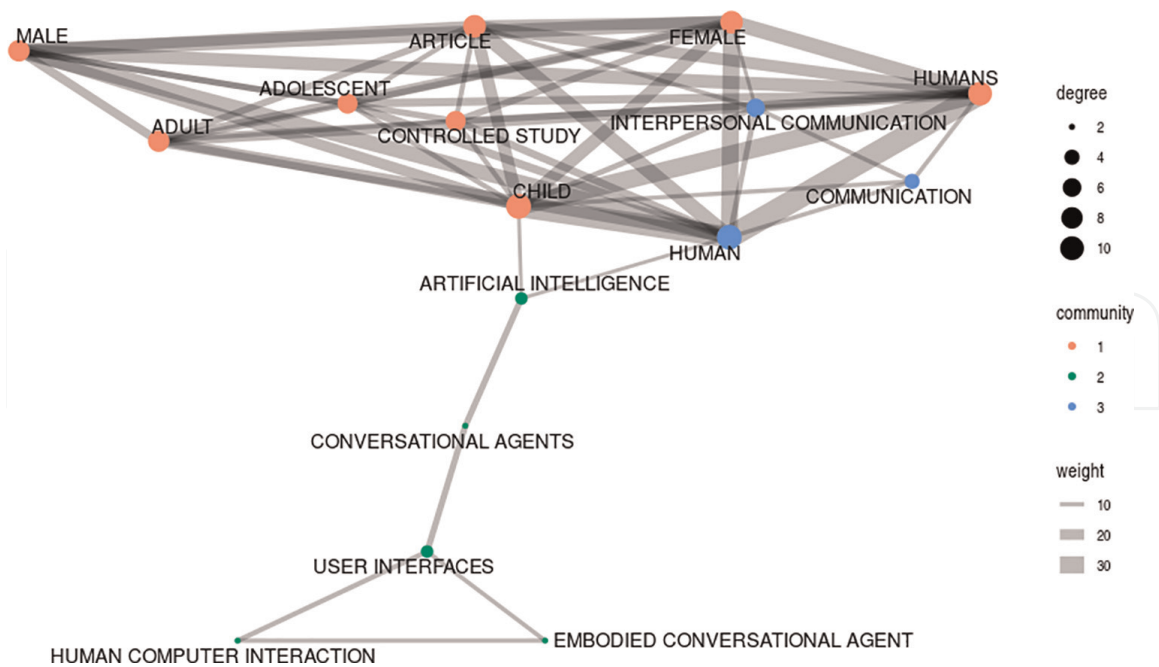


Figure 2. Results from our bibliometric analysis of 440 papers focusing on CAs and children: Most frequent terms, cluster and co-occurrence network. The thicker the link, the more weight the co-occurrence of words has. The size of the nodes indicates the frequency of the keyword (the larger the radius, the greater the use) and their color (red, blue or green) denotes which cluster they belong to.

Figure 2 shows the results of the keyword frequency, co-occurrence and clustering analysis carried out over the 440 papers related to children and CAs. The bibliometric algorithm identifies three main clusters of terms (nodes). The most populated cluster and the one with more co-occurrences of keywords, is the one represented with red nodes, having the term “controlled study” at the centre, strongly linked to “humans” and demographic aspects including “male”, “adult”, “adolescent”, “female” and “child”. This cluster is therefore likely to be related to a large amount of literature presenting in-lab (controlled) experiments, where “adults” frequently act as guardians. The second cluster, with blue nodes and very close to (almost contained in) the first cluster, presents fewer but highly influential keywords, including “communication” and “interpersonal communication”, which translates into studies analyzing the communicative—rather than or in addition to demographic—aspects of child-CA interaction in “controlled studies”. The third cluster, with green nodes, appears clearly separated from the other two, which is an interesting finding as it contains more technical words such as “artificial intelligence”, “user interfaces” and “(embodied) conversational agents”. This decoupling of studies focusing on behavioral analyses from the more technical ones might translate into a gap to be bridged in the research community: fostering multidisciplinary collaborations between social scientists and AI/human-machine-interaction researchers, designers and engineers. This is of particular importance in this new era where large language models will pave the way towards fully automated (and less “controlled”) human-CA interaction.

2.2.2 Risks and opportunities for children

Recent studies show that children actively use and explore CAs in home settings more than adults [21, 26, 27]. Thus, it is essential to consider the potential benefits and risks that these devices present for children when developing and implementing them.

In this context, we present and expand upon previous work [13, 28] regarding the risks, challenges and opportunities associated with these devices.

According to research, CAs bring several benefits to children, including:

- Engaging learning: by assisting in information search [29, 30], language learning [31] and school material learning [32, 33].
- Supporting health: by helping record treatments and track certain diseases [34] and chatting to reduce and control depression and anxiety [35].
- Enhancing accessibility: by enabling children with limited writing abilities (including toddlers), dyslexia or physical disabilities to interact with computers [36, 37].
- Improving social behavior: by promoting the use of persuasive strategies in games [38] and helping autistic children with social skills [39, 40].

However, as CAs are designed for a general population, their interactions with children may be affected by the unique characteristics and needs of children [41, 42]. Their language and communication abilities, as well as their particular rights, can pose challenges for the various components of the CA (**Figure 1**). The challenges of child-robot interaction have stimulated extensive research aimed at mitigating potential risks. Researchers have identified the following risks in the literature:

- Exclusion: children's different speech and understanding can lead to exclusion. Researchers have been working to enhance CA's performance for children, such as developing speech identification for babies [43] and identifying strategies for when the system does not understand a child [44].
- Over-trust: children may view CAs as friends [45, 46], which can lead to data disclosure risks. Transparent information, as demonstrated by Straten et al. [47], helps to reduce over-reliance on the system.
- Gender bias: chatbots are often designed with gender-specific cues and explicitly or implicitly conveyed as female [48].
- Unanticipated risks: the novelty of CAs makes it challenging to anticipate and prevent future problems, such as bullying experienced by girls named "Alexa" [49] due to Amazon's use of her name as an activation command for their device. Regular evaluation of these systems can help mitigate unanticipated risks.

Given those upcoming risks, guidelines are necessary to ensure the development of trustworthy CAs. These guidelines should take advantage of the benefits of these devices while assessing and minimizing potential risks and harm they may cause.

2.3 Trustworthy artificial intelligence

Several organizations worldwide have devoted efforts in the recent years to reflect on the ethical impact of artificial intelligence (AI) systems. The main goal of initiatives on AI and ethics is to raise awareness about the ethical considerations related to these

systems, deepen our understanding of them and minimize the potential risks of AI while maximizing their benefits.

For example, the high level expert group (HLEG) of the European Commission developed the ethical guidelines for trustworthy AI [50] having in mind the respect for fundamental rights in various contexts where AI systems are used. These guidelines put forward a set of seven key requirements that AI systems should meet in order to be deemed trustworthy, which are:

1. Human agency and oversight: AI systems should respect human autonomy and decision-making and should be overseen by humans.
2. Technical robustness and safety: AI systems should be accurate, reliable and safe and should adopt a preventative approach to risks.
3. Privacy and data governance: AI systems should protect privacy and have legitimate access to data.
4. Transparency: AI systems should have clear documentation and inform users about their decisions, capabilities and limitations.
5. Diversity, non-discrimination and fairness: AI systems should promote inclusion throughout their entire life cycle.
6. Societal and environmental well-being: AI systems should benefit society and the world.
7. Accountability: AI systems should have mechanisms in place to ensure responsibility for their development, deployment and use.

These ethical guidelines are complemented by an assessment list for trustworthy AI (ALTAI) [51], which is designed as a practical tool to help organizations self-assess the trustworthiness of their AI systems. ALTAI is a list of 69 self-evaluation questions grouped into the aforementioned seven requirements. Although these ethics guidelines are designed for general populations, it refers to children as a relevant vulnerable population, and it states the need to pay particular attention to them.

With a focus on children, UNICEF developed a policy guidance [52] that aims to raise awareness of children's rights in the context of AI systems. The guidance is based on nine requirements, including: (1) supporting children's development and well-being, (2) promoting inclusiveness for children, (3) prioritizing fairness and avoiding discrimination for children, (4) protecting children's data and privacy, (5) ensuring safety for children, (6) providing transparency, explainability and accountability for children, (7) empowering knowledge of AI and children's rights, (8) preparing children for present and future AI developments and (9) creating an enabling environment.

Previous research [13] has shown that most of the requirements in UNICEF's policy guidance for AI and children align with HLEG ALTAI (**Table 1**), except for requirement 8, which focuses on educational policies, of special relevance for children and broadly addressed by the HLEG in the context of jobs and skills. Despite this alignment, the two guidelines differ in their focus: while UNICEF's ones include policy considerations, HLEG places an emphasis on the development and evaluation of AI systems.

A recent report by the Joint Research Centre of the European Commission [28] recognized the need for a connection between existing research in the area of AI and children’s rights and the current policy initiatives and needs, proposing an integrated agenda for research and policy. In order to do that, the authors conducted a series of workshops with policymakers, researchers and children to gain insights into the interplay between different stakeholders, to connect scientific evidence with policymaking and to change the focus from the identification of ethical guidelines towards the definition of methods for practical future AI implementations. The report highlights the need for strategic and systemic choices to develop AI-based services that limit the use of AI to tasks that serve a valuable purpose. It emphasizes the importance of minimizing the environmental impact of AI technology, particularly in reducing CO₂ emissions from data centres. Developers must ensure that AI technology is child-friendly and free of discriminatory biases, while children must have control over their personal data. Transparency, explainability and accountability are critical to empowering young users of AI technology. The report also stresses the need for further research to understand how agency is developed in children when interacting with AI-based systems.

3. A scenario from the field of child-robot interaction

In this section, we present current scientific evidence regarding the impact of conversational agents on children. The objective is to emphasize the importance of

UNICEF AI for children	HLEG ALTAI						
	1	2	3	4	5	6	7
1. Supporting children’s development and well-being		xx	xx		xx	xxx	
2. Promoting inclusiveness for children					xxx		
3. Prioritizing fairness and avoiding discrimination for children					xxx		
4. Protecting children’s data and privacy	xx		xxx				
5. Ensuring safety for children	xx	xxx				xx	
6. Providing transparency, explainability and accountability for children	xxx	xx		xxx	xx		xxx
7. Empowering knowledge of AI and children’s rights				x		xxx	
8. Preparing children for present and future AI developments	x			x		x	
9. Creating an enabling environment	xx				xx	xx	

This table shows the correspondence between HLEG ALTAI’s seven requirements (1) Human agency and oversight, (2) Technical robustness and safety, (3) Privacy and data governance, (4) Transparency, (5) Diversity, non-discrimination, and fairness, (6) Societal and environmental well-being, and (7) Accountability, and the corresponding requirements in UNICEF’s AI for children guidelines (rows). The cells are marked with x, xx, or xxx to indicate the degree of correspondence between the related requirements, with low correspondence indicated by x, mid-level correspondence indicated by xx, and high correspondence indicated by xxx.

Table 1. Mapping between HLEG ALTAI and UNICEF AI for children requirements.

ethical guidelines, as outlined in Section 2.3, and to highlight the need for their adaptation in the context of CAs and children.

3.1 Motivation and rationale

As discussed in previous sections, conversational agents are present in various contexts and embodiments. For instance, a written interaction with an open-domain chatbot on a computer differs significantly from a voice interaction with a driving assistant in a car. We acknowledge that the embodiment and context of the conversational agent impact the user's perception and behavior, beyond the dialog. However, given the diverse range of conversational agents available and the findings of our literature review in Section 2.2.1, which showed the popularity of child-robot interaction studies, we have chosen to present a use case of a social robot in a controlled educational context.

To illustrate the potential impact of CAs on children, this use case reflects current research on social robots and builds upon previous research with educational focus [53–55]. This aligns with procedures suggested by UNICEF and provides a relevant scientific context of application of CAs.

Specifically, we discuss a large-scale experimental study with hybrid small-group settings, where we investigated the effects of certain robot behaviors on children's problem-solving processes, social dynamics and perceptions of the robot. This experimental study serves as a starting point for identifying emerging issues that could apply to other conversational agents. In the following subsections, we provide a brief overview of the methodology and results of the study. For a more detailed description, we direct the reader to ref. [56, 57].

3.2 Methodology

For this experimental study, 84 children (5–8 years old) were recruited and divided into pairs to collaborate with a conversational agent embodied in a table-top social robot in order to perform a Tower of Hanoi problem-solving task (see **Figure 3**). The study was structured as follows: In the preliminary session, children were introduced to the Hanoi tower game. In the robot intervention session, they collaborated with the robot to solve the task. The interview session involved a semi-structured interview and a picture task to capture children's perceptions the robot.

We manipulated two variables of the CA: its cognitive reliability and its expressivity, which are described as follows:

- In terms of cognitive reliability, the dialog management (DM) module allowed the CA to provide either optimal or non-optimal suggestions to solve the problem-solving task, depending on the selected behavior.
- In terms of expressive behavior, the natural language generation (NLG) module was configured to create two different behaviors: an expressive version of the system, which employed more emotive and engaging phrasings like “What do you think super-team? Do you feel like playing again?”; and a neutral version that used phrasings like “Would you like to repeat the game?”. In addition, these verbal expressions were coupled with different configurations of the text-to-speech (TTS) module, as outlined in **Table 2**, to control the expressivity of the robot's

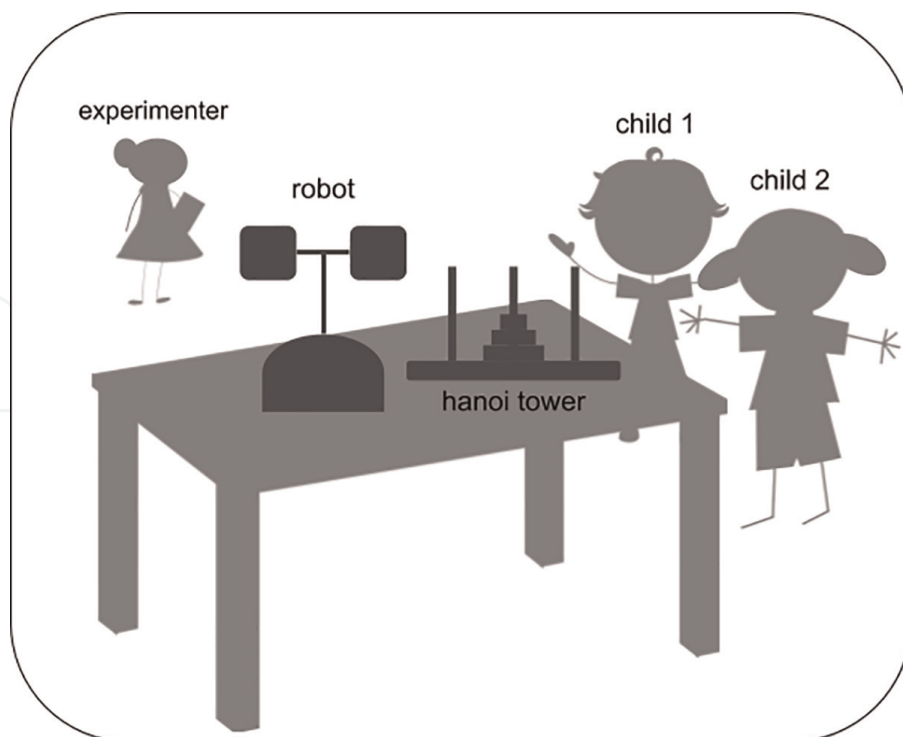


Figure 3. Experimental setup for studying collaborative problem-solving between two children and a table-top social robot using the tower of Hanoi logic task.

sentences. While in the neutral condition, the CA has used all the time the *neutral* configuration and the expressive condition changed between *original*, *calm* and *happy* configurations. We selected two different features, pitch and speed, and empirically set those parameters during the design phase of the study [56].

In the study, we measured the task performance and group dynamics during the interaction. We also interviewed children to understand both the influence of the different behaviors and children’s perception of the system.

3.3 Results

We summarize here the results of the study, with a focus on the link between the mentioned ethical guidelines and the communication features employed by the social robot during interactions.

The analysis of the recorded interviews showed that children generally find the social robot to be friendly, able to perceive them and not able to harm anyone (over 90% positive answers). The thematic analysis also revealed two important concepts:

Expressions	Mood	Speed	Pitch
Neutral	Serious	80%	-3
Expressive	Original	100%	0
	Calm	85%	0
	Happy	120%	+7

Table 2. TTS features in both expressive and neutral conditions.

perceived autonomy, meaning children's perceptions about the robot's ability to act on its own even when they know the robot has to follow its programming, and shared responsibility, which involves working together towards a common goal, fostering a sense of understanding and teamwork among the group members.

Additionally, children were affected by the different robot behaviors, including the ones related to the CAs described above.

- **Task performance.** A low performance by the robot in the tower of Hanoi task led to a decrease in the overall task performance and changed the dynamics of the group, as one child often took the lead. Children who perceived the robot's low performance recognized their own help to the robot, leading to a better-perceived help relationship between the child and the robot. This collaboration, combined with a probable lack of confidence in solving the task, resulted in children asking for the robot's assistance more quickly. On the other hand, children in the condition of a cognitive reliable robot who experienced high performance from the robot were more self-assured in solving the task and took longer to request help, even though they perceived more help from the robot.
- **Expressivity.** A high level of expressivity displayed by the robot during the interaction led to a stronger perception of the robot as a friend and less as a machine. Children also viewed the robot as more capable of assisting them with their homework, even when its task performance was low, emphasizing the social support it provided rather than its accuracy in the homework task. The researchers noted that children were able to distinguish the behavior of the expressive robot when compared with the non-expressive one; however, during the interaction, there was a tendency to view the robot as being expressive even when it was not.

Overall, this experiment provides valuable insights into children's perceptions of an embodied CA and the impact of the DM (task performance) and DM + NLG (expressivity) modules. The results indicate that children had a positive opinion about the CA, but a low task performance affects children's social dynamics and a high expressivity increases its perception as a friend and children's aim to do homework with it. These examples help us to illustrate the impact CAs have on children and help to understand the role of CAs in children's lives.

4. Towards trustworthy conversational agents for children

In previous sections, we have demonstrated the potential impact of CAs on children underscoring the need for developing trustworthy CAs that are suitable for them. Previous research on AI and children [28] has acknowledged the need to move from the identification of ethical guidelines to practical implementation. Building upon our experience in the study described in Section 3, in this section, we introduce an adaptation of ethical guidelines for CAs that consider children as potential users, along with their implementation in a practical system design.

4.1 Adaptation of ethical guidelines to CAs and children

A recent study aimed to adapt ethical guidelines for AI systems to the specific case of CAs and children [13]. A team of four experts in computer science, AI ethics and

children's rights evaluated each item of ALTAI (introduced in Section 2.3) in terms of relevance and particular considerations for CAs and children. A Delphi method approach [58] was followed to perform a risk level analysis [59] as follows. The experts rated each ALTAI item based on the likelihood of application and impact on children and CAs. Later on, the individual ratings were analysed to identify critical points and disagreements. Finally, the identified disagreements were discussed and resolved at an expert meeting in order to reach a consensus.

The quantitative analysis consisted of a risk level assessment, using the formula $Risk = Likelihood \times Impact$ to measure the level of risk of the different items and sections in ALTAI. A thematic analysis was also carried out on the annotated comments provided by the experts. The study's main findings are summarized below.

- **Stakeholders' involvement:** The experts emphasized the importance of involving children, teachers and parents as stakeholders in the design, use and testing of the system. Some experts emphasized the need for multiple stakeholders to be involved and for stakeholders to be taught to help oversee the system. All experts agreed that children's involvement should be done in a meaningful and entertaining way, as they should not be put in job-like conditions.
- **Risk management:** The importance of risk management for children, as a vulnerable population, was stressed by the experts during the CA development. They recommended implementing high measures for data storage to ensure privacy and security of personal data, avoiding access by third parties. To guarantee personal data privacy and security, the experts recommended implementing high measures for data storage and prevent access by third parties. They also suggested defining metrics and risk levels to monitor system performance and facilitate testing, evaluation and external audits. Allowing users to write reports about the system can also aid in identifying risks and errors. Finally, they emphasized the importance of transparency in addressing privacy concerns and minimizing children's data disclosure.
- **AI Awareness:** Experts underscored the importance of emphasizing the non-human aspect of CAs when designing them for children. This is important to minimize the attachment that children may form towards them and reduce their influence on the child. Maximizing the user's agency is also recommended to further reduce the impact. For example, providing multiple options when a recommendation is requested. To maintain transparency, constant access to the system's information, including its nature, functions and limitations, can be provided.
- **Age-appropriate behavior:** Inclusivity, as highlighted by the experts, is a vital aspect of children's education and development. Therefore, it is important to mitigate any technical difficulties that may arise when CAs interact with children or marginalized groups. In case of any breakdown, a reliable recovery strategy can help to continue the interaction. Additionally, guardians should take responsibility for children when they are using CAs. In the event of a problem, they should be contacted for assistance, and double consent mechanisms should be in place considering the guardian and the child. To enhance critical thinking

and self-regulation, transparency can be applied by using a language that is adapted to the user’s age.

- **Transparency.** According to the experts, transparency is an important factor that can help address several critical considerations mentioned before. The CA trustworthiness could be improved by providing information about the system’s nature, privacy and limitations in a language that is appropriate for the user’s age.

The findings of this study are largely consistent with previous work on ethical guidelines for AI [28, 51, 52] but provide a more in-depth perspective of the specific issues related to CAs and children. The risk level assessment revealed that although all the identified points are critical, some present higher risks than others (**Table 3**). Particularly, the concerns of children were rated as having a higher impact, while the likelihood of issues was higher for CAs, resulting in a higher overall risk for CAs. Based on these findings, “Privacy and data governance” and “Human agency and oversight” were identified as the two critical requirements that should be prioritized when developing CAs for children.

For more details on the methodology and results of this study, please refer to ref. [13]. The critical points identified in this study, which have been taken into account in subsequent work [60], were applied throughout an agent’s design process as illustrated in the subsequent section.

4.2 Proposal of a system

In this section, we introduce a previous study [60] that applied the presented guidelines (Section 4.1) to the development of a conversational agent intended to create a list of games and toys according to user preferences.

The design of the interaction is limited to one user at a time, and the system can ask questions to provide the user with a list of interesting items to choose from. The system can also ask for data such as hobbies, idols and cost limits to fill the user’s profile and determine the restrictions on the products that can be offered. The algorithm for the general CA is presented in Algorithm 1.

HLEG requirement	Risks to Children ^a	Risks to CAs ^a	Total Risk ^b
Human agency and oversight	5.21*	5.87*	30.59*
Technical robustness and safety	3.38	4.10	13.85
Privacy and data governance	4.99	6.96*	34.75*
Transparency	4.56	5.14*	23.43*
Diversity, non-discrimination and fairness	4.14	5.71*	23.63*
Societal and environmental well-being	2.59	3.00	7.78
Accountability	3.10	4.93	15.28

^aRisk = Likelihood × Impact.

^bTotal Risk = Risk on Children × Risk on CAs.

High risk values are marked with an* (5-9 for Risks to Children and Risks to CAs, and 18-81 for the Total Risk).

Table 3.

Risk assessment results based on expert evaluations of the likelihood and impact of questions in conversational agents and children.

The article proposed modifications consider children as possible users, including changes to the design, technology, interaction and post-interaction phases. These are summarized in **Table 4** and described in detail in the following paragraphs.

- Regarding the design phase, the “game recommender system” is redefined as a “games wish list” to improve children’s interest. A list of relevant stakeholders is defined to gather ideas on how to adapt the system to a children-friendly version. Stakeholders are involved in defining age ranges to be considered, as well as, in time limits for interaction, identification of positive game properties and evaluation of the system.

Algorithm 1: Pseudo-code for the proposed CA.

```
1 Welcome message
2 Ask if the user wants a specific item or a suggestion
3 if user input request to end then
4   | Start Goodbye
5   | Display list
6 else if user input request a specific item then
7   | Add item to the list
8 else if user input request a suggestion then
9   | Check if the profile information is enough for suggestions
10  | while information not enough do
11  |   | Ask information to the user to fill the profile
12  |   Provide suggestions
13  |   Go to line 3 (Ask to add any items to the list)
14 Ask information to the user to fill the profile
15 Go to line 3 (starts the interaction again)
```

- Regarding technology, the article suggests minimizing the use of standard search engines and black box approaches in the DM and NLG modules. A safety check module could improve the system’s reliability. The storage of personal data is minimized, and all data is stored in a secure server to prevent access by third parties. The ASR module is chosen with a good understanding rate for children and vulnerable populations to maximize the system’s inclusivity.
- Regarding the interaction phase, modifications are proposed, including guessing the user’s age range, using adapted vocabulary, splitting the welcome message into a guardian and a child welcome message and informing the user about the system’s non-human nature, confidentiality and algorithmic decisions. The system should incorporate a control mechanism when accessing online search engines, filter out non-adequate items for children, promote neutral games, or always suggest some toys associated with another gender, display at least three varied suggestions and have a good recovery strategy.
- Regarding the post-interaction phase, audits with access to the system’s metrics and reports will help identify critical and not critical problems.

General	Specific	Particular measures	Line of code
Stakeholders' involvement	Through the CA lifecycle	• Define features (e.g. age ranges and max interaction time).	-
		• Consult stakeholders throughout design, implementation and evaluation.	-
Risk management	Privacy measures	• Minimize the personal data to be stored.	11, 14
		• Do not allow additional usage/transfer of stored data.	-
	Security measures	• Reduce standard black boxes and search engine usage in DM and NLG.	-
		• Incorporate a control mechanism for online search.	12
		• Define trigger keywords for guardian involvement (e.g. weapons and sex).	2, 6, 11, 14
• Store data in a safe server with cybersecurity measures.	-		
• Define metrics for risk management, e.g. time spent with the user, times the system calls the guardian.	-		
Facilitate reports		• After the interaction, gather feedback from children and guardians.	After 5
		• Offer accessible error reporting and mention it in the welcome message.	1, after 5
AI awareness	Access to information	• Include concise relevant CA information in the welcome message and pointers to additional details.	1
		• Inform about the system's not-human and not-feeling nature.	1
		• Inform about CA's confidentiality and algorithmic decisions.	1
	Influence	Configure the system to display at least three suggestions.	12
Age approp. Behavior	Guardians	• Split welcome message: guardian and child. Need two consents.	1
		• Invoke guardians in security issues (e.g. dangerous requests or persistent breakdowns).	2, 11, 14
	Education and self-development	• Define toys-classification to benefit children's development. Consider them for suggestions.	12
		• Consider gender bias in recommended items.	12
		• Control and communicate the time spent on the interaction.	Before 14
Inclusivity		• Guess/ask for the user's age at the beginning of the interaction.	1
		• Define functionality as a "wish list" if a child is recognized.	1
		• Adapt the list of recommended items to age.	12
		• Adapt the vocabulary of the interaction to age.	1, 2, 4, 5, 11, 12, 14
		• Choose an inclusive ASR module.	3, 6, 8, 11, 14
• Minimize neutral responses in breakdowns.	3, 6, 8, 11, 14		

Table 4.

Recommendations to the design of a CA that generates a list of preferred toys/games for children.

Overall, the proposed modifications aim to improve the system's inclusivity and reliability, promote critical thinking and decrease overtrust and data disclosure.

5. Conclusions and challenges for future research

The field of conversational agents with children has made significant strides in recent years. Although extensive research has been conducted on the ethical development of artificial intelligence (AI), in general, there has been a growing emphasis on employing AI in interactions with children. Such work has emphasized the need for systems that enhance opportunities while mitigating risks.

However, the development and deployment of conversational agents for children come with specific challenges and ethical and social responsibility concerns. This chapter is dedicated to exploring the particular considerations conversational agents should have when interacting with children.

The chapter reviewed related research on conversational agents and children, identifying popular topics as well as opportunities and risks. Worldwide ethical guidelines on the development of trustworthy AI were presented as a framework for the ethical design and deployment of conversational agents with children. These guidelines emphasize the importance of the protection of user rights in the development and deployment of AI technologies. Additionally, a case study was presented that demonstrated the significant impact conversational agents can have on children's learning and social development.

An adaptation of previous AI ethical guidelines to the specific case of conversational agents and children was also presented, highlighting the importance of data protection and human agency. The application of ethical guidelines to the design of a conversational agent presented in this chapter served as an example of how these guidelines can be integrated into the development process of these systems.

It is important to note that the variety of conversational agent systems requires personalized study and application of these guidelines for each case. Furthermore, even state-of-the-art technology may not be able to address some proposed considerations, such as ASR modules that cannot understand little children's speech in less popular languages. Therefore, researchers in this area should continue to strive towards achieving new breakthroughs that enable the development of more ethically sound devices for the benefit of future generations.

In summary, the development and deployment of conversational agents with children require a careful balance between innovation and ethical responsibility. Ethical principles should guide research and development, and systems should be designed with the safety, privacy and well-being of children in mind. By doing so, conversational agents have the potential to be a powerful tool for enhancing children's learning and social development.

Acknowledgements

This work was carried out with the support of the Joint Research Centre of the European Commission in the framework of the Collaborative Doctoral Partnership Agreement No.35500.

Conflict of interest

The authors declare no conflict of interest.

IntechOpen

Author details


Marina Escobar-Planas^{1,2*}, Vicky Charisi², Isabelle Hupont²,
Carlos-D Martínez-Hinarejos¹ and Emilia Gómez²

1 Universitat Politècnica de València, Valencia, Spain

2 European Commission, Joint Research Center, Seville, Spain

*Address all correspondence to: marescplajob@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] McTear M. Conversational ai: Dialogue systems, conversational agents, and chatbots. *Synthesis Lectures on Human Language Technologies*. 2020; **13**(3):1-251
- [2] Lee K, M, Moon Y, Park I, Lee J-g. Voice orientation of conversational interfaces in vehicles. *Behaviour & Information Technology*. 2023:1-12. DOI: 10.1080/0144929X.2023.2166870
- [3] Garg R, Cui H, Seligson S, Zhang B, Porcheron M, Clark L, et al. The last decade of hci research on children and voice-based conversational agents. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Vol. 1–19. New York, NY, United States: Association for Computing Machinery; p. 2022
- [4] Singh S, Beniwal H. A survey on near-human conversational agents. *Journal of King Saud University-Computer and Information Sciences*. 2022;**34**(10): 8852-8866
- [5] Diederich S, Brendel AB, Morana S, Kolbe L. On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*. 2022;**23**(1): 96-138
- [6] van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. Chatgpt: Five priorities for research. *Nature*. 2023; **614**(7947):224-226
- [7] Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv*. 2023
- [8] Ngai EWT, Lee MCM, Luo M, Chan PSL, Liang T. An intelligent knowledge-based chatbot for customer service. *Electronic Commerce Research and Applications*. 2021;**50**:101098
- [9] Simpson J, Nalepka P, Kallen RW, Dras M, Reichle ED, Hosking SG, et al. Conversation dynamics in a multiplayer video game with knowledge asymmetry. *Frontiers in Psychology*. 2022;**13**: 1039431. DOI: 10.3389/fpsyg.2022.1039431
- [10] Ansar S, A, Jaiswal K, Aggarwal S, Shukla S, Yadav J, Soni N. Smart home personal assistants: Fueled by natural language processor and blockchain technology. In: *2022 Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS)*. IEEE; 2022. pp. 113-117
- [11] Sawad AB, Narayan B, Alnefaie A, Maqbool A, Mckie I, Smith J, et al. A systematic review on healthcare artificial intelligent conversational agents for chronic conditions. *Sensors*. 2022;**22**(7): 2625
- [12] Khosrawi-Rad B, Rinn H, Schlimbach R, Gebbing P, Yang X, Lattemann C, et al. Conversational agents in education—a systematic literature review. In: *Proceedings of the 30th European Conference on Information Systems (ECIS)*. Timis, Oara, Romania (Forthcoming). 2022. Available from: <https://aisel.aisnet.org/ecis2022/>. ISBN: 978-1-958200-02-5
- [13] Escobar-Planas M, Gómez E, Martínez-Hinarejos C-D. Guidelines to develop trustworthy conversational agents for children. 2022:342-360. Available from: <https://sites.utu.fi/ethicomp2022/wp-content/uploads/sites/1104/2022/09/Ethicomp-2022-Proceed>

ings_Corrected.pdf. ISBN: 978-951-29-8989-8

[14] Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv. 2018. DOI: 10.48550/arXiv.1810.04805

[15] Lipton Z, Li X, Gao J, Li L, Ahmed F, Deng L. Bbq-networks: Efficient exploration in deep reinforcement learning for task-oriented dialogue systems. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. Palo Alto, California USA: AAAI Press; 2018

[16] Gao J, Galley M, Li L. Neural approaches to conversational ai. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York, NY, United States: Association for Computing Machinery; 2018. pp. 1371-1374

[17] Sousa RG, Ferreira PM, Costa PM, Azevedo P, Costeira JP, Santiago C, et al. Ifetch: Multimodal conversational agents for the online fashion marketplace. In: Proceedings of the 2nd ACM Multimedia Workshop on Multimodal Conversational AI. New York, NY, United States: Association for Computing Machinery; 2021. pp. 25-26

[18] Wei W, Le Q, Dai A, Li J. AirDialogue: An Environment for Goal-Oriented Dialogue Research. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics; 2018. pp. 3844-3854. Available from: <https://aclanthology.org/D18-1>

[19] Eric M, Krishnan L, Charette F, Manning CD. Key-value retrieval

networks for task-oriented dialogue. In: Proceedings of the 18th annual SIGdial meeting on discourse and dialogue. Saarbrücken, Germany: Association for computational linguistics; 2017. pp. 37-49. DOI: 10.18653/v1/W17-5506

[20] Xu Y, Vigil V, Bustamante AS, Warschauer M. “elinor’s talking to me!”: Integrating conversational ai into children’s narrative science programming. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. Vol. 1–16. New York, NY, United States: Association for Computing Machinery; p. 2022

[21] Garg R, Sengupta S. 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. 2020;4(1):1-24

[22] Aria M, Cuccurullo C. Bibliometrix: An r-tool for comprehensive science mapping analysis. Journal of Informetrics. 2017;11(4):959-975

[23] Nasir J, Oppliger P, Bruno B, Dillenbourg P. Questioning wizard of oz: Effects of revealing the wizard behind the robot. In: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE; 2022. pp. 1385-1392

[24] Zou J, Gauthier S, Salvatore M, Anzalone, David Cohen, and Dominique Archambault. A wizard of oz interface with qtrobot for facilitating the handwriting learning in children with dysgraphia and its usability evaluation. In: Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proceedings, Part II. Springer; 2022. pp. 219-225. Available from:

<https://link.springer.com/book/10.1007/978-3-031-08645-8>

[25] Ganguli D, Askill A, Schiefer N, Liao T, Lukošiūtė K, Chen A, et al. The capacity for moral self-correction in large language models. arXiv. 2023. Available from: <https://arxiv.org/abs/2302.07459>

[26] Sciuto A, Saini A, Forlizzi J, Hong JI. “Hey alexa, what’s up?” a mixed-methods studies of in-home conversational agent usage. In: Proceedings of the 2018 Designing Interactive Systems Conference. New York, NY, United States: Association for Computing Machinery; 2018. pp. 857-868. DOI: 10.1145/3196709

[27] Lovato SB, Piper AM, Wartella EA. Hey google, do unicorns exist? Conversational agents as a path to answers to children’s questions. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children. New York, NY, United States: Association for Computing Machinery; 2019. pp. 301-313. DOI: 10.1145/3311927

[28] Charisi V, Chaudron S, Di Gioia R, Vuorikari R, Planas ME, Sanchez MJI, et al. Artificial Intelligence and the Rights of the Child: Towards an Integrated Agenda for Research and Policy. Technical report, Joint Research Centre of the European Commission. (JRC127564, Joint Research Centre (Seville site)); 2022

[29] Landoni M, Murgia E, Huibers T, Pera MS. You’ve got a friend in me: Children and search agents. In: Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization. New York, NY, United States: Association for Computing Machinery; 2020. pp. 89-94. DOI: 10.1145/3386392

[30] Downs B, French T, Wright KL, Pera MS, Kennington C, Fails JA. Children and search tools: Evaluation remains unclear. In: KidRec Workshop co-Located with ACM IDC 2019. 2019

[31] Gilani SN, Traum D, Merla A, Hee E, Walker Z, Manini B, et al. Multimodal dialogue management for multiparty interaction with infants. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction. New York, NY, United States: Association for Computing Machinery; 2018. pp. 5-13. DOI: 10.1145/3242969

[32] Law E, Ravari PB, Chhibber N, Kulic D, Lin S, Pantasdo KD, et al. Curiosity notebook: A platform for learning by teaching conversational agents. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY, United States; Association for Computing Machinery; 2020. pp. 1-9. DOI: 10.1145/3334480

[33] Xu Y, Warschauer M. “Elinor is talking to me on the screen!” integrating conversational agents into children’s television programming. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY, United States; Association for Computing Machinery; 2020. pp. 1-9. DOI: 10.1145/3334480

[34] Sezgin E, Noritz G, Elek A, Conkol K, Rust S, Bailey M, et al. Capturing at-home health and care information for children with medical complexity using voice interactive technologies: Multi-stakeholder viewpoint. *Journal of Medical Internet Research*. 2020;22(2):e14202

[35] Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to

young adults with symptoms of depression and anxiety using a fully automated conversational agent (woeBot): A randomized controlled trial. *JMIR mental health*. 2017;**4**(2):e7785

[36] Fabio Catania, Pietro Crovari, Eleonora Beccaluva, Giorgio De Luca, Erica Colombo, Nicola Bombaci, and Franca Garzotto. Boris: a spoken conversational agent for music production for people with motor disabilities. In *CHIItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter*, pages 1–5, 2021.

[37] Pradhan A, Mehta K, Findlater L. “Accessibility came by accident” use of voice-controlled intelligent personal assistants by people with disabilities. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, United States: Association for Computing Machinery; pp. 1, 2018-13

[38] Fraser J, Papaioannou I, Lemon O. Spoken conversational ai in video games: Emotional dialogue management increases user engagement. In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. New York, NY, United States; Association for Computing Machinery; 2018. pp. 179-184. DOI: 10.1145/3267851

[39] Ali MR, Razavi SZ, Langevin R, Al Mamun A, Kane B, Rawassizadeh R, et al. A virtual conversational agent for teens with autism spectrum disorder: Experimental results and design lessons. In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. New York, NY, United States: Association for Computing Machinery; 2020. pp. 1-8. DOI: 10.1145/3383652

[40] Zhang L, Weitlauf AS, Amat AZ, Swanson A, Warren ZE, Sarkar N.

Assessing social communication and collaboration in autism spectrum disorder using intelligent collaborative virtual environments. *Journal of Autism and Developmental Disorders*. 2020;**50**: 199-211

[41] Narayanan S, Potamianos A. Creating conversational interfaces for children. *IEEE Transactions on Speech and Audio Processing*. 2002;**10**(2):65-78

[42] Kennedy J, Lemaignan S, Montassier C, Lavalade P, Irfan B, Papadopoulos F, et al. Child speech recognition in human-robot interaction: Evaluations and recommendations. In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, United States: Association for Computing Machinery; 2017. pp. 82-90. DOI: 10.1145/2909824.3020229

[43] Lavechin M, Bousbib R, Bredin H, Dupoux E, Cristia A. An open-source voice type classifier for child-centered daylong recordings. *arXiv*. 2020

[44] Røyneland K. “It Knows How to Not Understand us!” a Study on What the Concept of Robustness Entails in Design of Conversational Agents for Preschool Children [Thesis]. Oslo, Norway: University of Oslo; 2019

[45] Kahn Jr PH, Kanda T, Ishiguro H, Freier NG, Severson RL, Gill BT, et al. “Robovie, you’ll have to go into the closet now”: Children’s social and moral relationships with a humanoid robot. *Developmental Psychology*. 2012;**48**(2): 303

[46] Druga S, Williams R, Breazeal C, Resnick M. 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. New York, NY, United States: Association for Computing Machinery; 2017. pp. 595-600. DOI: 10.1145/3078072
- [47] van Straten CL, Peter J, Kuhne R, Barco A. Transparency about a robot's lack of human psychological capacities: Effects on child-robot perception and relationship formation. *ACM Transactions on Human-Robot Interaction (THRI)*. 2020;9(2):1-22
- [48] Tolmeijer S, Zierau N, Janson A, Wahdatehagh JS, Leimeister JMM, Bernstein A. Female by default?—exploring the effect of voice assistant gender and pitch on trait and trust attribution. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, United States: Association for Computing Machinery; pp. 1, 2021-7. DOI: 10.1145/3411763
- [49] Johns T. Parents of Children Called Alexa Challenge Amazon. BBC. Available from: <https://www.bbc.com/news/technology-57680173>
- [50] HLEG AI. Ethics Guidelines for Trustworthy Ai. Brussels; 2019. Available from: <https://digital-strategy.ec.europa.eu/en/library/ethics-guideline-s-trustworthy-ai>
- [51] Ala-Pietilä P, Bonnet Y, Bergmann U, Bielikova M, Bonfeld-Dahl C, Bauer W, et al. The Assessment List for Trustworthy Artificial Intelligence (ALTAI). European Commission; 2020. p. 34. Available from: [https://books.google.es/books?id=cu8dEAAAQBAJ&dq=The+Assessment+List+for+Trustworthy+Artificial+Intelligence+\(ALTAI\).&lr=&hl=es&source=gbs_navlinks_s](https://books.google.es/books?id=cu8dEAAAQBAJ&dq=The+Assessment+List+for+Trustworthy+Artificial+Intelligence+(ALTAI).&lr=&hl=es&source=gbs_navlinks_s)
- [52] Dignum V, Penagos M, Pigmans K, Vosloo S. Policy Guidance on AI for Children. Communications of UNICEF; 2021. Available from: <https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children>
- [53] Charisi V, Gomez E, Mier G, Merino L, Gomez R. Child-robot collaborative problem-solving and the importance of child's voluntary interaction: A developmental perspective. *Frontiers in Robotics and AI*. 2020;7:15
- [54] Charisi V, Davison D, Wijnen F, Van Der Meij J, Reidsma D, Prescott T, et al. Towards a child-robot symbiotic co-development: A theoretical approach. In: *AISB Convention 2015. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB)*; 2015. Available from: <https://dspace.library.uu.nl/handle/1874/324608>
- [55] Charisi V, Dennis L, Fisher M, Lieck R, Matthias A, Slavkovik M, et al. Towards moral autonomous systems. *arXiv*. 2017. Available from: <https://arxiv.org/abs/1703.04741>
- [56] Charisi V, Merino L, Escobar M, Caballero F, Gomez R, Gómez, E. The effects of robot cognitive reliability and social positioning on child-robot team dynamics. In: *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE; 2021. pp. 9439-9445. Available from: https://ieeexplore.ieee.org/abstract/document/9560760?casa_token=hFdtPyqOh58AAAAA:ZLONvEfeRPancjL0fxjWGoMl97LqW8kafeUE8QDKgoaDPyzJz4XYGuDfhcvy-sJMWTgZJfq8sP2G-s
- [57] Escobar-Planas M, Charisi V, Gomez E. "That robot played with us!" children's perceptions of a robot after a child-robot group interaction. *Proceedings of the ACM on Human-*

Computer Interaction. 2022;**6**(CSCW2):
1-23

[58] Linstone HA, Turoff M, et al. The Delphi Method. Reading, MA: Addison-Wesley; 1975

[59] Kovačević N. Aleksandra Stojiljković, and Mitar Kovač application of the matrix approach in risk assessment. Operational Research in Engineering Sciences: Theory and Applications. 2019;**2**(3):55-64

[60] Marina Escobar-Planas, Emilia Gómez, Carlos-D Martínez-Hinarejos. Enhancing the Design of a Conversational Agent for an Ethical Interaction with Children. Proc. IberSPEECH 2022. 2022:171–175.
DOI: 10.21437/IberSPEECH.2022-35