IberSPEECH 2022
14-16 November 2022, Granada, Spain

Enhancing the Design of a Conversational Agent for an Ethical Interaction with Children

Marina Escobar-Planas[1,2], *Emilia Gómez*[2], *Carlos-D. Martínez-Hinarejos*[1]

[1] Pattern Recognition and Human Language Technologies Research Center, Universitat Politècnica de València, Camí de Vera, s/n, 46022, València, Spain
[2] European Commission, Joint Research Center, Seville, Spain

{marina.escobar-planas,emilia.gomez-gutierrez}@ec.europa.eu, cmartine@dsic.upv.es

Abstract

Conversational agents (CAs) have become one of the most popular applications of speech and language technologies in the last decade. Those agents employ speech interaction to perform several tasks, from information retrieval to purchase goods from on-line stores. However, these agents are defined to address a general sector of population, mainly adults without speech production problems, and then they fail to obtain a similar performance with specific groups, such as elderly or children. The case of children is particularly interesting because they naturally engage in interaction with those CAs and they have special needs in terms of technical and ethical considerations. Therefore, CAs must fulfil some conditions that could affect their general design in order to provide a trustworthy interaction with children. In this article we present how to improve a general CA design to fulfil the specific ethical needs of children interaction. We address the development of a CA devoted to complete a wish list of games using user preferences, and its improvements towards children.

Index Terms: conversational agents, children interaction, artificial intelligence ethics.

1. Introduction

A conversational agent (CA) or dialogue system is defined as an intelligent agent which employs speech or written interaction in natural language with human users in order to achieve a certain objective [1]. These agents have become extensively popular in the last decade, specially in the form of voice assistants such as Siri[1], Google Assistant[2], or Alexa[3]. However, many other systems that follow this paradigm have been developed in different research projects and commercial applications [2, 3].

All CAs have roughly the same structure, based on the following five modules: (1) Automatic speech recognition (ASR): it transforms speech input into a text transcription to be used in the next modules; (2) Natural language understanding (NLU); this module provides a semantic interpretation of the transcription, extracting semantic tokens for the next steps; (3) Dialogue manager (DM), or the core of the CA, it takes the semantic representation and the previous interactions with the user (dialogue history) to perform an action, i.e. an answer or other activities such as managing devices (screens, mechanical actors, . . .); (4) Natural language generation (NLG): it translates the DM's responses to natural language text; and (5) Text to speech (TTS), which generates a speech from the text provided by the NLG.

[1]https://www.apple.com/es/siri/
[2]https://assistant.google.com/
[3]https://developer.amazon.com/alexa

ASR and TTS are not present in text-based CAs. The design of all these modules, and specially the DM, is usually conducted according to the specific task that the CA must achieve. This task defines, among others, the vocabulary, the relevant semantic terms, and the possible system actions. In terms of users, CAs usually aim at broad populations, and consequently they are designed to fit to the most general features of the population; for example, the ASR module would try to cover the most extended languages, accents and pitches, or the DM would try to cover most of the actions expected by the target population.

This may cause the CA to fail when some groups of population use it, not only because of technical reasons (difficulty in speech recognition, adapted NLG, . . .) but also for not integrating ethical considerations that should affect the DM performance (access to inappropriate content, actions that require legal qualifications, . . .). This could be the case of children. Nowadays, children are common users of CAs such as the voice assistants named above [4], and consequently they are a population sector whose special features must be taken into account when using a CA. However, current CAs do not support specifically the needs for this sector of the population [5, 6].

From a technical point of view, a CA facing interactions with children has to deal with different challenges, e.g. the need for non-written communication of very young children, the improvement of ASR modules to understand the different characteristics of children's speech (pitch, prosody and timbre) [7, 8], or the adaptation of NLGs to allow children's comprehension (avoiding complex sentences and unusual words).

From an ethical point of view, CAs also face different challenges when dealing with children. Apart from children's right to inclusivity, which is related to the technical challenges mentioned above, since children are a vulnerable population, other aspects must be taken into account, such as the need to consider a guardian that must play a role in certain decisions, the explanation needed to face the different understanding of the world (for a child, a car can be an object and also have feelings and be their friend) or the high relevance of personal data protection.

In this article, we present the design of a general CA for creating a wish list of presents, more specifically games, from user petitions and preferences, along with its adaptation to children by applying a set of ethical guidelines. Section 2 presents the related scientific work, and then Section 3 defines the general system. After that, Section 4 summarizes the main ethical guidelines considered in this work, and Section 5 shows how we used the ethical guidelines to adapt the general system to a child-friendly one. Section 6 provides our conclusions and future work lines.

10.21437/IberSPEECH.2022-35

2. Related work

Currently, state of the art approaches addressing the different components of a CA are mostly based on Deep Learning techniques. This includes ASR[9], NLU (e.g.BERT [10]), DM (Reinforcement learning [11, 12]), NLG (e.g.GPT [13]) and TTS (neural systems [14]).

In this work, we are particularly interested in task-oriented CAs [15], which cover tasks such as clothes selection [16], flight booking [17], or driving assistance [18]. Some of these tasks are designed for children (e.g. science learning [19]), but CA oriented to adult-centered tasks are also used by children [20].

In terms of ethical considerations, in recent years, there has been an increased attention on the impact of Artificial Intelligence (AI) systems in people's lives. AI's respect for fundamental rights has been a mainstay of reports from international institutions. For instance, the European Commission (EC)'s High Level Expert Group on AI released an assessment list to self-evaluate if an AI is "trustworthy" [21]. In addition, UNICEF and the EC Joint Research Centre released reports providing recommendations on how to design AI systems so that they respect children's rights [22, 23]. This previous work was discussed by a group of experts in the specific case of CAs and children, and their considerations were taken into account to generate a guide for the development of trustworthy CAs for children [24]. Further information about this guidelines can be found in Section 4.

3. Algorithm general design

As a general design, we propose a CA that assists in generating a list of presents for the user (Algorithm 1). The interaction is limited to one user at a time, and the type of presents would be limited to games and toys, in order to avoid a strong task adaptation when considering children users.

The CA is based on the idea of obtaining a list of specific items. Those items can be directly obtained from the user when they give a specific proposal at any point of the interaction. However, the user might not have a clear idea on what item to ask for, hence, the CA would ask some questions in order to provide the user a list of interesting items, from which the user could choose a specific one. Once the specific item is chosen, its inclusion in the final list is confirmed and the process starts again unless the user wants to stop.

The system may ask the user for several data such as hobbies (sports, music, fashion, literature, etc.), idols or fan items (real or fictional characters, sagas from literature, cinema or TV, favourite styles of music or literature, etc.), cost limits, or other factors (such as product origin, trademark, delivery dates, etc.). This would allow the system to have a set of restrictions on the products that can be offered to the user. When getting information to determine the restrictions of the possible items to be offered, the process also determines if a minimal set of information slots is covered in order to make a proposal.

Based on these features, the algorithm for the general CA could be described as presented in Algorithm 1. The different variables are described as follows:

- *end* shows if the interaction needs to end.
- *mininf* is the minimum required data from the user in order to do a proposal. It would be configured according to the number of products the system can provide.
- *request* shows if the user requested a proposal from the system.
- *user input* stores the different aims from the user:
  `spec`. If the user proposes a specific item.
  `prop`. If the user requests a proposal.
  `conf`. If the user confirms to add the item to the list.
  `cont`. If the user wants to continue with the interaction.
  `info`. If the user provides requested info for the profile.

Algorithm 1: Pseudo-code for the proposed general CA. The comments state the considered user inputs.

```
User input: spec, prop, conf, cont or info
Welcome message, ask what to add to the list, end = False
Get user input                                  // spec, prop
while not end do
    while user input is not spec do
        while user input is not (spec or prop) do
            Ask user for information
            Get user input                      // info, spec
            if not spec then
                Get information and add to the profile
                if mininf then
                    Ask user if accepts proposal
                    Get user input              // prop
                    if user input is prop then
                        request = True
        if request then
            Show proposals
            Get user input                      // spec, prop
            if not spec then
                request = False
            else save item
    Ask for item confirmation
    Get user input                              // conf
    if conf then
        Add item to list
    Ask for continuation
    Get user input                              // cont
    if user input is not cont then
        end = True
Output: List of presents
```

System actions can directly affect the user's perception of the system in terms of trust and efficiency, even for children [25]. For instance, if the system considers hobbies as enough to determine the list of items, the user can perceive that the system is not interested to know more about their preferences and, as a consequence of this perception, they can reject the system proposal. In the same line, if the system presents a list of selected items that are of a single vendor, the user can suspect of bias to such vendor and abandon the system for future uses. This can be solved if proper explanations are given when showing the list [26].

For general users, we can consider that restrictions applied on these system actions are not so critical, as adults are supposed to have enough maturity to deal with the designed system behaviour. However, this is not the case for children, and we must ensure that ethical considerations are applied. Children can, for instance, overtrust the CA and disclose sensitive data or make choices proposed by the system that may not be suitable for them. CA interactions with children must guarantee that these situations do not happen, e.g. asking for a guardian supervision, or using interactions that discourage the sharing of sensitive data.

4. Ethical design for children

As mentioned in Section 2, previous work have adapted general ethical guidelines for AI systems to the specific case of CAs and children. With a keen eye on minimizing risks, a group of experts (in the field of computer science, AI ethics, and children's rights) evaluated how the ethical guidelines and assessment list for trustworthy AI (ALTAI) should be applied to CAs and children, respecting the seven requirements for trustworthy AI, which embraces a set of 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. On the one hand, the quantitative analysis of these discussions provided an order for addressing ALTAI requirements and items for this specific case. On the other hand, the qualitative analysis highlighted some critical points to consider when developing these devices for children. The recommendations of this study can be summarized as follows:

Stakeholders involvement. It is important to involve stakeholders (e.g. children, guardians, teachers) during the whole design process. It should be done in a meaningful way giving that, for instance, children cannot be consider a work force when developing a commercial product.

Risk management. Considering children as a vulnerable population, risk management should be highly considered during the CA development. High privacy and security measures regarding data storage is needed to ensure that personal data is not be accessible to third parties. In addition, metrics and risk levels should be defined to track the system performance, facilitating its testing and evaluation as well as external audits. In addition, users capability to write reports about the system can facilitate the identification of risks and errors. Transparency can also be used to inform about privacy concerns and diminish children's data disclosure.

AI Awareness. Due to the different perception of the world children may have, it is important to highlight the non-human nature of the CA in order to minimize children's attachment to it and CAs influence on the child. Regarding influence, maximising the user's agency will also be beneficial (e.g. providing more than one answer/option when answering/suggesting). Transparency can be used to provide constant access to the system's information including nature, functions and limitations.

Age appropriate behaviour. Improving inclusivity is very important for children education and development. It is therefore important to mitigate the technical difficulties that CAs may have when interacting with children and other minorities. A good recovery strategy, to continue interaction after a breakdown, may help. We should also address guardians as responsible for the children, implementing mechanisms for double consent, but also approaching them when a problem is encountered. Transparency can be applied using a language adapted to the user's age, in order to improve the user's critical thinking and self-regulation.

5. Applying recommendations

Following the guidelines presented in Section 4, we propose some modifications to the general algorithm design of Section 3, in order to consider children as possible users. The mapping from the guidelines to our modifications and lines of code to be applied, are summarized in Table 1 and further described as follows:

Design. Regarding functionality, if a child is detected we re-define the "game recommender system" to a "games wish list" in order to improve children's interest. We define a list of relevant stakeholders (little children, teenagers, parents, teachers, psychologist, CA developers and toys sellers) and suggest to gather ideas from them on the adaptation from an adult-centred to a children-friendly version. During the system design, stakeholders are a valuable resource to define different factors: age ranges to be considered (as the system needs to adapt its behaviour), time limit for the interaction (as children may need to learn self-regulation) or identification of positive game properties (as the system may promote culture, socialization or physical activities). In addition, they should also be involved in the evaluation and testing of the system.

Technology. Regarding the most suitable technology to be used by our child-friendly CA system, we would minimize the use of standard search engines and black box approaches in the DM and NLG modules —unless they are designed for children— in order to prevent uncontrolled outputs. Otherwise, a module to check the safety of the system's output could improve its reliability. In addition, the storage of personal data will be minimized (to age, gender and preferences) and all data would be stored in a secure server avoiding its access to third parties. Last but not least, we will choose an ASR module with a good understanding rate for children and vulnerable populations in order to maximise the system's inclusivity (e.g. the one developed by Kelly et al. [27]).

Interaction. The structure of the interaction remains the same (Algorithm 1): the user can add items to the list at any point, and in the meantime, the system will ask for information to suggest games. However, some modifications are proposed:

- *Starting*. At the beginning of the interaction, we would try to guess the user's age range (automatically or by asking). This will allow the system to adapt the interaction when a minor is recognized.
- *Adapted vocabulary.* Ideally, the system would speak in a different way to different age ranges. Vocabulary for different age ranges and ranges can be defined together with the stakeholders.
- *Welcome message.* When interacting with a child, the system should split the welcome message in two: guardian welcome message (asking to talk with the guardian, informing about the system, asking for consent and offering the possibility to adjust certain parameters such as children's age, price limit, delivery date limit, or interaction time limit) and a child welcome message (asking to talk with the minor, informing about the system, asking for consent and starting the interaction).
- *Informing about the system.* The system should inform about its non-human nature, with a keen eye on its non-living and non-feeling nature. In addition, it would be important to communicate about confidentiality (e.g.

Table 1: *Program adaptations to include children as possible users.*

General recommend.	Specific recommendations	Measures to apply in our particular system	Line of code
Stakeholders involvement	Consider stakeholders in all the CA lifecycle	–Define features (e.g. age ranges, max interaction time). -Consult stakeholders' opinion during design, implementation and evaluation.	- -
Risk management	Privacy measures	-Minimize the personal data to be stored. -Do not allow additional usages/transfer of stored data.	2, 7, 9, 12, 17, 21, 25 -
	Security measures	-Minimize the use of standard black boxes and search engines in the DM and NLG modules. - Incorporate a control mechanism for online search. -Define keywords that would trigger the involvement of a guardian (e.g. weapons, sex). -Store data in a safe server with cybersecurity measures. -Define metrics for risk management, e.g. time spent, times the system calls the guardian.	- 16 7 - -
	Facilitate reports	-At the end of the interaction, gather feedback from children and guardians -Provide access at any point to open a report. Inform about it in the welcome message.	After 26 1, new user input aim
AI awareness	Access to the system information	-Add to the welcome message concise relevant information about the CA and pointers to additional details. -Inform about the system's not-human, not-living and not-feeling nature. -Inform about the system's confidentiality and algorithmic decisions.	1 1 1
	Influence	-Configure the system to display at least 3 suggestions	16
Age approp. behavior	Guardians	-Split the welcome message into: guardian & child. Consider two consents. -Call the guardian when difficult security decisions arise (e.g. request of toys with dangerous keywords or assisting when an interaction breakdown persists).	1 2, 7, 9, 12, 17, 21, 25
	Education and self-development	-Define toys-classification to benefit children's development. Consider them for suggestions. -Consider gender bias in recommended items. -Control and communicate the time spent on the interaction.	16 16 After 4
	Inclusivity	-Guess/ask for age information at the beginning of the interaction. -Define functionality as "wish list" if a child is recognized. -Adapt the list of recommended items to age. -Adapt the vocabulary of the interaction to age. -Choose an inclusive ASR module. -Minimize neutral responses in breakdowns.	1 1 16 1, 6, 11, 16, 20, 24 2, 7, 9, 12, 17, 21, 25 2, 7, 9, 12, 17, 21, 25

how data is stored) and explain algorithmic decisions (e.g. how list items are proposed) in order to increase the user's agency, promote critical thinking and decrease overtrust and data disclosure. A key point summary during the welcome message, and the capability to later access further information at any point, can improve the interaction flow.

- *Game recommendations.* When interacting with a child, the system should incorporate a control mechanism when accessing online search engines, it could filter out recommendations of non adequate items for children. Furthermore, as for "education and personal development", we propose to work with the stakeholders in order to define different toy classification that may be beneficial for children (e.g. culture, socialization or physical activity). In addition, the system could fight gender stereotypes of toys [28], by promoting neutral games, or always suggesting some toys socially associated to another gender. Finally, the system should display at least three suggestions (varied in terms of gender and beneficial activities) to minimise the system's influence during the interaction.
- *Errors and risks.* Because of the inclusivity of our design, frequent breakdowns should be considered. Therefore, a good recovery strategy is fundamental. Previous work suggest to improve the device collaboration with the user by decreasing neutral responses (without information) and increasing specific responses (with information) [29]. In fact, with persisting breakdowns, the system can request to collaborate with the guardian. In addition, guardians can be consulted when the CA detects a possible risk.
- *Metrics.* Recording certain variables can help develop metrics for risk management. Stakeholders can help with this considerations, but we could consider the following examples: time spent with the user (for self-regulations purpose), times the system calls the guardian (to identify persistent problems and offer to fill an interaction report), how many recommendations are accepted and position of the selected recommendation (to track the system's influence), risk words such as weapons, drugs or sex (to determine the need of guardian involvement).
- *Finalising.* Asking the user (or guardian and child) to fill a quick interaction report at the end of their experience will help identify critical and not critical problems.

Post-Interaction. Audits with access to the system's metrics and reports will be beneficial to improve the CA design.

6. Conclusions and future work

In this paper, we have proposed a CA general design to generate a list of presents. Later on, we have presented ethical guidelines to consider children as possible users of the system. Finally, we have adapted the previous general CA to better fit children's needs.

Ideally, stakeholders should be involved from the first stage of the design. This would improve the level of specifications of the proposed system. However, this adaptation remains general, as we understand this is a first approach to the design, and further development and implementation of the system remains as future work.

7. References

[1] M. McTear, "Conversational ai: Dialogue systems, conversational agents, and chatbots," *Synthesis Lectures on Human Language Technologies*, vol. 13, no. 3, pp. 1–251, 2020.

[2] J. Miehle, I. Bagci, W. Minker, and S. Ultes, *A Social Companion and Conversational Partner for the Elderly*. Cham: Springer International Publishing, 2019, pp. 103–109. [Online]. Available: https://doi.org/10.1007/978-3-319-92108-2_12

[3] P.-T. Singamaneni, A. Mayima, G. Sarthou, Y. Sallami, G. Buisan, K. Belhassein, J. Waldhart, and C. Aurélie, "Guiding task through route description in the MuMMER project," in *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Cambridge, UK, 2020. [Online]. Available: https://hal.laas.fr/hal-02286050

[4] C. Biele, A. Jaskulska, W. Kopec, J. Kowalski, K. Skorupska, and A. Zdrodowska, "How might voice assistants raise our children?" in *Intelligent Human Systems Integration 2019*, W. Karwowski and T. Ahram, Eds. Cham: Springer International Publishing, 2019, pp. 162–167.

[5] J. Kennedy, S. Lemaignan, C. Montassier, P. Lavalade, B. Irfan, F. Papadopoulos, E. Senft, and T. Belpaeme, "Child speech recognition in human-robot interaction: evaluations and recommendations," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 82–90.

[6] J. H. Nilsen and K. Røyneland, ""it knows how to not understand us!" a study on what the concept robustness entails in design of conversational agents for preschool children," Master's thesis, 2019.

[7] R. Sinha and S. Shahnawazuddin, "Assessment of pitch-adaptive front-end signal processing for children's speech recognition," *Computer Speech Language*, vol. 48, pp. 103–121, 10 2018.

[8] V. Bhardwaj, M. T. Ben Othman, V. Kukreja, Y. Belkhier, M. Bajaj, B. S. Goud, A. U. Rehman, M. Shafiq, and H. Hamam, "Automatic speech recognition (asr) systems for children: A systematic literature review," *Applied Sciences*, vol. 12, no. 9, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/9/4419

[9] B. Jolad and R. Khanai, "Anns for automatic speech recognition—a survey," in *Expert Clouds and Applications*, I. Jeena Jacob, F. M. Gonzalez-Longatt, S. Kolandapalayam Shanmugam, and I. Izonin, Eds. Singapore: Springer Singapore, 2022, pp. 35–48.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.

[11] Z. Lipton, X. Li, J. Gao, L. Li, F. Ahmed, and L. Deng, "Bbq-networks: Efficient exploration in deep reinforcement learning for task-oriented dialogue systems," vol. 32, Apr. 2018. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/11946

[12] S.-Y. Su, X. Li, J. Gao, J. Liu, and Y.-N. Chen, "Discriminative deep Dyna-Q: Robust planning for dialogue policy learning," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3813–3823.

[13] B. Peng, C. Zhu, C. Li, X. Li, J. Li, M. Zeng, and J. Gao, "Few-shot natural language generation for task-oriented dialog," in *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 172–182.

[14] X. Tan, T. Qin, F. Soong, and T.-Y. Liu, "A survey on neural speech synthesis," 2021. [Online]. Available: https://arxiv.org/abs/2106.15561

[15] J. Gao, M. Galley, and L. Li, "Neural approaches to conversational ai," *Foundations and Trends® in Information Retrieval*, vol. 13, no. 2-3, pp. 127–298, 2019. [Online]. Available: http://dx.doi.org/10.1561/1500000074

[16] R. G. Sousa, P. M. Ferreira, P. M. Costa, P. Azevedo, J. P. Costeira, C. Santiago, J. Magalhaes, D. Semedo, R. Ferreira, A. I. Rudnicky *et al.*, "ifetch: Multimodal conversational agents for the online fashion marketplace," in *Proceedings of the 2nd ACM Multimedia Workshop on Multimodal Conversational AI*, 2021, pp. 25–26.

[17] W. Wei, Q. Le, A. Dai, and J. Li, "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, Oct.-Nov. 2018, pp. 3844–3854.

[18] M. Eric, L. Krishnan, F. Charette, and C. D. Manning, "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, Aug. 2017, pp. 37–49.

[19] Y. Xu, V. Vigil, A. S. Bustamante, and M. Warschauer, ""elinor's talking to me!": Integrating conversational ai into children's narrative science programming," in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–16.

[20] A. Sciuto, A. Saini, J. Forlizzi, and J. I. Hong, """ hey alexa, what's up?" a mixed-methods studies of in-home conversational agent usage," in *Proceedings of the 2018 designing interactive systems conference*, 2018, pp. 857–868.

[21] P. Ala-Pietilä, Y. Bonnet, U. Bergmann, M. Bielikova, C. Bonefeld-Dahl, W. Bauer, L. Bouarfa, R. Chatila, M. Coeckelbergh, V. Dignum *et al.*, *The assessment list for trustworthy artificial intelligence (ALTAI)*. European Commission, 2020.

[22] V. Dignum, M. Penagos, K. Pigmans, and S. Vosloo, *Policy guidance on AI for children*. Communications of UNICEF, 2021. [Online]. Available: https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children

[23] V. Charisi, S. Chaudron, R. Di Gioia, R. Vuorikari, M. Escobar-Planas, I. Sanchez, and E. Gomez, *Artificial intelligence and the rights of the child : towards an integrated agenda for research and policy*. Publications Office of the European Union, 2022. [Online]. Available: https://data.europa.eu/doi/10.2760/012329

[24] M. Escobar-Planas, E. Gómez, and C.-D. Martínez-Hinarejos, "Guidelines to develop trustworthy conversational agents for children," *arXiv*, 2022.

[25] M. Escobar-Planas, V. Charisi, and E. Gómez, ""that robot played with us!" children's perceptions of a robot after a child-robot group interaction," vol. 6, no. CSCW2, 2022. [Online]. Available: https://doi.org/10.1145/3555118

[26] F. Nothdurft and W. Minker, *Justification and Transparency Explanations in Dialogue Systems to Maintain Human-Computer Trust*. Cham: Springer International Publishing, 2016, pp. 41–50. [Online]. Available: https://doi.org/10.1007/978-3-319-21834-2_4

[27] A. C. Kelly, E. Karamichali, A. Saeb, K. Veselỳ, N. Parslow, A. Deng, A. Letondor, R. O'Regan, and Q. Zhou, "Soapbox labs verification platform for child speech." in *INTERSPEECH*, 2020, pp. 486–487.

[28] A. Raj and M. D. Ekstrand, "Fire dragon and unicorn princess; gender stereotypes and children's products in search engine responses," *arXiv preprint arXiv:2206.13747*, 2022.

[29] E. Beneteau, O. K. Richards, M. Zhang, J. A. Kientz, J. Yip, and A. Hiniker, "Communication breakdowns between families and alexa," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13. [Online]. Available: https://doi.org/10.1145/3290605.3300473