

Título y Resumen:

TOWARDS TRUSTWORTHY CONVERSATIONAL AGENTS FOR CHILDREN

Resumen: Los agentes conversacionales (AC) son una tecnología revolucionaria que trae muchas oportunidades a la sociedad. Sin embargo, están entrando en nuestras vidas a tal velocidad que es difícil predecir el impacto que pueden tener en la sociedad. Debido a esto, durante los últimos años, la Comisión Europea ha trabajado para definir un enfoque centrado en el ser humano para la Inteligencia Artificial (IA), que incluye los (AC).

Basándose en las recomendaciones de la Comisión Europea, es imprescindible obtener ACs de confianza. Esto es aún más importante para los niños, que se sienten más atraídos y son más vulnerables a esta tecnología estando además en un periodo de desarrollo socio-emocional y cognitivo. Mi tesis doctoral se centrará en la evaluación y desarrollo de agentes conversacionales fiables para niños, teniendo en consideración sus derechos, deberes y necesidades específicas.

El proyecto de doctorado profundizará en el concepto de sistemas fiables, definiendolos y evaluandolos. También llevaremos a cabo casos de estudio con usuarios. El proyecto contribuirá con ideas novedosas y producirá una serie de publicaciones científicas.

Índice

| | |
|---------------------------------------|---|
| Título y Resumen | 1 |
| Menciones y Lugar de desarrollo | 2 |
| Antecedentes y estado actual | 2 |
| Bibliografía | 4 |
| Hipótesis y Objetivos | 6 |
| Metodología a utilizar | 6 |
| Medios a utilizar | 7 |
| Planificación temporal | 7 |

Lugar de desarrollo:

Otra institución: European Commission, Joint Research Centre, Sevilla

Departamento: Unidad de Economía Digital

Instituto: Grupo de investigación HUMAINT

Descripción lugar de desarrollo: Sevilla

[Índice](#)

Antecedentes y estado actual:

Hoy día muchos estamos acostumbrados al uso de agentes conversacionales (AC). Su desarrollo computacional nos ha permitido tener pequeños ordenadores en dispositivos útiles: Google Assistant puede indicarnos el camino a casa en nuestros coches, Siri puede enviar mensajes a través de nuestros móviles y Alexa puede reproducir música en nuestro altavoz inteligente. La inteligencia artificial conversacional está ganando popularidad y mucha gente la está introduciendo en su día a día.

Tampoco podemos pasar por alto que los ACs son accesibles y populares entre los niños. Muchos estudios sobre el uso de los ACs en el hogar [1], [2] y [3], han identificado la relevancia de estos dispositivos para niños, y el hecho de que hacen un uso más extenso y exploran más sus capacidades.

La literatura ha identificado muchas oportunidades de los ACs para los niños:

- Mejora de la accesibilidad: para niños demasiado jóvenes para escribir, con dislexia o con discapacidad física [4] y [5].
- Incitación al aprendizaje: ayudando a la búsqueda de información [6] y [7], al aprendizaje de idiomas [8] y [9], o a la enseñanza de materia escolar [10] y [11].
- Promoción del comportamiento social: mejorando la persuasión [12] y [13], ayudando a niños autistas [14] y [15].
- Apoyo a la salud en el hogar: ayudando a niños a hacer seguimiento de sus tratamientos [16], reduciendo la ansiedad y la depresión en adolescentes [17].

Sin embargo, el diseño de los ACs está centrado en los adultos, generando ciertos desafíos a la hora de interactuar con niños [18], [19]. Se han identificado diferentes problemas:

- Las características del habla de los niños difieren de las adultas: tienen un rango de tono más amplio y a veces una prosodia diferente. Esto complica el trabajo del AC al transformar la entrada de audio a texto (para su posterior procesamiento).
- Las expresiones de los niños pueden diferir de las de los adultos: pueden tener errores gramaticales o de vocabulario o incluso utilizar palabras inventadas. Esto complica el trabajo del AC para comprender la entrada del usuario.

- Los derechos de los niños difieren de los adultos: como el derecho de compra o el de protección. Esto complica el trabajo del AC a la hora de decidir qué acciones realizar para el usuario.
- La comprensión de los niños difiere de la de los adultos: los niños pueden necesitar palabras o explicaciones más simples. Esto complica el trabajo del AC a la hora de traducir la intención de la máquina al lenguaje humano.

Debido a estas particularidades, la precisión de los ACs puede ser baja dependiendo de su demografía (edad, nivel socioeconómico o lenguaje). Este hecho afecta a sus derechos de participación e incrementa las desigualdades en nuestra sociedad. Precisamente por ello existen muchos estudios para la mejora de la interacción niño-ordenador [20] desarrolla un identificador del habla para bebés, y [21] y [22] identifican el uso de buenas estrategias a realizar cuando un sistema no ha entendido a un niño.

Otros riesgos que se han identificado, provienen de la desinformación sobre qué es un AC y cómo funciona, generando una sobreconfianza en el sistema. Los niños tienden a percibir a los ACs como amigos [23] y [24], que tienen una alta influencia en los niños y aumentan el riesgo de divulgación de datos. En [25] los autores luchan contra la sobreconfianza a través de la programación y la transparencia sobre los ACs.

Uno de los desafíos de la prevención de riesgos de estos sistemas, es que son nuevos y diferentes a cualquier experiencia previa. Nuestra inexperiencia dificulta la prevención de problemas futuros. Un ejemplo de problema inesperado es la situación en la que se han encontrado las niñas y niños llamados “Alexa”, que están sufriendo acoso y se están viendo obligados a cambiar de nombre y de colegio. Los padres de estos niños han terminado solicitando a Amazon que cambie la palabra de activación de su dispositivo por el impacto que está creando en sus vidas [26].

Al igual que con la inteligencia artificial (IA), los AC brindan muchas oportunidades y nuevas aplicaciones, pero también tienen riesgos potenciales. El grupo de expertos de alto nivel en IA (HLEG por el inglés High Level Expert Group), ha desarrollado una base para el desarrollo de IA fiable [27], basado en siete principios: agencia humana y supervisión, solidez técnica y seguridad, privacidad y control de datos, transparencia, diversidad y equidad, bienestar social y ambiental y responsabilidad. Siguiendo estas pautas queremos desarrollar y evaluar ACs fiables para que interactúen con nuestros hijos.

[Índice](#)

Bibliografia:

1. Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018, June). " Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp.857-868).
<http://doi.org/10.1145/3196709.3196772>
2. Garg, R., & Sengupta, S. (2020). He is just like me: a study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1), 1-24.
3. Lovato, S. B., Piper, A. M., & Wartella, E. A. (2019, June). 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* (pp. 301-313).
4. Pradhan, A., Mehta, K., & Findlater, L. (2018, April). " 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* (pp. 1-13).
5. Catania, F., Crovari, P., Beccaluva, E., De Luca, G., Colombo, E., Bombaci, N., & Garzotto, F. (2021, July). Boris: a Spoken Conversational Agent for Music Production for People with Motor Disabilities. In *CHIItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter* (pp. 1-5).
6. Downs, B., French, T., Landau Wright, K., Pera, M. S., Kennington, C., & Fails, J. A. (2019, June). Children and search tools: Evaluation remains unclear. In *KidRec Workshop co-located with ACM IDC 2019*.
7. Landoni, M., Murgia, E., Huibers, T., & Pera, M. S. (2020, July). 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* (pp. 89-94).
8. Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1-2), 61-84.
9. Nasihati Gilani, S., Traum, D., Merla, A., Hee, E., Walker, Z., Manini, B., ... & Petitto, L. A. (2018, October). Multimodal dialogue management for multiparty interaction with infants. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (pp. 5-13).
10. Xu, Y., & Warschauer, M. (2020, April). " 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* (pp. 1-8).
11. Law, E., Baghaei Ravari, P., Chhibber, N., Kulic, D., Lin, S., Pantasdo, K. D., ... & Dillen, N. (2020, April). Curiosity Notebook: A Platform for Learning by Teaching Conversational Agents. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-9).
12. Fraser, J., Papaioannou, I., & Lemon, O. (2018, November). Spoken conversational ai in video games: Emotional dialogue management increases user engagement. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (pp. 179-184).
13. Keizer, S., Guhe, M., Cuayáhuil, H., Efstathiou, I., Engelbrecht, K. P., Dobre, M., ... & Lemon, O. (2017, April). Evaluating persuasion strategies and deep reinforcement learning methods for negotiation dialogue agents. *ACL*.
14. Ali, M. R., Razavi, S. Z., Langevin, R., Al Mamun, A., Kane, B., Rawassizadeh, R., ... & Hoque, E. (2020, October). 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* (pp. 1-8).
15. Zhang, L., Weitlauf, A. S., Amat, A. Z., Swanson, A., Warren, Z. E., & Sarkar, N. (2020). Assessing social communication and collaboration in autism spectrum disorder using

- intelligent collaborative virtual environments. *Journal of autism and developmental disorders*, 50(1), 199-211.
16. Sezgin, E., Noritz, G., Elek, A., Conkol, K., Rust, S., Bailey, M., ... & Huang, Y. (2020). Capturing at-home health and care information for children with medical complexity using voice interactive technologies: multi-stakeholder viewpoint. *Journal of medical Internet research*, 22(2), e14202.
 17. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). 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*, 4(2), e7785.
 18. Narayanan, S., & Potamianos, A. (2002). Creating conversational interfaces for children. *IEEE Transactions on Speech and Audio Processing*, 10(2), 65-78., <http://doi.org/10.1109/89.985544>
 19. Kennedy J, Lemaignan S, Montassier C, Lavalade P, Irfan B, Papadopoulos F, Senft E, Belpaeme T. (2017) *Child Speech Recognition in Human-Robot Interaction: Evaluations and Recommendations*, <http://doi.org/10.1145/2909824.3020229>
 20. Lavechin M, Bousbib R, Bredin H, Dupoux E, Cristia A (2020), *An open-source voice type classifier for child-centered daylong recordings*, arXiv:2005.12656
 21. Røyneland, K. (2020). It knows how to not understand us!. *A study on what the concept of robustness entails in design of conversational agents for preschool children*. <https://www.duo.uio.no/handle/10852/69059>. Accessed, 21. Markus Schedl, Christine Bauer (2017) Online Music Listening Culture of Kids and Adolescents: Listening Analysis and Music Recommendation Tailored to the Young, KidRec 2017 <https://arxiv.org/abs/1912.11564>
 22. Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018). Why doesn't it work? voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 337-348).
 23. Kahn Jr, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., ... & Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology*, 48(2), 303.
 24. Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017, June). "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* (pp. 595-600).
 25. Straten, C. L. V., Peter, J., Kühne, R., & Barco, A. (2020). 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)*, 9(2), 1-22.
 26. <https://www.bbc.com/news/technology-57680173> last access 08/10/2021
 27. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> last access 10/10/2021

Hipótesis y Objetivos :

Hipótesis: Los agentes conversacionales (AC) traen muchas oportunidades a los niños. Pero también hay muchos retos y riesgos que deben ser minimizados. La hipótesis principal de esta tesis es la necesidad de metodologías prácticas para la evaluación y desarrollo de agentes conversacionales fiables que aseguren su beneficio.

Objetivos:

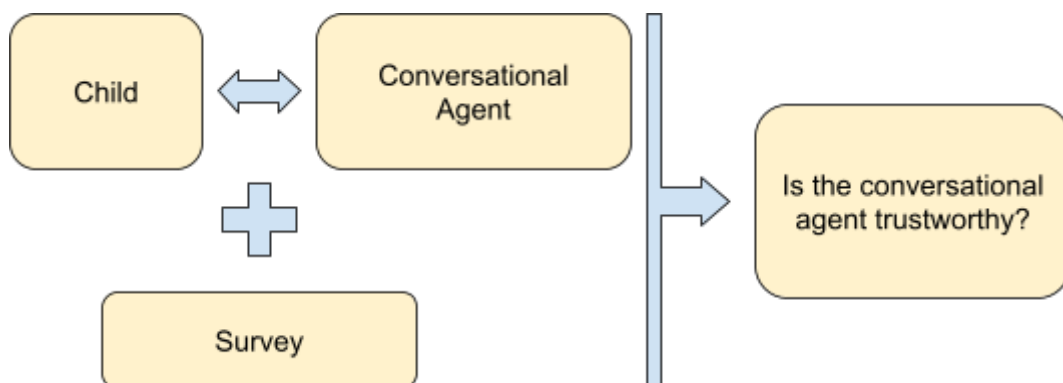
1. Comprender las oportunidades, riesgos y retos involucrados en el desarrollo de AC para niños.
2. Definición de metodologías de evaluación de los AC para niños y el diseño de sistemas fiables.
3. Explotación de estas metodologías en un set de experimentos o estudios de interacción niño-AC .

[Índice](#)

Metodología a Utilizar:

A través de la documentación estudiaré el estado de investigaciones previas en el campo que puedan serme de utilidad en mi propia investigación.

También realizaremos algunos experimentos usando agentes conversacionales con niños. Estos experimentos consistirán en la interacción del agente conversacional con el niño, seguido de una entrevista. Durante el experimento recopilaremos datos tanto cualitativos (como el grado de satisfacción del niño), como cuantitativos (como el número de veces que el sistema no ha logrado comprender al niño). Posteriormente, utilizaremos estos datos para el estudio de las dificultades que presenta el desarrollo de los ACs para niños/niñas y también para evaluar la fiabilidad de los mismos.



[Índice](#)

Medios a utilizar:

Realizaré experimentos utilizando diferentes plataformas y sistemas para el desarrollo de agentes conversacionales:

- Chatbots en internet: El JRC me puede proporcionar soporte técnico para su desarrollo.
- Asistentes del hogar (Alexa, Google Assistant, Siri, ...): Tengo acceso a algunos Alexas en mi casa, que me permiten crear "skills" que otros pueden descargarse en su propio dispositivo Alexa.
- Robots sociales (Pepper, Nao, Haru): El JRC me puede proporcionar acceso al robot Pepper y a su laboratorio. Además tengo contactos en las universidades de Sevilla y Pablo de Olavide con los que puedo iniciar colaboraciones usando los robots Nao y Haru.
- GPT-3: Open AI concede algunos accesos a GPT-3 a proyectos específicos que lo soliciten.

Durante el análisis de los datos, usaré Jupyter con python, pandas, nltk, matplotlib o wordcloud. También planeo usar spss para el análisis estadístico, teniendo R como segunda opción.

En caso de necesitar GPUs para el desarrollo de algún agente conversacional, tengo acceso tanto por parte del JRC como por parte de la UPV (con 28 clusters y algunos servidores de datos).

[Índice](#)

Planificación temporal y utilidad de la investigación:

Tareas:

1. Formación:
 - a. Como investigadora
 - b. Área de conocimiento
2. Planificación y desarrollo del plan de investigación
3. Estudios:
 - Estudio 1
 - Estudio 2
 - Estudio 3
4. Difusión de resultados
5. Escritura de la tesis

