

Compendio de Comunicaciones del XXV Congreso Internacional de Interacción Persona-Ordenador



Compendio de Comunicaciones

XXV Congreso Internacional de Interacción
Persona-Ordenador (INTERACCIÓN 2025)

Comité Organizador

Valladolid, España, del 3 al 5 de Septiembre de 2025

Editores:

María Jesús Rodríguez-Triana, Juan A. Muñoz-Cristóbal, Valentín Cardeñoso-Payo, Alejandra Martínez-Monés, Jesús Gallardo

ISBN: 978-84-09-75360-4

Interacción 2025
AIPO

Este congreso ha sido posible gracias al apoyo de los siguientes colaboradores y patrocinadores:



Diseño gráfico de portada del libro y de imagen de portada de capítulos: RQR, Comunicación
Maquetación: Juan A. Muñoz-Cristóbal

Copyright © 2025 de los artículos individuales, de la portada del libro, de los logos, y de la imagen de la portada de los capítulos pertenece a sus respectivos autores. Copyright © 2025 del libro como colección pertenece a sus editores. Este volumen, como conjunto, se publica bajo la licencia Creative Commons Atribución-NoComercial-SinDerivados 4.0 Internacional.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Los artículos contenidos en este volumen se publican bajo la licencia Creative Commons Atribución 4.0 Internacional.  <https://creativecommons.org/licenses/by/4.0/>

Índice general

Prefacio	6
Comités	8
Artículos largos y cortos	13
Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales — Hitos-García, Alberto; Albiol-Pérez, Sergio; Medina-Medina, Nuria; Paderewski-Rodríguez, Patricia; Bautista, Susana; Gutiérrez-Vela, Francisco Luis	14
Automatización del protocolo Thinking Aloud en evaluaciones de usabilidad — Li, Shuoshuo; Macías, José A.	23
Desarrollo de un módulo conversacional basado en IA para la interacción emocional con robots sociales — Mendoza, Liany; Cerezo, Eva; Arribas Mateo, Adrian; Baldassarri, Sandra	35
“Robots in the wild”: grupos intergeneracionales y robots sociales en espacios públicos — Lacuesta, Raquel; Cerezo, Eva; Matinero, Loreto; Baldassarri, Sandra	49
Evaluación de detectores de estrés durante la interacción con juegos serios que estimulan actividades cognitivas — Martínez-Limeres, Leticia; Condori-Fernández, Nelly; Catala, Alejandro	64
Evaluación de la comprensión de las pautas WCAG 2.1: análisis de ejemplos prácticos en tres universidades españolas — Pascual-Almenara, Afra; Pedrero-Esteban, Alberto; Bautista, Susana	80
Primeras experiencias para el aprendizaje de ciberseguridad a través del uso de robots sociales — Lacuesta, Raquel; Pérez, Daniel M.; Hervás, Eva; Azuara, Guillermo; Gallardo, Jesús	109
Robots sociales en museos – un enfoque centrado en el análisis de emociones — Lacuesta, Raquel; Barrado Lucía, Fernando; Guillén Florian, Alejandro; Gaya Morey, Francesc Xavier; Buades Rubio, José María; Perales López, Francisco José	118
SanbotPPT: presentador configurable para un robot social — Matinero, Loreto; Baldassarri, Sandra; Cerezo, Eva	127
Modelado de audiodescripción prosódica basado en narrativa semántica — Rozo-Torres, Alexander; Plazas, Cristian; Quintero, Christian	144

Conferencias invitadas	160
HCI en tiempos de IA: ética, diseño y regulación — Hupont, Isabelle	161
Interfaces cerebro-máquina: hacia una nueva era en la interacción persona-ordenador — López Bernal, Sergio	162
Demostraciones	163
AffeX: a real-time emotion monitoring and visualization tool — Méndez Flórez, Juan Camilo; Condori-Fernández, Nelly; Catala, Alejandro; Figueroa Vásquez, Salvador Alonso; Alonso-Moral, Jose M.	164
TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados — Lozano, María D.; Orozco, Luis; Gallud, Jose A.; Penichet, Víctor M. R.; Sebastián Rivera, Gabriel; Tesoriero, Ricardo; Garrido, Juan E.; Costa-Tebar, Felipe; Polo-Martínez, Lucía; Luna-Lorente, Belén; Navarro, Alejandro; Cerezo, Eva; Baldassarri, Sandra	168
SimpLee: an artificial intelligence agent for Easy Reading in support of demographic challenge — Cea-López, Margarita; Jiménez Sánchez, Alberto; Rodríguez Marcos, Clara; Manso-Conde, Javier; Grande-Pérez, Laura; Gil-González, Ana-Belén	172
Coloquio doctoral	176
Virtual interfaces and augmented reality for orthopedic surgery — Costa, Ánxela Pérez; Souto, Nieves Pedreira	177
Diseño de un modelo de empatía en diálogos de agentes sociales interactivos para la mejora de la aceptación tecnológica — Hitos-García, Alberto; Gutiérrez-Vela, Francisco Luis; Paderewski-Rodríguez, Patricia	181
Presentaciones de grupos o proyectos de investigación	185
EXPLainable Artificial INtelligence Methods and Evaluation for trustworthy, human- centric, and ethical AI (EXPLAINME) — Manresa-Yee, Cristina; Jaume-i-Capó, Antoni; Ghazel, Adel; Hupont, Isabelle; Leiva, Luis A.; MacKenzie, I. Scott	186
Artículos relevantes ya publicados	190
Resumen del artículo: un experimento empírico de un método de elicitación de requisitos de usabilidad para diseñar GUI basadas en entrevistas — Ormeño, Yeshica Isela; Panach, José Ignacio; Pastor, Oscar	191
Concurso TFG/TFM	195
Heurísticas de evaluación de usabilidad para plataformas de comercio electrónico — Vargas Muñoz, Alejandra Yicel; Collazos, César A.; Granollers i Saltiveri, Antoni	196
Entrando en Oz: simulación virtual y evaluación de robots sociales a través de técnicas de Mago de Oz — Cubero, Guillermo; Villa, Laura; Hervás, Ramón	199

Análisis y modelado del comportamiento de estudiantes en cursos en línea desde el enfoque de la Analítica del Aprendizaje Multimodal — Becerra, Alvaro; Cobos, Ruth	203
PlanTEA-WM: migración y mejora de una aplicación de planificación para personas con TEA — Lara, José; Molina, Ana I.; Lacave, Carmen	209
Índice de autores	214

Prefacio

Interacción es un evento anual promovido por la Asociación de Interacción Persona-Ordenador (AIPO) y por el capítulo español de ACM SIGCHI (*Special Interest Group on Computer-Human Interaction*, CHISPA), cuyo objetivo es fomentar y difundir los avances científicos en la disciplina de la Interacción Persona-Ordenador (IPO) entre la comunidad hispanohablante.

AIPO organiza anualmente el congreso *Interacción* desde el año 2000, con la única excepción del año 2020, en que se suspendió debido a la pandemia de COVID-19. Desde la primera edición, celebrada en Granada, hasta la más reciente hasta la fecha —en A Coruña, durante el CEDI 2024—, el congreso ha tenido lugar en diversas ciudades de España y Latinoamérica. Este año hemos tenido el honor de celebrar su vigésimo quinto aniversario, que refleja la consolidada trayectoria del congreso en la difusión de la investigación en IPO en el contexto latinoamericano.

Esta XXV edición del Congreso Internacional de Interacción Persona-Ordenador, *Interacción 2025*, se ha celebrado del 3 al 5 de septiembre de 2025 en la ciudad de Valladolid (España). Ha sido organizada por dos grupos de investigación de la Universidad de Valladolid: ECA-SIMM (Entornos de Computación Avanzada y Sistemas de Interacción Multimodal) y GSIC (Grupo de Sistemas Inteligentes y Colaborativos), centrados en los ámbitos de las tecnologías del habla y el aprendizaje apoyado por tecnología, desde los marcos conceptuales y metodológicos de la IPO. Ambos grupos se caracterizan por una aproximación interdisciplinar a la investigación, en la que los factores cognitivos, sociales y éticos se consideran esenciales en los procesos humanos que buscan potenciar mediante sus desarrollos tecnológicos.

En esta edición se han recibido un total de 34 contribuciones (artículos y demostraciones), de las cuales se seleccionaron 17 artículos, escritos en inglés, para su publicación en los “*Proceedings of the XXV International Conference on Human-Computer Interaction (INTERACCIÓN 2025)*”, publicados en CEUR-WS.org (<https://ceur-ws.org/Vol-4000/>). El presente volumen recoge un compendio del resto de trabajos aceptados para su presentación en el congreso: 10 artículos (largos o cortos), 3 demostraciones, 2 resúmenes de conferencias invitadas, 2 resúmenes de tesis doctoral, 1 presentación de proyecto de investigación, 1 resumen de artículo relevante ya publicado y 4 resúmenes de TFGs.

Una parte fundamental del trabajo de la comunidad consiste en reconocer y apoyar a quienes inician su camino en esta área de investigación. Por ello, el congreso incluye la entrega de premios a los mejores TFG y TFM, promovidos por AIPO, así como la celebración de un coloquio doctoral. Además, la comunidad está presente mediante presentaciones de grupos y proyectos de investigación en IPO, y mediante un taller sobre “Perspectiva de Género en la Investigación”.

Este último, coordinado por investigadoras destacadas en el área, busca compartir experiencias y propuestas para seguir impulsando una visión equitativa de la investigación en un ámbito donde la presencia femenina sigue siendo escasa.

El congreso ha contado con la participación de dos conferenciantes de reconocido prestigio internacional, que han contribuido a abrir nuevos horizontes en la disciplina. Isabelle Hupont, investigadora del *Joint Research Centre* de la Comisión Europea y asesora del Ministerio de Ciencia, Innovación y Universidades, impartió la conferencia “HCI en tiempos de IA: ética, diseño y regulación”, en la que compartió su visión sobre el papel de la ciencia y la tecnología en la formulación de políticas, con especial énfasis en la regulación de una inteligencia artificial confiable. Por su parte, Sergio López Bernal, investigador del *CyberDataLab* de la Universidad de Murcia, presentó la conferencia “Interfaces cerebro-máquina: hacia una nueva era en la interacción persona-ordenador”, donde abordó sus investigaciones en la intersección entre ciberseguridad, interfaces cerebro-máquina, inteligencia artificial y análisis de datos.

Este volumen refleja la evolución de la investigación y el desarrollo en IPO, destacando la colaboración entre universidades en trabajos que abordan temáticas de gran interés actual. Entre ellas se incluyen la interacción social con robots —considerando aspectos cognitivos y de accesibilidad—; el estudio de la relación entre aspectos cognitivos, variables clínicas y agentes inteligentes; así como investigaciones sobre interfaces avanzadas (realidad virtual y extendida) y cuestiones transversales a la IPO como evaluación, usabilidad y experiencia de usuario (UX). Los temas tratados muestran una disciplina cada vez más necesaria en un mundo donde la tecnología domina la interacción y configura incluso nuestra forma de pensar, lo que exige enfoques interdisciplinarios que integren aspectos éticos como la *explicabilidad* y la accesibilidad, también presentes en las comunicaciones de esta edición.

Interacción 2025 se ha celebrado en el Campus Miguel Delibes de la Universidad de Valladolid, sede de la Escuela de Ingeniería Informática de la UVA, donde desarrollan su actividad los grupos organizadores. Esta labor se lleva a cabo en estrecha colaboración con investigadores e investigadoras de otras facultades del campus, como la Facultad de Educación y Trabajo Social, la ETS de Ingenieros de Telecomunicación y la Facultad de Ciencias, así como de centros cercanos, como la Facultad de Filosofía y Letras.

Para la celebración de esta edición hemos contado con el apoyo en la gestión administrativa y económica de la Fundación Universidad de Valladolid, así como con el respaldo económico de la Escuela de Ingeniería Informática, el Departamento de Informática y el Centro de Inteligencia Artificial, financiado a su vez por IdeVA del Ayuntamiento de Valladolid.

Finalmente, queremos agradecer a la Asociación AIPO su confianza en nosotros para la organización de esta especial XXV edición de *Interacción*. Extendemos nuestro reconocimiento a todas las personas colaboradoras y colegas por su implicación y dedicación, esenciales para el éxito del evento. Esperamos que este compendio refleje el dinamismo y la riqueza del trabajo presentado, y que contribuya a enriquecer vuestra experiencia durante *Interacción 2025*.

¡Disfrutad, compartid e interactuad!

El Comité Organizador

Comités

Presidencia General

- Valentín Cardenoso Payo (Universidad de Valladolid)
- Alejandra Martínez-Monés (Universidad de Valladolid)
- Jesús Gallardo Casero (Universidad de Zaragoza)

Presidencia del Comité Organizador

- David Escudero Mancebo (Universidad de Valladolid)
- María Jesús Rodríguez-Triana (Universidad de Valladolid)
- Mario Corrales Astorgano (Universidad de Valladolid)

Comité Organizador

- Juan A. Muñoz-Cristóbal (Universidad de Valladolid)
- Valentín Cardenoso Payo (Universidad de Valladolid)
- Alejandra Martínez-Monés (Universidad de Valladolid)
- Jesús Gallardo Casero (Universidad de Zaragoza)
- Alejandro Ortega-Arranz (Universidad de Valladolid)

Presidencia del Comité de Programa

- Valentín Cardenoso Payo (Universidad de Valladolid)
- Alejandra Martínez-Monés (Universidad de Valladolid)
- Jesús Gallardo Casero (Universidad de Zaragoza)

Comité de Programa

- Silvia T. Acuña (Universidad Autónoma de Madrid)
- Rubén Alcaraz Martínez (Universitat de Barcelona)
- Francisco Álvarez (Universidad Autónoma de Aguascalientes)
- Myriam Arrue (Euskal Herriko Unibertsitatea)
- Sandra Baldassarri (Universidad de Zaragoza)
- Susana Bautista (Universidad Francisco de Vitoria)
- James Berrett (Swinburne University of Technology)
- Federico Botella (Universidad Miguel Hernández)
- David Cantarero Tomás (Universidad de Zaragoza)
- Valentín Cardeñoso-Payo (Universidad de Valladolid)
- Alejandro Catala (Universidade de Santiago de Compostela)
- Eva Cerezo (Universidad de Zaragoza)
- Cesar Collazos (Universidad del Cauca)
- Nelly Condori Fernández (Universidade de Santiago de Compostela)
- Mario Corrales Astorgano (Universidad de Valladolid)
- Rosanna Costaguta (Universidad Nacional de Santiago del Estero)
- Horacio Del Giorgio (Universidad Nacional de La Matanza)
- Antonio Diaz-Estrella (Universidad de Málaga)
- Rafael Duque (Universidad de Cantabria)
- Pedro Encarnação (Católica Lisbon School of Business & Economics)
- David Escudero Mancebo (Universidad de Valladolid)
- Ramon Fabregat (Universitat de Girona)
- Eduardo Fernández (Universidad de la República, INCO)
- Francisco Javier Galán Pérez (Universidad de Zaragoza)
- Jesús Gallardo Casero (Universidad de Zaragoza)

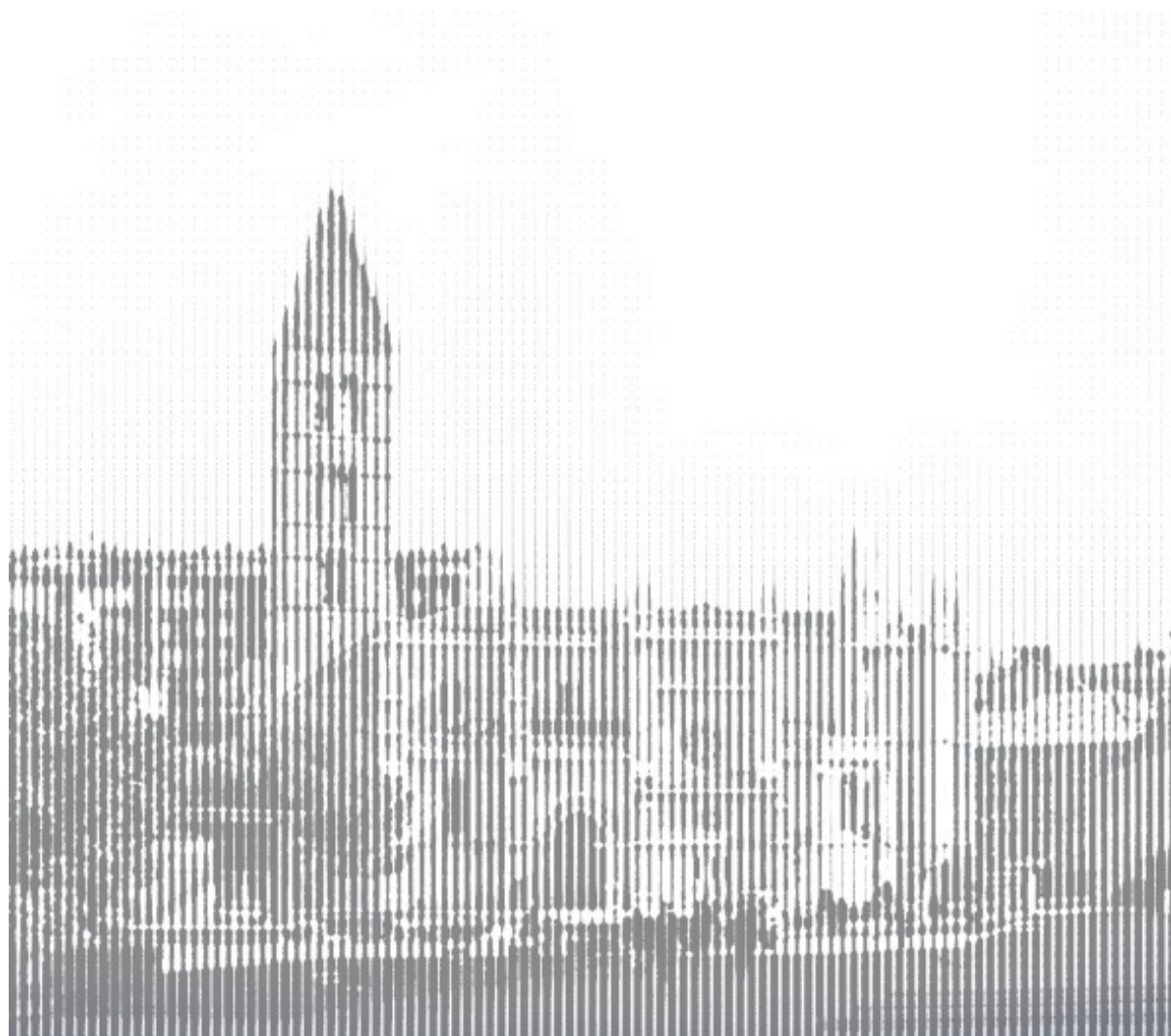
- Nestor Garay-Vitoria (Euskal Herriko Unibertsitatea)
- Alicia García-Holgado (Universidad de Zaragoza)
- Francisco José García-Peñalvo (Universidad de Zaragoza)
- Juan Enrique Garrido Navarro (Universidad de Castilla-La Mancha)
- Alberto Gaspar (Universitat de Valencia)
- Rosa Maria Gil (Universitat de Lleida)
- María Paula Gonzalez (CONICET / UNS)
- Pascual González (Universidad de Castilla-La Mancha)
- Mariano González (Universidad de Sevilla)
- Toni Granollers (Universitat de Lleida)
- Daniel Guasch (Universitat Politècnica de Catalunya)
- Francisco Luis Gutiérrez Vela (Universidad de Granada)
- Ganix Lasa (Mondragon Unibertsitatea)
- Silvia María Hernández Muñoz (Universidad de Zaragoza)
- Francisco Iniesto (Universidad Nacional de Educación a Distancia)
- Javier Jaen (Universitat Politècnica de València)
- Angel Jimenez-Fernandez (Universidad de Sevilla)
- Raquel Lacuesta (Universidad de Zaragoza)
- Luis A. Leiva (University of Luxembourg)
- Martín Llamas-Nistal (Universidad de Vigo)
- Juan Miguel López (Euskal Herriko Unibertsitatea)
- Víctor López-Jaquero (Universidad de Castilla-La Mancha)
- Begoña Losada (Euskal Herriko Unibertsitatea)
- Maria Dolores Lozano (Universidad de Castilla-La Mancha)
- Huizilopoztli Luna-García (Universidad Autónoma de Zacatecas)
- José Antonio Macías Iglesias (Universidad Autónoma de Madrid)

- Cristina Manresa-Yee (Universidad de Sevilla)
- Estefania Martin (Universidad Rey Juan Carlos)
- Paloma Martínez Fernández (Universidad Carlos III de Madrid)
- Alejandra Martínez-Monés (Universidad de Valladolid)
- Valéria F. Martins (Universidade Presbiteriana Mackenzie)
- Asier Marzo (Public University of Navarre)
- Ramon Mas (Universitat de les Illes Balears)
- Maitane Mazmela Etxabe (Mondragon Unibertsitatea)
- Nuria Medina Medina (University of Granada)
- António Mendes (University of Coimbra)
- Roi Méndez Fernández (Universidade de Santiago de Compostela)
- José Pascual Molina Massó (Universidad de Castilla-La Mancha)
- Luis Molina-Tanco (Universidad de Málaga)
- Francisco Montero (Universidad de Castilla-La Mancha)
- Lourdes Moreno (Universidad Carlos III de Madrid)
- Jaime Muñoz-Arteaga (Universidad Autónoma de Aguascalientes)
- Juan A. Muñoz-Cristóbal (Universidad de Valladolid)
- Manuel Ortega Cantero (Universidad de Castilla-La Mancha)
- Patricia Paderewski (Universidad de Granada)
- Jose Ignacio Panach Navarrete (Universitat de València)
- Alejandro Panagiotidis (Universidad de Alicante)
- Afra Pascual (Universitat de Lleida)
- Jose Antonio Pow-Sang Portillo (Pontificia Universidad Católica del Perú)
- Miguel Redondo (Universidad de Castilla-La Mancha)
- Ranci Ren (Universidad Autónoma de Madrid)
- Arcadio Reyes-Lecuona (Mondragon Unibertsitatea)

- Mireia Ribera (Universitat de Barcelona)
- Covadonga Rodrigo (Universidad Nacional de Educación a Distancia)
- Inmaculada Rodríguez (Universitat de Barcelona)
- María Jesús Rodríguez-Triana (Universidad de Valladolid)
- Cecilia Sanz (Universidad Nacional de La Plata)
- Wilson Javier Sarmiento (Universidad Militar Nueva Granada)
- Christian Sturm (Hamm-Lippstadt University of Applied Sciences)
- Rogério Taveira (University of Lisbon)
- Miguel A. Teruel (Universidad de Alicante)
- Susanna Tesconi (Universitat Oberta de Catalunya)
- Roberto Theron (Universidad de Salamanca)
- Jesús Torres (Universidad de La Laguna)
- Carlos Vaz de Carvalho (Instituto Politécnico do Porto)
- Pablo Vera (National University of La Matanza)
- José Luis Verdegay (Universidad de Granada)
- Markel Vigo (The University of Manchester)
- Sergio Zepeda (Universidad Autónoma Metropolitana)

Artículos largos y cortos

Como se indica en el [Prefacio](#), una selección de artículos largos y cortos escritos en inglés se ha publicado en los *proceedings* del congreso en CEUR-WS.org (<https://ceur-ws.org/Vol-4000/>). El resto de trabajos presentados en el congreso se incluyen en el presente Compendio de Comunicaciones.



Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales

Alberto Hitos-García^{1,*†}, Sergio Albiol-Pérez^{2,†}, Nuria Medina-Medina^{1,†},
Patricia Paderewski-Rodríguez^{1,†}, Susana Bautista-Blasco^{3,†} y
Francisco Luis Gutiérrez-Vela^{1,†}

¹Universidad de Granada, Dpto. de Lenguajes y Sistemas Informáticos - ETSIT, Granada, España

²Universidad de Zaragoza; Instituto Salud Aragón (IIS), Teruel, España.

³Universidad Francisco de Vitoria, Escuela Politécnica Superior, Madrid, España

Resumen

El uso de Asistentes Virtuales (AV) es cada día más frecuente en diferentes contextos, lo cual abre una oportunidad única de integración social y acceso a la información para los adultos mayores. Esto podría romper la brecha digital que existe en este colectivo, siempre y cuando se estudien y analicen las necesidades especiales de los adultos mayores y cómo se puede mejorar su interacción con esta tecnología para generar una experiencia de usuario más satisfactoria. Con ese objetivo, este trabajo presenta un primer estudio de guías de usabilidad para adultos mayores y AV, identificando los problemas de usabilidad de este colectivo, revisando las guías existentes y proponiendo algunas guías adicionales que ayudan a integrar mejor a los adultos mayores a la hora de interactuar con los AV.

Palabras clave

usabilidad, adultos mayores, asistentes virtuales, guías

1. Introducción

La población de adultos mayores está creciendo de manera significativa a nivel mundial, debido al aumento de la esperanza de vida y a la disminución de las tasas de natalidad. Según datos de la Organización Mundial de la Salud (OMS) [1], se estima que para el año 2030, una de cada seis personas en el mundo tendrá 60 años o más, lo que representará un aumento del grupo poblacional hasta alcanzar los 1,400 millones. Para 2050, esta cifra se podría casi duplicar. Además, se prevé que el número de personas mayores de 80 años se triplicará entre 2020 y 2050, alcanzando los 426 millones. Por tanto, resulta fundamental atender las necesidades de este grupo desde una perspectiva tecnológica, ya que la tecnología puede contribuir significativamente a mejorar su vida diaria y a promover un envejecimiento activo y de calidad.

La interacción de los adultos mayores con **Asistentes Virtuales (AV)** representa una oportunidad clave para mejorar su calidad de vida, promoviendo la independencia, la accesibilidad y la conexión social [2]. Sin embargo, este grupo enfrenta desafíos únicos al adoptar nuevas tecnologías, como pueden ser las limitaciones físicas, cognitivas y emocionales que dificultan el uso intuitivo de dispositivos digitales. Por ello, diseñar interfaces y funcionalidades adaptadas a sus necesidades específicas se convierte en una prioridad.

Los AV, como Alexa, Google Assistant y Siri, han demostrado ser herramientas valiosas para esta población [2, 3]. Ofrecen beneficios como recordatorios de medicamentos, asistencia en tareas diarias,

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Autor responsable.

†Estos autores han contribuido equitativamente.

✉ albertohitos@ugr.es (A. Hitos-García); salbiol@unizar.es (S. Albiol-Pérez); nmedina@ugr.es (N. Medina-Medina);
patricia@ugr.es (P. Paderewski-Rodríguez); susana.bautista@ufv.es (S. Bautista-Blasco); fgutierr@ugr.es
(F. L. Gutiérrez-Vela)

🆔 0009-0009-3136-5022 (A. Hitos-García); 0000-0002-6280-1474 (S. Albiol-Pérez); 0000-0002-6013-732X (N. Medina-Medina);
0000-0001-6626-9633 (P. Paderewski-Rodríguez); 0000-0003-1648-0208 (S. Bautista-Blasco); 0000-0001-6629-7597
(F. L. Gutiérrez-Vela)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

control de dispositivos del hogar y compañía mediante interacciones conversacionales. Además, su diseño basado en comandos de voz reduce barreras tecnológicas al eliminar la necesidad de interfaces complejas. Sin embargo, también es importante considerar las necesidades emocionales y sociales de esta población. Muchos adultos mayores sienten ansiedad o falta de confianza al interactuar con nuevas tecnologías, lo que puede llevar a la frustración o incluso al rechazo.

Para maximizar la utilidad de los AV es esencial incorporar guías de usabilidad específicas para adultos mayores. Esto incluye incorporar interfaces minimalistas, accesibles y personalizables, así como procesos de co-diseño que involucren directamente a los usuarios mayores en todas las etapas del desarrollo. Integrar estas estrategias no solo mejora la aceptación tecnológica, sino que también fomenta un envejecimiento activo y saludable.

El artículo está estructurado de la siguiente forma: tras la sección de introducción, en la sección 2 se presenta el estado del arte relacionado con el tema. A continuación, sección 3, se describe la propuesta del trabajo, abordando los problemas de usabilidad y analizando las guías existentes para, posteriormente, introducir las nuevas guías de usabilidad diseñadas específicamente para AV dirigidos a adultos mayores. Finalmente, se concluye, sección 4, presentando las conclusiones del estudio y se proponen líneas de acción futuras para continuar desarrollando este trabajo.

2. Estado del arte

Los AV permiten la interacción con los usuarios a través del habla, lo que facilita su uso de forma sencilla en actividades cotidianas de la vida diaria de los adultos mayores [4]. Esta tecnología, que tiene un gran potencial para este colectivo, representa una línea de investigación en auge, pero todavía existen pocos estudios que analicen específicamente su uso en la población objetivo [3].

Los principales hitos en el diseño de este tipo de tecnología están enfocados en intentar resolver las problemáticas inherentes en adultos mayores: 1) pérdida de audición. Este colectivo, por lo general, necesita del uso de audífonos, que junto con ruidos ambientales, dificultan la interacción con este tipo de tecnología; 2) falta de comprensión a la hora de obtener un producto capaz de satisfacer las necesidades del adulto mayor; y 3) barreras emocionales fruto de la soledad no deseada en adultos mayores.

Debido a este tipo de problemáticas, los estudios de investigación existentes hoy en día se han centrado en aspectos relacionados con la usabilidad [5], interfaces de usuario [6], confianza [7], soledad psicológica [8], enriquecimiento de la salud mental [9], y hasta incluso con felicidad en el adulto mayor [10].

Otros tipos de soluciones capaces de paliar las problemáticas en adultos mayores están enfocadas en el diseño y posterior uso de AV capaces de interactuar de forma correcta con el usuario final. En estos diseños se muestran formas humanas o animales, con la capacidad de generar y reconocer tanto estímulos verbales como no-verbales (movimientos oculares, rotaciones de cabeza, generación de expresiones, control postural, etc). Gracias a estos diseños, es factible la generación de turnos de palabra entre el AV y el adulto mayor, junto con retroalimentaciones visuales y auditivas [11]. Esta tecnología permite la captación de la atención del usuario, el entretenimiento, y el enriquecimiento de las relaciones sociales [12].

Existen estudios del uso de AV que incluyen agentes conversacionales que han sido validados con adultos mayores en múltiples campos de actuación: asistenciales [13], apoyo social [14], y en el enriquecimiento de las Actividades de la Vida Diaria (AVD) [15].

En la actualidad, existen pocos estudios enfocados a la evaluación de la usabilidad de AV en adultos mayores.

3. Propuesta de guías de usabilidad de AV para adultos mayores

En este artículo se exploran guías prácticas de usabilidad diseñadas para optimizar la interacción de los adultos mayores con AV que han sido recopiladas de diversos autores. Además, se presenta una propuesta para la creación de nuevas guías elaborada a partir de la experiencia adquirida, que se

consideran relevantes para atender las necesidades detectadas. Estas recomendaciones tienen como objetivo no solo reducir la brecha tecnológica, sino también empoderar a esta población para que aproveche plenamente los beneficios que los AV pueden ofrecer en su vida diaria.

3.1. Problemas de usabilidad identificados

En base al estudio bibliográfico realizado y su posterior análisis, se han detectado diferentes problemas significativos asociados al uso de los AV por parte de los adultos mayores. Estos problemas se pueden clasificar en cuatro grupos: Relacionados con las **características propias** de los adultos mayores (sobre todo debido a los problemas del envejecimiento), relacionados con el **contexto** en el que se usan los AV, en relación con el **diseño y la funcionalidad** de este tipo de dispositivos (sobre todo lo que afecta a la forma de interactuar con ellos) y en relación con los aspectos **sociales y de relación** con el AV.

Respecto a las **características propias** de los adultos mayores, existen los siguientes problemas:

- P1. Problemas por pérdida de atención y comprensión en mensajes largos.
- P2. Diferencias en las capacidades y preferencias.
- P3. Necesidad de compañía de los mayores (soledad no deseada).

En cuanto al **contexto** y al uso de los AV, se han detectado los siguientes problemas:

- P4. Dificultad para recordar el funcionamiento del AV.
- P5. Complejidad en la recolección de información que necesita el AV sobre el usuario.
- P6. Dificultad de los usuarios para identificar la información clave que proporciona el AV.
- P7. Dificultad en que el AV comprenda al usuario.
- P8. Dificultad para que el usuario comprenda al AV.

En el caso de cómo influye el **diseño y las funcionalidades propias** de estos dispositivos se encuentran los problemas:

- P9. Dificultad para activar el AV.
- P10. Dificultad para comprender el funcionamiento del AV.
- P11. Desmotivación o falta de comprensión por tiempos de respuesta del AV excesivamente largos.
- P12. Dificultades con las interacciones multimodales en especial con las interfaces gráficas de algunos de estos dispositivos.

Y, por último, en el caso de los **aspectos sociales y de relación** con el AV aparecen los problemas:

- P13. Uso incorrecto de los protocolos sociales de comunicación.
- P14. Poca integración del AV en el contexto social del usuario.
- P15. Falta de adaptación de la voz del AV a las emociones.
- P16. Falta de confianza de los mayores con los AV.
- P17. Necesidad de personalización de los AV a los usuarios.

3.2. Guías de usabilidad

Durante el análisis de los trabajos de diversos autores, se han identificado una serie de guías de usabilidad aplicadas en el diseño de la interacción entre AV y adultos mayores.

Basándose en el trabajo de Koebel et al. [2], se ha optado por agrupar las guías de usabilidad en categorías. Mientras que dichos autores proponen cuatro categorías —Contenido y estructura, Lenguaje, Diseño de la interacción e Interfaz y Voz—, en este artículo se presentan cinco: **Aprendizaje (AP)**, **Contenido y estructura (CE)**, **Diseño de la interacción e Interfaz (DI)**, **Lenguaje (LE)** y **Personalización (PE)**. La principal diferencia radica en la incorporación de las categorías **Aprendizaje** y **Personalización**, así como en la eliminación de la categoría Voz, la cual se considera integrada dentro de la de **Personalización**.

Aprendizaje (AP) se refiere a las guías orientadas a enseñar al adulto mayor a interactuar con el AV, mostrando el potencial de la aplicación y fomentando el desarrollo de sus habilidades de uso. **Contenido y estructura (CE)** abarca las recomendaciones relacionadas con la organización lógica de la información, facilitando el acceso a las funcionalidades y evitando la sobrecarga cognitiva. **Diseño de la interacción e Interfaz (DI)** incluye pautas para garantizar una experiencia de usuario fluida e intuitiva, teniendo en cuenta no solo la interacción por voz sino también los elementos visuales, táctiles y de navegación. **Lenguaje (LE)** engloba aspectos relacionados con la sintaxis y el vocabulario utilizados por el AV, procurando un discurso claro, comprensible y adaptado a las capacidades lingüísticas de los adultos mayores. Finalmente, **Personalización (PE)** se refiere a la posibilidad de adaptar el aspecto, el género, la voz y otros parámetros del AV según las preferencias individuales del usuario, promoviendo una experiencia más cercana y confortable.

La Tabla 1 presenta, para cada guía de usabilidad, los problemas que atiende (según los identificados en la sección 3.1), el momento en el que se debe aplicar y la categoría que le corresponde (según las presentadas anteriormente).

3.3. Propuesta para la creación de nuevas guías de usabilidad para AV y adultos mayores

Una vez analizadas las recomendaciones de diseño existentes en la literatura científica para mejorar la usabilidad de los AV cuando son utilizados por adultos mayores, se ha detectado que existen ciertos aspectos en los que se debe seguir trabajando. Se han categorizado estos aspectos en torno a cuatro tipos de problemas: 1) los **tiempos de la conversación no se ajustan a las necesidades** del adulto mayor e impiden que se pueda realizar la comunicación de manera correcta; 2) el **flujo de la conversación no es adecuado** para el adulto mayor, de manera que no es capaz de comprender la información que está siendo transmitida por el AV; 3) el **tiempo para completar una tarea con el AV es demasiado extenso**, lo que hace que el adulto mayor sea incapaz de completarla; 4) la **adaptación que se realiza del AV es insuficiente**, de manera que la satisfacción del adulto mayor durante la interacción con el mismo no es tan alta como cabría esperar.

A continuación, se proponen una serie de recomendaciones para crear nuevas guías que complementen las recopiladas de la literatura (que aparecen resumidas en la Tabla 1), con el objetivo de mejorar la interacción de los adultos mayores con los AV. Estas recomendaciones constituyen la base preliminar para definir las guías definitivas, que se establecerán tras un trabajo experimental con adultos mayores.

3.3.1. Guías para el diseño de los tiempos de la conversación

Durante la interacción del adulto mayor con el AV es posible identificar varios tiempos que deben ser cuidadosamente establecidos para que la conversación se realice de manera efectiva. Por un lado está el tiempo que el AV espera a que el adulto mayor responda su petición, una vez formulada. A este tiempo se le denominará “**tiempo de espera del AV (T_{AV})**” y se produce cuando el AV inicia la interacción. Si transcurrido dicho tiempo, el adulto mayor no ha hablado, el AV debería repetir la petición y ofrecer ayuda si no es la primera repetición. El T_{AV} no debe ser demasiado largo porque el adulto mayor podría pensar que se ha acabado la conversación, que el AV se ha desconectado, etc. Pero el T_{AV} tampoco puede ser demasiado corto, porque el adulto mayor necesita suficiente tiempo para procesar la petición y ser capaz de elaborar su respuesta.

Por otro lado, está el tiempo que el AV tarda en responder después de una petición del adulto mayor. A este tiempo se le denominará “**tiempo de espera del adulto mayor (T_{AM})**” y, en este caso, será el adulto mayor el que inicia la interacción. Este tiempo debe ser cuidadosamente establecido para que el adulto mayor no crea que se ha interrumpido la conversación, lo que se produciría si se fija un T_{AM} demasiado largo. Pero, al mismo tiempo, el T_{AM} no puede ser demasiado corto, ya que esto podría hacer que el adulto mayor no estuviera preparado para comprender la respuesta del AV o que incluso no hubiera terminado de responder.

Tabla 1
Listado de guías de usabilidad

Guía de usabilidad	Problemas que atiende	Cuándo se usa	Categoría
Proporcionar un tutorial inicial para explicar el uso del AV. [16, 17, 18, 19]	P4, P6, P8, P9, P10, P12	En la primera interacción con el AV.	AP
Incluir una sección de ayuda accesible donde el AV pueda explicar su funcionamiento bajo demanda o de forma proactiva. [16]	P4, P6, P8, P10, P12	Cuando se detecte una posible dificultad o cuando el usuario lo solicite.	AP
Estructurar y explicar de forma automática por niveles la información o los diálogos. [2, 18]	P1, P2, P4, P5, P6, P7, P8, P10	En cada interacción del usuario.	CE
Explicar el motivo de cada solicitud de información al usuario. [18]	P5, P16	Cada vez que el AV requiere información del usuario.	CE
Proporcionar retroalimentación visual y/o auditiva para indicar que el AV está escuchando. [16, 19]	P9	En cada interacción del usuario.	DI
Implementar respuestas empáticas según el estado emocional del usuario. [17, 20, 21]	P3, P13, P14, P15, P16, P17	Durante las respuestas del AV.	DI
Aplicar escucha activa con frases como "Te escucho", "Dime más". [5, 19, 20]	P3, P13, P16	En cada interacción del usuario.	DI
Dotar al AV de una alta tolerancia a fallos para la interacción por voz. [2, 19]	P2, P4, P5, P6, P7, P8, P10	Cada vez que el AV recibe una respuesta que no sabe procesar.	DI
Proporcionar opciones multimodales para facilitar la comprensión. (Combinar texto y voz). [2]	P1, P2, P4, P8, P10, P12	Cuando el AV cuenta con interfaz gráfica.	DI
Usar fuentes grandes, botones accesibles y colores con alto contraste. [2]	P2, P6, P10, P12	Cuando el AV cuenta con interfaz gráfica.	DI
Permitir que el usuario interrumpa al AV en cualquier momento. [18]	P1, P8, P13	Durante las respuestas del AV.	DI
Optimizar la generación de diálogos para reducir demoras. [19]	P2, P11	Cada vez que se usa el AV.	DI
El AV debe reconocer su turno en la conversación para no interrumpir al usuario. [19]	P7, P8, P10, P13, P14	Cada vez que se usa el AV.	DI
Incluir variaciones de tono y efectos de sonido según el contexto emocional. [19]	P15	Cuando el contenido lo requiera.	DI
Ajustar la interacción del AV según si está solo con el usuario o en un entorno social. [20]	P7, P8, P13, P14	Según el contexto de uso.	DI
Utilizar vocabulario sencillo y estructuras gramaticales simples, evitando ambigüedades. [2, 19, 22]	P1, P2, P5, P6, P8	Durante las respuestas del AV.	LE
Hacer preguntas simples y secuenciales. [18]	P1, P2, P5, P7, P8, P10	Cada vez que se solicita información al usuario.	LE
Adaptar el nivel de guía e interacción según el perfil del usuario. [18, 20]	P2, P4, P5, P6, P7, P8, P10, P12, P17	Durante las respuestas del AV.	PE
Adaptar la personalidad del AV según el contexto y la relación previa con el usuario. [20]	P16, P17	A medida que el usuario interactúa con el AV.	PE
Preferencia por agentes femeninos frente a los masculinos. [23]	P2, P17	Cada vez que se usa el AV.	PE
Diseñar el AV considerando aspectos emocionales y de identificación del usuario a nivel cultural. [23]	P2, P17	Cada vez que se usa el AV.	PE

- Problema a resolver: Tiempos de conversación no ajustados al adulto mayor, generando problemas de comprensión en la interacción.
- Recomendación: Se deben ajustar los tiempos, T_{AV} y T_{AM} , haciendo pruebas con usuarios reales del colectivo al que va dirigida la aplicación que se está desarrollando.
- Categoría: Diseño de la interacción
- Cuándo se usa: En cada interacción.

3.3.2. Guías para el diseño del flujo de la conversación

Durante una conversación entre el AV y el adulto mayor se realizan una serie de comunicaciones que fluyen en dos sentidos: a) desde el adulto mayor al AV y b) desde el AV al adulto mayor. Aunque algunos investigadores [2, 18, 19, 22] ya han señalado la necesidad de simplificar las conversaciones, se considera que es necesario especificar qué aspectos concretos se deben simplificar.

Para ello, es útil basarse en las máximas conversacionales de Grice [24], que establecen principios clave para una comunicación efectiva: la máxima de cantidad (proporcionar la información necesaria, ni más ni menos), la máxima de calidad (decir la verdad y evitar información falsa), la máxima de relevancia (ser relevante) y la máxima de modalidad (ser claro, ordenado y sin ambigüedades).

Con esta intención, se identifican dos tipos de simplificación, según impliquen: 1) el **número de comunicaciones** (N_c) o 2) el **tamaño de cada comunicación** (L_{ci}). Estas simplificaciones se alinean principalmente con la máxima de cantidad, que implica ajustar la cantidad de información para no saturar ni dejar vacíos, y con la máxima de modalidad, para asegurar que cada comunicación sea clara y concisa.

De esta manera, la simplificación se realiza con dos niveles de abstracción. El nivel más alto requiere reducir el número de comunicaciones en cada conversación (minimizar N_c para lograr el objetivo de la conversación). Esta actuación pretende evitar saturar o aburrir al adulto mayor con una conversación innecesariamente larga, garantizando al mismo tiempo la relevancia de cada mensaje.

En el nivel más bajo de abstracción, se propone establecer un límite de palabras en cada comunicación (ci) para que la longitud de la comunicación (L_{ci}) no sea excesiva, favoreciendo así la máxima de modalidad para que cada mensaje sea claro, ordenado y fácil de procesar.

Esta propuesta es similar a todo el movimiento de lectura fácil que se enmarca dentro del Procesamiento de Lenguaje Natural (PLN) para simplificación de textos y mejora de la comprensión de la información [25].

- Problema a resolver: Flujo de la conversación no ajustado al adulto mayor, generando problemas de comprensión en la interacción.
- Recomendación: Se debe establecer un límite bajo para N_c y L_{ci} . Estos límites deberían ser fijados haciendo pruebas con usuarios reales.
- Categoría: Diseño de la interacción
- Cuándo se usa: En cada interacción.

3.3.3. Guías para establecer tareas atómicas alcanzables

Una **tarea (petición) atómica** (T_a) es el conjunto de comunicaciones con entidad propia que no puede interrumpirse hasta que la petición haya sido resuelta. Las tareas atómicas pueden agruparse para construir **tareas más complejas** (T_c). De manera que una tarea siempre se puede descomponer en una o varias tareas atómicas ($T_c = \Sigma T_a$).

El diseño de las tareas atómicas debe realizarse de manera similar a las conversaciones, minimizando el **número de comunicaciones** (N_c) necesarias para completar la tarea. Sin embargo, en este caso también es necesario considerar el **tiempo total** (T) requerido para completar la tarea. Mientras que en una conversación no habría mayor problema si ésta se interrumpe a mitad, en el caso de una tarea atómica, si no se realizan todas las comunicaciones necesarias, ésta quedaría incompleta. Esta situación implica que el adulto mayor no ha logrado cumplir con el objetivo asociado a la tarea, lo que podría generar desmotivación o incluso provocar sentimientos de frustración.

- Problema a resolver: El adulto mayor no logra completar una tarea propuesta por el AV y esto genera emociones negativas, pudiendo llegar a desmotivar al adulto mayor para seguir usando el AV.
- Recomendación: Se deben diseñar tareas atómicas de corta duración, que minimicen el número de comunicaciones (N_c) necesarias para completar la tarea con éxito. El diseño de las tareas atómicas debería realizarse siguiendo un enfoque de Diseño Centrado en el Usuario, para garantizar que la mayoría de adultos mayores sean capaces de terminar la tarea. Además, el AV debería permitir almacenar los progresos del adulto mayor en una tarea, de manera que si la sesión se interrumpe hayan quedado almacenadas las tareas atómicas que ya habían sido completadas. Esto permitiría que tareas complejas fueran realizadas a lo largo de diferentes sesiones.
- Categoría: Diseño de la interacción
- Cuando se usa: En cada interacción que implique una tarea.

3.3.4. Guías para el diseño de la adaptación

Varios investigadores han subrayado el incremento de usabilidad que supone permitir personalizar diferentes aspectos del AV, como son: el tono, el ritmo, la voz, etc. [19, 20, 23]. No obstante, existen otros parámetros que también son susceptibles de ser adaptados al adulto mayor y no aparecen en la literatura científica. Por ejemplo, los tiempos y el flujo de la conversación (T_{AV} , T_{AM} , N_c , L_{ci}) podrían ser ajustados para cada usuario concreto, así como el tiempo establecido como máximo para ejecutar las tareas atómicas. Esto es importante para atender la diversidad de capacidades cognitivas que presentan los adultos mayores.

Por otro lado, algunos aspectos identificados como positivos en la literatura científica, como puede ser que el AV sea amigable o afectivo, deberían poder ser configurables, ya que no todos los adultos mayores pueden sentirse cómodos con esta simulación de empatía o afecto.

Sin embargo, para ser capaces de realizar una adaptación completa del AV, es necesario que el hardware y el software del propio AV permitan configurar los parámetros necesarios. En este sentido, cabe plantear la necesidad de crear guías de usabilidad no solo para el diseño de aplicaciones para AV, sino también para el diseño de los propios AV.

- Problema a resolver: Cada adulto mayor es diferente y tiene capacidades diferentes.
- Recomendación: Permitir a cada usuario configurar parámetros relacionados tanto con la voz o la personalidad del AV, como con los tiempos de la interacción.
- Categoría: Personalización
- Cuando se usa: Al usar el AV por primera vez o cuando el usuario desee modificar los parámetros proporcionados para la personalización del AV.

4. Conclusiones y trabajo futuro

El uso de AV en la interacción con adultos mayores es ya una realidad que representa una oportunidad valiosa para reducir la brecha digital y contribuir a una mejora significativa en su calidad de vida. Estos sistemas pueden facilitar el acceso a la información, fomentar la autonomía y promover la inclusión social en entornos digitales.

Este trabajo reúne un conjunto de guías de usabilidad orientadas a mejorar la interacción entre adultos mayores y AV. Para ello, se han identificado los principales problemas de usabilidad que enfrenta este colectivo, se han revisado guías existentes y se han formulado nuevas recomendaciones que atienden a las necesidades detectadas.

En concreto, los problemas identificados se han clasificado en cuatro grandes áreas y las guías de usabilidad, por su parte, se han organizado en cinco categorías clave. Toda esta información ha sido sintetizada en una tabla que muestra la relación entre las guías encontradas, los problemas que abordan, su contexto de uso y la categoría correspondiente. Gracias a este análisis, se han podido detectar

carencias en las guías actuales y se han propuesto algunas recomendaciones para crear nuevas guías que complementen y enriquezcan el panorama existente.

Actualmente, se están desarrollando experiencias de co-diseño con adultos mayores con el objetivo de validar estas guías y seguir avanzando en la construcción de propuestas de usabilidad más inclusivas y ajustadas a sus necesidades reales.

Por último, este artículo abre la puerta a futuras investigaciones sobre cómo las guías de usabilidad orientadas a adultos mayores pueden beneficiar también a otros grupos de usuarios. Aunque el enfoque de este estudio ha sido facilitar la interacción para esta población, muchas recomendaciones podrían aplicarse en contextos más amplios, por lo que un trabajo futuro relevante será evaluar su adaptación a distintos perfiles de usuario.

Agradecimientos

Este artículo está cofinanciado por el proyecto PLEISAR, Ref. PID2022-136779OB, con fondos del Ministerio de Ciencia e Innovación del Gobierno de España (MCIN/AEI/ 10.13039/501100011033, UE); por el proyecto C-ING-179-UGR23 financiado por la Consejería de Universidades, Investigación e Innovación y por el Programa FEDER Andalucía 2021-2027; y por el Gobierno de Aragón, Departamento de Industria e Innovación, y Fondo Social Europeo “Construyendo Europa desde Aragón”.

Declaración sobre IA generativa

Los autores declaran que no han usado ninguna herramienta de IA generativa.

Referencias

- [1] Organización Mundial de la Salud (OMS), Envejecimiento y salud, 2024. URL: <https://www.who.int/es/news-room/fact-sheets/detail/ageing-and-health>.
- [2] K. Koebel, M. Lacayo, M. Murali, I. Tarnanas, A. Çöltekin, Expert insights for designing conversational user interfaces as virtual assistants and companions for older adults with cognitive impairments, in: Chatbot Research and Design, CONVERSATIONS 2021, Springer, Online, 2021, pp. 23–38. doi:10.1007/978-3-030-94890-0_2.
- [3] F. Portet, M. Vacher, C. Golanski, C. Roux, B. Meillon, Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects, *Personal Ubiquitous Comput.* 17 (2013) 127–144. doi:10.1007/s00779-011-0470-5.
- [4] M. Morris, B. Adair, K. Miller, E. Ozanne, R. Hampson, A. Pearce, N. Santamaria, L. Viegas, M. Long, C. Said, Smart-home technologies to assist older people to live well at home, *Journal of Aging Science* 1 (2013) 101–110. doi:10.4172/2329-8847.1000101.
- [5] J. Kowalski, A. Jaskulska, K. Skorupska, K. Abramczuk, C. Biele, W. Kopeć, K. Marasek, Older adults and voice interaction: A pilot study with Google Home, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, ACM, Glasgow, UK, 2019, p. 1–6. doi:10.1145/3290607.3312973.
- [6] A. L. Smith, B. S. Chaparro, Smartphone text input method performance, usability, and preference with younger and older adults, *Human factors* 57 (2015) 1015–1028. doi:10.1177/0018720815575644.
- [7] B. Knowles, V. L. Hanson, Older adults' deployment of 'distrust', *ACM Transactions on Computer-Human Interaction (TOCHI)* 25 (2018) 1–25. doi:10.1145/3196490.
- [8] A. Pradhan, A. Lazar, L. Findlater, Use of intelligent voice assistants by older adults with low technology use, *ACM Transactions on Computer-Human Interaction (TOCHI)* 27 (2020) 1–27. doi:10.1145/3373759.

- [9] S. S. Bassuk, T. A. Glass, L. F. Berkman, Social disengagement and incident cognitive decline in community-dwelling elderly persons, *Annals of internal medicine* 131 (1999) 165–173. doi:10.7326/0003-4819-131-3-199908030-00002.
- [10] M. Matsangidou, T. Solomou, F. Frangoudes, K. Ioannou, P. Theofanous, E. Papayianni, C. S. Pattichis, Affective out-world experience via virtual reality for older adults living with mild cognitive impairments or mild dementia, *International Journal of Environmental Research and Public Health* 20 (2023) 2919–2934. doi:10.3390/ijerph20042919.
- [11] J. Cassell, Embodied conversational interface agents, *Communications of the ACM* 43 (2000) 70–78. doi:10.1145/3706598.3713456.
- [12] L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner, T. Bickmore, Designing relational agents as long term social companions for older adults, in: *Intelligent Virtual Agents: 12th International Conference, IVA 2012*, Springer, Santa Cruz, Ca, 2012, pp. 289–302. doi:10.1007/978-3-642-33197-8_30.
- [13] A. Malhotra, J. Hoey, A. König, S. van Vuuren, A study of elderly people’s emotional understanding of prompts given by virtual humans., in: *10th EAI International Conference on Pervasive Computing Technologies for Healthcare, PERVASIVEHEALTH 16*, ACM, Cancun, MX, 2016, pp. 13–16. doi:10.4108/eai.16-5-2016.2263327.
- [14] L. Ring, L. Shi, K. Totzke, T. Bickmore, Social support agents for older adults: longitudinal affective computing in the home, *Journal on Multimodal User Interfaces* 9 (2015) 79–88. doi:10.1007/s12193-014-0157-0.
- [15] M. Kramer, R. Yaghoubzadeh, S. Kopp, K. Pitsch, A conversational virtual human as autonomous assistant for elderly and cognitively impaired users? Social acceptability and design considerations, in: *Informatik angepasst an Mensch, Organisation und Umwelt, INFORMATIK 2013*, Gesellschaft für Informatik eV, Koblenz, Ge, 2013, pp. 1105–1119. URL: <https://dl.gi.de/handle/20.500.12116/20541>.
- [16] S. Kim, et al., Exploring how older adults use a smart speaker-based voice assistant in their first interactions: Qualitative study, *JMIR mHealth and uHealth* 9 (2021) e20427. doi:10.2196/20427.
- [17] M. Liu, C. Wang, J. Hu, Older adults’ intention to use voice assistants: Usability and emotional needs, *Heliyon* 9 (2023) e21932. doi:10.1016/j.heliyon.2023.e21932.
- [18] A. Mahmood, S. Cao, M. Stiber, V. N. Antony, C.-M. Huang, Voice Assistants for health self-management: Designing for and with older adults, *arXiv preprint arXiv:2409.15488* (2024). doi:10.1145/3706598.3713839.
- [19] T. Gama, A. Ferreira, F. Oliveira, Guidelines for the development of common ground-based relational agents for elderly, in: *International Conference on Human-Computer Interaction, HCI 2022*, Springer, Online, 2022, pp. 345–365. doi:10.1007/978-3-031-17902-0_25.
- [20] B. Irfan, S. Kuoppamäki, G. Skantze, Recommendations for designing conversational companion robots with older adults through foundation models, *Frontiers in Robotics and AI* 11 (2024). doi:10.3389/frobt.2024.1363713.
- [21] V. K. Jones, M. Hanus, C. Yan, M. Y. Shade, J. Blaskewicz Boron, R. Maschieri Bicudo, Reducing loneliness among aging adults: the roles of personal voice assistants and anthropomorphic interactions, *Frontiers in public health* 9 (2021). doi:10.3389/fpubh.2021.750736.
- [22] S. Sayago, B. B. Neves, B. R. Cowan, Voice assistants and older people: some open issues, in: *Proceedings of the 1st International Conference on Conversational User Interfaces, CUI 2019*, ACM, Dublin, Ir, 2019, pp. 1–3. doi:10.1145/3342775.3342803.
- [23] A. Esposito, T. Amorese, M. Cuciniello, M. T. Riviello, A. M. Esposito, A. Troncone, M. I. Torres, S. Schlögl, G. Cordasco, Elder user’s attitude toward assistive virtual agents: the role of voice and gender, *Journal of Ambient Intelligence and Humanized Computing* 12 (2021) 4429–4436. doi:10.1007/s12652-019-01423-x.
- [24] H. P. Grice, Logic and conversation, in: P. Cole, J. L. Morgan (Eds.), *Syntax and semantics. Vol 3: Speech acts*, Academic Press, New York, NY, USA, 1975, pp. 41–58.
- [25] S. Bautista, R. Hervás, P. Gervás, J. Rojo, An approach to treat numerical information in the text simplification process, *Universal Access in the Information Society* 16 (2017) 85–102. URL: <https://hdl.handle.net/20.500.14352/24411>.

Automatización del protocolo Thinking Aloud en evaluaciones de usabilidad*

Shuoshuo Li^{1,†} and José A. Macías^{1,†}

¹Universidad Autónoma de Madrid, Escuela Politécnica Superior, Tomás y Valiente 11, 28049 Madrid, Spain

Resumen

Este artículo presenta una propuesta para automatizar el protocolo Thinking Aloud en evaluaciones de usabilidad con usuarios. Para ello, se ofrece una revisión sistemática de la literatura que permita ir más allá de los enfoques manuales habituales para buscar soluciones más automatizadas. Se analizan tecnologías actuales como el reconocimiento de audio y video, y se identifican vacíos en la automatización del protocolo a través del análisis multimodal. Finalmente, se proponen líneas de trabajo futuro orientadas al desarrollo de herramientas automáticas para mejorar la eficiencia de este método mediante las tecnologías indagadas.

Palabras clave

pensando en voz alta, pruebas de usabilidad, métodos de automatización, detección de emociones y sentimientos, extracción de conocimiento, análisis de audio y vídeo

1. Introducción

Las pruebas de usabilidad son un medio fundamental para evaluar la experiencia del usuario y la facilidad de uso de los sistemas interactivos. El protocolo Thinking Aloud (TA) se considera uno de los métodos más utilizado para realizar pruebas de usabilidad [1]. En particular, este método se ha adoptado ampliamente debido a su capacidad para registrar de manera intuitiva los procesos de pensamiento y los problemas que enfrentan los usuarios al ejecutar tareas [1]. Durante una prueba con TA, los usuarios expresan en voz alta sus pensamientos mientras realizan una tarea específica. Este enfoque de "pensar en voz alta mientras se actúa" ayuda a identificar problemas de usabilidad ocultos en el sistema y a comprender los patrones cognitivos de los usuarios [2].

Sin embargo, la mayoría de las pruebas tradicionales con TA se desarrollan manualmente, a través de anotaciones por parte de observadores. En algunos casos, se utiliza también un proceso de grabación, transcripción y análisis manual, lo que no solo es un proceso que consume mucho tiempo y esfuerzo, sino que también dificulta la eliminación de juicios subjetivos en el análisis [2]. A medida que crece la demanda de automatización en el análisis del método TA, los investigadores también han estado explorando activamente enfoques tecnológicos para la predicción de emociones y la extracción automática de características. En los últimos años, se han propuesto y aplicado diversos métodos basados en aprendizaje profundo para la extracción automática de características en la predicción del estado emocional a partir de señales de voz [29]. Sin embargo, algunos estudios han señalado que estos enfoques aún presentan ciertas limitaciones, como una excesiva concentración en el método concurrente, sin considerar en profundidad las diferencias en la aplicación de otros enfoques [27]. Con los avances en el reconocimiento automático de voz (ASR), procesamiento de lenguaje natural (NLP), aprendizaje profundo y análisis de video, la automatización del análisis de grabaciones y videos de usuarios se ha convertido en una dirección clave para poder mejorar la eficiencia y objetividad de las evaluaciones mediante TA [3]. En ese sentido, la investigación actual se centra principalmente en cómo aprovechar las tecnologías existentes para convertir de manera

* Interacción '25: XXV International Conference on Human-Computer Interaction, September 03-05, 2025, Valladolid, Spain

† Corresponding authors: J. A. Macías

† These authors contributed equally.

✉ shuoshuo.li@estudiante.uam.es (Shuoshuo Li); j.macias@uam.es (J. A. Macías)

ORCID 0000-0001-5071-0076 (José A. Macías)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

efectiva los datos de voz, texto y video de los usuarios en indicadores cuantificables de emociones y comportamiento, un problema que está recibiendo una creciente atención [3].

Este artículo tiene como objetivo explorar cómo utilizar TA para lograr una evaluación de usabilidad automatizada. A través de la integración de herramientas avanzadas de detección de emociones y extracción de conocimiento, se busca desarrollar un sistema de evaluación eficiente, objetivo y basado en datos. Esto proporcionará un sólido respaldo técnico para mejorar la experiencia del usuario y el diseño de interacción. Además, se analizará el estado actual de las tecnologías existentes en el procesamiento de datos de voz, texto y video, así como sus limitaciones [3].

Este artículo se estructura de la siguiente forma. En la Sección 1 introduce el contexto de la investigación y se discute la importancia del TA en los estudios de usabilidad. En la Sección 2 se presenta el método utilizado para la de Revisión Sistemática de la Literatura (SLR), así como las preguntas de investigación planteadas. En la Sección 3 se proporciona los resultados obtenidos y la discusión correspondiente para cada una de las preguntas de investigación. En la Sección 4 se analizan las posibles amenazas a la validez identificadas durante el desarrollo del estudio. Finalmente, en la Sección 5 se presenta las conclusiones y las líneas de continuación de la investigación.

2. Revisión de la literatura

Para investigar sobre las posibilidades de automatización del protocolo TA, se llevará a cabo una SLR que permita recopilar y analizar datos utilizando un método repetible y analítico [4]. Este enfoque es adecuado para situaciones que requieren un análisis preciso y es una vía apropiada para identificar y examinar la evidencia existente. La idea es realizar una integración exhaustiva de la literatura relacionada con la automatización de las pruebas de usabilidad basadas en TA, la detección de emociones y la extracción de conocimiento. Esto proporcionará una base teórica sólida para la selección de tecnologías y el diseño experimental en investigaciones futuras.

2.1. Método

En términos generales, la metodología adoptada puede variar entre diferentes disciplinas. El método empleado en este estudio se basa en una versión simplificada de una SLR [5], con una adaptación específica del marco PICOC (Población, Intervención, Comparación, Resultados y Contexto) a nuestro dominio de investigación [6]. Esto nos permitirá comprender de manera integral los últimos avances en detección de emociones, extracción de conocimiento y análisis de video dentro de la evaluación automatizada de usabilidad mediante TA. El método SLR no solo garantiza la repetibilidad y sistematicidad, sino que también permite una recopilación y análisis exhaustivo y riguroso de la literatura existente. A continuación, se presentarán las preguntas de investigación, la cadena de búsqueda, los criterios de inclusión y exclusión, y el cribado y la selección de artículos.

2.2. Preguntas de investigación

Se plantean las siguientes preguntas de investigación para guiar la SLR y analizar los resultados deseados, en base a los objetivos marcados en la sección anterior:

- RQ1: ¿Qué propuestas existen para el análisis de emociones, principalmente a partir de grabaciones con usuarios, analizando video y audio?
- RQ2: ¿Qué métodos específicos existen para implementar el protocolo Thinking Aloud en estudios de usabilidad, considerando aquellos que puedan ser implementados de manera automática a través de métodos de extracción de conocimiento y de emociones?
- RQ3: ¿Qué tecnologías y herramientas de apoyo existen para la detección de emociones a partir de videos grabados con usuarios?

2.3. Selección de palabras clave y fuentes bibliográficas

Para poder encontrar literatura que respalde y responda las preguntas anteriores, se han seleccionado una serie de palabras clave que compondrán la cadena de búsqueda bibliográfica:

Situación Experimental y Entorno:	Thinking Aloud
Tratamientos:	AND (Design Thinking OR Usability OR User Experience OR UX OR User-Centered Design OR User-Centred Design)
Variables de Respuesta:	AND (Automatic OR Voice OR Video OR Speech Recognition OR Conversational OR Natural Language Processing OR Intelligent OR Sentiment Analysis OR Transcription)

Por otro lado, las bases de datos utilizadas para la búsqueda fueron las siguientes:

- (1) IEEE Xplore
- (2) ACM Digital Library
- (3) Scopus
- (4) Springer
- (5) Google Scholar

2.4. Criterios de selección de artículos

Para filtrar los artículos extraídos de las cinco bases de datos mediante la cadena de búsqueda se definieron los siguientes criterios:

- Criterios de inclusión: Los artículos seleccionados deben estar relacionados con evaluaciones de usabilidad que utilicen TA, e incluir procesos de automatización a través de la detección de emociones o la extracción de conocimiento. También, los artículos seleccionados deben estar en inglés o español, y deben haber sido publicados después del 2010 para asegurar que son trabajos recientes.
- Criterios de exclusión: Se descartarán aquellos artículos que no estén relacionados directamente con la temática definida, aquellos que sean antiguos, con contenido en curso, demasiado corto o poco desarrollado, y aquellos que se encuentren repetidos.

2.5. Cribado y selección de artículos

Inicialmente, se obtuvo una alta cantidad de artículos que posteriormente fueron filtrados mediante los criterios de inclusión y exclusión. En la Tabla 1 se muestra un resumen de estos procesamientos iniciales. Para aplicar los criterios de inclusión y exclusión se utilizaron los filtros automáticos que proporcionan las bases de datos. Además, se analizaron los resúmenes, palabras clave y contenido general de los artículos, descartando aquellos que no cumplieran con los criterios establecidos. Como podemos ver en la Tabla 1, la cantidad de artículos se redujo significativamente, obteniendo al final un subconjunto más reducido y relevante de artículos.

Tabla 1

Resultados de la búsqueda de artículos

Base de datos	Número inicial de artículos encontrados	Número de artículos después de aplicar los criterios de exclusión
(1) IEEE Xplore	382	25
(2) ACM Digital Library	1240	44
(3) Scopus	980	39
(4) Springer	1650	60
(5) Google Scholar	32000	95

Finalmente, fueron 15 los artículos considerados como primarios, es decir, aquellos relevantes y directamente relacionados con el tema de investigación propuesto, por lo que se utilizarán principalmente para la contestar a las diferentes preguntas de investigación planteadas.

Además, para ampliar el alcance de la revisión de la literatura, se utilizó el método de la *bola de nieve*. Para ello, tras seleccionar los 15 artículos primarios, se revisaron sus referencias bibliográficas, encontrando otros 6 artículos adicionales considerados como relevantes para la investigación, lo que aumentó el número de artículos a 21. Estos artículos adicionales enriquecieron la investigación, proporcionando más argumentaciones e información de interés para elaborar este artículo.

Por último, se seleccionaron 14 documentos más, a través de búsquedas menos estructuradas en Google. Estos documentos, denominados literatura gris, tratan sobre temas técnicos que son de interés para enriquecer la investigación. Esto permitió incrementar en 35 el número total de artículos de interés para el trabajo de investigación planteado.

Los artículos encontrados permitieron abordar cada una de las preguntas de investigación, aportando bibliografía básica para poner en contexto cada una de ellas. Por un lado, para abordar la RQ1 se encontramos artículos que discuten principalmente los métodos de reconocimiento de emociones en video y audio, así como el análisis multimodal de emociones. Para la RQ2 se encontraron artículos que exploran cómo utilizar tecnologías de reconocimiento automático de voz, procesamiento de lenguaje natural y aprendizaje profundo. Por otro lado, para la RQ3, se encontraron artículos que presentan principalmente tecnologías de reconocimiento facial, detección de emociones en tiempo real y herramientas de código abierto relacionadas.

3. Resultados y discusión

La revisión en profundidad de la literatura arrojó un resultado alentador para la investigación, y es que, hasta el momento, no se han encontrado soluciones ni propuestas concretas completamente automatizadas para la evaluación de la usabilidad utilizando TA a través de tecnologías de reconocimiento de audio y video para la captura y análisis del pensamiento del usuario. A pesar de los avances en la automatización de algunos aspectos, como el análisis de emociones a partir de audio y video, todavía no existe un sistema integrado que cubra las necesidades del TA de una manera completamente automatizada. Por lo tanto, este campo sigue siendo un área de investigación poco explorada, lo que subraya

la importancia de este estudio y su contribución para avanzar hacia una automatización más completa de las pruebas de usabilidad.

A continuación, se analiza cada una de las preguntas de investigación, discutiendo los trabajos encontrados y las soluciones más relevantes aportadas por la literatura.

(RQ1) ¿Qué propuestas existen para el análisis de emociones, principalmente a partir de grabaciones con usuarios, analizando video y audio?

Durante mucho tiempo, en la investigación sobre usabilidad ha habido una gran variedad de métodos utilizados para el análisis de emociones basados en grabaciones de usuarios (audio y video). Actualmente, se emplean principalmente los siguientes métodos:

1. Reconocimiento de emociones unimodales en voz: Los investigadores analizan las señales de voz y video de los usuarios, extrayendo características como el volumen, ritmo, energía y tono para detectar emociones. Klaus R. Scherer fue uno de los primeros académicos en investigar la relación entre las características de las señales de voz y la expresión emocional en 1986 [22]. En 2010, Rafael A. Calvo y Sidney D'Mello discutieron las teorías psicológicas de las emociones y revisaron los métodos tradicionales de detección de emociones (como la fisiología, el análisis facial y el análisis de voz), así como los sistemas multimodales emergentes [7]. Propusieron métodos de análisis de emociones basados en aprendizaje automático y modelos estadísticos. En 2011, El Ayadi y otros señalaron que las señales de voz son una forma rápida y efectiva de interacción hombre-máquina, enfocándose en el reconocimiento de emociones en la voz (SER) en aplicaciones como la traducción y asistentes inteligentes, y resumieron las técnicas más relevantes, como la extracción de características y los métodos de clasificación [10]. También se identificaron algunas observaciones sobre los patrones de comportamiento de los usuarios en diferentes condiciones. Por ejemplo, una investigación señala que los analistas emplearon diversos métodos para identificar problemas de usabilidad. En el modo de “trabajo en silencio”, los usuarios tendieron a depender más de la búsqueda en el sistema, mientras que en el modo de “pensar en voz alta”, este método mostró un carácter más activo y orientado a objetivos [28]. Este hallazgo ofrece una perspectiva valiosa para comprender las características cognitivas del método TA.
2. Métodos de reconocimiento de emociones multimodales: Comenzó en los años 90 y, con el desarrollo de la tecnología, la combinación de voz y video ha mejorado la precisión del reconocimiento de emociones. En 1995, la investigación de la profesora Rosalind Picard sentó las bases para este enfoque [9]. En 2017, Poria y otros, junto con Scherer y Ellgring, señalaron que la combinación de voz y video permite un reconocimiento de emociones más efectivo [11]. Por otro lado, en 2011, Mohammad Soleymani y otros investigaron la base de datos multimodal MAHNOB-HCI, que se utiliza para la investigación de la detección de emociones y etiquetado [23]. Realizaron experimentos utilizando datos de video, movimientos oculares, audio y señales fisiológicas, y demostraron que los métodos multimodales son más precisos que los unimodales. A lo largo del tiempo, los investigadores han comparado diferentes métodos de reconocimiento de emociones. En 2011, Schuller y otros resumieron los avances en el reconocimiento de emociones a partir de la voz [8]. En 2017, Scherer y Ellgring estudiaron la expresión multimodal de las emociones y descubrieron que las emociones están influenciadas por el contexto [11]. Estos métodos, combinados con técnicas de aprendizaje profundo y aprendizaje automático, han mejorado la precisión del reconocimiento.

(RQ2) ¿Qué métodos específicos existen para implementar el protocolo Thinking Aloud en estudios de usabilidad, considerando aquellos que puedan ser

implementados de manera automática a través de métodos de extracción de conocimiento y de emociones?

Para automatizar el TA, se pueden utilizar cinco métodos, cada uno basado en diferentes tecnologías que convierten la voz en texto y luego usan tecnologías avanzadas para realizar el análisis de emociones y la extracción de conocimiento, permitiendo el análisis automático de los informes verbales de los usuarios:

1. Transcripción automática de voz: Utiliza la tecnología de reconocimiento automático de voz (ASR) para convertir las descripciones verbales en texto. Un estudio preliminar realizado por Nguyen y otros (2017) respalda esta idea, al demostrar que la transcripción de protocolos de pensamiento en voz alta mediante ASR es viable y puede integrarse de manera efectiva en estudios de usabilidad [14]. En 2012, Geoffrey Hinton y otros señalaron que las redes neuronales profundas (DNN) son más eficaces que los métodos tradicionales, mejorando la precisión del reconocimiento de voz, y mencionaron las experiencias de cuatro equipos de investigación exitosos [12].
2. Comprensión del texto y análisis de emociones: Después de la transcripción automática de voz, el texto se introduce en un sistema de procesamiento de lenguaje natural (NLP), utilizando modelos de lenguaje pre-entrenados como BERT para el análisis. En 2019, Jacob Devlin y otros señalaron que BERT es un modelo de lenguaje basado en Transformer, que aprende representaciones del lenguaje mediante un entrenamiento bidireccional profundo, y es adecuado para diversas tareas de NLP [13]. En comparación con los modelos tradicionales, BERT mostró mejor rendimiento en varias tareas, pudiendo identificar eficazmente las inclinaciones emocionales de los usuarios, apoyando la detección de emociones. Además de estos modelos avanzados, Liu, Li y Wang (2016) propusieron un enfoque basado en procesamiento de lenguaje natural que puede ser usado para el análisis automatizado del protocolo TA, demostrando que es posible identificar patrones problemáticos en los informes verbales de los usuarios de manera automática [15].
3. Extracción de palabras clave y descubrimiento de conocimiento: El uso de tecnologías de aprendizaje profundo permite extraer palabras clave de manera rápida y precisa. En 2019, Wang y otros propusieron un método automático de extracción de palabras clave basado en aprendizaje profundo, que soporta la extracción de palabras clave de las transcripciones, lo que facilita la posterior extracción de conocimiento y el análisis de emociones [13]. En 2018 desarrollaron un enfoque específico, que podría ser utilizado para automatizar el TA, orientado a reducir la carga manual y mejorar la precisión del proceso [16].
4. Modelado de contexto y manejo de textos largos: En 2017, Vaswani y otros escribieron que Transformer puede utilizarse para manejar textos largos y modelar el contexto [24], lo que mejora la comprensión del proceso de pensamiento del usuario y la precisión en la detección de emociones.
5. Explicación del modelo y validación de resultados: En 2016, Ribeiro propuso el modelo LIME, que ayuda a automatizar el proceso de extracción de características emocionales y de conocimiento [25]. Este método utiliza redes neuronales profundas para la transcripción de voz, BERT para el análisis de emociones, aprendizaje profundo para la extracción de palabras clave, y Transformer para el procesamiento del contexto, mejorando así la precisión y la eficiencia de los datos.

Por otro lado, los métodos multimodales pueden ser también de utilidad para en análisis de gestos en el protocolo TA. Si bien la mayor parte del conocimiento se puede extraer a través de la voz, también pueden ser de interés las emociones expresadas por el usuario a través de gestos o expresiones faciales. En se sentido, la combinación de voz y video ha mejorado la precisión del reconocimiento de emociones, y prueba de ellos son las investigaciones que afirman que la combinación de voz y video permite un reconocimiento de emociones más efectivo [11], a través de

trabajos donde se realizaron experimentos utilizando datos de video, movimientos oculares, audio y señales fisiológicas, demostrando que los métodos multimodales son más precisos que los unimodales [23]. Otros trabajos descubrieron que las emociones están influenciadas por el contexto [11], y la combinación de técnicas de aprendizaje profundo y aprendizaje automático pueden ser la solución para un reconocimiento automático de emociones más completo en la evaluación automática de la usabilidad a través de TA.

Adicionalmente, en los últimos años, el desarrollo de las tecnologías de reconocimiento de emociones ha dado un nuevo impulso a las investigaciones en este ámbito. Por ejemplo, algunos estudios han demostrado que los métodos de reconocimiento emocional basados en expresiones faciales pueden alcanzar una tasa de precisión superior al 90 % en condiciones en tiempo real, lo que evidencia su viabilidad y eficacia en aplicaciones prácticas [30]. Además, la comunidad académica ha llevado a cabo revisiones sistemáticas sobre la aplicación de la computación afectiva en el reconocimiento de emociones psicológicas, lo cual aporta un sólido respaldo teórico para la comprensión de la percepción emocional en la interacción humano-computadora [31]. Otros trabajos han desarrollado programas de reconocimiento de emociones faciales basados en realidad virtual, dirigidos específicamente a la evaluación y tratamiento asistido de personas con esquizofrenia [32]. Al mismo tiempo, las tecnologías multimodales de reconocimiento emocional sin contacto también están evolucionando rápidamente, y la literatura existente ha explorado sus múltiples aplicaciones, desafíos técnicos, soluciones propuestas y perspectivas futuras [33].

De la misma forma, en los últimos años, los investigadores también han comenzado a explorar diferentes enfoques tecnológicos, como los métodos de reconocimiento de emociones en el habla basados en aprendizaje profundo [34], así como el uso de la realidad aumentada para la detección de emociones a partir de expresiones faciales [35]. Estos avances enriquecen aún más el contexto tecnológico en el que se enmarca este artículo.

(RQ3) ¿Qué tecnologías y herramientas de apoyo existen para la detección de emociones a partir de videos grabados con usuarios?

Las tecnologías actuales utilizadas para analizar videos de usuarios y realizar la detección de emociones se centran principalmente en el reconocimiento de expresiones faciales en los videos. Para ello, se lleva a cabo la extracción de características faciales de los fotogramas del video para analizar el estado emocional. Las tecnologías y herramientas principales se dividen en los siguientes cuatro enfoques:

1. Métodos de reconocimiento de emociones en video basados en aprendizaje profundo: En 2017, Li y otros propusieron la base de datos RAF-DB y optimizaron el aprendizaje de características mediante el modelo DLP-CNN, demostrando que este modelo supera a los métodos existentes [17]. Ese mismo año, Mollahosseini y otros crearon la base de datos AffectNet, que recopila más de un millón de imágenes faciales, descubriendo que los métodos de Redes Neuronales Profundas (DNN) son más efectivos que los métodos tradicionales. En una investigación previa, los mismos autores habían evidenciado que el uso de arquitecturas DNN más profundas podía mejorar significativamente la precisión en el reconocimiento de expresiones faciales, sentando así una base técnica sólida para sistemas de análisis emocional en video [18].
2. Tecnologías de detección de emociones mediante el reconocimiento en tiempo real de expresiones faciales: En 2017, Shuster, M., y otros investigaron la tecnología de reconocimiento de expresiones faciales en tiempo real. Esta tecnología permite capturar rápidamente los cambios emocionales de los usuarios durante su interacción, asegurando la precisión y la inmediatez de la detección de emociones [19]. En 2019, C. Jiang, Y. Qiu, H. Gao y otros propusieron una plataforma de retroalimentación para usuarios basada en computación de borde. Utilizando tecnología de aprendizaje profundo y video de alta velocidad, esta plataforma captura y analiza en tiempo real las expresiones faciales de los

usuarios para proporcionar retroalimentación emocional. Esto podría ayudar a los diseñadores a optimizar dinámicamente la interfaz de usuario y mejorar los resultados de las pruebas de usabilidad [20]. En general, estas tecnologías podrían ser adecuadas para pruebas de usabilidad en línea y otros escenarios altamente interactivos.

3. Herramientas y plataformas de código abierto: El paquete OpenFace, desarrollado en 2018 por T. Baltrušaitis y otros, es una herramienta de análisis facial de código abierto que puede detectar puntos clave faciales, postura de la cabeza, unidades de acción y seguimiento ocular, y soporta procesamiento en tiempo real. OpenFace tiene una alta precisión y puede ejecutarse en cámaras comunes, siendo aplicable en campos como la interacción humano-computadora, la computación emocional y el análisis médico [21].
4. Explicación del modelo y mejoras: En 2017, Selvaraju y otros propusieron el método Grad-CAM, que genera imágenes visuales utilizando información de gradientes, ayudando a entender las áreas clave en el reconocimiento de expresiones faciales [26]. Las tecnologías actuales, basadas en visión por computadora y aprendizaje profundo, pueden detectar en tiempo real el estado emocional del usuario, proporcionando soporte para las pruebas de usabilidad.

En general, todas estas tecnologías y herramientas parecen prometedoras para su utilización dentro de la interacción en tiempo real con el usuario, con el objetivo de poder detectar información con precisión a través de técnicas de aprendizaje profundo. Esto permitiría el automatizar en análisis de resultados en pruebas de usabilidad basadas en TA. A través de la integración de herramientas avanzadas de detección de emociones y extracción de conocimiento, se podría desarrollar un sistema de evaluación eficiente, objetivo y basado en datos. Esto proporcionará un sólido respaldo técnico para mejorar la experiencia del usuario y el diseño de interacción [36], aportando además una solución automatizada de calidad al proceso [39].

4. Amenazas a la validez

Con respecto a las amenazas a la validez [40] relacionadas con el estudio presentado, el sesgo de publicación y selección, junto con la arbitrariedad de la cadena de búsqueda, son las principales amenazas internas a tener en cuenta, mientras que la generalizabilidad puede entenderse como la amenaza externa a considerar en este contexto.

Por un lado, el sesgo de publicación se refiere a la tendencia de que los resultados positivos se publiquen con mayor frecuencia que los negativos, lo que afecta los resultados de la búsqueda de literatura. Para reducir este sesgo, se utilizaron múltiples bases de datos (como IEEE Xplore, ACM Digital Library, etc.) y motores de búsqueda (como Google Scholar), además de seleccionar algunos documentos considerados como *literatura gris*, asegurando que se pudiera realizar una búsqueda completa de investigaciones relacionadas con la automatización del protocolo TA.

Por otro lado, el sesgo de selección se refiere a la inclusión de literatura no relevante debido a criterios de selección demasiado amplios o no estrictos. Para evitar este sesgo, se establecieron criterios claros de inclusión: los artículos deben abordar temas relacionados con la automatización del protocolo TA, la detección de emociones y la extracción de conocimiento, y deben estar escritos en español o inglés, con una fecha de publicación posterior a 2010. Al mismo tiempo, se excluyeron estrictamente aquellos artículos que no se ajustaran al tema, que estuvieran repetidos o que tuvieran un contenido demasiado breve, asegurando así que los artículos seleccionados fueran altamente relevantes para el tema de la investigación.

En lo que se refiere a la arbitrariedad de la cadena de búsqueda, ciertamente el diseño de la misma tiene un impacto importante en los resultados obtenidos. Para reducir los problemas de omisión o redundancia causados por combinaciones inadecuadas de palabras clave, estas fueron seleccionadas por los dos autores del artículo, asegurando así que la expresión de búsqueda pudiera capturar con precisión todos los artículos clave relacionados con la automatización del protocolo TA a través de la detección de emociones.

Finalmente está el problema de la generalizabilidad, es decir, la posibilidad de que los resultados de la investigación no sean aplicables a un campo más amplio. Para reducir este problema se utilizó el método la *bola de nieve*, aumentando así el número de trabajos relacionados y asegurando que los artículos seleccionados representaran mejor el tema tratado.

Aunque podrían seguir existiendo amenazas a la validez, las acciones llevadas a cabo permiten minimizarlas, proporcionaron una base sólida de literatura para el estudio y análisis de la automatización del protocolo TA a partir de métodos modernos de aprendizaje y extracción automatizada de conocimiento y emociones, que es el objeto de estudio.

5. Conclusión y trabajo futuro

El protocolo TA es uno de los más utilizados en las pruebas de evaluación de la usabilidad con usuarios, ya que permite, de una forma expresiva, conocer las sensaciones de este mientras interactúa con la aplicación a evaluar, lo que facilita obtener información de primera mano a los evaluadores. La mayoría de las evaluaciones tradicionales mediante TA se desarrollan manualmente, a través de anotaciones que toman los observadores. En algunos casos, se utiliza también un proceso de grabación para su posterior procesamiento, o de transcripción y análisis manual, lo que consume tiempo y esfuerzo.

Con los avances en el reconocimiento automático de voz, procesamiento de lenguaje natural, aprendizaje profundo y análisis de video, la automatización del análisis de grabaciones y videos de usuarios es una solución interesante para mejorar la eficiencia y objetividad de las evaluaciones mediante TA, lo que permitiría un proceso más sistemático de análisis y obtención automática de resultados, logrando pautas para integrar de manera satisfactoria la inteligencia artificial en la interacción con el usuario [37, 38].

Según la literatura existente, no existen todavía propuestas concretas en ese sentido, pero sí hay tecnologías y herramientas que ayudarían a implementar soluciones que permitirían la automatización del protocolo TA y su posterior análisis, especialmente a través del análisis de emociones y de la extracción automática de conocimiento. Esto incluye el reconocimiento automático de voz mediante redes neuronales profundas, que ha permitido mejorar la precisión de la transcripción de voz, haciendo el proceso más eficiente. Por otro lado, el análisis de emociones y procesamiento de lenguaje natural tiene cabida a través de herramientas (como BERT) que permiten analizar el texto transcrito e identificar automáticamente las emociones del usuario. Además, el reconocimiento de emociones multimodales, que permite combinar datos de voz y video, permite capturar con mayor precisión el estado emocional del usuario. Todo ello, sumado al reconocimiento de expresiones faciales mediante aprendizaje profundo, permitiría analizar en tiempo real las expresiones faciales en los videos para asistir en el análisis emocional.

Como líneas futuras para continuar esta investigación, se propone la construcción y testeo de una herramienta que permita el análisis automático de evaluaciones de usabilidad a través de TA, no como método único, sino para la toma de decisiones, de forma que los resultados puedan ser comparados con la percepción de otros observadores humanos. Esto permitiría cubrir las carencias existentes en este ámbito y desarrollar un sistema automatizado completamente integrado en tiempo real que combine datos multimodales y proporcione retroalimentación en tiempo real en evaluaciones de usabilidad basadas en TA. Para ello, es necesario mejorar la integración entre módulos y herramientas y optimizar la capacidad de procesamiento en tiempo real, así como fortalecer la interpretación y validación de modelos. Sin duda, la utilización de estos métodos, técnicas y herramientas permitirían aumentar la eficiencia, reducir la intervención humana y mejorar la objetividad. Sin embargo, existen algunos retos a mejorar, como la posible falta de precisión en entornos ruidosos y las dificultades técnicas en el análisis emocional en tiempo real cuando se procesan grandes volúmenes de datos. Aunque se han logrado avances significativos, aún falta un sistema completamente integrado para la automatización de la evaluación de usabilidad a través del protocolo Thinking Aloud, lo que deja margen para futuras investigaciones.

Agradecimientos

Esta investigación ha sido subvencionada a través de los proyectos de investigación TED2021-129381B-C21 y PID2021-122270OB-I00, ambos de la Agencia Estatal de Investigación.

Declaración de IA Generativa

Los autores no han utilizado ninguna herramienta de IA Generativa.

Referencias

- [1] T. Boren, J. Ramey, Thinking Aloud: reconciling theory and practice, *IEEE Transactions on Professional Communication* 43 (2000) 261–278. doi:10.1109/47.867942.
- [2] M. Hertzum, K. D. Holmegaard, Thinking Aloud influences perceived time, *Hum Factors* 57 (2015) 101–109. doi:10.1177/0018720814549709.
- [3] B. Pang, L. Lee, Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval* 2 (2008) 1–135. URL: <https://doi.org/10.1561/15000000011>. doi:10.1561/15000000011.
- [4] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering - a systematic literature review, *Information and Software Technology* 51 (2009) 7–15. URL: <https://doi.org/10.1016/j.infsof.2008.09.009>. doi:10.1016/j.infsof.2008.09.009.
- [5] B. Kitchenham, S. Charters, Guidelines for performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE 2007, EBSE, 2007. URL: <https://www.durham.ac.uk/media/durham-university/departments-/computer-science/research/technical-reports/Guidelines-for-Performing-Systematic-Literature-Reviews-in-Software-Engineering-2007.pdf>, technical Report.
- [6] A. G. Padua, Propuesta de un proceso de revisión sistemática de experimentos en ingeniería del software, in: *CIbSE 2010 - Actas - 13th Conferencia Iberoamericana en Software Engineering*, E.T.S. de Ingenieros Informáticos (UPM), 2010, pp. 313–318.
- [7] R.A. Calvo, S. D'Mello, Affect detection: an interdisciplinary review of models, methods, and their applications, *IEEE Transactions on Affective Computing* 1.1 (2010): 18–37. doi:10.1109/T-AFFC.2010.7.
- [8] B. Schuller, A. Batliner, S. Steidl, D. Seppi, Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge, *Speech Communication* 53 (2011) 1062–1087. doi:10.1016/j.specom.2011.01.011.
- [9] S. Poria, E. Cambria, R. Bajpai, A. Hussain, A review of affect analysis: from unimodal analysis to multimodal fusión, *Information Fusion*, 37 (2017) 98–125. doi:10.1016/j.inffus.2017.02.003.
- [10] M. El Ayadi, M. Kamel, F. Karray, Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern Recognition* 44.3 (2011): 572–587. doi:10.1016/j.patcog.2010.09.020.
- [11] K. R. Scherer, H. Ellgring, Multimodal expression of emotion: affect programs or componential appraisal patterns? *Emotion* 7 (2007) 158–171. doi:10.1037/1528-3542.7.1.158.
- [12] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, B. Kingsbury, Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups, *IEEE Signal Processing Magazine*, 29.6 (2012): 82–97. doi:10.1109/MSP.2012.2205597.
- [13] Y. Sun, H. Qiu, Y. Zheng, Z. Wang, C. Zhang, SIFRank: A New Baseline for Unsupervised Keyphrase Extraction Based on Pre-Trained Language Model, *IEEE Access* 8 (2020) 10896–10906. doi:10.1109/ACCESS.2020.2965087.
- [14] K. Kuhn, V. Kersken, B. Reuter, N. Egger, G. Zimmermann, Measuring the Accuracy of Automatic Speech Recognition Solutions, *ACM Trans. Access. Comput.* 16 (2024) Article 25, 23 pages. doi:10.1145/3636513.

- [15] J. Zhang, C. Borchers, V. Alevan, R. S. Baker, Using Large Language Models to Detect Self-Regulated Learning in Think-Aloud Protocols, in: Proceedings of the 17th International Conference on Educational Data Mining, International Educational Data Mining Society, 2024, pp. 157-168. doi:10.5281/zenodo.12729790.
- [16] S. Li, X. Huang, T. Wang, et al., Using text mining and machine learning to predict reasoning activities from think-aloud transcripts in computer assisted learning, *J Comput High Educ* 37 (2025) 477–496. doi:10.1007/s12528-024-09404-6.
- [17] S. Li, W. Deng, Deep Facial Expression Recognition: A Survey, *IEEE Transactions on Affective Computing* 13 (2022) 1195-1215. doi:10.1109/TAFFC.2020.2981446.
- [18] A. Mollahosseini, B. Hasani, M.H. Mahoor, Going deeper in facial expression recognition using deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2016), 2016, pp. 1–6.
- [19] M. S. Bartlett, G. Littlewort, I. Fasel, J. R. Movellan, Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction, in: 2003 Conference on Computer Vision and Pattern Recognition Workshop, Madison, WI, USA, 2003, pp. 53-53. doi:10.1109/CVPRW.2003.10057.
- [20] C. Jiang, Y. Qiu, H. Gao, T. Fan, K. Li, J. Wan, An edge computing platform for intelligent operational monitoring in internet data centers, *IEEE Access* 7 (2019) 133375-133387. doi: 10.1109/ACCESS.2019.2939614.
- [21] T. Baltrušaitis, P. Robinson, L P. Morency, OpenFace: an open source facial behavior analysis toolkit, in IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 2016, pp. 1-10, doi: 10.1109/WACV.2016.7477553.
- [22] K.R. Scherer, Vocal communication of emotion: a review of research paradigms, *Speech Communication* 5.1-2 (1986) 1-49. doi:10.1016/0167-6393(86)90070-X.
- [23] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, A multimodal database for affect recognition and implicit tagging, in *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, Jan.-March 2012. doi: 10.1109/T-AFFC.2011.25.
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems* (2017) 5998–6008.
- [25] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you?": explaining the predictions of any classifier, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 1135–1144. doi:10.1145/2939672.2939778.
- [26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626. doi:10.1109/ICCV.2017.74.
- [27] S. McDonald, H. M. Edwards, T. Zhao, Exploring Think-Alouds in usability Testing: an international survey, *IEEE Transactions on Professional Communication*, 55.2 2012: 2-19. <https://doi.org/10.1109/TPC.2011.2182569>.
- [28] S. McDonald, G. Cockton, A. Irons, The impact of Thinking-Aloud on usability inspection, in: *Proceedings of the ACM on Human-Computer Interaction*, 4, 2020, 1-22. doi:10.1145/3397876.
- [29] R. Jahangir, Y. Teh, H. Wah, F. Faiqa, G. Mujtaba, Deep learning approaches for speech emotion recognition: state of the art and research challenges, *Multimedia Tools and Applications* 80.16 2021: 23745-23812. doi: 10.1007/s11042-020-09874-7.
- [30] F. Abdat, C. Maaoui, A. Pruski, Human-computer interaction using emotion recognition from facial expresión, in: *UKSim 5th European Symposium on Computer Modeling and Simulation*, Madrid, Spain, 2011, pp. 196-201. doi: 10.1109/EMS.2011.20.
- [31] V. S. Bakkialakshmi, T. Sudalaimuthu, A survey on affective computing for psychological emotion recognition, in: *5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECOT)*, Mysuru, India, 2021, pp. 480-486. doi: 10.1109/ICEECOT52851.2021.9707947.

- [32] T. Souto, H. Silva, A. Leite, A. Baptista, C. Queirós, A. Marques, Facial emotion recognition: virtual reality program for facial emotion recognition—a trial program targeted at individuals with schizophrenia, *Rehabilitation Counseling Bulletin*, 63.2 (2019): 79-90. doi: 10.1177/0034355219847284.
- [33] U.A. Khan, Q. Xu, Y. Liu, A. Lagstedt, A. Alamäki J. Kauttonen, Exploring contactless techniques in multimodal emotion recognition: insights into diverse applications, challenges, solutions, and prospects, *Multimedia Systems* 30.115 (2024). doi: 10.1007/s00530-024-01302-2.
- [34] A. Bhavan, M. Sharma, M. Piplani, P. Chauhan, S. Hitkul, R.R. Shah, Deep learning approaches for speech emotion recognition, in: Agarwal, B., Nayak, R., Mittal, N., Patnaik, S. (eds) *Deep Learning-Based Approaches for Sentiment Analysis. Algorithms for Intelligent Systems*. Springer, Singapore, 2020, doi: 10.1007/978-981-15-1216-2_10.
- [35] V. Bhardwaj, A. Joshi, G. Bajaj, V. Sharma, A. Rushiya, S. S. Bharghavi, Emotion detection from facial expressions using augmented reality, in: *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2023, pp. 1-5. doi: 10.1109/ICIRCA57980.2023.10220824.
- [36] L.A. Rojas, J.A. Macías, An agile information-architecture-driven approach for the development of user-centered interactive software, in: *Proceedings of the XVI International Conference on Human Computer Interaction*. Article No.: 50, 2015, pp 1-8. doi: 10.1145/2829875.2829919.
- [37] J.A. Macías, Intelligent assistance in authoring dynamically generated web interfaces, *World Wide Web* 11 (2008): 253-286. doi:10.1007/s11280-008-0043-3.
- [38] J.A. Macías, Enhancing interaction design on the semantic web: A case study, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.6 (2012): 1365-1373. doi: 10.1109/TSMCC.2012.2187052.
- [39] C. Quintal, J.A. Macías, Measuring and improving the quality of development processes based on usability and accessibility, *Universal Access in the Information Society* 20.2 (2021): 203-221. doi: 10.1007/s10209-020-00726-7.
- [40] L.A. Rojas, J.A. Macías, Toward collisions produced in requirements rankings: A qualitative approach and experimental study, *Journal of Systems and Software* 158 (2019): 110417. doi: 10.1016/j.jss.2019.110417.

Desarrollo de un módulo conversacional basado en IA para la interacción emocional con robots sociales^{*}

Development of an AI-based conversational module for emotional interaction with social robots

Liany Mendoza^{1,*†}, Eva Cerezo^{2,*†}, Adrián Arribas^{1,†} and Sandra Baldassarri^{2,†}

¹ Escuela de Ingeniería y Arquitectura, Universidad de Zaragoza, C/Maria de Luna 1, 50018 Zaragoza, España

² Instituto de Investigación en Ingeniería de Aragón (I3A), C/Mariano Esquillor s/n, 50018, Zaragoza, España

Resumen

Para lograr una interacción más personalizada y emocional entre el usuario y el robot, se ha desarrollado un sistema que combina interfaces gráficas, movimientos, luces y voz, respaldadas por inteligencia artificial y APIs externas, con el fin de enriquecer la experiencia del usuario. El sistema fue desarrollado en Android, utilizando el SDK del robot Sanbot Elf para el manejo de sus componentes físicos y expresiones faciales. El corazón de dicho sistema lo conforma el Módulo Conversacional que utiliza un modelo de lenguaje grande (LLM) para orquestar el diálogo inteligente con el usuario y el Módulo Emocional que traduce estados afectivos en manifestaciones físicas (expresiones faciales, movimientos físicos y patrones de iluminación LED). Esta sinergia permite que cada respuesta verbal vaya acompañada de expresiones emocionales mejorando la coherencia y la fluidez de la interacción. Una primera evaluación llevada a cabo con usuarios ha permitido valorar el funcionamiento del sistema, su usabilidad, aceptación y el impacto de la personalización y los aspectos emocionales en la interacción. Los resultados obtenidos destacan el potencial de esta solución para soportar la interacción multimodal persona-robot, si bien ponen de manifiesto la necesidad de manejar adecuadamente la complejidad de las interfaces de configuración y el aumento de los tiempos de procesamiento.

Abstract

To achieve a more personalized and emotional interaction between the user and the robot, a system has been developed that combines graphic interfaces, movements, lights and voice, supported by artificial intelligence and external APIs, in order to enrich the user experience. The system was developed in Android, using the SDK of the Sanbot Elf robot to manage its physical components and facial expressions. The heart of the system is the Conversational Module that uses a large language model (LLM) to orchestrate the intelligent dialogue with the user and the Emotional Module that translates affective states into physical manifestations (facial expressions, physical movements and LED lighting patterns). This synergy allows each verbal response to be accompanied by emotional expressions, improving the coherence and fluidity of the interaction. A first evaluation carried out with users has made it possible to assess the functioning of the system, its usability, acceptance and the impact of personalization and emotional aspects in the interaction. The results obtained highlight the potential of this solution to support multimodal human-robot interaction, although they highlight the need to adequately manage the complexity of the configuration interfaces and the increase in processing times.

Palabras clave

robots sociales, computación afectiva, LLM

1. Introducción y estado del arte

Con los avances en Inteligencia Artificial (IA) e Interacción Persona-Robot (HRI), los sistemas robóticos inteligentes se van a ir haciendo poco a poco omnipresentes en la vida humana. Más allá

^{*}Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

^{1*} Corresponding author.

[†] These authors contributed equally.

✉ lmendoza@unizar.es (L. Mendoza); ecerezo@unizar.es (E. Cerezo); 795593@unizar.es (A. Arribas); sandra@unizar.es (S. Baldassarri)

 0009-0008-5391-6270 (L. Mendoza); 0000-0003-4424-0770 (E. Cerezo); 0000-0002-9315-6391 (S. Baldassarri)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

de la asistencia en tareas industriales que requieren gran precisión y exactitud, los robots sociales se van a ir convirtiendo en parte integrante de nuestra vida cotidiana en forma de asistentes, tutores e incluso compañeros [1].

Si bien las definiciones de lo que es un robot social son heterogéneas [2], Sarrica et al [3] identificaron en su estudio algunos rasgos comunes: los robots sociales son agentes físicamente (corpóreos) que tienen cierta autonomía (o total) y participan en interacciones sociales con humanos, comunicándose, cooperando y tomando decisiones. La componente social es esencial para que los robots proporcionen apoyo físico y emocional a las personas, participando y manteniendo interacciones a largo plazo con ellos [4].

Aunque aún se está lejos de que los robots sociales ofrezcan las mismas oportunidades que los humanos para las interacciones sociales [5], ellos pueden ofrecer valiosas oportunidades cuando se introducen en contextos específicos de forma cuidadosa y éticamente responsable [6], contribuyendo a mejorar la calidad de vida de las personas [7]. Una base de pruebas cada vez mayor documenta cómo los robots sociales podrían funcionar como herramientas autónomas de apoyo a las intervenciones cognitivas [8], fisioterapia y salud física [9-11], intervenciones psicosociales y *well-being* [12, 13], el entretenimiento y la educación [14,15].

Precisamente en el contexto educativo se ha observado [14] que los estudiantes esperan que los robots educativos puedan entender y expresar emociones como alegría, tristeza, miedo y vergüenza. Esto incluye la capacidad de mostrar emociones a través de expresiones faciales, tono de voz y movimientos de manos, también prefieren que los robots tengan una personalidad distintiva y que puedan interactuar de manera similar a un amigo humano, incluyendo la capacidad de hablar y reaccionar al tacto. Así, en la interacción entre robots y niños se ha visto [15] que las expresiones faciales de felicidad y enojo de los robots influyen en la percepción de confianza, simpatía y competencia por parte de los niños durante los primeros encuentros.

Pero esto no se da solo en contextos educativos, según un estudio de De Graaf, Allouch y Van Dijk [16], la capacidad de un robot para mantener un diálogo, analizar el estado de ánimo de su usuario y expresar emociones son las características claves necesarias para que un robot social sea aceptado. También es muy importante que la emoción expresada por un robot esté más orientada hacia los demás que hacia sí mismo: la empatía [17] constituye un aspecto clave en la interacción entre humanos y robots [18,19], ya que favorece una relación más cómoda y natural. Para lograr estos objetivos, es fundamental el estudio e incorporación de componentes emocionales [20], dado que los robots que pueden manifestar emociones mediante expresiones faciales o gestos tienden a ser percibidos como más cercanos y familiares. Además, la capacidad expresiva del robot facilita la comprensión de su comportamiento, estado emocional, motivación y razonamiento [21], lo que, a su vez, contribuye a que sea visto como un agente confiable y empático.

Por tanto, aunque es esencial mejorar la funcionalidad de los robots sociales, también es fundamental que el robot sea capaz de comunicarse de una forma que se considere socialmente aceptable [16]. En una interacción comunicativa, el receptor suele participar activamente a través de señales no verbales, como el contacto visual, los gestos y otros indicios que evidencian su atención hacia el emisor. Esta combinación de elementos verbales y no verbales [22] contribuye a la fluidez y naturalidad en la comunicación entre humanos. Por esta razón, en el diseño de la interacción humano-robot resulta esencial la implementación de técnicas multimodales, que integren diversos canales de comunicación.

Por otro lado, los modelos lingüísticos grandes (LLM) son excelentes para mantener una conversación fluida debido a su capacidad para comprender el contexto e interpretar comandos de lenguaje natural [23]. En este sentido, se están configurando como una herramienta fundamental de soporte a la conversación humano-robot [24,25]. Sin embargo, como se ha visto, no basta con

que el robot sea capaz de sostener una conversación verbal, sino que también debe emplear mecanismos de comunicación no verbal, tales como la proximidad, el contacto visual y la expresión facial [26]. Además, se espera que los robots sociales sean sensibles a las diferencias individuales (debidas a la cultura, la edad, el sexo o la personalidad, entre otros factores), ofreciendo una experiencia de interacción natural y atractiva personalizada para cada usuario [27,28].

En este trabajo se presenta un módulo conversacional que hace uso de un modelo LLM (gpt-4o-mini) y de las posibilidades de interacción multimodal de un robot (Sanbot Elf) que soporte la interacción natural, emocional y personalizada humano-robot. En los siguientes apartados se detallan las características del robot, de las interfaces y del módulo conversacional desarrollado, una primera evaluación llevada a cabo con usuarios, así como las conclusiones obtenidas y el trabajo futuro.

2. Interacción conversacional emocional humano-robot

2.1. El robot Sanbot Elf

El robot Sanbot Elf, ofrece una serie de funcionalidades claves para el desarrollo del módulo conversacional, aunque también presenta algunas limitaciones. Entre sus características destacadas se encuentra el control del habla, que incluye funciones de *Text to Speech* (TTS) para convertir texto en audio con ajustes de velocidad y entonación, y *Speech to Text* (STT) para reconocer y transformar el habla del usuario en texto, facilitando la interpretación del lenguaje natural. En cuanto al control del movimiento, Sanbot Elf (mide 90,2 cm) puede mover su cabeza, brazos y ruedas omnidireccionales, características fundamentales para expresar emociones mediante gestos y movimientos, así como controlar LEDs de color en la cabeza, brazos y base (ver Figura 1).

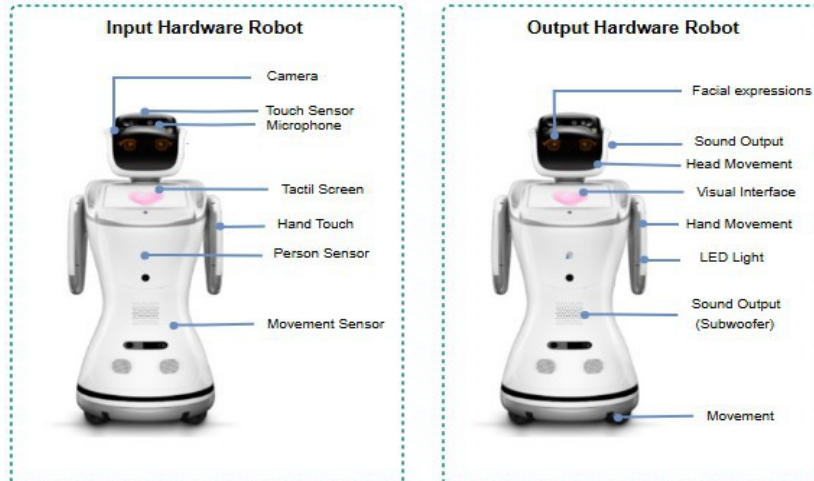


Figura 1: Componentes físicos del robot Sanbot Elf: entradas (izquierda) y salidas (derecha).

Por otro lado, el control del audio permite ajustar el volumen de los altavoces y gestionar el micrófono, optimizando la interacción verbal. Además de la interacción verbal, Sanbot permite la interacción táctil a través de una pantalla de 10.1 pulgadas que posee en su frontal y dispone de un gran número de sensores de tacto en distintas partes de su cuerpo (ver Figura 1), además de dos cámaras que abren la posibilidad a reconocer expresiones faciales y de rostros. Las mayores limitaciones vienen de su movilidad, ya que no dispone de módulos de navegación.

El robot también dispone de una pequeña pantalla, a modo de cara, que permite mostrar 18 expresiones faciales diferentes (ver Figura 2). En cuanto al software, los fabricantes implementaron

un SDK, llamado QihanOpenSDK², el cual fue utilizado para el desarrollo de la herramienta conversacional que se presenta en este documento.

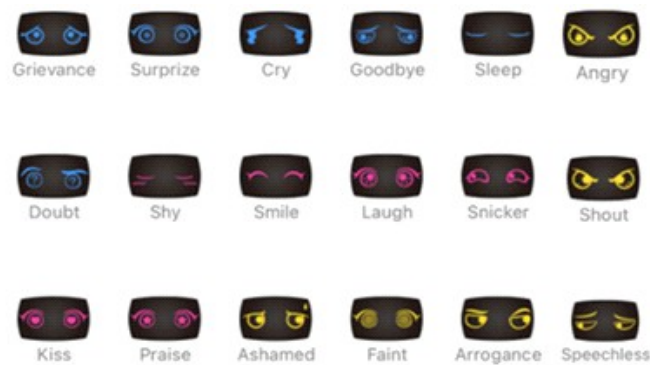


Figura 2: Expresiones faciales del robot Sanbot Elf

2.2. Arquitectura

En la Figura 3 se muestra la arquitectura a alto nivel del sistema desarrollado, que consta de tres bloques. El primer bloque corresponde a las Interfaces de Usuario, donde se incluyen pantallas para la interacción con el robot, como la pantalla conversacional, la de configuración y personalización. El segundo bloque es el Módulo Conversacional, que constituye el núcleo de la interacción y está compuesto por dos elementos clave: el Módulo Emocional, encargado de gestionar las emociones del robot, asegurando una respuesta más natural y expresiva; y el Módulo de Diálogo, responsable de la gestión de la conversación, construyendo los mensajes intercambiados entre el usuario y el robot. El tercer bloque engloba componentes externos utilizados en la implementación, como el Módulo de IA, encargado del procesamiento del lenguaje natural mediante reconocimiento de voz y generación de respuestas, las APIs Externas, que permiten realizar consultas y acceder a funciones adicionales; y el Control del Robot, que gestiona hardware como el movimiento de extremidades, el habla y el sistema en general.

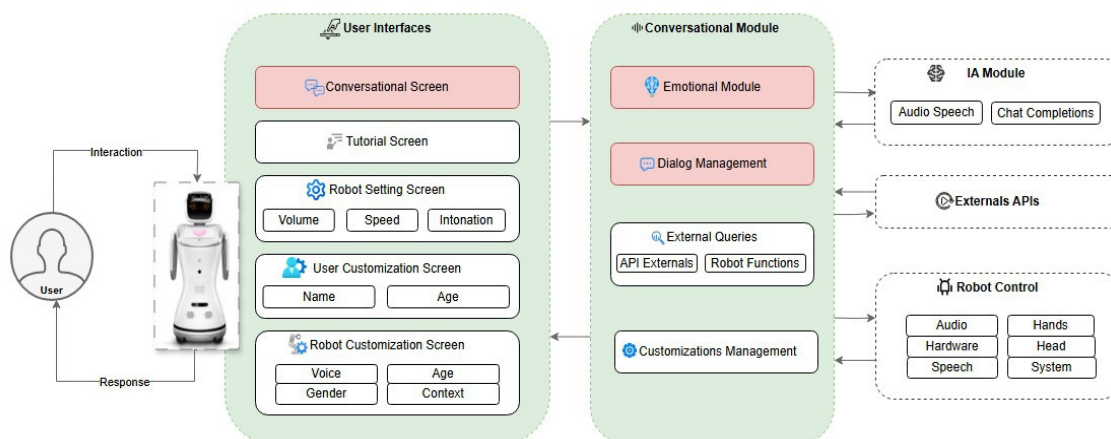


Figura 3: Arquitectura del sistema conversacional desarrollado (sombreados en color rojo los componentes claves de la herramienta).

A continuación, se comentan los componentes principales.

1. Interfaces de Usuario

Consta de tres pantallas principales (ver Figura 4):

²<http://en.sanbot.com/support/>

- Pantalla Conversacional (*Conversational Screen*): Es la interfaz principal donde el usuario interactúa y visualiza la conversación mantenida con el robot. El usuario puede iniciar conversaciones, repetir lo que el robot ha dicho o detener su habla.
- Pantalla Personalización del Usuario (*User Customization Screen*): El usuario puede introducir información personal como nombre y edad, datos que se utilizarán durante la conversación para personalizar la interacción, mencionar que esta pantalla es opcional ya que el robot comienza la conversación preguntando dichos datos en caso de que no hayan sido introducidos anteriormente.
- Pantalla Personalización del Robot (*Robot Customization Screen*): Permite al usuario ajustar características del robot: el tipo de voz, género simulado y grupo de edad, lo que influye en cómo el robot interactúa con el usuario.

Además, existen otras dos pantallas, del Tutorial (*Tutorial Screen*) y la de ajuste generales del Robot (*Robot Setting Screen*) que permite al usuario modificar parámetros como el volumen, la velocidad y la entonación de la voz del robot.



Figura 4: Pantallas: (1) Personalización del usuario (2) Conversación (3) Personalización del robot

2. Módulo Conversacional

Destacan dos componentes:

- Gestor de Diálogo (*Dialogue Management*): Gestiona la interacción del chat entre el usuario y el modelo LLM (*Large Language Model*).
- Módulo Emocional (*Emotional Module*): Módulo que gestiona las emociones del robot y las interacciones emocionales con el usuario. A partir de las emociones detectadas por la IA y, en el futuro, de otros posibles módulos (de reconocimiento emocional facial, reconocimiento emocional en voz, etc.) gestiona la salida emocional del robot. Para ello utiliza controles del robot (manos, cabeza, hardware y sistema) para expresar emociones a través de movimientos, luces LED y expresiones faciales (ver Figura 2). El módulo emocional se encarga de hacer el mapeo entre las emociones detectadas en el discurso del usuario y las posibles salidas del robot. Como el modelo emocional usado es el de Plutchik [29] ello implica un trabajo de agrupación de las emociones para su mapeo al catálogo de expresiones faciales disponibles. En relación con los colores LED del robot, se tomó como referencia el código de colores mencionado por Son, E. [30], referenciando la película de Disney *Inside Out*. Se consideró que este esquema cromático sería fácilmente reconocible para muchas personas, facilitando así la interpretación emocional del robot.

Además, el gestor de Consultas Externas (*External Queries*) permite consultar APIs externas y el gestor de Personalizaciones (*Customizations Managements*), que gestiona todos los parámetros de personalización del usuario y el robot.

3. Integración con otros componentes

- **Módulo de Inteligencia Artificial (*IA Module*):** Procesa las solicitudes del usuario a través de peticiones HTTP, utilizando un modelo LLM para entender y responder adecuadamente. El módulo integra dos componentes principales relacionados con los *endpoints* de la API de OpenAI, el primero es *ChatCompletions*, que permite interactuar con modelos de lenguaje basados en inteligencia artificial, para generar respuestas en conversaciones estructuradas. El segundo componente es *AudioSpeech*, que permite la conversión de texto a voz y de voz a texto. Cada una con diferente tonalidad que, subjetivamente, se pueden considerar de un género concreto para poder realizar diversas pruebas. El modelo emocional usado para la detección del estado emocional del usuario es el de Plutchik [29]. Para ello, se solicita a *ChatCompletions* que analice y categorice cada entrada del usuario según ese modelo y de una respuesta coherente en función de la emoción detectada intentando empatizar lo máximo posible con el usuario: no buscando imitarlo sino animarlo si está triste, calmarlo si está enfadado, etc. En el prompt se solicita también que en la respuesta se tenga también en cuenta la edad y el nombre de la persona.
- **Control del Robot (*Robot Control*):** Utiliza las funcionalidades del SDK para gestionar acciones físicas como el movimiento de extremidades y la gestión del habla, así como el módulo de voz propio del robot.

2.3. Ejemplo de interacción conversacional

Una vez presentada la arquitectura y explicados los distintos módulos, se mostrará en este apartado un ejemplo de flujo conversacional entre la persona y el robot. El proceso comienza cuando el usuario inicia una conversación, ya sea mediante voz o a través de la interfaz visual (pantalla conversacional) utilizando el teclado:

Entrada del usuario: "¡Hola! Me llamo Lucia y tengo 5 años"

El módulo conversacional recibe esta entrada, y se la envía a OpenAI mediante peticiones HTTP, generando la posible respuesta:

Respuesta del robot: [(1)/(24)] ¡Hola Lucia! Es muy bonito conocerte. ¿Qué te gustaría hacer en tu día libre? Me encantaría saber más sobre lo que te divierte. <Lucia,5>

Como se puede observar, la respuesta está compuesta por tres partes:

1. **Interpretación emocional:** Los números entre corchetes, como (1)/(24), representan emociones. El primer número (1) corresponde a la emoción del usuario, que en este caso sería "Alegría". El segundo número (24) es la emoción que el robot debe transmitir en su respuesta, en este caso sería "Interés". La emoción del usuario se utiliza para adaptar la conversación en el módulo conversacional, mientras que la emoción del robot se procesa en el módulo emocional para definir cómo debe expresarse físicamente.
2. **Respuesta textual:** Es el texto generado y verbalizado por el modelo, que en este caso es: "¡Hola Lucia! Es muy bonito conocerte. ¿Qué te gustaría hacer en tu día libre? Me encantaría saber más sobre lo que te divierte."

3. Datos del usuario: En este apartado se incluye información personal proporcionada por el usuario, como su nombre y edad, en este caso: <Lucia,5>.

El Módulo Emocional transforma el dato sobre la emoción que debe transmitir el robot en acciones de los actuadores del robot, es decir en la expresión física de la emoción. Eso incluye expresión facial (ver Figura 2) y expresión corporal, en la forma de movimientos de brazos y cabeza, y activación de luces LED de diferentes colores.

Es el Módulo Conversacional el que combina la respuesta textual con las respuestas físicas para comunicarse con el usuario. En el ejemplo dado, el robot respondería, mientras muestra una expresión facial de interés y realiza movimientos que refuerzan esta emoción. Este flujo conversacional permite que el robot no solo responda de manera coherente, sino que también exprese emociones de manera natural.

Los datos del usuario se utilizan para personalizar la conversación, ya que en momentos de inactividad el robot cada cierto tiempo le pregunta al usuario si desea continuar la conversación, llamándole por su nombre. Además, es importante manejar la edad como un dato esencial para que las respuestas del robot estén enfocadas al grupo generacional del usuario.

3. Evaluación con usuarios

El objetivo fundamental de la evaluación era comprobar el correcto funcionamiento del módulo, valorar su usabilidad, aceptación y el impacto en la interacción de los aspectos emocionales y de personalización. Asimismo, se consultó sobre el tipo de voz a usar poniendo especial énfasis en la consideración de la variable de género. En la evaluación participaron 18 usuarios en total, 7 hombres y 11 mujeres con edades entre 18 y 83 años. Las sesiones se llevaron a cabo en el laboratorio del grupo de investigación, a lo largo de 6 días, con una duración por sesión de aproximadamente 30 minutos con la siguiente estructura:

- cuestionario inicial de caracterización: utilizado para conocer los datos sobre el usuario participante, como su edad, capacidades sensoriales al utilizar dispositivos electrónicos o frecuencia de uso de la tecnología.
- prueba de voces: se mostraban varias voces a las personas, cinco de las voces eran sacadas del endpoint Audio Speech de OpenAI y la voz restante era la propia del robot Sanbot; las voces eran tres masculinas y tres femeninas y se preguntaba al usuario por su preferencia.
- conversación con el robot: Con el objetivo de medir el impacto de la componente emocional, se utilizó la técnica Testing A/B que consiste en dividir a los usuarios en dos grupos, cada grupo evaluaría una versión del sistema desarrollado [31]: el grupo A interactuó con el sistema que tenía capacidad de realizar consultas a OpenAI y que contaba con sus voces, además incluía el análisis emocional del usuario y la expresión emocional del robot mediante gestos, expresión facial o luces LED, mientras que el grupo B interactuó con el sistema que únicamente tenía la capacidad de conversar con el robot, pero con la voz original del robot Sanbot sin análisis emocional ni expresión emocional. El usuario mantenía dos conversaciones con el robot, la primera con un tema libre que elegía el usuario y la segunda con un tema cerrado que elegía el evaluador. El objeto de esa segunda conversación era explotar la capacidad de crear conversaciones sobre situaciones –ficticias– en las que se pudieran encontrar usuario y robot (safari, viaje espacial, descubrimiento de un planeta...).
- contestación de dos cuestionarios: el POST VAVA-Q [32], cuestionario generado en el grupo de investigación para evaluar la aceptación tecnológica (ver Tabla 1) y un

cuestionario de usabilidad con siete preguntas extraídas del SUS [33] y nueve preguntas específicas sobre el funcionamiento del módulo conversacional (ver Tabla 2).

Tabla 1

Cuestionario POST VAVA-Q. Respuestas del tipo Likert

Preguntas del cuestionario	Respuestas
Creo que con los recursos y conocimientos que tengo, seguir las instrucciones para usar el robot Sanbot Elf, me ha sido...	(1-Fácil a 5-Difícil)
Si se me propusiera otra vez, tengo la intención de volver a utilizar el robot Sanbot Elf en los próximos días.	(1-Falso a 5-Verdadero)
Valorando lo que me puede aportar utilizar el robot Sanbot Elf, creo que es una idea...	(1-Inútil a 5-Útil)
La posibilidad de contarles a las personas importantes para mí que estoy utilizando el robot Sanbot Elf, es una idea que...	(1-Me gusta a 5-No me gusta)
Al utilizar el robot Sanbot Elf, resolver los retos que se han presentado ha sido...	(1-Fácil a 5-Difícil)
Quiero volver a utilizar el robot Sanbot Elf en los próximos días.	(1-Falso a 5-Verdadero)
Utilizar el robot Sanbot Elf, ha sido una idea...	(1-Aburrida a 5-Divertida)
Si las personas que son importantes para mí supieran que estoy utilizando el robot Sanbot Elf, yo creo que...	(1-Les gustaría a 5-No les gustaría)
Yo creo que si me lo proponen, encontraré el tiempo para utilizar de nuevo el robot Sanbot Elf en los próximos días.	(1-Falso a 5-Verdadero)

Tabla 2

Cuestionario de Usabilidad. Respuestas del tipo Likert (1-Totalmente en desacuerdo a 5-Totalmente de acuerdo)

Preguntas del cuestionario	
Preguntas SUS	Creo que me gustaría utilizar este robot con frecuencia. Creo que necesitaría el apoyo de otra persona para poder utilizar este robot. Me imagino que la mayoría de la gente aprendería a utilizar este robot muy rápidamente. Encontré el robot innecesariamente complicado de usar. Me sentí muy seguro usando el robot. Encuentro el uso del robot raro. Siento que tengo que aprender muchas cosas antes de

		empezar a usar el robot.
Preguntas	interacción	He entendido bien lo que me decía el robot.
voz		Me pareció que al robot le costaba entender lo que le decía.
		Considero que la aparición de la respuesta en la pantalla me ha ayudado durante la conversación.
		Considero que la conversación que he tenido con el robot ha seguido un hilo coherente.
		Considero que el robot ha entendido correctamente mis consultas y ha sabido darles una respuesta con sentido.
		Considero que la voz del robot es agradable.
		Considero que la voz del robot suena artificial.
		Considero que la voz del robot es fácil de entender.
		Considero que la respuesta del robot es rápida.

3.1. Resultados

Durante la prueba de **Preferencia de Voces** se pudo comprobar que el 72.22% de las personas eligieron una voz que coincidía en género con el propio, aunque un mayor número de hombres (el 57% de ellos) y de mujeres (el 91% de ellas) eligieron una voz femenina.

En cuanto a la **Usabilidad (Preguntas SUS)** los resultados fueron muy satisfactorios, pudiendo destacar las puntuaciones a preguntas como el uso frecuente del robot o la poca complicación del módulo conversacional (ver Figura 6).

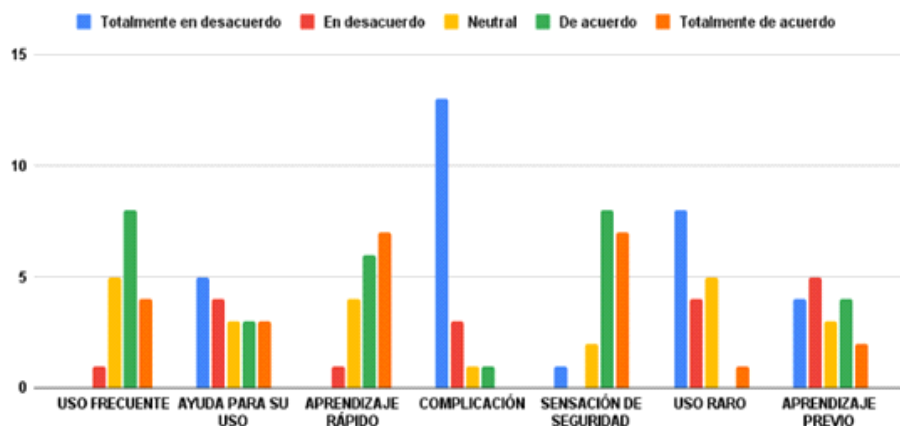


Figura 6: Diagrama resultados generales del test usabilidad preguntas SUS.

Si se hace un **análisis de género**, se puede observar en la Figura 7 que existen algunas diferencias: las mujeres tienen más interés en usar el robot, mientras que los hombres califican en mayor medida el uso del robot como raro, pero valoran su aprendizaje a la hora de usarlo como más rápido.

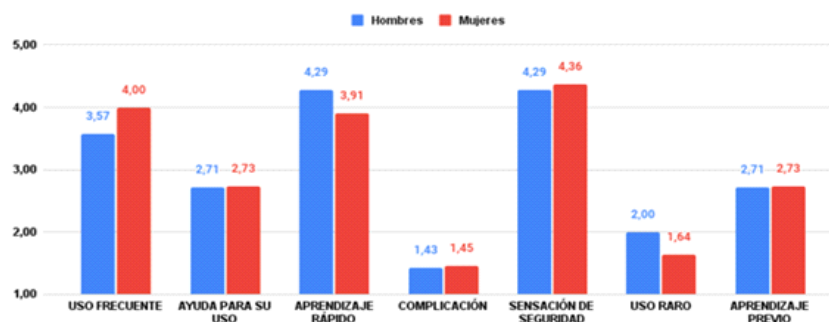


Figura 7: Diagrama resultados test usabilidad preguntas SUS según género.

En cuanto al **análisis por grupo (A/B)**, a pesar de que no existan diferencias significativas, el grupo A considera que necesitarían más ayuda para utilizar el robot, que su uso es más complicado y raro y que necesitan aprender muchas cosas antes de iniciar el módulo conversacional. Se cree que esto puede ser debido a la necesidad de pasar por las pantallas de personalización tanto del usuario como del robot antes de poder llevar a cabo la conversación con el robot. Ello pone de manifiesto la necesidad de mantener las interfaces sencillas al aumentar la complejidad del procesamiento.

También se analizó la **Usabilidad (Preguntas interacción voz)**, donde el resultado del análisis general arrojó que están muy satisfechos por la comprensión y entendimiento de la voz, tanto por parte del usuario como del robot, la coherencia de las conversaciones, la respuesta rápida del módulo o la aparición de la respuesta en pantalla (ver Figura 8). El grupo B ha comprendido mejor las respuestas del robot y consideran que la voz es más entendible, probablemente debido a los problemas de pronunciación de las voces OpenAI (nativamente inglesas), aunque opinaron que el robot entendía peor las consultas debido a que en ese grupo no se aplicaba ninguna personalización.

Por último, el grupo A consideró las respuestas del robot más lentas, seguramente debido a que la petición a OpenAI contenía más datos de tipo emocional, personalización o similares.

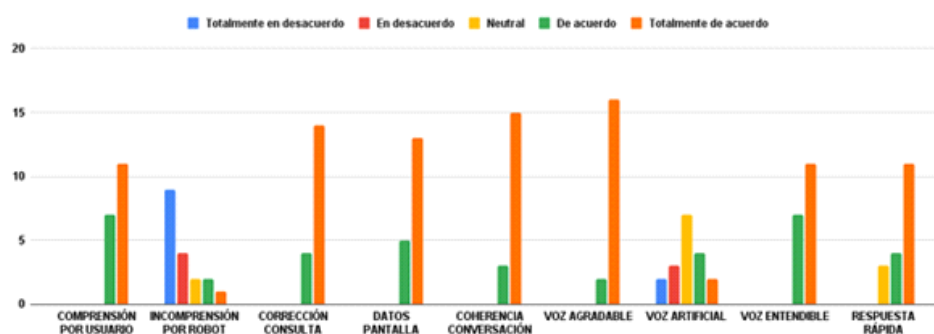


Figura 8: Resultados generales del test de usabilidad, preguntas de interacción por voz.

Finalmente, se analizó la **Aceptación**, donde los resultados fueron muy satisfactorios, pudiendo observar que los usuarios apenas consideraron el módulo difícil de utilizar a pesar de la existencia de diversas edades o experiencias con la tecnología, y que muchos de ellos tenían intención de usarlo nuevamente por su utilidad en base a sus necesidades. También se pudo observar que las mujeres encuentran más útil el uso del robot, sin embargo, los hombres opinaron que tienen mayor intención de usarlo nuevamente. En general, consideraron que la prueba fue muy divertida (ver Figura 9).

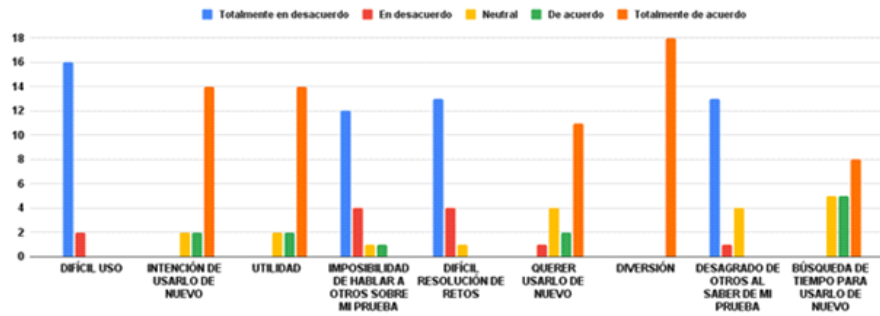


Figura 9: Resultados generales del test de aceptación.

Analizando los resultados en su conjunto, se puede concluir que todos los participantes calificaron el uso del robot como fácil o muy fácil, y la prueba como divertida. Todos ellos estuvieron muy de acuerdo o de acuerdo con que entendieron bien lo que les decía el robot y que la conversación siguió un hilo coherente, de forma que el robot había entendido la pregunta y había sabido contestar adecuadamente. Además, encontraron de gran ayuda el que la respuesta se mostrara también por pantalla. Si bien aquellos que probaron el modelo emocional completo opinaron que el robot entendía mejor sus consultas, calificaron al robot como más lento y más difícil de usar. El segundo punto puede estar relacionado con la necesidad de pasar, en ese caso, por varias pantallas de personalización, tanto del robot como del usuario, por lo que habría que reconsiderarlas o rediseñarlas.

4. Conclusiones y trabajo futuro

Se ha presentado un módulo conversacional desarrollado para dar soporte a la interacción natural y emocional persona-robot. Se ha llevado a cabo una primera evaluación con un número limitado de usuarios (18). En ella se han obtenido resultados positivos, pero también se han puesto de manifiesto algunas dificultades, como el uso de voces no pensadas para la interacción en español, el aumento de tiempos de procesado cuando se incluyen aspectos emocionales y la necesidad de mantener las interfaces de configuración sencillas a pesar de la creciente complejidad de los módulos de personalización. Una solución a este problema sería la incorporación de un sistema de identificación de usuarios a través de perfiles, lo que permitiría personalizar la experiencia de interacción y adaptar el comportamiento del robot en función del usuario sin la necesidad constante de configuración. Otra línea de trabajo futura es la integración de módulos de reconocimiento facial y de expresiones emocionales, que se han desarrollado para otros contextos de uso, para valorar su adecuación al contexto HRI. Asimismo, la expansión del módulo de peticiones externas, con el fin de mejorar la integración con servicios y bases de datos externas, permitiría al robot acceder a información en tiempo real y enriquecer sus respuestas con datos actualizados, todo ello buscando una interacción con el robot lo más efectiva y natural posible. En cuanto a futuras evaluaciones con usuarios, además de aumentar el número de participantes, se considerará la utilización de protocolos estándar en HRI, así como de instrumentos específicos para medir cuestiones como la empatía.

Agradecimientos

El trabajo está parcialmente financiado por el Ministerio de Ciencia e Innovación y Universidades (MCIU), la Agencia Estatal de Investigación (AEI) y la UE (FEDER) a través del contrato PID2022-136779OB-C31 y por el Gobierno de Aragón (Grupo T60_23R).

Declaración sobre la IA Generativa

Durante la preparación de este trabajo, los autores utilizaron GPT-4 para la corrección gramatical y ortográfica. Tras utilizar estas herramientas/servicios, revisaron y editaron el contenido según fuera necesario y asumen la plena responsabilidad del mismo.

Referencias

- [1] E. Van Oost, D. Reed, Towards a sociological understanding of robots as companions, in: International conference on human-robot personal relationship, Springer, 2010, pp. 11–18. doi:10.1007/978- 3- 642- 19385- 9_2.
- [2] A. Henschel, G. Laban, E. S. Cross, What makes a robot social? a review of social robots from science fiction to a home or hospital near you, Current Robotics Reports 2 (2021) 9–19. doi:10.1007/s43154- 020- 00035- 0.
- [3] M. Sarrica, S. Brondi, L. Fortunati, How many facets does a “social robot” have? a review of scientific and popular definitions online, Information Technology & People 33 (2020) 1–21. doi:10.1108/ITP- 04- 2018- 0203.
- [4] N. Churamani, S. Kalkan, H. Gunes, Continual learning for affective robotics: Why, what and how?, in: 2020 29th IEEE international conference on robot and human interactive communication (RO-MAN), IEEE, 2020, pp. 425–431. doi:10.1109/RO- MAN47096.2020.9223564.
- [5] E. S. Cross, R. Hortensius, A. Wykowska, From social brains to social robots: applying neurocognitive insights to human–robot interaction, 2019. doi:10.1098/rstb.2018.0024.
- [6] R. Wullenkord, F. Eyssel, Societal and ethical issues in hri, Current Robotics Reports 1 (2020) 85–96. doi:10.1007/s43154- 020- 00010- 9.
- [7] G.-Z. Yang, B. J. Nelson, R. R. Murphy, H. Choset, H. Christensen, S. H. Collins, P. Dario, K. Goldberg, K. Ikuta, N. Jacobstein, et al., Combating covid-19– the role of robotics in managing public health and infectious diseases, 2020. doi:10.1126/scirobotics.abb5589.
- [8] F. Alnajjar, S. Khalid, A. A. Vogan, S. Shimoda, R. Nouchi, R. Kawashima, Emerging cognitive intervention technologies to meet the needs of an aging population: a systematic review, Frontiers in Aging Neuroscience 11 (2019) 291. doi:10.3389/fnagi.2019.00291.
- [9] M. Assad-Uz-Zaman, M. Rasedul Islam, S. Miah, M. H. Rahman, Nao robot for cooperative rehabilitation training, Journal of rehabilitation and assistive technologies engineering 6 (2019) 2055668319862151. doi:10.1177/2055668319862151.
- [10] Y. Chen, S. Garcia-Vergara, A. M. Howard, Effect of feedback from a socially interactive humanoid robot on reaching kinematics in children with and without cerebral palsy: a pilot study, Developmental neurorehabilitation 21 (2018) 490–496. doi:10.1080/17518423.2017.1360962.
- [11] A. Mohebbi, Human-robot interaction in rehabilitation and assistance: a review, Current Robotics Reports 1 (2020) 131–144. doi:10.1007/s43154- 020- 00015- 4.
- [12] N. L. Robinson, T. V. Cottier, D. J. Kavanagh, Psychosocial health interventions by social robots: systematic review of randomized controlled trials, Journal of medical Internet research 21 (2019) e13203. doi:10.2196/13203.

- [13] A. A. Scoglio, E. D. Reilly, J. A. Gorman, C. E. Drebing, Use of social robots in mental health and well-being research: systematic review, *Journal of medical Internet research* 21 (2019) e13322. doi:10.2196/13322.
- [14] G. Kyprianou, A. Karousou, N. Makris, I. Sarafis, A. Amanatiadis, S. A. Chatzichristofis, Engaging learners in educational robotics: Uncovering students' expectations for an ideal robotic platform, *Electronics* 12 (2023) 2865. doi:10.3390/electronics12132865.
- [15] N. Calvo-Barajas, G. Perugia, G. Castellano, The effects of robot's facial expressions on children's first impressions of trustworthiness, in: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, 2020, pp. 165–171. doi:10.1109/RO- MAN47096.2020.9223456.
- [16] M. M. de Graaf, S. Ben Allouch, J. A. Van Dijk, What makes robots social?: A user's perspective on characteristics for social human-robot interaction, in: *International Conference on Social Robotics*, Springer, 2015, pp. 184–193. doi:10.1007/978- 3- 319- 25554- 5_19.
- [17] A. Paiva, I. Leite, H. Boukricha, I. Wachsmuth, Empathy in virtual agents and robots: A survey, *ACM Transactions on Interactive Intelligent Systems (TiIS)* 7 (2017) 1–40. doi:10.1145/2912150.
- [18] S. Brave, C. Nass, K. Hutchinson, Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent, *International journal of human-computer studies* 62 (2005) 161–178. doi:10.1016/j.ijhcs.2004.11.002.
- [19] I. Leite, A. Pereira, S. Mascarenhas, C. Martinho, R. Prada, A. Paiva, The influence of empathy in human–robot relations, *International journal of human-computer studies* 71 (2013) 250–260. doi:10.1016/j.ijhcs.2012.09.005.
- [20] M. S. Gou, V. Vouloutsi, K. Grechuta, S. Lallée, P. F. Verschure, Empathy in humanoid robots, in: *Biomimetic and Biohybrid Systems: Third International Conference, Living Machines 2014*, Milan, Italy, July 30–August 1, 2014. *Proceedings 3*, Springer, 2014, pp. 423–426. doi:10.1007/978- 3- 319- 09435- 9_50.
- [21] J. Xu, J. Broekens, K. Hindriks, M. A. Neerinx, Mood contagion of robot body language in human robot interaction, *Autonomous Agents and Multi-Agent Systems* 29 (2015) 1216–1248. doi:10.1007/s10458- 015- 9307- 3.
- [22] F. Babel, J. Kraus, L. Miller, M. Kraus, N. Wagner, W. Minker, M. Baumann, Small talk with a robot? the impact of dialog content, talk initiative, and gaze behavior of a social robot on trust, acceptance, and proximity, *International Journal of Social Robotics* 13 (2021) 1485–1498. doi:10.1007/s12369- 020- 00730- 0.
- [23] C. Wang, S. Hasler, D. Tanneberg, F. Ocker, F. Joublin, A. Ceravola, J. Deigmoeller, M. Gienger, Lami: Large language models for multi-modal human-robot interaction, in: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–10. doi:10.1145/3613905.3651029.
- [24] C. Y. Kim, C. P. Lee, B. Mutlu, Understanding large-language model (llm)-powered human-robot interaction, in: *Proceedings of the 2024 ACM/IEEE international conference on human-robot interaction*, 2024, pp. 371–380. doi:10.1145/3610977.3634966.

- [25] N. Cherakara, F. Varghese, S. Shabana, N. Nelson, A. Karukayil, R. Kulothungan, M. A. Farhan, B. Nasset, M. Moujahid, T. Dinkar, et al., Furchat: an embodied conversational agent using llms, combining open and closed-domain dialogue with facial expressions, arXiv preprint arXiv:2308.15214 (2023). doi:10.18653/v1/2023.sigdial- 1.55.
- [26] A. Bonarini, Communication in human-robot interaction, *Current Robotics Reports* 1 (2020) 279–285. doi:10.1007/s43154- 020- 00026- 1.
- [27] M. K. Lee, J. Forlizzi, S. Kiesler, P. Rybski, J. Antanitis, S. Savetsila, Personalization in hri: A longitudinal field experiment, in: *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 319–326. doi:10.1145/2157689.215780.
- [28] M. Dziergwa, M. Kaczmarek, P. Kaczmarek, J. Kędzierski, K. Wadas-Szydłowska, Long-term cohabitation with a social robot: a case study of the influence of human attachment patterns, *International Journal of Social Robotics* 10 (2018) 163–176. doi:10.1007/s12369- 017- 0439- 2.
- [29] R. Plutchik, The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, *American scientist* 89 (2001) 344–350. URL: <https://www.jstor.org/stable/27857503>.
- [30] E. Son, Visual, auditory, and psychological elements of the characters and images in the scenes of the animated film, *inside out*, *Quarterly Review of Film and Video* 39 (2022) 225–240. doi:10.1080/10509208.2021.1959815.
- [31] R. Kumar, Data-driven design: Beyond a/b testing, in: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 2019, pp. 1–2. doi:10.1145/3295750.3300046.
- [32] E. Cerezo, R. Lacuesta, J. Gallardo, A. Aguelo, Exploring the use of voice assistants in nursing homes, *International Journal of Human-Computer Interaction* (2025) 1–17. doi:10.1145/3295750.3300046.
- [33] P. W. Jordan, B. Thomas, I. L. McClelland, B. Weerdmeester, *Usability evaluation in industry*, CRC press, 1996.

“Robots in the wild”: grupos intergeneracionales y robots sociales en espacios públicos^{*}

“Robots in the wild”: intergenerational groups and social robots in public spaces

Raquel Lacuesta^{1,2,†}, Eva Cerezo^{1,2,*†}, Loreto Matinero^{2,†} y Sandra Baldassarri^{1,2,†}

¹ Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, Zaragoza, España

² Departamento de Informática e Ingeniería de Sistemas, Universidad de Zaragoza, Zaragoza, España

Resumen

Cuando en el ámbito de la interacción humano-robot (HRI) se habla de robots “in the wild” se refiere a robots que interactúan con personas y realizan tareas en entornos reales, cotidianos y no controlados. El número de estudios de ese tipo en la literatura es muy limitado, más aún cuando se refiere a interacciones grupales en espacios públicos. Este artículo presenta una experiencia lúdica llevada cabo en un espacio público, dirigida a grupos familiares que participan de forma espontánea. En la experiencia se hizo uso de un juego, El misterio de Arquímedes, diseñado para ser jugado por un grupo familiar intergeneracional que hubo que adaptar al nuevo contexto de uso. El juego combina elementos analógicos y digitales en dinámicas colaborativas que integran la interacción con un robot social bajo la guía de una persona que hace el rol de mediador y gestor del grupo. A partir de la experiencia se presentan recomendaciones que tienen que ver con el diseño inclusivo, las dinámicas grupales, la gestión del entorno, las narrativas, el fomento de la colaboración, la flexibilidad y el rol esencial del mediador. La experiencia ha puesto de manifiesto el potencial de los robots sociales para enriquecer las dinámicas grupales intergeneracionales en espacios públicos.

Abstract

In Human Robot Interaction (HRI), robots “in the wild” refers to robots that interact with people and perform tasks in real, every day, uncontrolled environments. The number of such studies in the literature is very limited, even more so when referring to group interactions in public spaces. This article presents a play experience carried out in a public space, aimed at spontaneously participating family groups. The experience made use of a game, Archimedes' Mystery, designed to be played by an intergenerational family group that had to be adapted to the new context of use. The game combines analogue and digital elements in collaborative dynamics that integrate interaction with a social robot under the guidance of a person who plays the role of mediator and manager of the group. Based on the experience, recommendations are presented that have to do with inclusive design, group dynamics, management of the environment, narratives, fostering collaboration, flexibility and the essential role of the mediator. The experience has highlighted the potential of social robots to enrich intergenerational group dynamics in public spaces.

Palabras clave

Robots in the wild, robots sociales, actividades intergeneracionales, interacción grupal

1. Introducción

En el campo de la interacción humano-robot (HRI) lo habitual es llevar a cabo evaluaciones en laboratorios en las que las personas suelen interaccionar una a una o en grupos pequeños, llevando a cabo tareas bien definidas en un entorno totalmente controlado. Si bien está claro que hay mucho

^{*}Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

¹ Corresponding author.

[†] These authors contributed equally.

✉ lacuesta@unizar.es (R. Lacuesta); ecerezo@unizar.es (E. Cerezo); lmatinero@unizar.es (L. Matinero); sandra@unizar.es (S. Baldassarri)

ORCID 0000-0002-4773-4904 (R. Lacuesta); 0000-0003-4424-0770 (E. Cerezo); 0009-0004-8101-1023 (L. Matinero); 0000-0002-9315-6391 (S. Baldassarri)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

que aprender de ese tipo de evaluaciones y éstas deberán seguir siendo el pilar de los estudios sobre HRI, creemos que es también necesario llevar a cabo otro tipo de experiencias en las que los investigadores se aventuran con sus robots en lugares de trabajo, hogares y entornos públicos, en los que la gente se relacionará cada vez más con los robots [1]. En estos entornos, la interacción rara vez será individual, por lo que ese hace necesario hacer estudios que examinen el efecto de los robots en contextos multipersonales como equipos de trabajo, familias o grupos en entornos públicos para comprender la influencia de un robot más allá de la interacción directa de una sola persona que interactúa directamente con él.

Así, el término "in the wild" (al natural) se usa para describir interacciones en las que las personas usan los sistemas o dispositivos en entornos no controlados, cotidianos, y complejos, lejos de laboratorios controlados. La consideración de este tipo de entornos es esencial en el caso de los robots sociales, ya que permite abordar interacciones sociales más complejas, en las que hay múltiples personas, roles e interacciones simultáneas. Estas interacciones involucran diferentes dinámicas grupales, influencias mutuas y contextos sociales variados. Por ejemplo, en grupos familiares un robot podría actuar como mediador o facilitador de interacciones, ayudando a que diferentes generaciones se comuniquen mejor [2, 3]; en equipos de trabajo, los robots pueden participar en reuniones, influir en la toma de decisiones del grupo o apoyar tareas colaborativas, como sucede en hospitales o fábricas [4]; en espacios públicos los robots podrían actuar como guías en museos o centros comerciales influyendo en cómo las personas interactúan entre sí y creando nuevas normas sociales [5]. Este enfoque no solo amplía las capacidades de los robots, sino que también plantea interrogantes sobre su impacto en las dinámicas humanas y su convivencia con las personas en escenarios complejos.

En este artículo se presenta una experiencia lúdica con grupos intergeneracionales y en espacio público soportada por un robot social. La experiencia se llevó a cabo durante la Semana Europea de la Robótica (ERW2024) organizada por EU-Robotics en Zaragoza del 13 al 15 de noviembre de 2024. En el contexto de esa semana se organizó AIR2024, un festival de robótica e IA abierto al público en un espacio público en el centro de la ciudad. En la Figura 1 se muestra la información usada para la difusión de la actividad en el contexto del festival. El espacio estuvo abierto toda la tarde de forma que familias, sin la necesidad de apuntarse previamente, pudieran entrar e interactuar con los robots. La experiencia que se presenta aquí es "Ven a jugar con Lola", basada en una adaptación de un juego con robot social diseñado para una única familia. A lo largo de la tarde pasaron por la sala más de 20 familias, además de un grupo de niños y niñas con problemas de audición con sus maestras.

El robot utilizado en esta actividad fue el Sanbot Elf. Se trata de un robot social que mide 90,2 cm dotado de control del habla, que incluye funciones de Text to Speech (TTS) y Speech to Text (STT), control del movimiento (cabeza, brazos y ruedas omnidireccionales), control de luz (LEDs de color en la cabeza, brazos y base). El robot también dispone de una pequeña pantalla, a modo de cara, que permite mostrar 18 expresiones faciales diferentes y una pantalla táctil de 10.1 pulgadas. Posee un SDK, QihanOpenSDK, que es el que ha sido utilizado para la implementación del juego.

El artículo se organiza de la siguiente forma: en el apartado 2 se pone el estudio en contexto, abordando las dificultades que presenta la interacción con los robots en contextos grupales y su estudio al natural, en el apartado 3 se describe el juego utilizado y las adaptaciones llevadas a cabo, en el apartado 4 se detalla la experiencia, su planteamiento, desarrollo y resultados, para en el apartado 5 abordar la discusión y presentar recomendaciones; en el apartado 6 se presentan las conclusiones.

SALA DE CREATIVIDAD

Ven a jugar en familia con el robot social Sanbot LOLA!



El juego

Ven con tu familia a jugar con nuestro robot Sanbot LOLA. Forma un equipo con tus padres y abuelos (¡no los dejéis en casa!) y enfréntate a otras familias resolviendo retos ayudados por LOLA. Estaréis acompañados en todo momento por nuestro equipo de investigadores del proyecto PLEISAR, que buscan cómo hacer de los robots compañeros de juegos y experiencias en grupo y os explicarán todo lo que queráis saber

Organizador



Universidad Zaragoza grupo de investigación Affective Lab

Figura 1: Información acerca de la actividad realizada durante la Semana Europea de La Robótica (ERW2024) organizada por EU-Robotics en Zaragoza.

2. Estado del arte

En los últimos años, el desarrollo de robots sociales y colaborativos ha evolucionado rápidamente, planteando retos de diseño respecto a la forma en que los humanos interactúan con estas tecnologías. Los robots se han incorporado en contextos diferentes, ayudando tanto a la mejora de la productividad como en el fomento de actividades lúdicas. En concreto, el uso de robots en entornos grupales abre nuevas posibilidades por ejemplo para la realización de tareas de mediación social y/o fomento de las relaciones interpersonales.

Los trabajos que exploran las interacciones con robots “in the wild” son escasos. En [1] se discuten las dificultades que plantean dichos estudios que incluyen la necesidad de comprender el contexto social de la interacción, y el impacto de la interacción individual en el contexto social. En [6] se presentan una serie de evaluaciones in-situ de prototipos de robots en un entorno laboral. Como señaló Rogers [7], en este tipo de estudios el foco de traslada del experimentador al usuario, lo que los hace especialmente vulnerables a factores externos que van desde los fallos técnicos a la disposición (o no) a interactuar de los usuarios, a las interferencias de otras personas o las expectativas poco realistas con las que los usuarios se dispongan a la interacción. Incluso en un contexto de evaluación individual el trabajo pone también de manifiesto las dificultades encontradas en la recolección de datos al ser difícil la medida de las percepciones y comportamientos de las personas en esos contextos “naturales” [6]. En [8] se presentan unas evaluaciones de robots “in the wild” con niños hospitalizados. El trabajo repasa las dificultades y retos que plantea ese tipo de evaluación, que incluyen aspectos legales y administrativos, de uso de espacios y de planificación, desde la forma de presentar a los robots a las medidas a tomar ante fallos técnicos, con el foco puesto en hacer lo más satisfactoria posible la experiencia a los niños participantes.

Si se pasa a considerar interacciones grupales los retos y dificultades se disparan, empezando por las dificultades para gestionar simultáneamente interacciones con grupos de personas grandes, debido a limitaciones en el procesamiento de los robots. En estos contextos de interacción con múltiples usuarios, el entorno impredecible puede limitar fuertemente la capacidad del robot para responder de manera adecuada y natural. Se requiere, por lo tanto, la inclusión de sistemas avanzados de percepción y aprendizaje que permitan que el robot se adapte a estos entornos desafiantes [9], existiendo numerosos desafíos de percepción y generación de comportamiento, así como factores relacionados con el entorno, el grupo y las capacidades del robot [10, 11]. Por otro lado, deben ajustar su comportamiento en función de los objetivos, el estado mental de los usuarios [12] y su rendimiento [13]. Además, la respuesta tiene que seguir siendo fluida, ya que cuando no lo es (por ejemplo, pausas largas o respuestas inadecuadas), los usuarios tienden a perder interés rápidamente, afectando la confianza en los robots [12]. Así, en escenarios con múltiples usuarios, los robots deben gestionar los turnos y adherirse a las normas sociales para reducir los errores conversacionales y mejorar la eficacia de la comunicación. Este enfoque mejora la percepción del usuario sobre las habilidades comunicativas del robot [14]. Además, si los robots no interpretan señales sociales en situaciones dinámicas o en contextos con interrupciones frecuentes se limita su efectividad [15]. Estudios recientes destacan la importancia de personalizar las interacciones humano-robot, adaptándolas a las dinámicas específicas de cada contexto y evaluando su impacto en el comportamiento grupal, la cohesión social y los resultados colaborativos [16, 17].

Respecto al análisis de la literatura referente al diseño de experiencias de juego con múltiples interacciones sociales cuando se interacciona con niños, un ejemplo destacado es el trabajo en [18] donde se estudió cómo los robots pueden interactuar con grupos de niños en un entorno escolar, utilizando actividades como juegos colaborativos para fomentar la interacción prolongada. El estudio mostró que los robots pueden facilitar dinámicas grupales, pero enfrentan desafíos como mantener la atención de todos los participantes, la atención a largo plazo y adaptarse a diferentes niveles de interacción social. El diseño de los juegos intergeneracionales implica retos adicionales. Los estudios realizados en [18], sugieren la importancia de la narrativa y las dinámicas modulares de interacción para maximizar la participación. Siguiendo los resultados de estudios como [19], donde los adultos asumen un rol de apoyo crítico, el diseño de los retos fomenta el apoyo por parte de los adultos complementados por la figura de un mediador. En [20] se destaca cómo los robots sociales pueden generar conexiones emocionales significativas con los participantes, lo que abre la puerta a su consideración en la valoración de las experiencias.

En conclusión, se observa que en el diseño de actividades grupales e intergeneracionales surgen varios desafíos clave que merecen atención. Estos desafíos destacan la necesidad de enfoques de diseño cuidadosos y adaptativos que puedan abordar la diversidad y la complejidad de grupos grandes, reduciendo las limitaciones en la interacción con el robot de múltiples usuarios con diferentes intereses, interactuando simultáneamente, así como su colaboración y fomento de la participación de todos los grupos implicados. En el caso de plantear dichas interacciones “in the wild” a estos retos se le añade la necesidad de adaptar la experiencia al contexto inmediato de uso (número y características de los participantes, tiempo disponible, etc.) y las dificultades en la toma de datos.

A continuación, se presenta el juego intergeneracional utilizado en la experiencia, en cuyo diseño se han seguido las pautas marcadas en la literatura y que debió ser adaptado para su uso en la experiencia “in the wild”.

3. El misterio de Arquímedes

El juego es un juego basado en retos, que hay que ir superando, guiado por una narrativa [21] que se centra en la historia del origen y la creación del robot. En la historia, Arquímedes, su creador, lo

ha cifrado y borrado su memoria para protegerlo. El objetivo del juego es resolver los tres enigmas que plantea el robot, para así despertarlo y ayudarlo a descubrir el nombre de la persona que lo creó. Una vez descubierto el nombre de su creador el propio robot, dotado de un proyector, muestra una presentación sobre su creador (Arquímedes). Así, el juego está diseñado para que participen y colaboren personas de diversas generaciones, lo que se ve facilitado cuando estas pertenecen al mismo entorno familiar.

El juego se había diseñado para que participara una única familia en la experiencia, por lo que hubo de ser adaptado para que pudieran jugar varios grupos familiares. Ello implicó modificar la dinámica del juego. Así, cada familia deberá intentar resolver el enigma por separado para lo que se les entregará una hoja con el enunciado del problema y lápices para que puedan realizar pruebas; se acercarán al robot una vez obtengan la solución, evitando de esta forma la aglomeración de personas alrededor del robot. No se plantea que ninguna familia sea ganadora ni perdedora, sino que se esperará a que cada familia intente resolver el enigma; una vez obtenida la solución, la familia podrá ayudar al resto planteando pistas que les puedan guiar hacia la solución, fomentando así la colaboración intergrupala. Se decidió establecer un máximo de tres grupos familiares jugando simultáneamente. Teniendo en cuenta que una familia puede estar compuesta de entre dos y cinco miembros, se podría llegar así hasta un total de 15 participantes.

Una segunda adaptación tuvo que ver con reforzar el papel de mediador del juego, una persona que coordine la experiencia, y gestione el acceso al robot y a su pantalla ya que por su tamaño (1280x800 px), la visualización e interacción es difícil con muchas personas a su alrededor. El papel del mediador es fundamental, ya que se va a encargar de organizar los equipos y modular la participación de los mismos.

Una tercera adaptación llevada a cabo fue la consideración de niveles de dificultad en los retos más complejos, teniendo en cuenta que no se iba a tener información previa sobre la edad de los participantes. El mediador será el encargado de, a la vista de los equipos, plantear el nivel de dificultad y de controlar el tiempo de juego, incluso decidiendo saltar alguno de los enigmas, con objeto de que pueda hacer varias sesiones de juego y más familias participando. El tiempo de juego se estableció en unos 20-30 minutos por sesión.

A continuación, se detalla el funcionamiento del juego.

El juego comienza con el robot dormido y un texto en su pantalla. El texto que aparece ayuda a poner en contexto a los participantes, ya que es un mensaje escrito por el creador del robot. En él se muestra la emoción del creador por haber conseguido crear un “autómata que hace tareas humanas”, pero debido a unos invasores de la época, decide que lo mejor es ocultar esta tecnología tan avanzada, borrar su memoria y bloquearlo, para que así solo personas capaces de resolver enigmas puedan despertar el robot y además descubrir el nombre de la persona que lo creó.

Una vez se ha leído o se ha explicado a los participantes esta primera introducción, uno de ellos debe pasar a la siguiente pantalla. En ella se plantea el primer enigma del juego, el cual si se resuelve conseguirá despertar al robot.

El primer enigma (ver Figura 2 izqda.) consiste en unas series según los meses del año. Los meses son representados por la primera letra de su nombre, y la letra que se obtiene es la siguiente a la serie planteada, por lo que SON sería Septiembre, Octubre, Noviembre, y la letra obtenida D sería Diciembre, EFM sería Enero, Febrero, Marzo y la letra obtenida A sería Abril, pudiendo deducir entonces que MJJ es Mayo, Junio, Julio, por lo que su continuación sería el mes de Agosto, cuya letra inicial es la A.

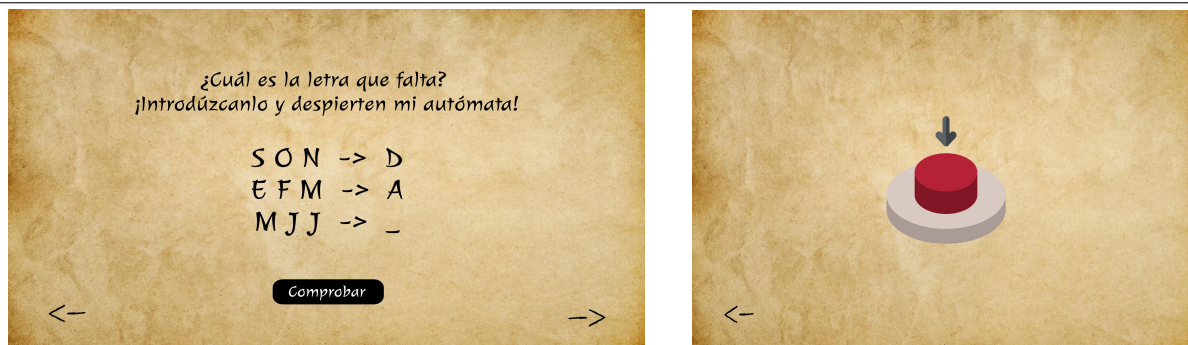


Figura 2: Primer enigma (izquierda) y botón de paso al siguiente (derecha).

Las familias trabajan el enigma de forma independiente y pueden ir probando las repuestas en la pantalla táctil del robot: para facilitar esta interacción y permitir participar a todos de forma equitativa, se ha permitido la introducción de múltiples respuestas correctas o incorrectas, para que así los usuarios puedan comprobar su respuesta. Por otro lado, si el mediador de la experiencia detecta que los participantes no están encontrando la solución, este les propondrá hasta un máximo de tres pistas.

Si una familia resuelve el reto, se le insta a que ayude a las demás, ya que no se podrá pasar al siguiente enigma hasta que todos hayan llegado a la solución. Cuando todas las familias participantes han dado con la solución, se continúa con la experiencia. En este momento, pasando de pantalla se muestra un gran botón (ver Figura 2 dcha.), el cual una vez pulsado despierta al robot.

A partir de este punto de la experiencia, es el mismo robot quien se encarga de dirigirla. A pesar de esto, la figura del mediador sigue siendo necesaria para ayudar a mantener el orden, resolver dudas o asistir al robot en caso de que fuera necesario.

Durante el despertar, el robot sigue una serie de pasos. En primer lugar, este se presenta y explica cómo debe realizarse la interacción con él, ya que para avanzar en la resolución de los enigmas es importante conocer cómo se habla con el robot, para solicitar pistas o comunicar la respuesta por voz. Una vez explicada la interacción, el robot cuenta que necesita que se resuelvan una serie de enigmas, para así recuperar su memoria y con ello recordar el nombre de la persona que lo creó.

A continuación, el robot introduce el segundo enigma (ver Figura 3), y pide ayuda a los participantes para resolverlo. Este segundo enigma consiste en completar los círculos con los números que faltan, teniendo en cuenta que son los números del 1 al 9, que no se pueden usar más de una vez, y que la multiplicación de los números en vertical y horizontal debe dar el mismo resultado. Este enigma en principio es realizado por las familias, pero si el mediador considera que puede llegar a ser muy complejo para el grupo, o no se dispone del tiempo necesario, este puede ser omitido.

La dinámica de juego es la misma que en el anterior enigma. Cuando los participantes creen obtener la respuesta, pueden acercarse e introducirla en la interfaz que ofrece el robot. A diferencia del anterior enigma, será el robot quien si pasa un largo tiempo desde que un participante ha interactuado con él, pregunté a los participantes si necesitan una pista.

Una vez todos los participantes obtengan la respuesta, deberán acercarse al robot y comunicarle el resultado de la multiplicación para poder pasar al siguiente enigma.

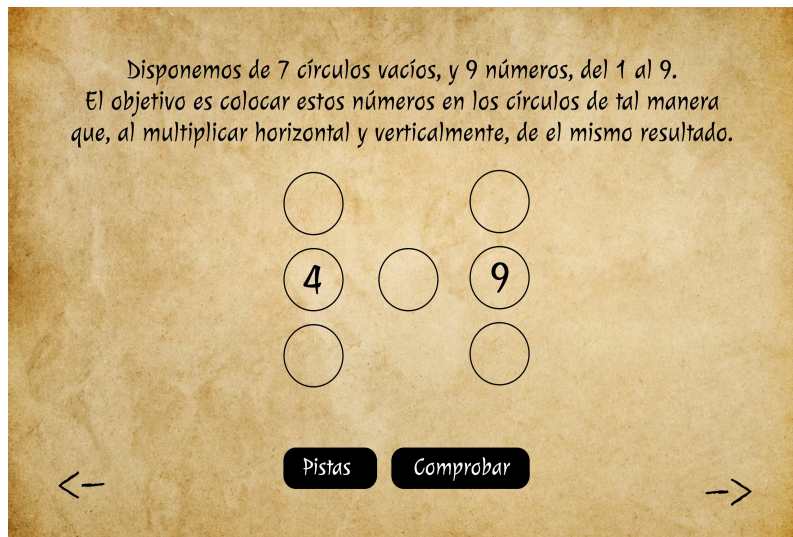


Figura 3: Segundo enigma.

Por último, el tercer enigma será el que permita a los participantes descubrir el nombre del creador del robot. Para este enigma se han preparado tres variantes, y el mediador será el encargado de decidir qué enigma es el más apropiado según los miembros del grupo. Esto se ha realizado así, ya que se pretende que disfruten de la experiencia desde niños pequeños hasta personas mayores, y estos puedan colaborar y trabajar juntos para resolver los enigmas. Las tres variantes son las siguientes:

La primera variante (ver Figura 4) consiste en una frase cifrada, en la que se indica al usuario que debe buscar la equivalencia dentro del alfabeto, dejando caer a los participantes que se está usando una especie de cifrado. A pesar de que sea complejo en un inicio, los participantes pueden solicitar al robot hasta un máximo de tres pistas que les ayuden a encontrar la equivalencia.

En la segunda variante (ver Figura 5), el usuario cuenta con la misma frase y las mismas pistas, pero a diferencia del caso anterior, se muestra una parte de la frase ya descifrada, facilitando así la resolución del enigma.

En los dos casos el desplazamiento del cifrado es de tres letras hacia atrás, obteniendo finalmente la frase “mi nombre es Arquímedes”.

Descifra el código secreto: busca la equivalencia en el alfabeto

ol proeuh hv dutxlohghv

Descifra el código secreto: busca la equivalencia en el alfabeto

ol proeuh hv dutxlohghv

mi nombre es _____

Figura 4: Tercer enigma primera variante (izquierda) y tercer enigma segunda variante (derecha).

En la tercera variante se entrega a los usuarios una sopa de letras, en la que deberán buscar el nombre de la persona que creó el robot. En la sopa de letras se han incluido varios nombres de personajes importantes, por lo que cada vez que un participante crea haber encontrado el nombre correcto, debe ir al robot y comprobar su respuesta. Este enigma es el único que no cuenta con pistas, ya que el mismo nombre a introducir se encuentra en la sopa de letras.

Al igual que en el resto de enigmas, se deberá esperar a que todas las familias den con la solución correcta. Una vez obtenida, estos se podrán acercar al robot y juntos comunicarle el

nombre de su creador, para finalmente pasar a la presentación final, donde se explica quién fue Arquímedes y algunos de sus experimentos.

4. Evaluación

4.1. Participantes

Las sesiones incluyeron familias con diversas composiciones, entre ellos, padre y madre con uno, dos o tres hijos, abuelos con uno, dos o tres nietos, y madres o padres solos acompañados de sus niños. También participó un grupo de niños y niñas con problemas de audición acompañados por dos docentes. En total participaron 68 personas divididas en 6 sesiones de juego. Durante la actividad diseñada, los participantes interactuaron con el robot para llevar a cabo la actividad presentada. A continuación, se analizan aspectos relacionados con la dinámica de juego, la resolución de retos y la interacción con el robot.

4.2. Procedimiento

Las familias participantes ingresaban al espacio y se les asignaba una mesa por unidad familiar. En el caso de familias pequeñas (de 2 miembros), se unían dos familias en una misma mesa. Se dispusieron tres mesas en total, ubicadas estratégicamente en el espacio y separadas del robot, para facilitar el trabajo con papel y bolígrafo en la resolución de los retos. El robot, por su parte, se colocó en un extremo de la sala.

La dinámica del juego comenzaba con todas las familias reunidas alrededor del robot, donde se realizaba la presentación de la actividad y del primer reto. Como se ha explicado, a lo largo de la sesión las familias si pensaban que tenían ya la solución al reto se acercaban al robot a comprobar en la pantalla táctil si la respuesta era la correcta. También podían acercarse al robot en cualquier momento a solicitarle pistas sobre el reto. Al finalizar, todas las familias se reunían nuevamente alrededor del robot para la presentación final y la resolución conjunta del desafío.

La actividad contó con la participación de una mediadora, encargada de explicar el juego, la interacción con el robot y coordinar las interacciones con éste. Además, tres evaluadoras se encargaban de dar apoyo y tomar notas.

Las evaluadoras usaron una hoja de observación para analizar los siguientes aspectos:

- Relación de ayuda: Interacciones verbales y no verbales entre adultos, niños y el robot.
- Resolución de problemas: Estrategias empleadas por los participantes y dinámicas de liderazgo.
- Expresión emocional: Reacciones ante las respuestas del robot y durante las dinámicas grupales.
- Uso de tecnología: Interacciones técnicas con el robot y mediación por parte de los adultos.

A lo largo de la sesión los evaluadores registraron de manera cualitativa observaciones relevantes relacionadas con estas variables, proporcionando un análisis detallado de las interacciones.

4.3. Resultados

En las figuras 6 al 11 se muestran diversos momentos de las experiencias. En la Figura 6, se puede observar como las familias tratan de resolver los retos en las mesas de trabajo. En la Figura 7 vemos la distribución del grupo frente al robot, en la Figura 8 se observa al mediador y a un niño

interaccionando. En la Figura 9 se observa como un niño se ofrece voluntario para interactuar con el robot. En la Figura 10 vemos el aspecto competitivo, una niña corre a interactuar con el robot para pedir una pista o resolver el reto. En la Figura 11 vemos la distribución del grupo donde los niños se posicionan delante.



Figura 5: Participantes en su mesa de trabajo agrupadas por familias (izquierda) e interacción del grupo con el robot (derecha).



Figura 6: Distribución física de los participantes con niños delante (izquierda) y mediador moderando la interacción grupal (derecha).



Figura 7: Una niña de una familia corre a probar una respuesta o a pedir una pista al robot, el resto de los componentes de la familia permanecen en la mesa de trabajo (izquierda) y niños interaccionan con el robot (derecha).

A continuación, se presentan los resultados obtenidos a partir de las notas de observación de las evaluadoras de acuerdo a cada variable de análisis planteadas.

En general, todos los participantes mostraban gran **interés y concentración** en la resolución del reto, reflejado en sus gestos y actitudes (seriedad, mirada en el folio del problema, hablando en todo momento acerca de cómo se resolvería el reto con el resto de componentes del equipo). También se observó un componente competitivo en las interacciones: “Corre, vamos, prueba”, decían para motivarse mutuamente. Sin embargo, cuando otro grupo lograba resolver un reto antes, o el propio grupo quedaba atascado, se notaba más seriedad en las expresiones. Ante esta situación, los adultos animaban con frases como “Vamos a pedir otra pista”, los niños cuando conocían la dinámica también sugerían “Vamos a pedir más pistas”. Referente a las **dinámicas familiares**, las familias celebraban los éxitos con sonrisas y alegría y mantenían una actitud reflexiva ante los fracasos. Los adultos motivaban a los niños a continuar, minimizando el desánimo. En dos casos fue necesario establecer turnos porque había en cada sesión un niño demasiado activo que no paraba de tocar (no hubo niñas hiperactivas). En la mayoría de los casos los adultos tenían el papel del reto, sino el hermano mayor. Algunos pedían folios para resolver en paralelo. En cuanto a la **resolución de problemas**, cuando los niños eran mayores, asumían parte del liderazgo en las actividades, especialmente si están acompañados de abuelos. Sin embargo, en casos de duda o cuando los niños eran pequeños, los adultos tomaban un rol dominante. La dinámica grupal variaba entre colaboración y competencia. Cuando otros grupos progresaban rápidamente, algunos adultos adoptaban actitudes competitivas, tomando mayor control en la actividad.

Respecto al **uso de afirmaciones y preguntas**, en general los adultos realizaron preguntas reflexivas para fomentar el pensamiento crítico y organizar la interacción de los niños con el robot, por ejemplo, decían “*si tenemos estos datos y estas pistas, ¿qué números podríamos utilizar?*” Los niños también preguntaban sobre las posibles soluciones preguntado por ejemplo “*¿podría ser esta opción?*” Los adultos utilizaban afirmaciones para proporcionar explicaciones científicas o matemáticas, confirmar avances o guiar la resolución de retos, “*tenemos los siguientes datos, por lo*

tanto una posible solución sería la siguiente". Para coordinar la interacción con el robot, tanto adultos como niños empleaban instrucciones directas, asegurándose de que todos tuvieran su turno *"ahora le toca a tu hermano"*, *"me toca a mí"*. Los niños utilizaban afirmaciones cuando avanzaban en la resolución del problema *"He encontrado un nombre, voy a probarlo"*. Respecto a las **expresiones de apoyo y ánimo**, los adultos brindaron apoyo a los niños, especialmente en momentos de confusión o dificultad. Los hermanos mayores apoyaban también a los pequeños. Respecto a la **expresión emocional** con el robot, los niños mostraron curiosidad, entusiasmo y alegría, especialmente cuando el robot los felicitaba. Experimentaban signos de frustración momentánea (mayor seriedad) si el robot no comprendía sus comandos. Durante la interacción, los niños se acercaban al robot con curiosidad y, tras familiarizarse con su funcionamiento, se sentían motivados e ilusionados al interactuar con él. En el caso de las niñas mostraban signos de afecto, acariciando al robot, llamándola guapa y despidiéndose de ella. En general, los adultos permitían que los niños lideraran la interacción, interviniendo sólo cuando los niños no sabían cómo proceder o la interacción no era fructífera. Aunque más de dos intentos fallidos podían causar cierto desánimo, la dinámica grupal ayudaba a mantener una actitud positiva, *"vamos a probar todos juntos"*. Las niñas parecían ser más tímidas y más respetuosas de las indicaciones y órdenes dadas por el mediador. En cuanto a la interacción con el robot los adultos actuaron como apoyo y mediación entre hermanos, asegurando turnos equitativos entre los niños que querían interactuar.

En la mayoría de los casos, los adultos se quedaban sentados y los que iban a interactuar con el robot eran los niños. En dos casos los adultos se quedaron sentados también cuando el robot explicaba o hacía cosas, como si la actividad fuera solo para los niños. En el caso de una abuela, fue muy activa y casi probaba ella antes que los niños, pero no fue la dinámica general. Le gustó que el robot fuera Lola, femenino, y sabía de la historia de Arquímedes. En algún caso se fueron un poco antes de acabar la actividad y no vieron las otras actividades que se presentaban. Cuando se juntaron varias familias en una mesa la interacción fue diversa, algunos hacían cosas juntos y otros no, dependía también de la edad de los niños. En cuanto a la **relación espacial**, generalmente, los niños se posicionaban delante del robot, mientras los adultos supervisaban desde atrás. Los niños se acercaban al robot de manera espontánea, y al comprender el mecanismo, interactuaban con confianza. Las **dificultades técnicas fueron mínimas**. Ocasionalmente, el robot tenía problemas para reconocer voces bajas o comandos mal pronunciados. En estos casos, los adultos intervenían para guiar a los niños, recordándoles activar correctamente la función de entrada de voz del robot: *"Tócale la cabecita, y cuando la luz esté verde, vuelve a intentarlo"*. De hecho, se vio necesario reforzar la explicación de que era necesario activar la cabeza del robot para hablar, poniendo de manifiesto lo conveniente de hacer ejercicios de interacción con el robot antes de comenzar la experiencia en sí.

Como se ha mencionado, participó un grupo de niños y niñas sordos, que se integraron sin problemas en las dinámicas con otros grupos, ejerciendo sus profesoras de traductoras a lenguaje de signos en español. Sin embargo, el robot hablaba demasiado rápido para ello, lo que nos hizo ver que debimos haber considerado a priori medidas de inclusión/accesibilidad, como activar la salida simultánea por pantalla o poder modular la velocidad del habla, por ejemplo.

En el apartado siguiente resumimos lo aprendido en la experiencia planteando una serie de recomendaciones.

5. Recomendaciones

Los resultados sugieren que el uso de robots sociales, en concreto el uso del robot Sanbot Elf, favorece la interacción familiar en espacios públicos, promoviendo la colaboración y el aprendizaje, siendo una herramienta efectiva para fomentar dinámicas de grupo y familiares.

A partir de la experiencia llevada a cabo y de las adaptaciones que han debido ser abordadas se proponen a continuación unas recomendaciones a la hora de plantear experiencias basadas en robots sociales “In the wild” en espacios públicos para grupos intergeneracionales.

1. Inclusión y accesibilidad

Formación de usuarios: Es necesario proporcionar una explicación interactiva al inicio del juego, adaptada a personas mayores con poca experiencia tecnológica y niños que desconocen cómo interactuar con el robot. Esto facilita un uso autónomo y fomenta la comodidad en la interacción.

Mediación inclusiva: es necesario aplicar técnicas de diseño inclusivo de la experiencia para que el mediador pueda adaptar la interacción a las personas que lo requieran de forma que no se sientan excluidas de la misma.

Adaptabilidad técnica: Incorporar elementos analógicos en el juego que fomenten la participación de los mayores, junto con los elementos digitales. Introducir interacción multimodal, por ejemplo táctil y por voz. Esto ayuda a superar las brechas tecnológicas generacionales y asegura que los participantes con diferentes niveles de familiaridad tecnológica puedan participar de forma equitativa.

2. Dinámicas grupales

Escalabilidad de participantes: Diseñar actividades que permitan la participación de pocos usuarios hasta de grupos grandes dependiendo del día y del contexto. Esto mejora la adaptabilidad del juego para distintos entornos y grupos intergeneracionales y en diferentes situaciones.

Igualdad en la interacción grupal: Introducir elementos que faciliten la interacción de todos los usuarios, por ejemplo, garantizar que se pueda introducir más de una vez la respuesta (correcta o incorrecta), introducir pistas para garantizar que todos los usuarios puedan avanzar, gestionando así diferencias cognitivas y minimizando que la gente se quede bloqueada en la experiencia grupal.

Fomento de colaboración y aprendizaje mutuo: Implementar mecánicas que promuevan la interacción constructiva entre familias y generaciones, intentando minimizar la competitividad excesiva.

3. Gestión del entorno

Gestión del espacio: Planificar el espacio de interacción para evitar aglomeraciones, posibilidad de usar sitios físicos de trabajo. Esto optimizaría la experiencia grupal (si fueran muchos usuarios) al habilitar varios espacios de trabajo.

4. Narrativa y motivación

Narrativa colaborativa: Desarrollar una historia atractiva que fomente el compromiso y establezca una meta común entre los participantes. Este enfoque puede incluir elementos culturales y educativos para enriquecer la experiencia.

5. Flexibilidad y adaptación

Modularidad de las pruebas: Diseñar desafíos que puedan adaptarse a las habilidades de los participantes o limitarse según el tiempo disponible. Esto asegura una experiencia

personalizada y flexible para grupos diversos. La variabilidad en las pruebas, permite que el mediador adapte/seleccione/pase aquellas actividades que sean o no adecuadas para la diversidad generacional y los niveles de habilidad, asegurando la flexibilidad de acuerdo al tiempo disponible o la complejidad percibida.

Duración controlada: Establecer una duración limitada de la actividad que no canse a los usuarios para mantener el interés de participantes de diferentes edades, evitando la fatiga y el cansancio en eventos más amplios.

6. Rol esencial del mediador

Rol activo del mediador: Garantizar la presencia de un mediador que guíe la experiencia, adapte las pruebas según los grupos, resuelva dudas técnicas y, de esta forma, reduzca las limitaciones del robot en situaciones grupales complejas. El mediador asegurará turnos, facilitará la interacción y ajustará la experiencia según las habilidades de los participantes, mejorando la interacción en grupos grandes o con necesidades específicas.

6. Conclusiones

Los resultados del estudio muestran la potencialidad de los robots sociales, para fomentar la interacción y el aprendizaje en entornos grupales intergeneracionales y en contextos públicos no controlados. Las dinámicas observadas confirman que los robots sociales pueden ser herramientas de apoyo para fomentar la interacción intergeneracional de manera enriquecedora para todos los miembros. Así, en las dinámicas se observó a todos los participantes con entusiasmo en la resolución del reto, los mayores o adultos asumían roles de liderazgo y apoyo fomentando la interacción grupal, los niños fueron en mayor medida los que interaccionan con el robot. Algunos aspectos importantes han sido: la generación de dinámicas colaborativas, la definición del rol de mediador, la incorporación de herramientas analógicas y digitales, que permiten reducir brechas tecnológicas entre las diferentes generaciones y un diseño adaptable que permite ayudar a los usuarios que lo necesiten o incluso adaptar el esquema del juego al tipo de grupo, tiempo disponible, etc.

Todos estos aspectos se han plasmado en unas recomendaciones a la hora de llevar a cabo experiencias intergeneracionales con robots “in the wild”. Este tipo de experiencias permiten complementar los hallazgos de aquellas otras más académicas y controladas y explorar el uso de los robots sociales en espacios reales con público numeroso, diverso e intergeneracional.

Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos PID2022-136779OB-C31 (Ministerio de Ciencia e Innovación) y TED2021-130374B-C22 (MCIN/AEI/10.13039/501100011033/ y por la Unión Europea NextGenerationEU/PRTR) y por el Grupo T60_23R (Gobierno de Aragón).

Declaración de IA Generativa

Los autores no han empleado ninguna herramienta de IA Generativa.

Referencias

- [1] M. Jung, P. Hinds, Robots in the wild: A time for more robust theories of human-robot interaction, *J. Hum.-Robot Interact.* 7 (2018). doi:10.1145/3208975.

- [2] S. Joshi, S. Šabanović, Robots for inter-generational interactions: implications for nonfamilial community settings, in: *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction, HRI '19*, IEEE Press, 2020, pp. 478–486.
- [3] S. Šabanović, W.-L. Chang, Socializing robots: constructing robotic sociality in the design and use of the assistive robot paro, *AI SOCIETY* 31 (2016) 537–551. doi:10.1007/s00146- 015- 0636- 1.
- [4] S. You, L. Robert, Teaming up with robots: An imoi (inputs-mediators-outputs-inputs) framework of human-robot teamwork, *International Journal of Robotic Engineering* 2 (2017) 1–7. doi:10.35840/2631- 5106/4103.
- [5] X. Yu, Y. Wang, T. T. M. Tran, Y. Zhao, J. S. Berrio Perez, M. Hoggenmüller, J. Humphry, L. Loke, L. Masuda, C. Parker, M. Tomitsch, S. Worrall, Robots in the wild: Contextually-adaptive human-robot interactions in urban public environments, in: *Proceedings of the 35th Australian Computer-Human Interaction Conference, OzCHI '23*, Association for Computing Machinery, New York, NY, USA, 2024, pp. 701–705. doi:10.1145/3638380.3638440.
- [6] S. Sabanovic, M. Michalowski, R. Simmons, Robots in the wild: observing human-robot social interaction outside the lab, in: *9th IEEE International Workshop on Advanced Motion Control, 2006.*, 2006, pp. 596–601. doi:10.1109/AMC.2006.1631758.
- [7] Y. Rogers, Interaction design gone wild: striving for wild theory, *Interactions* 18 (2011) 58–62. doi:10.1145/1978822.1978834.
- [8] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, C. Pozzi, Child-robot interaction in the wild: advice to the aspiring experimenter, in: *Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI '11*, Association for Computing Machinery, New York, NY, USA, 2011, pp. 335–342. doi:10.1145/2070481.2070545.
- [9] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, W. Woodall, Robot operating system 2: Design, architecture, and uses in the wild, *Science Robotics* 7 (2022). doi:10.1126/scirobotics.abm6074.
- [10] M. Nigro, E. Akinrintoyo, N. Salomons, M. Spitale, Social group human-robot interaction: A scoping review of computational challenges, *arXiv preprint arXiv:2412.16093* (2024).
- [11] M. F. Jung, S. Šabanović, F. Eyssel, M. Fraune, Robots in groups and teams, in: *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW'17 Companion*, Association for Computing Machinery, New York, NY, USA, 2017, p.401–407. doi:10.1145/3022198.3022659.
- [12] F. Cantucci, R. Falcone, Collaborative autonomy: Human–robot interaction to the test of intelligent help, *Electronics* 11 (2022). doi:10.3390/electronics11193065.
- [13] A. Andriella, C. Torras, G. Alenyà, Short-term human–robot interaction adaptability in real-world environments, *International Journal of Social Robotics* 12 (2020). doi:10.1007/s12369- 019- 00606- y.
- [14] M. Żarkowski, Short-term human–robot interaction adaptability in real-world environments multi-party turn-taking in repeated human–robot interactions: An interdisciplinary evaluation, *International Journal of Social Robotics* 11 (2019) 693–707. doi:10.1007/s12369- 019- 00603- 1.

-
- [15] C. Bartneck, T. Belpaeme, F. Eysel, T. Kanda, M. Keijsers, S. Šabanović, Human-robot interaction: an introduction, Cambridge University Press, 2020. doi:10.1017/9781108676649.
- [16] S. Hopko, J. Wang, R. Mehta, Human factors considerations and metrics in shared space human-robot collaboration: A systematic review, *Frontiers in Robotics and AI* 9 (2022). doi:10.3389/frobt.2022.799522.
- [17] H. Javed, N. Jamali, Modeling group dynamics for personalized robot-mediated interactions, 2023. URL: <https://arxiv.org/abs/2306.16633>. ArXiv:2306.16633.
- [18] T. Kanda, T. Hirano, D. Eaton, H. Ishiguro, Interactive robots as social partners and peer tutors for children: A field trial, *Human Computer Interaction (Special issues on human-robot interaction)* 19 (2004) 61–84. doi:10.1207/s15327051hci1901&2_4.
- [19] M. Fridin, Kindergarten social assistive robot: First meeting and ethical issues, *Comput. Hum. Behav.* 30 (2014) 262–272. doi:10.1016/j.chb.2013.09.005.
- [20] K. Wada, T. Shibata, Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house, *IEEE Transactions on Robotics* 23 (2007) 972–980. doi:10.1109/TRO.2007.906261.
- [21] J. Navarro, E. Cerezo, Robots Sociales en Experiencias Intergeneracionales: Diseño de una Experiencia, Master’s thesis, Universidad de Zaragoza, 2024.

Evaluación de detectores de estrés durante la interacción con juegos serios que estimulan actividades cognitivas

Evaluation of stress detectors during interaction with serious games that stimulate cognitive activities

Leticia Martínez-Limeres^{1,†}, Nelly Condori-Fernández^{2,3,†} and Alejandro Catala^{2,3†}

¹ *Universidade de Santiago de Compostela, Santiago de Compostela, Spain*

² *Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15782 Santiago*

³ *Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain*

Resumen

Este trabajo compara dos métodos de detección automática de estrés: uno basado en señales fisiológicas mediante la pulsera Empatica EmbracePlus y otro en expresiones faciales utilizando la herramienta Kopernica. El estudio se llevó a cabo en un entorno controlado, donde los participantes interactuaron con un juego cognitivo como un medio para inducir estrés. Se observó que el juego genera un aumento en los niveles de estrés, siendo el método fisiológico más sensible a variaciones leves. El análisis mostró una mayor coherencia entre los valores de EDA y el indicador *stress intens*, mientras que la percepción subjetiva del estrés difirió notablemente. La resolución de captura de vídeo influyó en la precisión del análisis facial, siendo 640x360 la resolución mínima aceptable. Los resultados evidencian fortalezas y limitaciones de ambos métodos, sugiriendo mejoras en el procesamiento de datos y la necesidad de investigar otras emociones y poblaciones para estudiar el estrés en diferentes etapas de la vida.

Abstract

This paper compares two methods of automatic stress detection: one based on physiological signals using the Empatica EmbracePlus wristband and the other on facial expressions using the Kopernica tool. The study was conducted in a controlled environment, where participants interacted with a cognitive game as a mean to induce stress. The game was observed to generate an increase in stress levels, with the physiological method being more sensitive to slight variations. The analysis showed a greater consistency between EDA values and the *stress intens* indicator, while the subjective perception of stress differed notably. Video capture resolution influenced the accuracy of facial analysis, with 640x360 being the minimum acceptable resolution. The results show strengths and limitations of both methods, suggesting improvements in data processing and the need to investigate other emotions and other populations to study stress at different stages of life.

Keywords

detección de estrés, señales fisiológicas, expresiones faciales, EDA, juegos serios

1. Introducción

El estrés es una respuesta natural del organismo ante situaciones percibidas como amenazantes o desafiantes [1]. Su detección temprana es fundamental para prevenir efectos negativos en la salud mental y física. Existen diversas metodologías propuestas para monitorizar y gestionar el estrés, que incluyen técnicas basadas en señales fisiológicas y en el análisis de expresiones faciales [2]. Sin embargo, los estudios realizados hasta la fecha muestran una gran variabilidad en las tecnologías

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

* Corresponding author.

† These authors contributed equally.

leticia.martinez.limeres@rai.usc.es (L. Martínez-Limeres); ✉ n.condori.fernandez@usc.es (N. Condori-Fernández); alejandro.catala@usc.es (A. Catala)

 0000-0002-1044-3871 (N. Condori-Fernández); 0000-0002-3677-672X (A. Catala)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

(como los sensores fisiológicos), en los métodos aplicados, en los perfiles de los participantes, y en los contextos experimentales, lo que afecta directamente en los resultados obtenidos.

Esta heterogeneidad dificulta la comparación directa entre diferentes herramientas de detección de estrés y, por consiguiente, impide determinar con claridad cuál metodología resulta más efectiva en términos de precisión y confiabilidad.

Aunque se han desarrollado numerosos métodos de detección de estrés [2], la mayoría de los estudios se centran en evaluar soluciones individuales. Las comparaciones entre enfoques suelen basarse en resultados aislados de distintos trabajos, lo que impide valorar su desempeño en igualdad de condiciones. Esta falta de análisis comparativos controlados limita la posibilidad de extraer conclusiones sobre la precisión y aplicabilidad. En este trabajo, se propone una evaluación empírica bajo un protocolo experimental, utilizando un juego serio digital como inductor de estrés. El presente estudio evalúa dos métodos de detección de estrés: uno basado en señales fisiológicas, presentado en [3], que utiliza la actividad electrodérmica (EDA, por sus siglas en inglés Electrodermal Activity) colectada con la pulsera Empatica EmbracePlus [4]; y otro basado en el análisis de expresiones faciales, utilizando la herramienta Kopernica Human [5] desarrollado por la empresa Neurologica [6].

El objetivo es no solo evaluar su efectividad en condiciones controladas, sino también proporcionar información sobre su capacidad para gestionar el estrés, teniendo en cuenta las variaciones tecnológicas y fisiológicas entre diferentes metodologías.

El documento se organiza de la siguiente manera: la Sección 2 presenta la revisión del estado del arte sobre el reconocimiento y detección automática del estrés. La Sección 3 describe la metodología empleada en el estudio, detallando el procedimiento experimental, las fases del experimento, los participantes y otros aspectos relevantes. La Sección 4 presenta y discute los resultados obtenidos. Finalmente, la Sección 5 ofrece las conclusiones y las principales aportaciones del trabajo, así como propuestas de futuras líneas de investigación o posibles mejoras.

2. Estado del arte

El avance de la computación afectiva ha permitido el desarrollo de dispositivos capaces de detectar el estrés en tiempo real, empleando sensores fisiológicos para medir respuestas del sistema nervioso y/o técnicas de visión por ordenador para analizar expresiones faciales. Ambas aproximaciones han demostrado ser útiles de forma individual o combinada, aunque presentan ventajas y limitaciones.

La Tabla 1 resume algunos de los estudios más representativos, destacando para cada estudio: el nombre de los autores, los clasificadores empleados, las señales de estrés analizadas, el tipo de estresores utilizados durante el experimento, el número de participantes implicados, la precisión (accuracy) alcanzada por los modelos y, finalmente, el número de clases en las que se categoriza el nivel de estrés (por ejemplo: estrés vs. no estrés, o niveles bajo, medio y alto). Algunos trabajos, como Kim et al. (2004) [7], que emplearon señales fisiológicas como la actividad electrodérmica, la temperatura de la piel y el ritmo cardíaco para detectar el estrés con un 78,4% de precisión, mientras que otros, como el trabajo de Womack y Hansen [8], se centraron en el análisis del habla. De forma más reciente, Sabour et al. (2021) [9] combinaron datos fisiológicos y faciales, obteniendo una precisión del 85,48%, aunque sin detallar el desempeño individual. Trabajos recientes (p. ej [10, 11, 12]) destacan el uso de señales fisiológicas y dispositivos portátiles (wearables), permitiendo recopilar datos sin necesidad de procedimientos invasivos para la detección de estrés en contexto cotidianos.

A pesar de los avances, la mayoría de los estudios presentan limitaciones: evaluaciones centradas en un método de detección, condiciones experimentales heterogeneas, resultados difíciles de comparar debido a la variabilidad en estresores, señales y perfiles de participantes. Este trabajo aporta un enfoque diferencial al comparar dos métodos de detección de estrés bajo un protocolo experimental común. Para ello, se empleó como estresor una versión digital del ejercicio Paper

Folding, un juego serio de carga cognitiva moderada y disponible públicamente [13], lo que facilita la reproducibilidad del experimento y su aplicación en contextos reales.

Tabla 1

Comparativa de trabajos centrados en el uso de detectores basados en EDA y/o expresiones faciales

Estudio	Clasificadores	Señal de estrés	Estresores	Participantes	Accuracy	Clases
Kim et al. (2004) [7]	SVM	EDA, SKT, ECG	Vídeos, Imágenes, Sonidos	50	78,4 %	3
Lee et al. (2004) [14]	MLP, GRNN, ANFIS	EDA, SKT, IIRV	SCWT	80	96,67 %	2
Healey et al. (2005) [15]	LDA	ECG, EMG, EDA, RSP	Tarea de conducción	24	97,3 %	3
Zhai et al. (2006) [16]	SVM	EDA, BVP, PD, SKT	Paced Stroop Test	32	90,1 %	3
Katsia et al. (2008) [17]	SVM, ANFIS	EMG, ECG, EDA, RSP	Tarea de conducción	30	79,3 %	4
Sierra et al. (2010) [18]	KNN, FDA	EDA-PPG	Hiperventilación y preparaci	19	90,7 %	2
Setz et al. (2010) [19]	LDA, SVM-RBF, NCC	EEG, BVP, EDA, IIRV, RR MIST		33	82,8 %	3
Hoscini et al. (2011) [20]	SVM, ENN	EEG, BVP, EDA	IAPS	15	84,6 %	2
Wijsman et al. (2011) [21]	Bayes, KNN, mínimos de Fisher	ECG, EMG, EDA	Tareas de aritmética, rompe	80	80 %	3
Hernández et al. (2011) [22]	SVM	EDA	Atender a llamadas	20	98,9 %	2
Giakoumis et al. (2012) [23]	LDA	EDA, ECG	SCWT	21	96,6 %	2
Kurniawan et al. (2013) [24]	k-means, GMM, SVM, Árbol de	Expresiones faciales	SCWT, TSST	10	92,6 %	2
Pedrotti et al. (2014) [25]	NN	Habla, EDA	Prueba de cambio de carril	33	79,2 %	2
Sharma et al. (2014) [26]	GA, SVM, ANN	EEG, EDA, SKT	Actividad de meditación	13	99,6 %	2
Liapis et al. (2015) [27]	LDA	EDA	Tareas de interacción persor	31	98,9 %	3
Ghadery et al. (2015) [28]	SVM, kNN	EDA, EMG, ECG	Tarea de conducción	20	98 %	3
Abouelenien et al. (2016) [29]	Árbol de decisión	PPG, EDA, Respiración,	Detección de mentiras	50	73 %	2
Giannakakis et al. (2017) [30]	kNN, SVM, Naive Bayes, AdaBo	Expresiones faciales	IAPS	23	91,68 %	3
Mozos et al. (2017) [31]	kNN, SVM, AdaBoost	EDA, PPG, Habla, Acele	TSST	18	94 %	2
Anusha et al. (2018) [32]	QDA, kNN	EDA, ECG, SKT	SCWT, MAT, TSST	34	97,13 %	3
Airij (2018) [33]	Lógica difusa, kNN	IIR, EDA, SKT	MAT	35	96,19 %	3
Yekta et al. (2019) [34]	RF, MLP	PPG, EDA, ACC	Concurso	21	97 %	2
Suni et al. (2020) [3]	ADWIN	EDA	Alarma fuego, SCWT	12	79,17 %	2
Indikawati and Winiarti (2020) [35]	Random Forest	PPG, EDA, SKT	TSST	15	96,68 %	2
Sabour et al. (2021) [9]	SVM, RBF, KNN, Regresión Logis	EDA, BVP, PRV, PPG, Ex	TSST	25	85,48 %	3
Mohammadi et al. (2022) [36]	STAIT	ECG, EDA, SKT	MAT, SCWT	18	94,4 %	2
Rodrigues and Correia (2023) [37]	Regresión Logística, SVM, Árbol	IIRV, PERCLOS, Expresio	Cuestionarios	28	86,6 %	2

3. Metodología

En esta sección se profundiza en las técnicas, herramientas y métodos empleados para la ejecución completa del experimento. Se exponen las preguntas de investigación, las hipótesis, las variables involucradas y las validaciones contempladas en el desarrollo de este estudio. Asimismo, se describen los pasos del tratamiento y procesamiento de los datos efectuados.

La sección ha sido elaborada conforme a los criterios establecidos para una adecuada planificación experimental, según lo indicado en [15]. Con el fin de garantizar la validez del experimento, se han considerado diversos factores clave, como la selección de dispositivos de medición fisiológica, la elección de un inductor de estrés adecuado y la definición de un entorno controlado que asegure la comodidad de los participantes. Además, se ha establecido un protocolo

experimental detallado que especifica cada paso del procedimiento, asegurando la uniformidad y consistencia en la ejecución de las pruebas.

3.1. Dispositivos de medición fisiológica

Para la medición de las señales fisiológicas de los participantes, se emplearon dos dispositivos principales. Por un lado, la pulsera Empatica EmbracePlus [4], seleccionada por su capacidad para registrar de manera no invasiva la actividad electrodérmica, un indicador directo de la activación del sistema nervioso simpático, que está estrechamente relacionado con la respuesta al estrés. Por otro lado, se empleó la herramienta Kopernica Human [5], desarrollada por Neurologica [6], para analizar las expresiones faciales de los participantes. Esta herramienta utiliza inteligencia artificial basada en redes neuronales para interpretar los estados emocionales a través de visión por computadora, análisis de lenguaje natural y microexpresiones faciales (pequeños movimientos musculares involuntarios que ocurren en el rostro durante fracciones de segundo y que reflejan estados emocionales auténticos, incluso cuando se intenta ocultarlos conscientemente). En el presente estudio, se emplearon dos indicadores generados por Kopernica Human: *stress* y *stress intens*, cuyo detalle se describen en la sección de variables.

3.2. Inductor de estrés

Para inducir estrés en los participantes, se utilizó el juego serio “Paper Folding - a Spatial Reasoning Game” [13] debido a su capacidad para generar carga cognitiva y estrés. Este juego se enfoca en el razonamiento espacial a través de la papiroflexia, desafiando a los jugadores a replicar patrones de doblado en un papel virtual.

En una segunda parte del mismo, los participantes deben identificar la cantidad y ubicación de las perforaciones en el papel virtual, el cual ha sido previamente doblado y luego desplegado. Deben seleccionar la respuesta correcta entre cinco opciones en un tiempo limitado, lo que incrementa la presión cognitiva. Además, la dificultad del juego se ajusta dinámicamente según el rendimiento del usuario, asegurando un desafío constante y adaptado a su nivel.

Más allá del diseño del propio juego, otros factores como la limitación temporal, los estímulos visuales del entorno y la experiencia previa de los jugadores pueden influir en el nivel de estrés percibido. Para validar la efectividad del juego como inductor de estrés, se realizaron pruebas funcionales y no funcionales que garantizaron su fiabilidad, usabilidad y capacidad de adaptación de la dificultad del mismo. Las pruebas funcionales se enfocaron en validar los componentes del juego, asegurando que el software respondiera adecuadamente a las acciones del usuario y que todos los elementos de la interfaz funcionaran según lo previsto. Por su parte, las pruebas no funcionales se centraron en varios aspectos: pruebas de usabilidad para evaluar la facilidad de uso del juego y la claridad de las instrucciones; pruebas de seguridad para verificar la seguridad de los datos guardados, recogidos de forma anónima e identificados mediante un identificador proporcionado; pruebas de compatibilidad para comprobar el funcionamiento correcto del juego en el dispositivo con el que se va a interactuar; pruebas de fiabilidad para garantizar que los datos recogidos durante la interacción con el juego se registrarán adecuadamente, evitando confusiones entre, por ejemplo, aciertos y fallos; y las pruebas de adaptabilidad para evaluar la capacidad del juego para ajustar su nivel de dificultad según las acciones del usuario y la retroalimentación recibida.

3.3. Preguntas e hipótesis

3.3.1. Preguntas de investigación

A partir del análisis previo y de los objetivos definidos en este trabajo, se establecen las siguientes preguntas de investigación, que guían el desarrollo y la validación del experimento:

- RQ1: ¿El juego cognitivo diseñado actúa como un estresor capaz de influir en el nivel de estrés del usuario, incrementando los indicadores registrados por los dispositivos de medición?
- RQ2: ¿En qué medida los dos métodos de detección de estrés empleados son congruentes entre sí y cómo se relacionan con el incremento de dificultad en las tareas del juego?
- RQ3: ¿Existe una relación entre las características extraídas a través de los sistemas automáticos de detección y el nivel de estrés percibido por los propios participantes?
- RQ4: ¿Cómo afecta la reducción de resolución del vídeo al rendimiento del sistema de análisis de expresiones faciales basado en imagen, y qué impacto tiene sobre el tiempo de procesamiento?
- RQ5: ¿Cuál es el umbral de reducción de resolución en el que se observa una disminución significativa en la precisión de detección de los indicadores de estrés?

3.3.2. Variables e hipótesis

En base al protocolo experimental diseñado, se formulan las siguientes hipótesis, estrechamente vinculadas a las preguntas de investigación mencionadas:

- Hipótesis 1 (H1): el juego de razonamiento cognitivo actúa como un estresor y, por consiguiente, su exposición elevará significativamente el nivel de estrés del usuario. Esta respuesta se reflejará en un aumento de los indicadores registrados tanto por la pulsera Empatica EmbracePlus como por el software Kopernica Human. Relacionada con RQ1.
- Hipótesis 2 (H2): los métodos de detección de estrés analizados serán congruentes entre sí y tendrán una relación positiva con el aumento de la dificultad de las tareas del juego, reflejando un incremento en los indicadores de estrés a medida que se añade complejidad al juego. Esto se manifestará en una elevación progresiva de los indicadores de estrés a medida que incrementa la complejidad de las pruebas. Hipótesis vinculada a RQ2.
- Hipótesis 3 (H3): se anticipa que los participantes tenderán a subestimar su nivel de estrés en comparación con los datos registrados por los sistemas de medición, lo que pondría de manifiesto una posible discrepancia entre el estrés real y la percepción subjetiva de los usuarios. Esta hipótesis da respuesta a la RQ3.
- Hipótesis 4 (H4): se plantea que una reducción en la resolución del vídeo impactará en el rendimiento del sensor de detección de expresiones faciales basado en imagen. Este cambio podría traducirse en una disminución tanto en la precisión de los resultados obtenidos como en el tiempo de procesamiento, que permita determinar una resolución óptima que equilibre la precisión y el coste computacional. Relacionada con RQ4 y RQ5.

Para validar estas hipótesis y responder de manera rigurosa a las preguntas de investigación planteadas, se han definido las siguientes variables independientes y dependientes:

- Variables 1: la variable independiente (VI1) es la fase o etapa experimental (Training, Test y Master), mientras que la variable dependiente (VD1) son los indicadores de estrés registrados por los sistemas de detección utilizados. Estas variables están asociadas a las hipótesis H1 y H2.
- Variables 2: la variable independiente (VI2) es el nivel de estrés registrado por los dispositivos de medición automática. En el caso de Kopernica, se utilizaron dos indicadores que la propia herramienta etiqueta como *stress* y *stress intens*, que se sustentan en la evaluación psicométrica para cuantificar la experiencia del estrés agudo. La variable dependiente es la percepción subjetiva de estrés (VD2) reportada por el propio participante. Esta relación corresponde a la hipótesis H3.

- Variables 3: la variable independiente (VI3) es la resolución del vídeo y la dependiente es la precisión del método de detección de estrés (VD3) a partir de expresiones faciales asociado al sistema Kopernica Human. Están relacionadas con la hipótesis H4.

3.4. Entorno e instrumentación del experimento

El experimento se lleva a cabo en el laboratorio, diseñado para ofrecer un entorno controlado y confortable. Como se muestra en la Figura 1, el espacio está dividido en dos zonas: una destinada a la interacción de los participantes con el juego en un ordenador táctil y otra desde donde se supervisa y registra el desarrollo del estudio.



Figura 1: Entorno del experimento.

Para la recolección de datos, se emplean dos dispositivos: la pulsera inteligente Empatica EmbracePlus para medir la conductancia de la piel y una cámara de grabación conectada a un ordenador con procesador Intel Core i7-10750H, utilizado también para análisis posteriores. Estos dispositivos permiten registrar indicadores clave de estrés, proporcionando información esencial para el estudio.



Figura 2: Dispositivos del experimento.

3.5. Procedimiento experimental

El experimento tiene como objetivo comparar dos detectores de estrés en un entorno controlado, a través de la interacción de los participantes con un juego cognitivo. El estudio se desarrolla en varias fases claramente definidas.

En primer lugar, se recibe al participante y se le acomoda en el espacio destinado a la prueba. Se le coloca la pulsera Empatica EmbracePlus en la mano no dominante y se ajusta la cámara para asegurar una correcta grabación de sus expresiones faciales. A continuación, se le proporcionan las instrucciones necesarias para realizar la actividad y se le entrega el documento de consentimiento informado, que debe leer y firmar antes de continuar.

Una vez completada esta etapa, se inicia un breve período de relajación mediante la proyección de un vídeo² de aproximadamente dos minutos. Esto permite establecer una línea base en un estado de calma, minimizando la influencia de posibles factores externos de estrés previos al experimento.

Tras la relajación, el participante interactúa con el juego cognitivo, dividido en tres modos de dificultad creciente: Training, Test y Master. Durante esta fase, se recopilan datos tanto fisiológicos, a través de la pulsera, como de expresiones facial, mediante la cámara. Finalizada la interacción con el juego, se detienen las mediciones automáticas y el participante completa un cuestionario³ en el que evalúa su nivel de estrés percibido durante la sesión.

La duración aproximada de la interacción con el juego oscila entre 30 y 45 minutos, a lo que se suman unos minutos adicionales para la preparación inicial, el consentimiento informado, el período de relajación y la cumplimentación del cuestionario final. En total, la duración estimada de cada sesión experimental es de entre 45 y 60 minutos.

Tabla 2

Tiempos estimados para fase del experimento

Fase experimento	Tiempo
Recepción del sujeto	3 minutos
Instrucciones y consentimiento informado	5 minutos
Periodo de relajación	2 minutos
Interacción con el juego	30-45 minutos
Cuestionario de estrés percibido	5 minutos

El estudio involucra a 28 voluntarios, mayoritariamente estudiantes jóvenes de diversas disciplinas como ingeniería, matemáticas, farmacia, economía, y educación primaria, además de opositores. La muestra se caracteriza por un equilibrio casi igualitario entre géneros, con 15 hombres y 13 mujeres, con el objetivo de asegurar una muestra representativa de la comunidad universitaria.

3.5.1. Validación del experimento

Para garantizar la validez y fiabilidad del experimento se consideraron distintos aspectos clave. En cuanto a la validez interna, se procuró que los niveles de estrés medidos por la pulsera Empatica EmbracePlus y la herramienta Kopernica Human fueran consecuencia directa de la exposición al juego de razonamiento cognitivo, minimizando la influencia de factores externos mediante un período de relajación previo. Sin embargo, el uso de un juego como inductor podría no generar respuestas consistentes en todos los participantes. La validez externa se ve limitada por el perfil de la muestra, compuesta mayoritariamente por estudiantes, lo que dificulta la generalización de los resultados. Respecto a la validez de constructo, se evaluó la precisión y consistencia de los dispositivos empleados, aunque persisten riesgos como fallos en la captura de datos, influencia del

²Video del periodo de relajación: <https://nextcloud.citius.gal/s/rw6Z9kttkaGmYt>

³Cuestionario estrés percibido: <https://nextcloud.citius.gal/s/9pwtEXLqRqyiRgq>

entorno de laboratorio o respuestas no sinceras en los cuestionarios. Por último, se prestó especial atención a la validez de conclusión, controlando posibles factores de sesgo que pudieran distorsionar las relaciones observadas entre las variables y afectar la interpretación de los resultados.

3.6. Tratamiento de los datos

El tratamiento de los datos se estructuró en tres etapas fundamentales, diseñadas para asegurar su calidad y su idoneidad para el análisis posterior (ver Figura 3):

Recolección de los datos: los datos fueron obtenidos siguiendo el Consentimiento Informado y la aprobación del Comité de Bioética de la Universidad, asegurando que los participantes estuvieran plenamente informados sobre los objetivos del estudio, los procedimientos involucrados y el uso de su información. Para preservar la privacidad y evitar la identificación personal, los datos registrados fueron exclusivamente numéricos, eliminando cualquier información sensible que pudiera comprometer la confidencialidad. La recolección de la señal fisiológica de actividad electrodérmica se realizó mediante la pulsera Empática, que midió la conductancia de la piel en microsiemens. Este dispositivo permitió un monitoreo continuo durante las sesiones experimentales, generando datos en tiempo real que fueron posteriormente analizados para identificar patrones de activación fisiológica. Simultáneamente, se evaluaron las expresiones faciales mediante Kopernica Human, una herramienta de reconocimiento facial que tradujo las microexpresiones en datos numéricos. Este sistema procesó los videos de los participantes, identificando cambios sutiles en la musculatura facial asociados a diferentes estados emocionales. La combinación de estas dos fuentes de datos permitió obtener una visión integral de las reacciones fisiológicas y conductuales de los participantes, facilitando un análisis más completo del estrés en diversas condiciones experimentales.

Preparación de los datos: en esta fase, los datos fueron convertidos a formato CSV para facilitar su manejo y análisis. Se trabajó con tres conjuntos principales: los datos del juego, que fueron exportados automáticamente a CSV, registrando las interacciones de los participantes en tres modos: entrenamiento, test y master, y se verificó la ausencia de valores nulos y duplicados; las señales fisiológicas, cuyos datos de la pulsera Empatica EmbracePlus fueron convertidos de AVRO a CSV mediante un script en Python, manteniendo las columnas de marca de tiempo, EDA y frecuencia de muestreo; y los videos, que fueron convertidos de MKV a MP4 y recodificados en distintas resoluciones (1920x1080, 720x406, 640x360 y 320x180) para su análisis. Se generó un archivo CSV por participante y resolución, asegurando la compatibilidad con la herramienta Kopernica Human. Durante todo el proceso, se realizaron verificaciones y sincronizaciones para garantizar la coherencia y la integridad de los datos.

Procesamiento de los datos: en esta etapa, se aplicó un detector de estrés a las señales EDA, transformando los valores en un rango de 0 a 5 para clasificar los niveles de estrés. Los valores 4 y 5 indicaron presencia de estrés, mientras que los valores restantes representaron ausencia de estrés. Además, las expresiones faciales se procesaron mediante Kopernica Human, que generó dos métricas de estrés distintas: stress y stress intens. Ambos valores fueron normalizados para su análisis comparativo. Con los datos procesados, se llevaron a cabo las pruebas experimentales que se exponen a continuación.

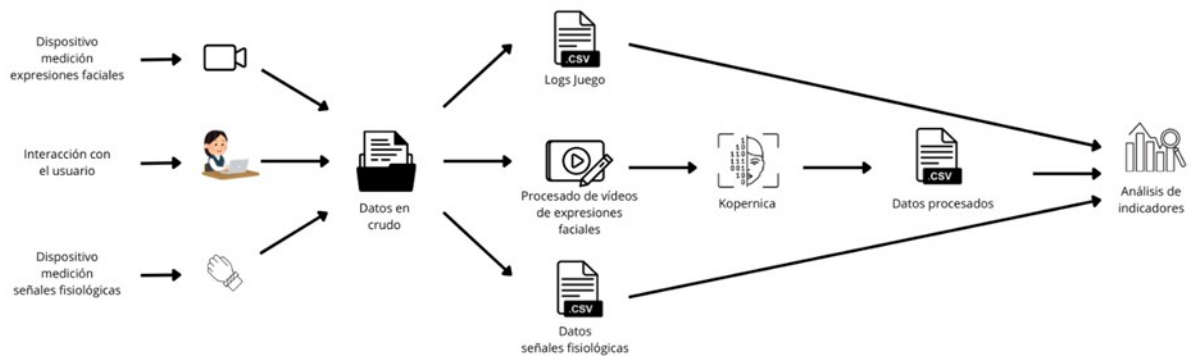


Figura 3: Tratamiento de datos del experimento.

4. Resultados experimentales

Se llevaron a cabo cuatro pruebas principales con los datos recolectados durante el experimento:

1. La primera prueba analiza si la interacción con el juego serio incrementa los niveles de estrés de los participantes (RQ1 y H1).
2. La segunda prueba evalúa el impacto de la dificultad de las tareas sobre los indicadores de estrés de ambos detectores, así como si esta influencia es similar en ambos sistemas (RQ2 y H2).
3. La tercera prueba examina la relación entre los indicadores de los detectores y la percepción subjetiva que los participantes tienen de su propio estrés (RQ3 y H3).
4. Finalmente, la cuarta prueba investiga cómo la resolución del vídeo afecta el procesamiento, la calidad de los datos y el rendimiento de los detectores de estrés (RQ4, RQ5 y H4).

Prueba 1. Influencia del juego serio en los niveles de estrés

Se analizó si el juego incrementa los indicadores de estrés (EDA y expresiones faciales) en los usuarios, comparando los niveles de estrés en distintas fases del juego con respecto al inicio. Los resultados mostraron un aumento del estrés, especialmente al comenzar y durante el modo de entrenamiento. A través de mapas de calor (Figura 4 y Figura 5), se validaron los datos y se analizó la respuesta individual de los participantes. De acuerdo a Figura 4, se observa que varios participantes experimentaron estrés a media que la dificultad del juego se incrementaba. Así como, no estrés en la fase de relajación.

Aplicando la prueba de McNemar para analizar si existían diferencias significativas en la detección automática de estrés entre las distintas fases del experimento. Por ejemplo, en la comparación entre las fases de Inicio + Relajación y Master no arrojó resultados estadísticamente significativos, a pesar de observarse cambios individuales en ambos sentidos. Del mismo modo, la comparación entre las fases de Entrenamiento y Master también resultó no significativa. Estos resultados sugieren que, aunque el número de detecciones de estrés varió entre fases a nivel individual, no se produjo un cambio sistemático en la clasificación binaria de estrés que permita afirmar diferencias significativas entre fases en el conjunto de la muestra, lo que no corrobora la hipótesis H1 en cuanto a la medición mediante EDA.

En cuanto a los resultados obtenidos a partir de las expresiones faciales (Ver Figura 5), observamos que la mayoría de los participantes experimentaron un aumento notable en los indicadores de estrés en todas las fases.

De acuerdo a los mapas, se destaca la sensibilidad de la EDA para detectar estrés, especialmente en las fases activas del juego, mientras que las expresiones faciales reflejan principalmente un nivel medio de estrés. Sin embargo, la ausencia de diferencias significativas entre fases observada en la prueba de McNemar puede deberse a múltiples factores; entre ellos, se encuentra la posible adaptación progresiva de los participantes al entorno del experimento, el carácter moderado del estresor utilizado (Paper Folding) y las limitaciones inherentes a la detección automática de estrés, en especial cuando se utilizan expresiones faciales como fuente principal. Además, la codificación

binaria utilizada en el análisis podría haber simplificado excesivamente las variaciones emocionales sutiles entre fases.

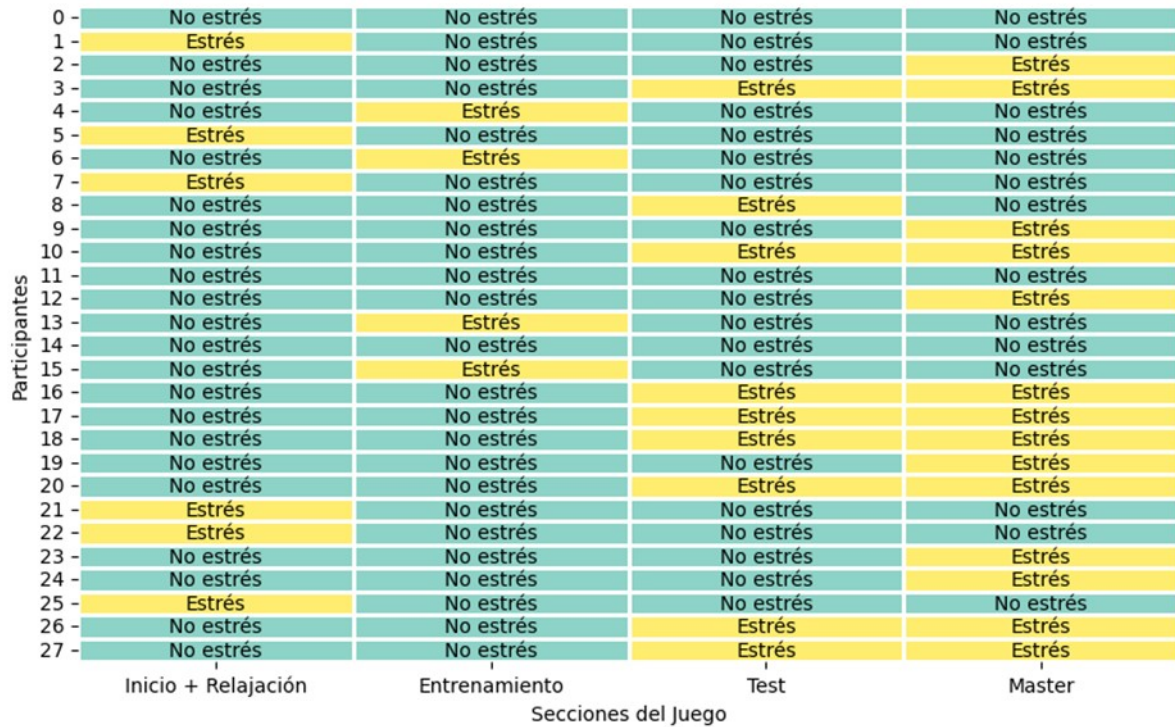


Figura 4: Mapa de calor binario que representa presencia de estrés detectado automáticamente a partir de EDA en cada fase del experimento (relajación y fases del juego).

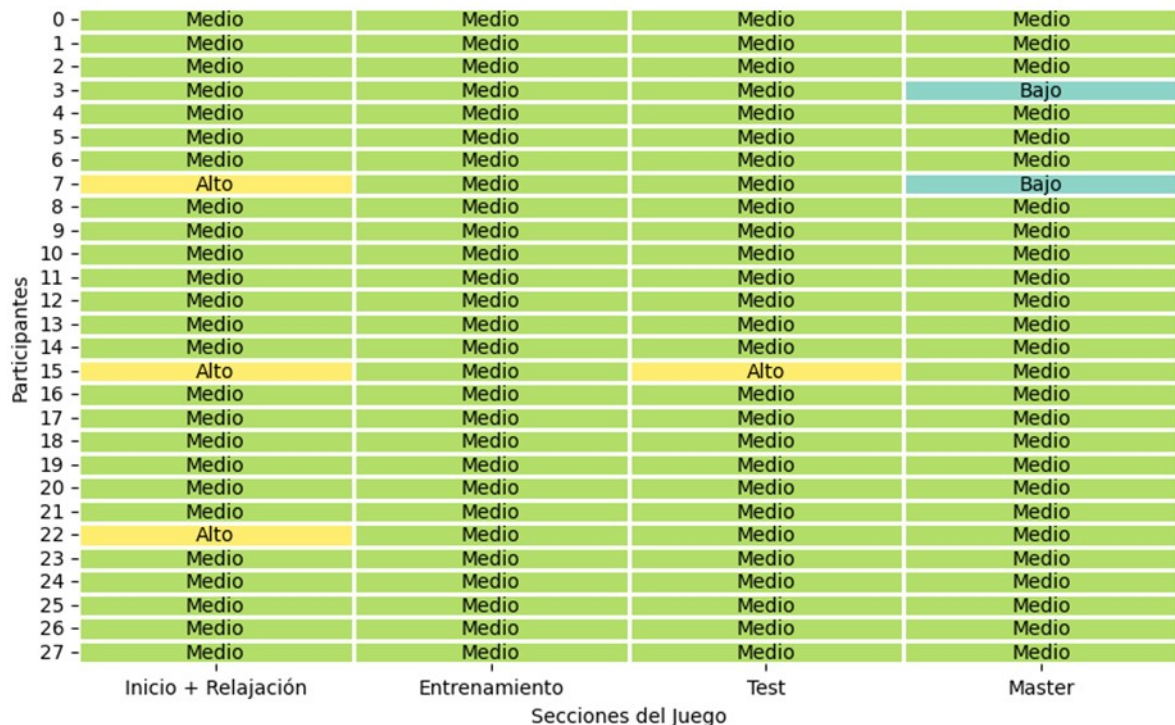


Figura 5: Mapa de calor que representa nivel de estrés detectado a partir de expresiones faciales en cada fase del experimento (relajación y fases del juego).

Prueba 2. Impacto de la dificultad de las tareas

Se estudió si la complejidad de las tareas, según el número de pliegues requeridos, incrementa el estrés, clasificando las tareas en nivel fácil, medio y difícil, y analizando los valores medios de EDA y expresiones faciales en cada nivel. Tal como se observa en Figura 6 (lado izquierdo), los resultados indican que el estrés aumenta con la dificultad de las tareas. Al analizar la relación entre la dificultad y el estrés medido por EDA, se observó una correlación positiva (0.31 en la correlación de Spearman). Sin embargo, mientras que no se encontró una relación significativa en las expresiones faciales (Figura 6, lado derecho), que mostraron una correlación negativa débil (-0.07). En conclusión, la dificultad de las tareas tiene un impacto sutil en los niveles de estrés detectados por EDA, pero no en las expresiones faciales, lo que resalta la mayor sensibilidad de la EDA a los cambios fisiológicos relacionados con el estrés en comparación con las expresiones faciales.

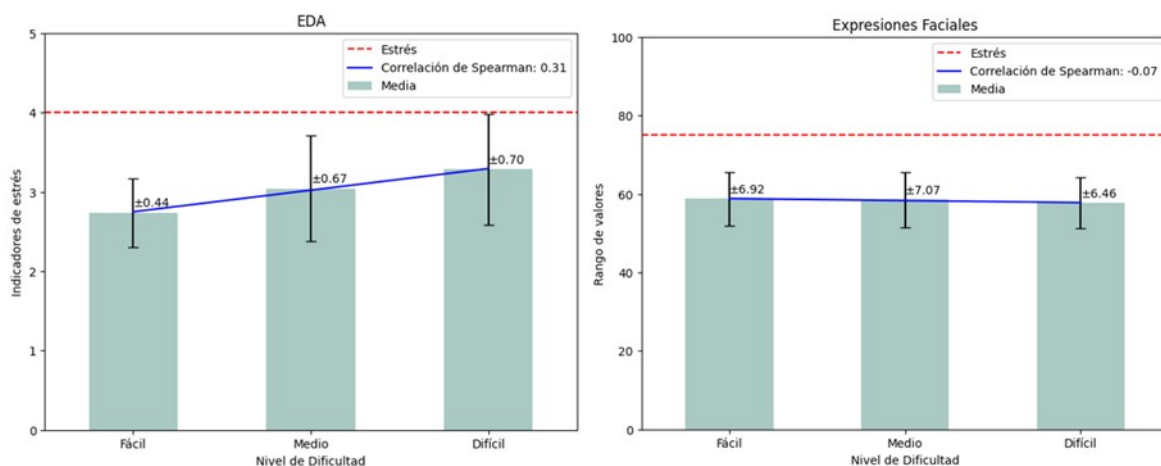


Figura 6: Relación entre complejidad de las tareas y nivel medio de los indicadores de los detectores.

Prueba 3. Relación entre detectores y la percepción de los usuarios

La Figura 7 muestra la tendencia promedio de los valores normalizados obtenidos por los distintitos métodos de detección de estrés para cada participante. Se comparan el nivel de estrés percibido por los propios usuarios (cuestionarios), nivel de estrés derivado a partir la señal EDA capturada por la pulsera, y los dos indicadores generados por Kopernica (*stress* y *stress intens*).

Se observa una notable dispersión entre el estrés percibido y los valores de estrés obtenidos con los métodos automáticos, con una mayor variabilidad en la percepción subjetiva (línea azul). En cuanto a los valores proporcionados por Kopernica, el indicador de *stress intens* (línea roja) muestra una evolución más constante y se mantiene más cercano a los valores derivados de las señales de EDA (línea naranja), lo que sugiere que esta métrica podría estar captando una activación fisiológica acumulativa similar. En cambio, el valor de *stress* (línea verde) presenta valores más bajos y mayor distancia respecto a las otras señales, con un patrón menos correlacionado.

En consecuencia, la hipótesis H3 no se confirma plenamente, aunque se identifican patrones interesantes que invitan a considerar métricas como *stress intens* y los obtenidos a partir de la señal EDA como posibles indicadores más sensibles a la experiencia emocional interna que los valores subjetivos (estrés percibido).

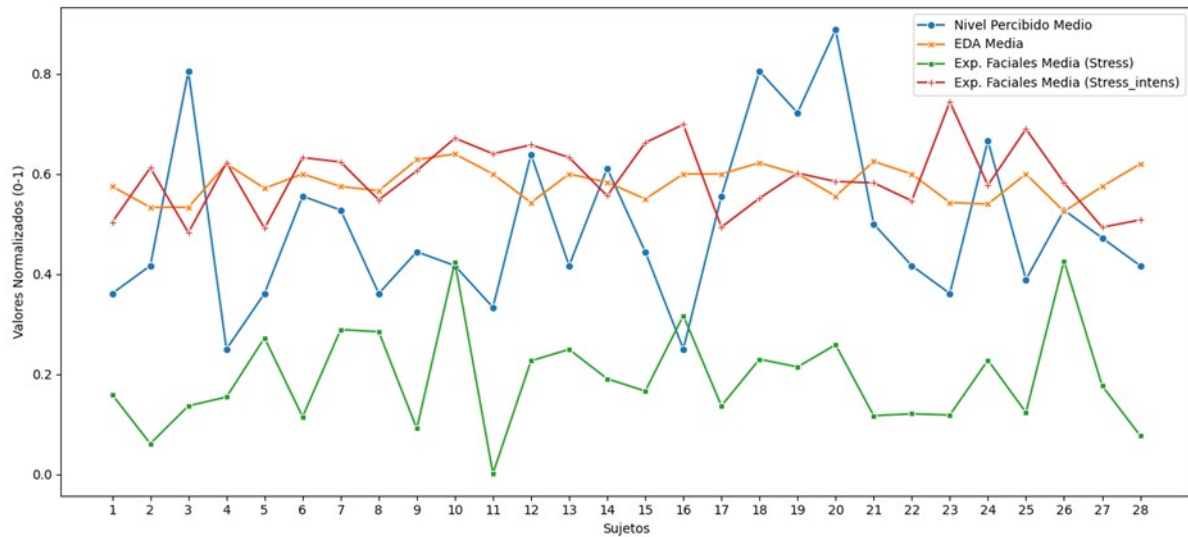


Figura 7: Tendencia de la relación entre los indicadores de ambos detectores y la percepción de los usuarios.

Prueba 4. Impacto de la resolución en el procesamiento y rendimiento.

Se analizó el impacto de la resolución de los vídeos en la fiabilidad del análisis de expresiones faciales como indicador de estrés, comparando las correlaciones de Spearman entre diferentes resoluciones. Tal como se observa en Figura 8, los resultados mostraron correlaciones positivas en todos los casos, lo que indica que la resolución no afecta significativamente la estabilidad de los indicadores. Aunque la resolución más baja (320x180) mantuvo una correlación significativa en la detección de estrés, se observó un aumento considerable en los tiempos de procesamiento al utilizar resoluciones más altas, como 1920x1080.

A pesar de la pérdida de precisión asociada a resoluciones más bajas, las correlaciones en las métricas de estrés se mantuvieron estables, lo que sugiere que las resoluciones reducidas no comprometen de manera significativa la capacidad de detectar estrés. De hecho, estas resoluciones más bajas pueden ser preferibles cuando se busca reducir los tiempos de procesamiento. Estos hallazgos responden a la pregunta de investigación RQ4, confirmando que la reducción de la resolución influye tanto en la precisión de los detectores como en los tiempos de procesamiento, tal como se planteó en la hipótesis H4.

Finalmente, se concluye que el umbral crítico identificado en la pregunta de investigación RQ5 se encuentra en 640x360, ya que esta resolución mínima permite mantener una precisión adecuada en la detección de indicadores de estrés. Resoluciones inferiores a esta conllevan una disminución considerable de la precisión.

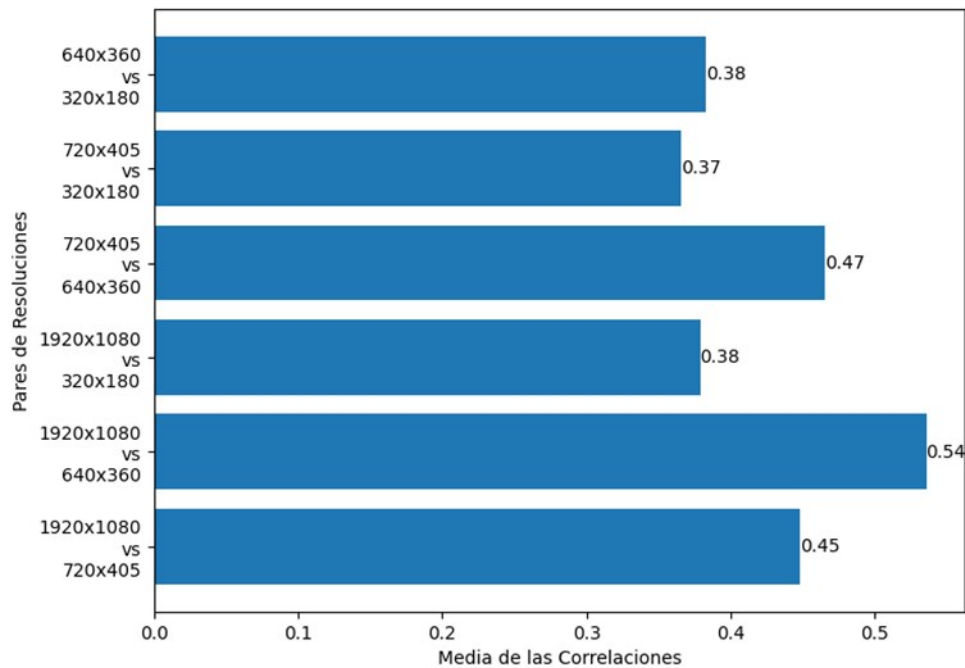


Figura 8: Media de las correlaciones de Spearman entre diferentes resoluciones para las expresiones faciales.

5. Conclusiones y posibles ampliaciones

Este estudio comparó dos métodos de detección automática de estrés: uno basado en la actividad electrodérmica (EDA) y otro en las expresiones faciales, utilizando un detector basado en señales fisiológicas desarrollado previamente y la herramienta industrial Kopernica Human. El objetivo fue analizar la eficacia de ambos métodos durante la interacción con un juego cognitivo.

Aunque los dos métodos automáticos no lograron replicar con precisión el estrés percibido por los usuarios, se identificó que el indicador *stress intens* mantiene una tendencia más coherente con la activación fisiológica (EDA), lo que sugiere su valor potencial como aproximación indirecta al estado emocional del usuario.

Los resultados también mostraron que la EDA es más sensible a los cambios en los niveles de estrés inducidos por el juego. Sin embargo, a pesar de no encontrarse diferencias significativas entre las fases del juego mediante la prueba de McNemar [40], los resultados podrían explicarse por la naturaleza moderada del estresor, la posible adaptación de los participantes al entorno experimental, o la sensibilidad limitada de los detectores, especialmente en el análisis facial.

Para mejorar la detección facial, se sugiere aplicar técnicas de filtrado propias de series temporales, como el filtro de media móvil, el filtro de Kalman o la transformada de Fourier, que permitan reducir la alta frecuencia y la variabilidad de los datos.

Una limitación importante del estudio fue la discrepancia entre el estrés percibido por los participantes y el registrado por los detectores, lo que sugiere que las personas tienden a subestimar su nivel de estrés. Además, la muestra utilizada consistió principalmente en estudiantes jóvenes, lo que limita la generalización de los resultados a otras poblaciones. También se identificó una pérdida de datos EDA (falta de los últimos tres minutos completos de señal), reduciendo la duración total del análisis.

La sincronización temporal presentó un desafío debido a la falta de un protocolo automatizado, lo que obligó a realizar una revisión manual de los videos. Para futuras implementaciones, se recomienda la inclusión de fotogramas iniciales con marcas de tiempo para facilitar esta tarea. Asimismo, la calidad de los videos, como la resolución mínima de 640x360 y una adecuada iluminación y posición, es crucial para mejorar el procesamiento.

La herramienta Kopernica Human, diseñada no solo para detectar estrés sino también otros estados emocionales, podría ser útil en estudios sobre esfuerzo cognitivo. Sin embargo, contar con el procesamiento de vídeos completos, sin necesidad de segmentarlos en intervalos cortos, sería beneficioso para conducir estudios mas largos, y siempre que se pueda tener dicho procesamiento de manera local.

A pesar de estas limitaciones, el estudio ofrece una comparación controlada de dos métodos de detección de estrés. Los resultados muestran la necesidad de seguir perfeccionando ambos métodos para su aplicabilidad en contextos reales.

Declaración de uso de IA Generativa

Los autores no han empleado herramientas de IA Generativa para la escritura de este manuscrito.

Agradecimientos

Se reconoce el apoyo de la Consellería de Educación, Universidades y Formación Profesional de la Xunta de Galicia y del Fondo Europeo de Desarrollo Regional, FEDER Una manera de hacer Europa (ayudas “Centro de investigación de Galicia con acreditación 2024-2027 ED431G-2023/04” y “Grupo de Referencia Competitiva con acreditación 2022-2025 ED431C 2022/19”). Este trabajo también cuenta con el apoyo del Ministerio de Ciencia, Innovación y Universidades de España (MCIN/AEI/10.13039/501100011033/) con las ayudas PID2021-123152OB-C21 y CNS2024-154915.

Los autores quieren expresar su gratitud al equipo de Neurologyca por proveer la herramienta Kopernica Human, así como el entrenamiento en su utilización. También nuestro agradecimiento va hacia las personas participantes en el experimento por su valiosa colaboración.

Referencias

- [1] G. Fink, Stress: Definition and history, *Stress science: neuroendocrinology* 3:9 (2010) 3-14. doi: 10.1016/B978-008045046-9.00076-0
- [2] S. Nandan, S. Mandal, P. Ghosal, Stress Detection and Monitoring: A Systematic Review, in: 2024 IEEE International Symposium on Smart Electronic Systems (iSES), New Delhi, India, 2024, pp. 309-314. doi: 10.1109/iSES63344.2024.00072.
- [3] F. Suni Lopez, N. Condori-Fernandez, A. Catala, Towards Real-Time Automatic Stress Detection for Office Workplaces, Springer, Cham, volume 898, pp. 273–288, 2019. DOI: 10.1007/978-3-030-11680-4_27
- [4] EmbracePlus, Empatica. Consultado el 30 de junio de 2025. URL: <https://www.empatica.com/en-eu/embraceplus/>
- [5] Kopernica, Neurologyca. Consultado el 30 de junio de 2025. URL: <https://kopernica.io/>
- [6] Neurologyca, Neurologyca. Consultado el 30 de junio de 2025. URL: <https://neurologyca.com/>
- [7] K.H. Kim, S.W. Bang, S. R. Kim, Emotion recognition system using short-term monitoring of physiological signals, *Med. Biol. Eng. Comput.* 42 (2004) 419–427. doi: 10.1007/BF02344719
- [8] B. D. Womack, J.H.L. Hansen, Classification of speech under stress using target driven features, *Speech Communication*, 20:1 (1996) 131-150. doi: 10.1016/S0167-6393(96)00049-0
- [9] R. M. Sabour, Y. Benezeth, P. De Oliveira, J. Chappé, F. Yang, UBFC-Phys: A Multimodal Database For Psychophysiological Studies of Social Stress, *IEEE Transactions on Affective Computing* 14:1 (2023) 622-636. doi: 10.1109/TAFFC.2021.3056960
- [10] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, M. Tsiknakis, Review on Psychological Stress Detection Using Biosignals, *IEEE Transactions on Affective Computing* 13:1 (2022) 440-460. doi: 10.1109/TAFFC.2019.2927337
- [11] Y. S. Can, B. Arnrich, C. Ersoy, Stress detection in daily life scenarios using smart phones and wearable sensors: A survey”, *Journal of Biomedical Informatics* 92 (2019) 103139. doi: 10.1016/j.jbi.2019.103139

- [12] M. Bolpagni, S. Pardini, M. Dianti, S. Gabrielli, Personalized Stress Detection Using Biosignals from Wearables: A Scoping Review, *Sensors* 24:10 (2024) 3221. doi: 10.3390/s24103221
- [13] Paper Folding, Spatial Reasoning Games. Consultado el 30 de junio de 2025. URL: <https://spatialreasoninggames.weebly.com/paper-folding.html>
- [14] M-H. Lee, G. Yang, H-K. Lee, S. Bang, Development stress monitoring system based on personal digital assistant (PDA), in: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004, pp. 2364-2367. doi: 10.1109/IEMBS.2004.1403685
- [15] J. A. Healey, R. W. Picard, Detecting stress during real-world driving tasks using physiological sensors, *IEEE Transactions on Intelligent Transportation Systems* 6:2 (2005) 156-166. doi: 10.1109/TITS.2005.848368
- [16] J. Zhai, A. Barreto, Stress Detection in Computer Users Based on Digital Signal Processing of Noninvasive Physiological Variables, *International Conference of the IEEE Engineering in Medicine and Biology Society* (2006), 1355-1358. doi: 10.1109/IEMBS.2006.259421
- [17] C. D. Katsis, N. Katertsidis, G. Ganiatsas, D. I. Fotiadis, Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 38:3 (2008) 502-512. doi: 10.1109/TSMCA.2008.918624
- [18] A. de Santos Sierra, C. S. Ávila, J. Guerra Casanova, G. Bailador del Pozo, V. Jara Vera, Two Stress Detection Schemes Based on Physiological Signals for Real-Time Applications, in: *Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Darmstadt, 2010, pp. 364-367. doi: 10.1109/IIHMSP.2010.95
- [19] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, U. Ehlert, Discriminating Stress From Cognitive Load Using a Wearable EDA Device, *IEEE Transactions on Information Technology in Biomedicine* 14:2 (2010) 410-417. doi: 10.1109/TITB.2009.2036164
- [20] S. A. Hosseini, M. B. Naghibi-Sistani, Classification of Emotional Stress Using Brain Activity, *Applied Biomedical Engineering, InTech*, 2011. doi: 10.5772/18294
- [21] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, J. Penders, Towards mental stress detection using wearable physiological sensors, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 1798-1801. doi: 10.1109/IEMBS.2011.6090512
- [22] J. Hernandez, R. R. Morris, R. W. Picard, Call Center Stress Recognition with Person-Specific Models", Springer, volume 6974, 2011, pp. 125-134. doi: 10.1007/978-3-642-24600-5_16
- [23] D. Giakoumis, A. Drosou, P. Cipresso, D. Tzovaras, G. Hassapis, A. Gaggioli, Using Activity-Related Behavioural Features towards More Effective Automatic Stress Detection, *PLoS ONE* 7:9 (2012). doi: 10.1371/journal.pone.0043571
- [24] H. Kurniawan, A. Maslov, M. Pechenizkiy, Stress detection from speech and galvanic skin response signals", in: 26th IEEE Conference on Computer-Based Medical Systems, 2013, pp. 209-214. doi: 10.1109/CBMS.2013.6627790
- [25] M. Pedrotti, M. A. Mirzaei, A. Tedesco, J. R. Chardonnet, F. Mérienne, S. Benedetto, T. Baccino, Automatic Stress Classification With Pupil Diameter Analysis, *International Journal of Human-Computer Interaction* 30:3 (2014) 220-236. doi: 10.1080/10447318.2013.848320
- [26] N. Sharma, T. Gedeon, Modeling a stress signal, *Applied Soft Computing* 14 (2014) 53-61. doi: 10.1016/j.asoc.2013.09.019
- [27] A. Liapis, C. Katsanos, D. Sotiropoulos, M. Xenos, N. Karousos, Stress recognition in human-computer interaction using physiological and self-reported data: a study of gender differences, in: *Proceedings of the 19th Panhellenic Conference on Informatics (PCI '15)*. Association for Computing Machinery, New York, NY, USA, 2015, 323-328. doi: 10.1145/2801948.2801964
- [28] A. Ghaderi, J. Frounchi, A. Farnam, Machine learning-based signal processing using physiological signals for stress detection, in: 22nd Iranian Conference on Biomedical Engineering (ICBME), 2015, pp. 93-98. doi: 10.1109/ICBME.2015.7404123

- [29] M. Abouelenien, M. Burzo, R. Mihalcea, Human Acute Stress Detection via Integration of Physiological Signals and Thermal Imaging, in: Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '16), Association for Computing Machinery, 2016, pp. 1–8. doi: 10.1145/2910674.2910705
- [30] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. G. Simos, K. Marias, M. Tsiknakis, Stress and anxiety detection using facial cues from videos, *Biomedical Signal Processing and Control* 31 (2017) 89-101. doi: 10.1016/j.bspc.2016.06.020
- [31] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, J. M. Ferrandez, Stress detection using wearable physiological and sociometric sensors, *International Journal of Neural Systems* 27:2 (2017) 1650041. doi: 10.1142/S0129065716500416
- [32] A. Anusha, J. Jose, S. Preejith, J. Jayaraj, et al., Physiological signal-based work stress detection using unobtrusive sensors, *Biomedical Physics & Engineering Express* 4:6 (2018) 065001. doi: 10.1088/2057-1976/aadbd4
- [33] A. G. Airij, R. Sudirman, U. U. Sheikh, “GSM and GPS Based Real-Time Remote Physiological Signals Monitoring and Stress Levels Classification”, 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS), pp. 130-135, 2018. DOI: 10.1109/ICBAPS.2018.8527406
- [34] Y. S. Can, N. Chalabianloo, D. Ekiz, C. Ersoy, Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study, *Sensors* 19:8 (2019) 1849. doi: 10.3390/s19081849
- [35] F. Indikawati, S. Winiari, Stress Detection from Multimodal Wearable Sensor Data, in: IOP Conference Series: Materials Science and Engineering, volume 771, no. 1, 2020, pp. 012028. doi: 10.1088/1757-899X/771/1/012028
- [36] A. Mohammadi, M. Fakharzadeh and B. Baraeinejad, An Integrated Human Stress Detection Sensor Using Supervised Algorithms, *IEEE Sensors Journal* 22:8 (2022) 8216-8223. doi: 10.1109/JSEN.2022.3157795
- [37] M. K. Moser, B. Resch, M. Ehrhart, An Individual-Oriented Algorithm for Stress Detection in Wearable Sensor Measurements, *IEEE Sensors Journal* 23:19 (2023) 22845-22856. doi: 10.1109/JSEN.2023.3304422
- [38] F. Rodrigues, H. Correia, Semi-supervised and ensemble learning to predict work-related stress, *Journal of Intelligent Information Systems* 62 (2024) 77-90. doi: 10.1007/s10844-023-00806-z
- [39] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén, *Experimentation in Software Engineering*, Springer, volume 6, 2000, pp. 119-126. doi: 10.1007/978-1-4615-4625-2_10
- [40] M.W. Fagerland, S. Lydersen, P. Laake, The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional, *BMC Med Res Methodol* 13 (2013) 91. doi: 10.1186/1471-2288-13-91.

Evaluación de la comprensión de las pautas WCAG 2.1: análisis de ejemplos prácticos en tres universidades españolas

Evaluation of the comprehension of WCAG 2.1 guidelines: analysis of practical examples in three Spanish universities

Afra Pascual-Almenara^{1,†}, Alberto Pedrero-Esteban^{2*,†} and Susana Bautista^{3,†}

¹ Departamento de Informática y Diseño Digital. Escuela Politécnica Superior. Universitat de Lleida, Lleida, Spain

² Facultad de Informática Universidad Pontificia de Salamanca, Salamanca, Spain

³ Escuela Politécnica Superior. Universidad Francisco de Vitoria. Madrid, Spain

Resumen

Entender las pautas WCAG puede resultar complicado cuando se empieza a trabajar en accesibilidad digital. Este trabajo de investigación presenta un análisis de la comprensión de las pautas WCAG 2.1 (niveles A y AA) en estudiantes de carreras técnicas pertenecientes a tres universidades (Universidad de Lleida - UdL, Universidad Francisco de Vitoria- UFV y Universidad Pontificia de Salamanca - UPSA) durante un curso académico (2024/25). Todos los participantes realizaron la misma actividad práctica en la que, de manera estructurada, recopilaron ejemplos accesibles e inaccesibles de cada criterio de las pautas WCAG 2.1. Este enfoque permitió evaluar las dificultades en la interpretación y aplicación de los principios de accesibilidad digital. Los resultados revelan que los estudiantes presentan una alta variabilidad en la interpretación de las pautas, así como dificultad en la comprensión de algunos criterios, especialmente cuando están redactados en lenguaje complejo. También se observó que los criterios que pueden evaluarse con una herramienta de evaluación directa fueron comprendidos con mayor claridad por los estudiantes. Al analizar los datos, se observó que el perfil del estudiantado —definido por el grado y la titulación que cursaba— influía significativamente en las estrategias empleadas para recopilar y/o crear un ejemplo adecuado en determinados criterios. Como resultado de la experiencia, está previsto ampliar el estudio a las pautas WCAG 2.2 e implementar en un futuro próximo una plataforma colaborativa en formato wiki para la difusión de los resultados. Esta iniciativa tendrá como propósito facilitar la comprensión y aplicación de los principios de accesibilidad digital mediante la sistematización y difusión de los ejemplos generados durante la investigación.

Abstract

Understanding WCAG guidelines can be complicated when starting to work on digital accessibility. This paper presents an analysis of the understanding of WCAG 2.1 guidelines (levels A and AA) in technical students from three universities (Universidad de Lleida - UdL, Universidad Francisco de Vitoria- UFV and Universidad Pontificia de Salamanca - UPSA) during an academic year (2024/25). All participants carried out the same practical activity in which, in a structured way, they collected accessible and inaccessible examples of each criterion of the WCAG 2.1 guidelines. This approach made it possible to assess difficulties in interpreting and applying the principles of digital accessibility. The results reveal that students show a high variability in the interpretation of the guidelines, as well as difficulty in understanding some criteria, especially when they are written in complex language. It was also observed that the criteria that can be assessed with a direct assessment tool were more clearly understood by the students. In analysing the data, it was observed that the student profile - defined by the degree and

^{1*} Corresponding author.

[†] These authors contributed equally.

✉ afra.pascual@udl.cat (A. Pascual-Almenara); apedreroes@upsa.es (A. Pedrero-Esteban); susana.bautista@ufv.es (S. Bautista)

ORCID 0000-0002-2368-755X (A. Pascual-Almenara); 0000-0002-1222-0882 (A. Pedrero-Esteban); 0000-0003-1648-0208 (S. Bautista)

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

qualification being studied - had a significant influence on the strategies used to collect and/or create a suitable example on certain criteria. As a result of the experience, it is planned to extend the study to WCAG 2.2 guidelines and to implement in the near future a collaborative wiki platform for the dissemination of results. This initiative aims to facilitate the understanding and application of digital accessibility principles through the systematization and dissemination of the examples generated during the research.

Palabras clave

evaluación, pautas WCAG, accesibilidad digital, diseño web

1. Introducción

Las pautas WCAG son el estándar global más reconocido para garantizar la accesibilidad web, aportando beneficios como mayor usabilidad, mejor posicionamiento en buscadores, reducción de riesgos legales y un alcance más amplio del público. El creciente auge en la accesibilidad digital resalta la importancia de comprender y aplicar correctamente las pautas WCAG, aunque su extensa información puede representar un desafío para quienes recién comienzan su estudio [1,2,3]. La complejidad técnica y el lenguaje abstracto de las WCAG dificultan su comprensión y aplicación práctica [4], especialmente para quienes no tienen formación en programación. Esto resalta la necesidad de enfoques educativos accesibles, con materiales didácticos claros y ejemplos concretos que faciliten el aprendizaje de estas pautas.

Este trabajo presenta los resultados de una iniciativa orientada a analizar la comprensión de las pautas WCAG, en la que participaron un total de 97 estudiantes: 33 de la UdL, 32 de la UFV y 32 de la UPSA.

El propósito del trabajo fue analizar el grado de comprensión de las pautas WCAG por parte del estudiantado de las tres universidades. Para ello, se propuso la recopilación de ejemplos accesibles y no accesibles asociados a distintos criterios. Posteriormente, se analizaron y validaron los ejemplos generados para determinar su grado de adecuación a las especificaciones de cada pauta WCAG.

El artículo se organiza de la siguiente manera: tras la introducción, la sección 2 presenta los trabajos relacionados y se incluye una sección relacionada con las dificultades en el aprendizaje de las pautas WCAG; a continuación, la sección 3 presenta el contexto del estudio donde se indican los criterios seleccionados para crear los ejemplos de las pautas WCAG y una presentación de las universidades, grados y asignaturas a las que pertenecían al estudiantado participantes en el estudio. En la sección 4 se presenta la metodología que se ha seguido para recopilar los ejemplos y validar su adecuación a cada uno de los criterios analizados. En la sección 5 se presentan los resultados divididos en dos secciones, resultados cuantitativos y resultados generales. En la sección 6 se presenta la discusión de los datos desde una perspectiva cuantitativa y se analiza la validez de ejemplos desde la perspectiva de cada universidad. Finalmente, el artículo termina con la conclusión y trabajo futuro.

2. Trabajo relacionado

Diversos estudios han puesto en evidencia la necesidad de revisar en profundidad las estrategias pedagógicas empleadas para integrar la accesibilidad la enseñanza. El trabajo de Lewthwaite [5], plantea una revisión sistemática sobre cómo se estructura la enseñanza de la accesibilidad digital en entornos universitarios, y subraya la falta de una pedagogía consolidada y la escasa sistematización en los métodos didácticos utilizados. Por otro lado, el trabajo de Moreno [6] propone integrar la accesibilidad no como un componente añadido, sino como un criterio transversal al diseño

didáctico y tecnológico de los materiales educativos. Esta perspectiva promueve un enfoque más holístico en el que accesibilidad, usabilidad y calidad educativa convergen, lo que resulta especialmente relevante para el desarrollo de recursos inclusivos sostenibles en el tiempo.

En esta línea, existen programas formativos en el ámbito de las TIC que han empezado a incorporar contenidos sobre accesibilidad universal y diseño para todos [7,8,9]. Sin embargo, persiste una falta de consenso en cuanto a los contenidos específicos y las metodologías de enseñanza. Si bien las pautas WCAG se reconocen como un conocimiento fundamental, no siempre se abordan explícitamente en los planes de estudio ni se profundiza en las dificultades asociadas a su aprendizaje [10,11].

Algunas experiencias educativas en entornos universitarios [12,13,14,15], evidencian que, para asegurar una aplicación sostenida de los conocimientos en accesibilidad a lo largo del tiempo, es necesaria una mayor carga docente dedicada a esta temática [16,17].

El estudio presentado por Ferrari [18] analiza el impacto de la integración de la accesibilidad digital en los cursos universitarios, destacando la efectividad del aprendizaje experiencial —como simulaciones, visitas a centros especializados y proyectos de aprendizaje-servicio— en la enseñanza de la accesibilidad digital, señalando además el uso frecuente de las directrices WCAG como recurso formativo central.

2.1. Dificultades encontradas en el aprendizaje de las pautas WCAG

A pesar de la disponibilidad de numerosos recursos formativos, tanto diferentes plataformas educativas como DequeUnivertisy², A11y Project³ o WebAIM⁴ como en los propios materiales del W3C⁵, el aprendizaje de las WCAG presenta importantes desafíos, especialmente en etapas iniciales. La comprensión e implementación de sus principios fundamentales: Perceptible, Operable, Comprensible y Robusto (POUR), las pautas no verificables con herramientas automáticas y los numerosos criterios de conformidad, puede resultar complejos, en particular para estudiantes procedentes del ámbito del diseño de interacción con conocimientos limitados en tecnologías web. Estos estudiantes, además, suelen enfrentarse al reto de integrar los requisitos de accesibilidad sin comprometer la creatividad visual, y a la dificultad de traducir conceptos abstractos en prácticas que respondan a las necesidades reales de personas con discapacidad. La Tabla 1 lista algunas de las dificultades comentadas por el estudiantado en sus primeras aproximaciones a las pautas.

Para facilitar el aprendizaje de los principios de accesibilidad propuestos por las WCAG, se pueden considerar distintos enfoques complementarios:

- El uso y análisis de ejemplos accesibles y no accesibles del mundo real permite a los estudiantes comprender las implicaciones prácticas de las directrices [19,20,21].
- La evaluación activa de sitios web con herramientas especializadas proporciona experiencia aplicada en la identificación de barreras de accesibilidad concretas [22].
- La incorporación de actividades que fomenten la empatía, como simulaciones de discapacidad, contribuye a una comprensión más profunda de las necesidades y desafíos que enfrentan los usuarios con diversidad funcional [23,24].

²Deque University: <https://dequeuniversity.com/>

³A11yproject: <https://www.a11yproject.com/>

⁴WebAIM: <https://webaim.org/>

⁵W3C - Digital Accessibility Foundations Free Online Course: <https://www.w3.org/WAI/courses/foundations-course/>

- El empleo de listas de comprobación y hojas de resumen accesibles actúa como un recurso de consulta rápida que facilita la validación de los distintos criterios de éxito y sus respectivos requisitos [25].

Estos enfoques no solo permiten una mejor comprensión técnica de las pautas WCAG, sino que también responden a la necesidad, señalada en el trabajo realizado por Ballesteros [26], de concebir la accesibilidad no únicamente como un reto tecnológico, sino como un cambio filosófico y social que requiere una transformación en la forma en que se diseñan e implementan las tecnologías digitales.

Tabla 1

Lista de retos para afrontar el conocimiento de las pautas WCAG.

Retos	Descripción
Uso del lenguaje	Uso de terminología especializada en la documentación de las WCAG que puede resultar difícil de entender para personas sin conocimientos técnicos.
Pautas abstractas	Dificultad para traducir los criterios generales de las WCAG en acciones específicas de diseño y desarrollo.
Volumen de información	La amplitud de la documentación de las WCAG y el número de criterios de éxito pueden resultar abrumadores para los estudiantes.
Equilibrio entre estética y accesibilidad	Conflicto entre la creación de diseños visualmente atractivos y muy interactivos y el cumplimiento de los requisitos de accesibilidad.
Comprensión de las necesidades de los usuarios	Dificultad para empatizar y comprender las diversas necesidades y retos a los que se enfrentan los usuarios con distintas discapacidades.
Aplicación al contenido dinámico	Retos a la hora de garantizar la accesibilidad de elementos interactivos y contenidos que cambian dinámicamente en una página web.
Falta de ejemplos prácticos	Dificultad para encontrar y comprender ejemplos concretos de cómo aplicar los criterios de éxito de las WCAG en situaciones reales.

3. Contexto del estudio

A continuación, se introducen los criterios de las pautas WCAG seleccionados, así como se introducen las universidades, grados, y asignaturas que cursan los estudiantes que han participado en el estudio.

3.1. Criterios seleccionados para crear los ejemplos de las pautas WCAG

Las pautas WCAG (2.1) [3] contienen trece directrices organizadas en torno a cuatro principios clave que describen las características ideales de los sitios web accesibles: 1. Perceptible (la información debe ser accesible a los sentidos), 2. Operable (los usuarios deben poder interactuar con la interfaz de manera efectiva), 3. Comprensible (el contenido y la interacción deben ser fáciles de entender), y 4. Robusto (el contenido debe ser compatible con tecnologías actuales y futuras). Cada directriz tiene criterios de éxito que pueden ser evaluados.

Los criterios de éxito se clasifican en tres niveles de conformidad que van desde el nivel A (el estándar mínimo de accesibilidad), el nivel AA (el estándar de facto requerido por la mayoría de las normativas mundiales [27], y el nivel AAA (el nivel ideal que representa la accesibilidad más alta posible). Con el objetivo de abordar de manera más detallada los criterios individuales definidos en las pautas WCAG —algunos de los cuales se presentan de forma muy genérica y carecen de directrices prácticas claras para su evaluación—, se recurrió a la documentación elaborada por el Observatorio de Accesibilidad Web (OAW)⁶, la cual organiza y desarrolla estos criterios en distintos documentos organizados por los cuatro principios fundamentales de la accesibilidad: perceptible [28], operable [29], comprensible [30] y robusto [31]. Estos materiales proporcionaron a los estudiantes instrucciones claras y estructuradas, que incluían tanto una descripción comprensible de cada criterio como una guía práctica para su evaluación en sitios web reales. Asimismo, cada criterio presentado en las guías estaba vinculado directamente con las técnicas de las pautas WCAG, lo que permitió a los estudiantes ampliar y complementar la información cuando lo consideraron necesario para comprender y ejemplificar adecuadamente su aplicación. En total se incluyeron en el estudio 94 subcriterios divididos en los cuatro principios: Principio 1 (52 subcriterios), Principio 2 (25 subcriterios), Principio 3 (14 subcriterios) y Principio 4 (3 subcriterios).

Todos los criterios seleccionados en el estudio correspondían al nivel AA de las pautas WCAG 2.1.

3.2. Universidades, grados y asignaturas participantes

El estudio se ejecutó de forma paralela en carreras técnicas pertenecientes a tres universidades españolas: Universidad de Lleida (UdL), Universidad Francisco de Vitoria (UFV) y Universidad Pontificia de Salamanca (UPSA) durante un curso académico 2024/25. La Tabla 2 muestra el listado de universidades, asignaturas, estudiantes que participaron en el estudio junto a la cantidad de subcriterios de las pautas WCAG que se evaluaron por cada alumno.

El objetivo de la asignatura de *Diseño Web* [32] de segundo curso del grado de Diseño Digital y Nuevas Tecnologías de la Universitat de Lleida (UdL) es proporcionar a los estudiantes los conocimientos y habilidades necesarios para aplicar los principios de experiencia de usuario (UX) en el diseño web, evaluando la usabilidad y accesibilidad, y desarrollando interfaces interactivas que sean tanto usables como accesibles para todos los usuarios.

La asignatura de *Human Computer Interaction* [33] del tercer curso del grado en Ingeniería Informática de la Universidad de Francisco de Vitoria se imparte en inglés y tiene como objetivo principal aplicar los conocimientos metodológicos necesarios para cubrir el ciclo de vida de una interfaz gráfica de usuario, incluyendo la consideración de factores humanos y de accesibilidad.

La asignatura de *Diseño, evaluación y desarrollo de interfaces* [34] del cuarto curso del grado en Ingeniería Informática, de la Universidad Pontificia de Salamanca (UPSA) tiene entre sus objetivos

⁶Observatorio de Accesibilidad Web: Ministerio para la Transformación Digital y de la Función Pública: <https://administracionelectronica.gob.es/ctt/oaw>

conocer las normativas de accesibilidad y los principios del diseño para todos y aplicarlos en el desarrollo de software y a la evaluación de proyectos software.

Tabla 2

Lista de universidades, asignatura y estudiantes que han participado en el estudio. Se incluye también la cantidad de criterios asignados a cada estudiante.

Universidad	Grado	Asignatura	Estudiantes	Criterios
Universidad de Lleida	Grado de Diseño Digital y Nuevas Tecnologías	Diseño Web	42 estudiantes matriculados(solo 35 asistentes a clase)	3 por estudiante
Universidad Francisco de Vitoria	Grado en Ingeniería Informática	Human computer interaction	36 estudiantes matriculados(solo 32 asistentes a clase)	2 por estudiante
Universidad Pontificia de Salamanca	Grado en Ingeniería Informática	Diseño, evaluación y desarrollo de interfaces	34 estudiantes matriculados(solo 32 asistentes a clase)	3 por estudiante

Un aspecto relevante a tener en cuenta en el análisis de las entregas de los estudiantes es que la actividad fue distribuida a la totalidad del estudiantado con asistencia regular. No obstante, se evidenció que no todos los participantes completaron la tarea asignada. En consecuencia, en las tablas con la valoración de los ejemplos presentados (Tablas 4 a 7), los campos en blanco corresponden a criterios para los cuales no se dispone de información, debido a la ausencia de entrega por parte del estudiante.

4. Descripción de la actividad y validación de datos

A continuación, se describe la actividad propuesta a los estudiantes de las tres universidades participantes, orientada a la recopilación de ejemplos de los criterios de las Pautas WCAG 2.1. Posteriormente, se detalla el procedimiento seguido para el análisis de los ejemplos, con el objetivo de garantizar un análisis de datos riguroso y objetivo.

4.1. Actividad de creación de ejemplos

Se diseñó una actividad formativa y evaluativa basada en la recopilación de ejemplos accesibles (ACC) y no accesibles (NOA), de acuerdo con los criterios establecidos en las Pautas de Accesibilidad al Contenido en la Web (WCAG) 2.1. Esta actividad fue implementada en paralelo en tres universidades como parte del desarrollo curricular de asignaturas vinculadas a Diseño Web (UdL), Interacción Persona Ordenador (UFV) y Diseño, evaluación y desarrollo de interfaces (UPSA). Se asignó a cada estudiante de dos a tres criterios concretos de las pautas WCAG 2.1, en función del número de participantes en cada grupo de clase (véase Tabla 2). A continuación, se explica las fases en las que se estructuró la actividad:

1. **Exploración de información:** antes de empezar la actividad, se proporcionó a los estudiantes acceso al material del Observatorio de Accesibilidad Web que explica con

detalle cada uno de los principios de las pautas WCAG 2.1 [28,29,30,31]. Estos documentos fueron acompañados de una explicación guiada en clase sobre cómo consultar y comprender dicha documentación. Además, se ofrecieron otros recursos complementarios (por ejemplo, artículos especializados , guías interactivas , wikis técnicas) para favorecer una comprensión más amplia y contextualizada de los criterios.

2. **Presentación de la actividad:** cuando los estudiantes ya estaban familiarizados con los fundamentos teóricos de las pautas WCAG, se publicó el enunciado de la actividad. Ésta consistía en proporcionar un ejemplo accesible y no accesible de los criterios asignados a cada estudiante. Además, se les mostraba un ejemplo del criterio 1.1.1-A para que los estudiantes consultaran de forma libre con el fin de observar el nivel de detalle esperado en las entregas.
3. **Realización de la actividad:** en base a la información compartida y el ejemplo de referencia, cada participante debía localizar dos ejemplos -uno accesible y no accesible- por cada criterio asignado de las pautas WCAG asignado. A continuación, se muestra la información a incluir para cada criterio:
 - a) Título del criterio
 - b) Ejemplo accesible (ACC)
 - c) Ejemplo no accesible (NOA)
 - d) Procedimiento para evaluar el criterio
 - e) Referencia WCAG 2.1 relacionada con el criterio

Se invitó al estudiantado que seleccionaran casos reales de sitios web para presentar un ejemplo accesible (ACC) y otro ejemplo no accesible (NOA) según el criterio a desarrollar y debían cumplimentar el apartado b y c del documento a entregar. Los apartados d y e se completaban directamente utilizando el texto del material de referencia del OAW.

4. **Entrega:** la actividad se formalizó con la presentación de un documento PDF en el cual se incluía la información solicitada de cada criterio (apartados del a al d). El documento también incluía las capturas de pantalla de ambos ejemplos y una justificación argumentada que explicaba porqué dichos ejemplos se consideraban accesibles o no accesibles según el criterio al que pertenecían.

4.2. Validación de ejemplos creados

Para validar de manera objetiva los ejemplos compartidos por el estudiantado en relación con cada uno de los criterios de las pautas WCAG 2.1, se diseñó un instrumento de evaluación estructurado en formato de hoja de cálculo (Excel). Este documento se organizó según los cuatro principios de accesibilidad (Perceptible, Operable, Comprensible y Robusto) asignando una pestaña independiente a cada uno de ellos. Dentro de cada hoja del documento Excel, se incluía el listado de criterios de accesibilidad organizados en columnas que incluían: el código del criterio, su título, el nivel de conformidad correspondiente únicamente a los niveles A o AA, el tipo de evaluación requerida (por ejemplo, alternativa textual, subtítulo, audiodescripción, código HTML, orden de lectura, uso del color, contraste, entre otros), y un breve texto guía de comprobación. Este texto de comprobación breve se sintetizó mediante el uso de herramientas de inteligencia artificial generativa (como ChatGPT) a partir de las descripciones de los criterios de las pautas WCAG 2.1.

Su fin fue facilitar la comprensión y evaluación de los ejemplos por parte de los revisores, y servir como base para un protocolo de validación uniforme y replicable.

Para evaluar los ejemplos propuestos por el estudiantado de cada universidad participante (UdL, UFV, UPSA) se incluyeron dos columnas específicas: ACC (para incluir el resultado del evaluador al validar el ejemplo accesible) y NOA (para incluir el resultado del evaluador al validar el ejemplo no accesible). En cada celda, se registró una valoración del grado de adecuación de cada ejemplo según el criterio correspondiente, utilizando una rúbrica ordinal de tres niveles:

- 0 - el ejemplo no corresponde al criterio,
- 1- el ejemplo está relacionado con el criterio, pero presenta errores o ambigüedades,
- 2- el ejemplo se adecúa totalmente al criterio.

Junto a esta puntuación, se incluyó un campo de comentarios destinado a justificar la valoración asignada y ofrecer observaciones adicionales relevantes sobre el ejemplo evaluado, tanto en su versión accesible como no accesible. Esta estrategia permitió integrar una perspectiva cuantitativa en el proceso de validación.

El proceso de revisión de validación de los ejemplos se realizó por un experto en accesibilidad digital. Se inició con un análisis preliminar orientado a identificar la existencia o ausencia de ejemplos presentados por el estudiantado de cada universidad participante. Esta información fue registrada en las correspondientes tablas de validación. Luego, se accedió al conjunto de documentos entregados por el estudiantado, compuesto por un total de 95 archivos en formato PDF y que cada uno contenía entre 2 y 3 propuestas de los criterios asignados. Se procedió a examinarlos detalladamente revisando los ejemplos accesibles y no accesibles propuestos para cada criterio. Posteriormente, se completó el documento de validación (en formato Excel), registrando una puntuación conforme a la rúbrica previamente establecida: (0, 1, 2). El criterio con el que se validaba cada uno de los ejemplos evaluados fue determinar el grado de adecuación y representatividad en relación con el criterio WCAG que pretendía explicar, atendiendo tanto a su precisión técnica como a su claridad explicativa.

Aunque algunos estudiantes no completaron la entrega de todos los ejercicios asignados —lo cual se reflejó en la tabla de validación mediante celdas en blanco—, se procuró que cada criterio fuera representado y evaluado por, al menos, dos de las tres universidades participantes. Esta estrategia tuvo como finalidad garantizar una muestra representativa y aumentar la fiabilidad de los resultados obtenidos en el proceso de validación. (Consultar Tablas 4 a 7).

5. Resultados

La presentación de resultados se ha dividido en dos partes, para abordar la explicación de manera más estructurada: análisis cuantitativo y análisis detallado de los datos.

5.1. Análisis cuantitativo

A partir de los resultados de las validaciones del evaluador incluidas en el documento Excel, se generaron los resultados numéricos que se presentan en las Tablas 4 a 7. Para hacer más comprensible los datos numéricos se han incluido también distintas gráficas (Gráficas 1 a 7).

5.1.1. Principio 1. Perceptible

El principio 1 es muy extenso, se presenta en la Tabla 4 y en los Gráficos 1, 2, 3, y 4. La pauta 1.1 muestra que solo el criterio 1.1.1-E del ejemplo no accesible (NOA) se ha podido validar en las tres universidades correctamente. Se observa que no se ha creado ningún ejemplo correcto en los criterios 1.1.1-B, 1.1.1-C, 1.1.1-D, 1.1.1-F de los ejemplos no accesibles (NOA), y no existe ejemplo ni accesible ni no accesible (ACC y NOA) en los subcriterios 1.1.1-G y 1.1.1-H. En la Pauta 1.2, se observa poca comprensión en los criterios 1.2.1 A no accesibles (NOA) y 1.2.1-B accesible (ACC). La pauta 1.3 también presenta poca comprensión de los criterios 1.3.1-P, 1.3.2-A, 1.3.2-B, 1.3.2-C al presentar ejemplos no accesibles (NOA) válidos y de 1.3.3.-A. en ambos tipos de ejemplos, accesibles y no accesibles (ACC y NOA). La Pauta 1.4, se observa ejemplos no adecuados en los criterios 1.4.1 -C y 1.4.11-A al presentar ejemplos accesibles (ACC).

Al analizar el conjunto de datos, se observa que en el Principio 1, el 25% de los criterios presentan ejemplos válidos accesibles (ACC), mientras que el 28,85% corresponde a ejemplos válidos no accesibles (NOA).

5.1.2. Principio 2. Operable

Los resultados numéricos del principio 2 se presentan en la Tabla 5 y en el Gráfico 5. Los subcriterios 2.1.2-A y 2.4.4-A no presentan propuestas válidas para los ejemplos no accesibles (NOA). El subcriterio 2.5.2-A no presenta un resultado válido en ninguno de los ejemplos accesibles y no accesibles (ACC y NOA). El subcriterio 2.5.4-A no presenta un resultado valido en el ejemplo accesible (ACC).

Al analizar el conjunto de datos, se observa que en el Principio 2, el 20% de los criterios presentan ejemplos válidos accesibles (ACC), mientras que solo el 8% corresponde a ejemplos válidos no accesibles (NOA).

5.1.3. Principio 3. Comprensible

El principio 3 se resume en la Tabla 6 y el Gráfico 6. Los subcriterios 3.2.1-A, 3.2.2-A, 3.3.2-A y 3.3.4-A no presentan ejemplos accesibles y no accesibles (ACC y NOA) validos. El subcriterio 3.2.4-A presenta solo un ejemplo no accesible (NOA) válido.

Al analizar el conjunto de datos, se observa que en el Principio 3, el 21,43% de los criterios presentan ejemplos válidos accesibles (ACC), mientras que el 14,29% corresponde a ejemplos válidos no accesibles (NOA).

5.1.4. Principio 4. Robusto

Los datos del principio 4 se presentan en la Tabla 7 y el Gráfico 7. Se observa que todos los criterios presentados presentan algún ejemplo válido en su modo accesible y no accesible. Únicamente el criterio 4.1.1 presenta ejemplos validos en todos los ejemplos accesibles y no accesibles (ACC y NOA)

Al analizar el conjunto de datos, se observa que solo el 33,33% de los ejemplos presentados han sido validos tanto accesibles (ACC) como no accesibles (NOA).

5.2. Análisis detallado de los datos

A continuación, se presenta una interpretación de los resultados obtenidos en la validación de los ejemplos propuestos.

5.2.1. Principio 1. Perceptible

Pauta 1.1 Texto alternativo, las imágenes han de tener una alternativa textual. El criterio 1.1.1 se ha dividido en diversos subíndices (A, B, C.. etc), tal y como propone el documento guía del OAW, porque evalúa diversos supuestos de forma más individual: imágenes decorativas, imágenes funcionales, imágenes de gráficos, imágenes en hojas de estilo CSS, texto en elementos multimedia....

Se observa que hay mucha diversidad en los resultados, en cuanto a adecuación de los errores en cada tipo (ACC y NOA). Los datos indican que el estudiantado sabe que las imágenes han de presentar un texto alternativo (el criterio 1.1.1-A es válido en las tres universidades), pero todos los estudiantes han tenido dificultades al presentar propuestas de ejemplos para los distintos tipos de imágenes (imágenes decorativas, imágenes en hojas de estilo CSS, imágenes funcionales, imágenes acompañadas de un enlace, contenido audiovisual). Esto puede indicar que la pauta 1.1.1 presenta muchas ambigüedades en su evaluación, y no es porque el estudiantado no sepan cómo evaluarla (visualizar ALT) sino porque no saben si el texto de la imagen es correcto según la imagen presentada.

Pauta 1.2 Contenido tempodependiente, el contenido multimedia debe presentar alternativas. El criterio 1.2.1 presenta que el contenido solo audio (sin imágenes) o solo video (sin audio) necesita una alternativa. Sin embargo, el estudiantado no ha identificado correctamente un ejemplo con audio solo o video solo. Una posible explicación es la complejidad del texto del criterio.

En el criterio 1.2.2-A todo el estudiantado tienen claro que los videos han de presentar subtítulos. Sin embargo, cuando se presentan los criterios 1.2.3-A, 1.2.4-A, 1.2.5-A se observa más diversidad en la adecuación de los ejemplos tanto accesibles como no accesibles (ACC y NOA). Puede ser debido a que las diversas alternativas de accesibilidad que han de incluirse en los videos (en directo o cuando es necesario incluir alternativas como audiodescripción o transcripción) son opciones más complejas de comprender o que los estudiantes no las conocen.

Pauta 1.3. Adaptable, todo el contenido ha de presentar una codificación adecuada para que se adapte a cualquier presentación. El criterio 1.3.1 presenta hasta 17 subcriterios (A- Q) que están muy relacionados con la adecuada codificación HTML de tablas, listas, encabezados, formularios... Se observa en los resultados que los estudiantes conocen cómo implementar código HTML accesible en elementos sencillos como los encabezados (1.3.1-A y 1.3.1-B), pero tienen más dificultades para presentar ejemplos accesibles y no accesibles (ACC y NOA) para otros elementos que pueden ser más complejos de implementar como tablas, formularios (1.3.1-C hasta 1.3.1-Q).

El criterio 1.3.2 ha obtenido muy pocos ejemplos válidos. Una posible explicación es la ambigüedad del texto de los subcriterios en la comprensión de términos relacionados con tablas de maquetación o dirección de lectura del texto. En el criterio 1.3.3, ningún estudiante ha podido plantear un ejemplo válido accesible y no accesible (ACC y NOA). El término “características sensoriales de los elementos” es ambiguo y puede causar confusión a los estudiantes. En el criterio 1.3.4-A, la mayoría han sabido cómo crear ambos tipos de ejemplos (ACC y NOA) al estar relacionado con la orientación del dispositivo, algo que puede evaluarse de forma directa. Así como en el criterio 1.3.5 donde los ejemplos presentados son también correctos porque la evaluación corresponde a observar si en el código HTML se incluye la etiqueta “autocomplete” en el formulario, algo que es fácil de verificar.

Pauta 1.4 Distinguible, debe poder ser fácil escuchar el contenido incluida la separación entre primer plano y fondo. El criterio 1.4.1 muestra resultados diversos en cuanto a la validez de los ejemplos presentados tanto accesibles como no accesibles (ACC y NOA). Una posible explicación es que el texto relacionado con este criterio no se muestra de forma adecuada para que el estudiantado conozca fácilmente cómo puede evaluarlo.

Del criterio 1.4.2 ha sido difícil encontrar un ejemplo real y los estudiantes han creado un código HTML propio para representar un ejemplo correcto accesible y no accesible (ACC y NOA). En el criterio 1.4.3 se han creado ejemplos adecuados. Se ha comprendido cómo aplicar el contraste a los elementos textuales. Por otro lado, el criterio 1.4.11 ha sido complejo y no se han presentado ejemplos válidos ni accesibles ni no accesibles (ACC y NOA). Una posible explicación es que el término “radio mínimo en colores adyacentes” es difícil de comprender. Para los criterios 1.4.4 y 1.4.10 relacionados con diseño responsive, los estudiantes han incluido ejemplos correctos accesibles y no accesibles (ACC y NOA). El criterio 1.4.5 solo presenta un ejemplo válido accesible en el ejemplo no accesibles (NOA). El enunciado es un poco complejo de comprender al plantear “conseguir una presentación visual deseada usando solo textos y estilos”. El criterio 1.4.12 dispone de un plugin directo y todos los estudiantes han presentado propuestas válidas en los ejemplos accesibles y no accesibles (ACC y NOA). El criterio 1.4.13 presenta únicamente un ejemplo accesible y no accesible (ACC y NOA) correctos, si bien en este caso es importante destacar que faltaban entregas de los estudiantes.

5.2.2. Principio 2. Operable

Pauta 2.1 Accesible por teclado, todas las funcionalidades han de estar disponibles por teclado. Los criterios 2.1.1 y 2.1.4 han sido válidos de identificar en todas las propuestas de los estudiantes. Una posible explicación es que el enunciado y el proceso de validar estos criterios son identificables de forma directa. El criterio 2.1.2 ha sido claro de identificar para el caso del ejemplo accesible (ACC), pero ninguna propuesta de ejemplo no accesible (NOA) ha sido adecuada. Puede ser debido al término “trampa para el foco del teclado” que es complejo de comprender.

Pauta 2.2 Tiempo suficiente, proporcionar el tiempo suficiente para leer y usar el contenido. El criterio 2.2.1 presenta mucha diversidad en la presentación de los ejemplos. Por el término “redirección”, y porque los ejemplos presentados no hacían referencia directa a actualización del contenido, sino en algunos casos los estudiantes incluían información para parar elementos en movimiento y el ejemplo no se correspondía exactamente al criterio presentado. En el criterio 2.2.2 hay diversidad en la validez de los ejemplos accesibles y no accesibles (ACC y NOA). Una posible explicación es que en algunos casos los estudiantes no comprenden el término “actualización automática” y se confunden con elementos en movimiento.

Pauta 2.3 Ataques epilépticos y reacciones físicas, no incluir contenido que puede causar ataques epilépticos y reacciones físicas. El criterio 2.3.1 todos los estudiantes han podido encontrar un ejemplo válido accesible y no accesible (ACC y NOA). En este caso, el enunciado del criterio es sencillo y es fácil de verificar.

Pauta 2.4 Navegación, proporcionar diferentes opciones de navegación, ubicación y encontrar contenido. En el criterio 2.4.1 los estudiantes se confundieron al identificar el ejemplo como accesible (ACC) porque hacía referencia a un ejemplo no accesible (NOA). Una posible explicación es que los estudiantes pueden tener confusión con el término “evitar contenido repetido” porque no saben a qué hace referencia. El criterio 2.4.2 el atributo *title* puede estar presente en varios elementos (enlaces, imágenes, botones, iframes,... etc) y los estudiantes han confundido la identificación del atributo *title* en el código HTML que ha de estar incluido en la cabecera de la página, no en cualquier otro elemento HTML. En el criterio 2.4.3 los ejemplos han sido muy diversos, porque ha sido difícil determinar correcta o incorrectamente el orden del foco en los

elementos interactivos. En algunos casos, sobre todo en los ejemplos no accesibles (NOA) los ejemplos no se han podido incluir correctamente. El criterio 2.4.4 las propuestas accesibles (ACC) y no accesibles (NOA) han sido diversas. Puede explicarse con la complejidad de la comprensión del término “propósito de los enlaces”. Para el criterio 2.4.5, las propuestas presentadas en los ejemplos accesibles (ACC) y no accesibles (NOA) han sido diversas en cuanto a su validez. Los estudiantes se han confundido en la presentación del ejemplo accesible y no accesible. Y esto indica poca comprensión del criterio. Para el criterio 2.4.6, hay diversos ejemplos accesibles (ACC) y no accesibles (NOA). Una posible explicación es que algunos estudiantes pueden confundirse con el criterio 1.3.1 que indica la jerarquía de los títulos en código HTML; sin embargo, este criterio solo hace referencia al texto del título que ha de ser adecuado al contenido de la sección. Para el criterio 2.4.7 solo existe una propuesta válida presentada y es correcto en ambos casos, ejemplo accesible (ACC) y no accesible (NOA).

Pauta 2.5 Modalidades de entrada, facilitar la activación de funcionalidades con el teclado y otros métodos. Para el criterio 2.5.1 se presentan propuestas de ejemplos accesibles variables, y se incluyen ejemplos accesibles de webs de ejemplo o referencia. Para el criterio 2.5.2 no se han incluido ejemplos válidos ni accesibles ni no accesibles. Esto puede explicarse por la dificultad de comprensión del término “acciones accidentales” y cómo se evalúa el criterio en el código HTML. En el criterio 2.5.3, los ejemplos que se han incluido accesibles y no accesibles (ACC y NOA) válidos se han cumplimentado con ejemplos de código HTML. El criterio 2.5.4, es complejo de comprender por el término “activación de movimiento” y no se han incluido casi ejemplos válidos accesibles o no accesibles.

5.2.3. Principio 3. Comprensible

Pauta 3.1 Legible, el contenido textual ha de ser legible. El criterio 3.1.1 ha sido fácil de comprender por todos los estudiantes por presentar instrucciones claras para evaluarlo. El criterio 3.1.2 ha presentado más ambigüedad en los ejemplos porque el enunciado se presenta más complejo y tampoco los estudiantes están habituados a encontrar contenidos en dos idiomas.

Pauta 3.2 Previsible, las páginas web han de comportarse de forma previsible. Los criterios 3.2.1 y 3.2.2 no presentan ejemplos válidos accesibles y no accesibles (ACC y NOA). Una posible explicación es la complejidad del enunciado en cuanto al término “cambio de contexto”. El criterio 3.2.3 y 3.2.4 presentan mejores resultados en cuanto a la validez de ejemplos. Esto es debido a que hacen referencia a la navegación e identificación consistente de elementos, conceptos comprensibles por los estudiantes.

Pauta 3.3 Ayuda en la entrada, evitar errores. El criterio 3.3.1 presenta diversos resultados en cuanto a la validez de los ejemplos accesibles y no accesibles (ACC y NOA). Una posible explicación podría ser la dificultad del texto del enunciado y la complejidad en la evaluación. En el criterio 3.3.2 el resultado de la validez de los ejemplos ha sido muy diversa porque algunas propuestas de ejemplos accesibles y no accesibles (ACC y NOA) no se adecuaban correctamente. Una razón podría ser la división de este criterio en varios subcriterios (A-D) y que cada uno hace referencia a aspectos concretos que los estudiantes no saben identificar, principalmente por la complejidad de evaluación y la confusión de mostrar el elemento concreto de la evaluación realizada. El criterio 3.3.3 presenta en general ejemplos accesibles y no accesibles válidos. Los estudiantes identificaron correctamente cuando en un formulario se presentaban los errores de forma comprensible. En el criterio 3.3.4, ningún estudiante ha podido plantear un ejemplo accesible o no accesible (ACC y NOA) válido. Una posible razón es que el enunciado del criterio se plantea de un modo ambiguo “asegurar que los procesos con errores importantes se puedan corregir antes de enviar los datos” y la evaluación puede ser demasiado técnica.

5.2.4. Principio 4. Robusto

Pauta 4.1 Compatible, el contenido ha de ser compatible con los agentes de usuario y ayudas técnicas. El criterio 4.1.1 dispone del evaluador automático de la gramática HTML y todos los estudiantes han sabido como evaluarlo correctamente e incluir ejemplos accesibles y no accesibles adecuados. En cuanto a los criterios 4.1.2 y 4.1.3 se observa diversidad al presentar los ejemplos tanto accesibles como no accesibles (ACC y NOA). Una posible explicación es la complejidad del texto del enunciado de cada uno de los criterios y la dificultad de evaluación muy relacionada con buscar etiquetas de WAI-ARIA en el código HTML que los estudiantes no conocen en profundidad porque son muy específicas de accesibilidad.

6. Discusión

A continuación, se presenta una discusión de los resultados en cuanto a análisis de datos cuantitativos y detallados de todos los criterios. Luego se presenta una discusión de los resultados observados para cada universidad. Posteriormente, se incluyen las limitaciones del estudio y las implicaciones de los resultados para mejorar las formaciones relacionadas con la accesibilidad

6.1. Análisis cuantitativo detallado

En cuanto al análisis global de todos ejemplos de los criterios analizados, los resultados cuantitativos presentan un bajo porcentaje de propuestas correctas (en ningún caso superando el 33%), tanto en los ejemplos accesibles (ACC) y no accesibles (NOA). Estos datos sugieren que solo una minoría de estudiantes logró representar de manera válida los ejemplos de los criterios de accesibilidad web en ambos tipos de ejemplos, lo cual evidencia la necesidad de fortalecer los procesos formativos vinculados a la interpretación y aplicación práctica de dichas pautas.

6.2. Análisis cuantitativo y detallado por universidad

La Tabla 3 presenta el detalle porcentual de ejemplos válidos (accesibles y no accesibles – ACC y NOA) desarrollados por los estudiantes de cada universidad, agrupados por los cuatro principios de accesibilidad de las WCAG 2.1. Los datos permiten observar diferencias relevantes en los niveles de adecuación alcanzados por cada universidad.

Se observa que la Universidad Pontificia de Salamanca (UPSA) presentó el mayor porcentaje de ejemplos válidos, con un 39,69% de adecuación total. Le sigue la Universitat de Lleida (UdL) con un 34,09%, y por último la Universidad Francisco de Vitoria (UFV) con un 28,46%.

Observando los resultados, se puede observar que la UdL y la UPSA presentan porcentajes similares (38,46% y 40,38%) y superiores a la UFV (23,08%), lo que puede indicar que existe una mejor comprensión de aspectos relacionados con la presentación visual y sensorial del contenido web. Por otro lado, en el Principio 2, se observa que la UdL y UFV presentan el mismo porcentaje (36%) mientras que la UPSA muestra un porcentaje bastante inferior (16%), esto puede ser debido a posibles dificultades en la interpretación de criterios relativos a navegación y funcionalidad del contenido. Respecto el Principio 3, se observan unos porcentajes bajos de adecuación de ejemplos. Una posible explicación podría ser una menor familiaridad de los estudiantes con los aspectos cognitivos y lingüísticos de la accesibilidad digital. Finalmente, el principio 4. Robusto la UPSA destaca notablemente con un 66,67% de ejemplos válidos, posiblemente debido a una mayor competencia técnica en el uso correcto de tecnologías web compatibles con agentes de usuario. Estos resultados ponen de manifiesto la variabilidad en la comprensión y aplicación de los principios de accesibilidad, lo cual señala la necesidad de mejorar la formación específica en criterios técnicos y prácticos de las WCAG 2.1 en el ámbito de los grados universitarios.

Tabla 3

Análisis crítico de los aciertos/dificultades.

	UdL	UFV	UPSA
Principio1. Perceptible	38,46%	23,08%	40,38%
Principio2. Operable	36,00%	36,00%	16,00%
Principio3. Comprensible	28,57%	21,43%	35,71%
Principio4. Robusto	33,33%	33,33%	66,67%
Total	34,09%	28,46%	39,69%

6.2.1. Criterios válidos presentados en una sola universidad

A continuación, se listan los criterios en los cuales cada universidad ha sido la única en presentar correctamente un ejemplo accesible y un ejemplo no accesible válido.

- UPSA ha creado los ejemplos correctos para los criterios 1.1.1 (Contenido no textual), subcriterios I y K, 1.2.4. (Subtitulado), 1.2.5 (audiodescripción), 1.3.1 (Información y relaciones) en los subcriterios J, N y Q; el criterio 2.4.6 (Encabezados y etiquetas), 3.3.1 (Identificación de errores), 3.3.3 (Sugerencias ante errores) y 4.1.3 (Mensajes de estado).

Todos estos criterios están más relacionados con aspectos multimedia y de codificación de código HTML, aspecto que encaja con los conocimientos adquiridos por los estudiantes la participar en la asignatura Diseño, evaluación y desarrollo de interfaces del Grado en Ingeniería Informática.

- UdL ha creado los ejemplos correctos para los criterios 1.4.1 (uso del color), 1.4.2 (control de audio), 1.4.13 (Contenido al pasar el puntero o al recibir el foco), 2.2.1 (tiempo ajustable), 2.4.2 (titulado de páginas), 2.4.7 (foco visible), 2.5.3 (etiqueta en el nombre), 3.3.1 (identificación de errores).

Todos estos criterios están más relacionados con diseño visual, aspecto que encaja con los conocimientos adquiridos por los estudiantes la participar en la asignatura Diseño Web del Grado de Diseño Digital y Nuevas Tecnología.

- UFV ha creado ejemplos correctos accesibles y no accesibles en los criterios 1.1.1 (contenido no textual) en el subcriterio J, 2.2.2(Pausar, detener, ocultar) en el subcriterio B, 2.4.1 (Evitar Bloques), 2.4.6 (Encabezados y etiquetas) en el subcriterio B y 3.3.1 (Identificación de errores) en el subcriterio B.

Criterios con una tendencia más relacionada con interacción digital, aspecto que encaja con los conocimientos adquiridos por los estudiantes la participar en la asignatura Human Computer Interacción del Grado en Ingeniería Informática.

6.2.2. Criterios válidos presentados en las tres universidades

Se observa en las Tablas 4, 5, 6 y 7 que las tres universidades han podido presentar ejemplos válidos accesibles y no accesible (ACC y NOA) de los siguientes criterios: 1.3. (Información y relaciones) en el subcriterio A, B y H, 1.4.4 (Redimensión del texto), 1.4.10 (Reflujo), 1.4.12 (Espaciado de texto), 2.1.1 (Teclado), 2.3.1 (Destellos), 3.1.1 (Idioma de la página) y 4.1.1 (Procesamiento).

Esto evidencia que los estudiantes, al menos en estos casos específicos, han logrado interiorizar aspectos clave de la accesibilidad web. Esto es un aspecto importante ya que muchos de los criterios seleccionados, como 1.3.1 (Información y relaciones), 1.4.12 (Espaciado del texto), 2.1.1 (Teclado) o 4.1.1. (Procesamiento) requieren no solo comprensión conceptual sino también de un dominio técnico mínimo para poder presentar ejemplos adecuados tanto accesibles como no accesibles de cada criterio.

En cuanto a los principios, destaca especialmente que la cantidad más elevada de ejemplos accesibles y no accesibles pertenecen al Principio 1. Perceptible. Esto puede ser debido a que los estudiantes conocen buenas prácticas relacionadas con la estructuración semántica del contenido (por ejemplo, el uso correcto de encabezados o listas) y su presentación visual adaptable. Además, sugiere que los estudiantes tienden a comprender más fácilmente aspectos relacionados con la visibilidad y organización del contenido, posiblemente debido a su carácter más tangible o visual y los estudios técnicos a los que pertenecen. Respecto a los criterios del Principio 2. Operable, necesitaban un nivel de comprensión más elevada los criterios relacionados con la navegación por el teclado y los destellos en el contenido fueron más fáciles de identificar. En el caso del Principio 3. Comprensible, los estudiantes identificaron correctamente el idioma de la página, mostrando una fácil comprensión de criterios que son fáciles de evaluar. Y en el caso del Principio 4. Robusto, los estudiantes aplicaron correctamente la evaluación de la gramática HTML, lo que refleja una competencia técnica completamente adquirida.

6.2.3. Criterios no válidos presentados en las tres universidades

Se observa en las Tablas 4, 5, 6 y 7 que las tres universidades no crearon ejemplos válidos accesibles y no accesible (ACC y NOA) de los siguientes criterios: 1.1.1 (Contenido no textual) en los subcriterios B, C, D, F y G, 1.3.3 (Características sensoriales) en el subcriterio A, 3.2.1 (Al recibir el foco), 3.2.2 (Al ingresar los datos), 3.3.2 (Etiquetas o instrucciones), 3.3.4 (Prevención de errores), lo que sugiere una dificultad generalizada para interpretar correctamente su aplicación. Los subcriterios vinculados al uso adecuado de textos alternativos para imágenes (todos relacionados con el criterio 1.1.1) registraron un alto nivel de dificultad entre los estudiantes. Estos subcriterios exigen una comprensión más profunda y detallada de situaciones particulares en las que el uso del atributo “alt” no es suficiente o no se aplica de forma convencional. Por ejemplo, el criterio 1.1.1-B se refiere a la necesidad de incluir descripciones más extensas cuando el contenido visual es complejo, algo que requiere identificar adecuadamente el contexto de uso y conocer alternativas como descripciones enlazadas fuera del “alt”. El subcriterio 1.1.1-C, indica cómo incluir alternativas a elementos visuales no tradicionales como código ASCII, emoticonos o lenguaje *leetspeak*, lo cual introducen un nivel adicional de abstracción que no siempre es evidente para los estudiantes sin experiencia previa en accesibilidad cognitiva o semiótica visual. En cuanto al criterio 1.1.1-D, centrado en imágenes embebidas en hojas de estilo, y el 1.1.1-G, que trata la agrupación de elementos visuales con significado conjunto, requieren conocimientos técnicos específicos sobre el diseño web y una lectura interpretativa sobre la función comunicativa de los recursos gráficos. Finalmente, el criterio 1.1.1-F demanda saber cuándo y cómo proporcionar alternativas textuales extensas para gráficos informativos, lo que conlleva una capacidad crítica para juzgar la densidad semántica de una imagen y su necesidad de descripción extendida.

La aplicación de estos criterios exige habilidades técnicas y analíticas que van más allá del uso básico del atributo “alt”, en el que se necesita una comprensión profunda del contexto comunicativo y del diseño web. La falta de familiaridad con estos escenarios, junto con la ambigüedad en la interpretación de algunos criterios, dificultó la creación de ejemplos adecuados. Esto pone de manifiesto la necesidad de una mayor mediación pedagógica y de materiales de apoyo más accesibles que traduzcan los requerimientos normativos en orientaciones prácticas comprensibles.

En cuanto al subcriterio 1.3.3-A exige evitar referencias sensoriales como único medio de transmisión de información, lo que requiere un análisis semántico detallado y abstracto que dificulta su representación clara por los estudiantes.

Los criterios 3.2.1 y 3.2.2, centrados en evitar cambios de contexto inesperados al recibir foco o modificar controles, exigen simular interacciones dinámicas difíciles de encontrar en un entorno web real y reproducir en modo ejemplo generado por el estudiante. El criterio 3.3.2, requiere proporcionar etiquetas e instrucciones adecuadas en formularios, suele presentar dificultades por ser poco comprensible su enunciado. Finalmente, el criterio 3.3.4 implica la representación de escenarios críticos (legales, financieros o evaluativos) que requieren flujos de validación y corrección de errores que sobrepasan las competencias técnicas básicas.

6.3. Análisis de los resultados para mejorar la enseñanza de accesibilidad.

La observación de errores sistemáticos en criterios como los del grupo 1.1.1 (Contenido no textual), 1.3.3 (Referencias sensoriales), o aquellos relativos a cambios de contexto: 3.2.1 (Al recibir el foco) y 3.2.2 (Al ingresar datos), sugieren la necesidad de repensar las estrategias docentes para abordar estos contenidos. La ambigüedad del lenguaje normativo puede dificultar el proceso de aprendizaje. Por tanto, se hace necesario un mayor esfuerzo de mediación didáctica para traducir los requerimientos técnicos en propuestas pedagógicas claras y visuales. La disparidad encontrada en los resultados entre universidades revela que el tratamiento de la accesibilidad web en la formación universitaria no es suficientemente homogéneo tanto en metodología como curricularmente. Por ello, se establece una necesidad de integrar en los grados universitarios no solo fundamentos teóricos de accesibilidad, sino también prácticas de ejercicios guiados, simulaciones interactivas y retroalimentación por los docentes especializados.

Los resultados del estudio muestran que es necesario adaptar los materiales y estrategias didácticas según el nivel de complejidad de los distintos criterios WCAG. Algunos criterios, como los relacionados con la apariencia visual del contenido, son más fáciles de entender y aplicar. En cambio, otros criterios que implican aspectos más abstractos, como la interacción del usuario, los cambios de contexto o el significado del contenido, requieren conocimientos más avanzados y, en muchos casos, combinar conocimientos de distintas áreas como la programación, el diseño, la usabilidad o la semiótica.

En este contexto, se propone desarrollar itinerarios formativos que aborden la accesibilidad de forma progresiva, desde los conceptos más básicos hasta los más complejos, integrándola de manera transversal a lo largo de toda la formación universitaria. Esta aproximación permitiría una mejor preparación del estudiantado y una incorporación más efectiva de la accesibilidad en los futuros proyectos profesionales.

6.4. Amenazas y limitaciones

El presente estudio presenta una serie de amenazas y limitaciones metodológicas que deben ser tenidas en cuenta al interpretar los resultados:

- Ausencia de validación experta múltiple: Dado que se trata de una aproximación inicial de carácter exploratorio, la validación de los resultados se realizó únicamente con un experto en accesibilidad digital. Esta limitación reduce la solidez metodológica del análisis y restringe el alcance y la generalización de las conclusiones obtenidas.
- Ausencia de hipótesis y preguntas de investigación claramente formuladas: Aunque el objetivo general era identificar qué criterios de las pautas WCAG 2.1 resultan más difíciles de interpretar para el estudiantado, no se establecieron preguntas de investigación específicas ni se formularon hipótesis contrastables, lo cual restringe la posibilidad de aplicar análisis estadísticos rigurosos.
- Diseño limitado a un único tipo de actividad: La actividad práctica utilizada —la recolección de ejemplos accesibles e inaccesibles—, si bien estaba correctamente estructurada, puede no ser suficiente para evaluar en profundidad la comprensión de los criterios por parte de los estudiantes. Hubiera sido deseable incorporar instrumentos adicionales como entrevistas, rúbricas de evaluación o pruebas de transferencia de conocimiento.
- Muestra no representativa: La participación se limitó a estudiantes de titulaciones técnicas de tres universidades concretas, por lo que los resultados no pueden generalizarse a otros perfiles académicos o contextos formativos distintos.
- Análisis centrado en la WCAG 2.1: Aunque pertinente en el momento del estudio, el análisis no incorpora criterios de la versión 2.2 de las pautas WCAG, actualmente vigente, lo que limita su actualización futura y su aplicabilidad a nuevos estándares.

7. Conclusiones y trabajo futuro

El presente estudio ha permitido analizar el grado de comprensión y las principales dificultades que enfrentan los estudiantes de titulaciones técnicas al interpretar y aplicar los criterios de las pautas de accesibilidad WCAG 2.1 (niveles A y AA). A través de una actividad práctica común —centrada en la observación y elaboración de ejemplos accesibles e inaccesibles de cada pauta— se constató una notable variabilidad en la interpretación de los criterios, especialmente en aquellos formulados con un lenguaje técnico complejo o ambiguo. En cambio, los criterios susceptibles de verificación mediante herramientas automáticas fueron más fácilmente comprendidos y aplicados por el estudiantado.

El estudio se enmarca en una experiencia docente desarrollada durante los cursos académicos 2023–2024 y 2024–2025, en la que participaron 97 estudiantes procedentes de tres universidades españolas: Universitat de Lleida (UdL), Universidad Francisco de Vitoria (UFV) y Universidad Pontificia de Salamanca (UPSA). El objetivo principal fue fomentar la comprensión crítica de los 94 criterios que conforman las pautas WCAG 2.1 (niveles A y AA), mediante la aplicación práctica en el análisis de ejemplos reales.

Los resultados permiten identificar patrones significativos en la forma en que los estudiantes comprenden cada principio, pauta y criterio. En particular, se observó que los criterios con redacción clara o cuya evaluación puede apoyarse en herramientas semiautomáticas tienden a generar menos ambigüedad. Por el contrario, los criterios que implican juicios más interpretativos, contextuales o ligados a aspectos cognitivos, perceptivos o interactivos —tales como la comprensión del contenido, la orientación espacial o la navegación— provocan mayores dificultades, lo que sugiere la necesidad de reforzar la enseñanza con recursos específicos y metodologías pedagógicas adaptadas a su complejidad.

Asimismo, el tipo de ejemplos generados mostró una correlación directa con el perfil académico y el enfoque disciplinar de los participantes. En el caso de la UPSA, los estudiantes del Grado en Ingeniería Informática elaboraron ejemplos centrados en la codificación, en consonancia con los contenidos de la asignatura “Diseño, evaluación y desarrollo de interfaces”. Los estudiantes de la UdL, inscritos en el Grado en Diseño Digital y Nuevas Tecnologías, generaron ejemplos principalmente vinculados al diseño visual, lo que refleja la orientación de la asignatura “Diseño Web”. Por su parte, en la UFV, los ejemplos presentaron un enfoque más centrado en la interacción digital, alineado con los contenidos de la asignatura “Human-Computer Interaction”, donde se abordan conceptos como usabilidad, diseño centrado en el usuario y comportamiento interactivo.

De cara al futuro, se plantea ampliar el alcance del estudio abordando algunas de sus limitaciones metodológicas. En particular, se prevé incorporar un diseño basado en preguntas de investigación explícitas, hipótesis contrastables e inferencias estadísticas. Asimismo, se contempla diversificar la muestra incluyendo estudiantes de disciplinas como diseño, comunicación o educación, así como aplicar metodologías mixtas (cuantitativas y cualitativas) que permitan no solo identificar los criterios más problemáticos, sino también comprender las causas subyacentes. También se planea establecer mecanismos de validación externa —mediante revisión de expertos y triangulación de resultados— y evaluar el impacto de diferentes estrategias pedagógicas como el aprendizaje basado en proyectos, gamificación o simulaciones, en la adquisición de competencias en accesibilidad digital.

Finalmente, se prevé extender el análisis a los nuevos criterios incorporados en las pautas WCAG 2.2 y publicar los ejemplos más representativos en una wiki colaborativa de acceso abierto.

Agradecimientos

Agradecemos sinceramente a todos los estudiantes que participaron en este estudio por su implicación y compromiso. En particular, queremos destacar la valiosa contribución de aquellos que colaboraron activamente en la elaboración de la actividad de creación de ejemplos accesibles y no accesibles. Su participación ha sido clave para el desarrollo de este proyecto.

- Estudiantes de 2º curso del Grado en Diseño Digital y Tecnologías Creativas del curso 2024/25 de la Universitat de Lleida (UdL), que han cursado la asignatura de Diseño Web.
- Estudiantes de 3º curso del Grado en Ingeniería Informática del curso 2024/25 de la Universidad Francisco de Vitoria (UFV) que han cursado la asignatura de Human Computer Interaction
- Estudiantes de 4º curso del Grado en Ingeniería Informática, curso 2024/25 de la Universidad Pontificia de Salamanca (UPSA) que han cursado la asignatura Diseño, evaluación y desarrollo de interfaces.

Declaración de uso de IA Generativa

Durante la preparación de este trabajo, los autores utilizaron GPT-4 en versión libre para: Revisión gramatical y ortográfica. Tras utilizar este servicio, los autores revisaron y corrigieron el contenido. Los autores se responsabiliza plenamente del contenido de la publicación.

Referencias

- [1] W3C Web Accessibility Initiative, WCAG 2 overview, World Wide Web Consortium, 2025. URL: <https://www.w3.org/WAI/standards-guidelines/wcag/> (recuperado en marzo de 2025).

- [2] W3C, WCAG 2.1: Understanding the Web Content Accessibility Guidelines, 2024. URL: <https://www.w3.org/WAI/WCAG21/Understanding/intro> (recuperado en marzo de 2025).
- [3] W3C, Web Content Accessibility Guidelines (WCAG) 2.1, 2024. URL: <https://www.w3.org/TR/WCAG21/> (recuperado en marzo de 2025).
- [4] Y. Yesilada, G. Brajnik, M. Vigo, S. Harper, Exploring perceptions of web accessibility: a survey approach, *Behaviour & Information Technology* 34 (2015) 119–134. doi:10.1080/0144929X.2013.848238.
- [5] S. Lewthwaite, S. Horton, A. Coverdale, Researching pedagogy in digital accessibility education, *SIGACCESS Accessibility and Computing*, no. 134, Art. 2, Oct. 2022 (pub. 2023). doi:10.1145/3582298.3582300.
- [6] L. Moreno, A. M. Fernández-Pampillón, A. Sarasa, C. Rodrigo, J. García-Villalobos, Y. González, R. García-Mata, How to interweave accessibility with didactic and technological quality of digital educational materials, *J. Access. Des. All*, vol. 9, no. 2, pp. 141–168, 2019. doi:10.17411/jaccess.v9i2.193.
- [7] A. Lazar, J. Lazar, A. Pradhan, Using modules to teach accessibility in a user-centered design course, in: *The 21st International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '19*, Association for Computing Machinery, New York, NY, USA, 2019, pp. 554–556. doi:10.1145/3308561.3354632.
- [8] K. Shinohara, S. Kawas, A. J. Ko, R. E. Ladner, Who Teaches Accessibility?: A Survey of U.S. Computing Faculty, in: *Proceedings of the 49th ACM Technical Symposium on Computer Science Education, SIGCSE '18*, Association for Computing Machinery, New York, NY, USA, 2018, pp. 197–202. doi:10.1145/3178628.3178644.
- [9] ACM, Computer science curricula 2023, 2023. URL: <https://csed.acm.org/> (recuperado en marzo de 2025).
- [10] A. N. Kumar, R. K. Raj, Computer Science Curricula 2023 (CS2023): The final report, in: *Proceedings of the 55th ACM Technical Symposium on Computer Science Education, SIGCSE '24*, ACM Press, Portland, OR, USA, 2024, p. 2. doi:10.1145/3626253.3633405.
- [11] S. Lewthwaite, S. Horton, A. Coverdale, Workplace approaches to teaching digital accessibility: establishing a common foundation of awareness and understanding, *Frontiers in Computer Science* 5 (2023) 1155864. doi:10.3389/fcomp.2023.1155864.
- [12] W3C Web Accessibility Initiative, WAI Curricula on Web Accessibility, W3C, 2022. URL: <https://www.w3.org/WAI/curricula/> (recuperado en marzo de 2025).
- [13] AAP Nordic, IWAC – WAC: The integration of Web Accessibility Courses in ICT programmes, 2025. URL: <http://www.iaapnordic.org/projekt/iwac/> (recuperado en marzo de 2025).
- [14] Carleton University, READi – Research and Education in Accessibility, Design, and Innovation, 2025. URL: <https://carleton.ca/readi/> (recuperado en marzo de 2025).
- [15] Fundación ONCE, Real Patronato sobre Discapacidad, Formación curricular en accesibilidad y diseño para todas las personas en titulaciones técnicas relacionadas con las TIC (Ref. 461/2022), Fundación ONCE/Vía Libre, 2022. URL: <https://web.archive.org/web/20230320040010/https://biblioteca.fundaciononce.es/publicaciones/otras-editoriales/coleccion-accesibilidad/formacion-curricular-en-accesibilidad-y> (recuperado en marzo de 2025).
- [16] A. G. Gay, Open curriculum for teaching digital accessibility, *Frontiers in Computer Science* 5 (2023). doi:10.3389/fcomp.2023.1113936.
- [17] Q. Zhao, V. Mande, P. Conn, S. Al-khazraji, K. Shinohara, S. Ludi, M. Huenerfauth, Comparison of methods for teaching accessibility in university computing courses, in: *Proceedings of the 22nd ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*, Art. 6, pp. 1–12, 2020. doi:10.1145/3373625.3417013.
- [18] C. Ferrari, D. Kletenik, K. Sonka, D. Sturm, A. Hurst, Evaluating instructor strategy and student learning through digital accessibility course enhancement, in: *Proceedings of the 21st ACM SIGACCESS Conference on Computers and Accessibility*, pp. 377–388, 2019. doi:10.1145/3308561.3353795.
- [19] Washington University, Accessible University, 2025. URL: <https://www.washington.edu/accesscomputing/AU/index.html> (recuperado en marzo de 2025).

- [20] W3C WAI, Good and Bad Web Content Examples, 2012. URL: <https://www.w3.org/WAI/demos/bad/> (recuperado en marzo de 2025).
- [21] GRIHO, University of Lleida, Web accessible y no accessible del congreso interacción 2023, 2023. URL: <https://interaccion2023.github.io/interaccion2023/> (recuperado en marzo de 2025).
- [22] A. Pascual-Almenara, T. Granollers-Saltiveri, Combining two inspection methods: usability heuristic evaluation and WCAG guidelines to assess e-commerce websites, in: P. H. Ruiz, V. Agredo-Delgado, A. L. S. Kawamoto (Eds.), Human-Computer Interaction. HCI-COLLAB 2021, Cham: Springer, 2021, vol. 1478. doi:10.1007/978-3-030-92325-9_1.
- [23] A. Pascual-Almenara, J. H. Elich, T. G. Saltiveri, Past, present and future of user-centred design methodologies that empathise with the user and generate better accessible experiences: from theory to practice, in: S. Sayago, M. Sendín, M. Teixidó (Eds.), Future Perspectives on Human-Computer Interaction Research, Springer, 2025. doi:10.1007/978-3-031-71697-3_2.
- [24] A. Pascual-Almenara, J. Humanes-Elich, T. Granollers-Saltiveri, MPIu + aX, user-entered design methodology that empathizes with the user and generates a better accessible experience (from theory to practice), in: Proceedings of the XXIII International Conference on Human Computer Interaction (Interacción '23), Art. 11, pp. 1–3, 2023. doi:10.1145/3612783.3612795.
- [25] W3C, How to meet WCAG (Quick Reference), 2025. URL: <https://www.w3.org/WAI/WCAG22/quickref> (recuperado en marzo de 2025).
- [26] E. Ballesteros, M. Ribera, A. Pascual, Reflections and proposals to improve the efficiency of accessibility efforts, Universal Access in the Information Society 14 (2015) 583–586. doi:10.1007/s10209-014-0356-1.
- [27] W3C, Web Accessibility Laws & Policies, 2022. URL: <https://www.w3.org/WAI/policies/> (recuperado en marzo de 2025).
- [28] Observatorio de Accesibilidad Web, Guía técnica del principio Perceptible, 2021. URL: https://administracionelectronica.gob.es/pae_Home/dam/jcr:ec04c677-64b1-4be2-a51f-1352144e4556/Guia_Tecnica_-_Principio_Perceptible.pdf (recuperado en marzo de 2025).
- [29] Observatorio de Accesibilidad Web, Guía técnica del principio Operable, 2021. URL: https://administracionelectronica.gob.es/pae_Home/dam/jcr:ffb08d21-f9bf-4258-b111-246504fca67e/Guia_Tecnica_-_Principio_Operable.pdf (recuperado en marzo de 2025).
- [30] Observatorio de Accesibilidad Web, Guía técnica del principio Comprensible, 2021. URL: https://administracionelectronica.gob.es/pae_Home/dam/jcr:0218c154-89cf-4a2f-937d-74b57f0a41e9/Guia_Tecnica_-_Principio_Comprensible.pdf (recuperado en marzo de 2025).
- [31] Observatorio de Accesibilidad Web, Guía técnica del principio Robusto, 2021. URL: https://administracionelectronica.gob.es/pae_Home/dam/jcr:2c13789f-3f56-463b-89f6-9ac75a2ffb45/Guia_Tecnica_-_Principio_Robusto.pdf (recuperado en marzo de 2025).
- [32] Universitat de Lleida, Guía docente de la asignatura Diseño Web del Grado Diseño Digital y nuevas tecnologías, 2024-25. URL: https://guiadocent.udl.cat/es/html/2024-25_102184 (recuperado en marzo de 2025).
- [33] Universidad Francisco de Vitoria, Guía docente de la asignatura Human Computer Interaction, 2025. URL: https://notas.ufv.es/documentos/gd/5633_p.pdf (recuperado en marzo de 2025).
- [34] Universidad Pontificia de Salamanca, Guía docente de la asignatura Diseño, evaluación y desarrollo de interfaces, 2025. URL: https://adminpro.upsa.es/uploads/Diseno_Interfaces_Guia_Docente_Informatica_21_22_a6a5ba40ef.pdf recuperado en marzo de 2025

A. Apéndice. Tablas de datos y gráficos

A continuación, se muestran las tablas de datos recogidas durante la experimentación y gráficos resultantes

Tabla 4

Resultados de validación de los criterios del principio 1. Perceptible. La puntuación de la tabla indica: 0 - el ejemplo no corresponde al criterio, 1- el ejemplo está relacionado con el criterio, pero presenta errores o ambigüedades y 2- el ejemplo se adecúa totalmente al criterio.

Universidad	UdL	UdL	UFV	UFV	UPSA	UPSA	Media	Media
Criterio	ACC	NOA	ACC	NOA	ACC	NOA	ACC	NOA
1.1.1-A	2	2	0	2	2	0	1,3	1,3
1.1.1-B	1	0	1	0	0	0	0,7	0
1.1.1-C	1	0	0	0	0	0	0,3	0
1.1.1-D	1	0	0	0	0	0	0,3	0
1.1.1-E	2	2	0	2	2	2	1,3	2
1.1.1-F	0	0	0	0	0	0	0	0
1.1.1-G	0	0	0	0	0	0	0	0
1.1.1-H	2	0	-	-	0	0	1	0
1.1.1-I	0	0	0	0	2	2	0,7	0,7
1.1.1-J	-	-	2	2	0	0	1	1
1.1.1-K	1	1	0	1	2	2	1	1,3
1.2.1-A	0	0	1	0	2	0	1	0
1.2.1-B	-	-	-	-	0	2	0	2
1.2.2-A	2	2	2	2	2	1	2	1,7
1.2.3-A	2	2	0	0	2	2	1,3	1,3
1.2.4-A	0	0	0	0	2	2	0,7	0,7
1.2.5-A	-	-	-	-	2	2	2	2
1.3.1-A	2	2	2	2	2	2	2	2
1.3.1-B	2	2	2	2	2	2	2	2
1.3.1-C	0	2	-	-	2	1	1	1,5
1.3.1-D	2	2	2	2	2	1	2	1,7
1.3.1-E	2	2	0	-	0	2	0,7	2
1.3.1-F	-	-	2	2	1	2	1	1,3
1.3.1-G	2	2	0	0	2	1	1,3	1
1.3.1-H	2	2	2	2	2	2	2	2
1.3.1-I	0	0	2	0	2	2	1,3	0,7
1.3.1-J	1	1	0	1	2	2	1	1,3
1.3.1-K	-	-	0	-	1	1	0,5	1
1.3.1-L	0	2	0	0	2	0	0,7	0,7
1.3.1-M	0	0	2	0	2	2	1,3	0,7
1.3.1-N	0	0	0	0	2	2	0,7	0,7
1.3.1-O	2	2	2	2	0	0	1,3	1,3
1.3.1-P	2	0	0	-	2	0	1,3	0
1.3.1-Q	0	0	-	-	2	2	1	1
1.3.2-A	2	0	0	0	-	-	1	0
1.3.2-B	0	0	0	0	1	0	0,3	0
1.3.2-C	0	0	2	0	2	0	1,3	0
1.3.3-A	0	0	0	0	0	0	0	0
1.3.4-A	2	2	-	-	2	2	2	2
1.3.5-A	2	-	-	-	2	2	2	2

Tabla 5 (continuación)

Resultados de validación de los criterios del principio 1. Perceptible. La puntuación de la tabla indica: 0 - el ejemplo no corresponde al criterio, 1- el ejemplo está relacionado con el criterio, pero presenta errores o ambigüedades y 2- el ejemplo se adecúa totalmente al criterio.

Universidad	UdL	UdL	UFV	UFV	UPSA	UPSA	Media	Media
Criterio	ACC	NOA	ACC	NOA	ACC	NOA	ACC	NOA
1.4.1-A	2	2	0	0	1	1	1	1
1.4.1-B	1	2	0	0	2	2	1	1,3
1.4.1-C	0	0	0	2	0	2	0	1,3
1.4.2-A	2	2	0	0	-	-	0,7	0,7
1.4.3-A	2	2	-	-	2	2	2	2
1.4.3-B	2	2	2	2	-	-	2	2
1.4.4-A	2	2	2	2	2	2	2	2
1.4.5-A	-	-	1	2	0	0	0,5	1
1.4.10-A	2	2	2	2	2	2	2	2
1.4.11-A	0	0	0	1	0	0	0	0,3
1.4.12-A	2	2	2	2	2	2	2	2
1.4.13-A	2	2	0	-	-	-	1	2

Tabla 6

Resultados de validación de los criterios del principio 2. Operable. La puntuación de la tabla indica: 0 - el ejemplo no corresponde al criterio, 1- el ejemplo está relacionado con el criterio, pero presenta errores o ambigüedades y 2- el ejemplo se adecúa totalmente al criterio.

Universidad	UdL	UdL	UFV	UFV	UPSA	UPSA	Media	Media
Criterio	ACC	NOA	ACC	NOA	ACC	NOA	ACC	NOA
2.1.1-A	2	2	2	2	2	2	2	2
2.1.2-A	2	0	2	0	2	0	2	0
2.1.4-A	2	2	2	2	2	0	2	1,3
2.2.1-A	2	2	0	1	0	0	0,7	1
2.2.1-B	0	0	2	2	2	0	1,3	0,7
2.2.1-C	2	2	0	0	0	2	0,7	1,3
2.2.2-A	2	2	0	0	2	2	1,3	1,3
2.2.2-B	0	0	2	2	0	0	0,7	0,7
2.3.1-A	2	2	2	2	2	2	2	2
2.4.1-A	0	0	2	2	0	0	0,7	0,7
2.4.2-A	2	2	0	0	-	-	1	0,7
2.4.3-A	2	0	2	0	0	2	1,3	0,7
2.4.3-B	0	0	2	0	0	1	0,7	0,3
2.4.4-A	0	0	1	0	-	-	0,5	0
2.4.5-A	0	0	2	1	-	-	1	0,5
2.4.5-B	1	2	2	0	1	2	1,3	1,3
2.4.5-C	2	2	2	2	-	-	2	2
2.4.6-A	2	1	-	-	2	2	1,3	1
2.4.6-B	0	0	2	2	0	0	0,7	0,7
2.5.1-A	0	1	2	2	2	0	1,3	1
2.5.2-A	0	0	0	0	-	-	0	0
2.5.3-A	2	2	0	0	1	1	1	1
2.5.4-A	-	-	0	0	0	2	0	1

Tabla 7

Resultados de validación de los criterios del principio 3. Comprensible. La puntuación de la tabla indica: 0 - el ejemplo no corresponde al criterio, 1- el ejemplo está relacionado con el criterio, pero presenta errores o ambigüedades y 2- el ejemplo se adecúa totalmente al criterio.

Universidad	UdL	UdL	UFV	UFV	UPSA	UPSA	Media	Media
Criterio	ACC	NOA	ACC	NOA	ACC	NOA	ACC	NOA
3.1.1-A	2	2	2	2	2	2	2	2
3.1.2-A	0	0	2	0	2	1	1,3	0,3
3.2.1-A	0	0	0	0	0	0	0	0
3.2.2-A	0	0	0	0	0	0	0	0
3.2.3-A	2	2	0	0	2	2	1,3	1,3
3.2.4-A	2	0	2	0	2	0	2	0
3.3.1-A	0	0	0	0	2	2	0,7	0,7
3.3.1-B	0	0	2	2	0	0	0,7	0,7
3.3.2-A	0	0	0	0	0	0	0	0
3.3.2-B	0	0	2	0	2	2	1,3	0,7
3.3.2-C	2	2	0	0	0	0	0,7	0,7
3.3.2-D	2	2	2	2	-	-	2	2
3.3.3-A	1	2	2	0	2	2	1,7	1,3
3.1.2-A	0	0	2	0	2	1	1,3	0,3

Tabla 8

Resultados de validación de los criterios del principio 4. Robusto. La puntuación de la tabla indica: 0 - el ejemplo no corresponde al criterio, 1- el ejemplo está relacionado con el criterio, pero presenta errores o ambigüedades y 2- el ejemplo se adecúa totalmente al criterio.

Universidad	UdL	UdL	UFV	UFV	UPSA	UPSA	Media	Media
Criterio	ACC	NOA	ACC	NOA	ACC	NOA	ACC	NOA
4.1.1-A	2	2	2	2	2	2	2	2
4.1.2-A	0	0	0	0	2	1	0,7	0,3
4.1.3-A	0	0	1	1	2	2	1	1

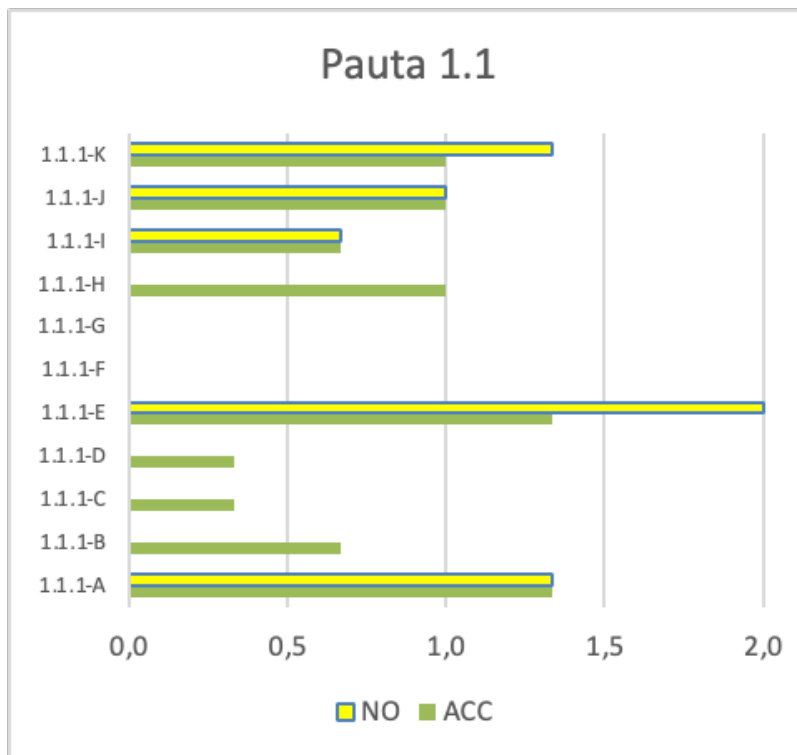


Gráfico 1: Principio 1 Perceptible. Pauta 1.1 Alternativas textuales

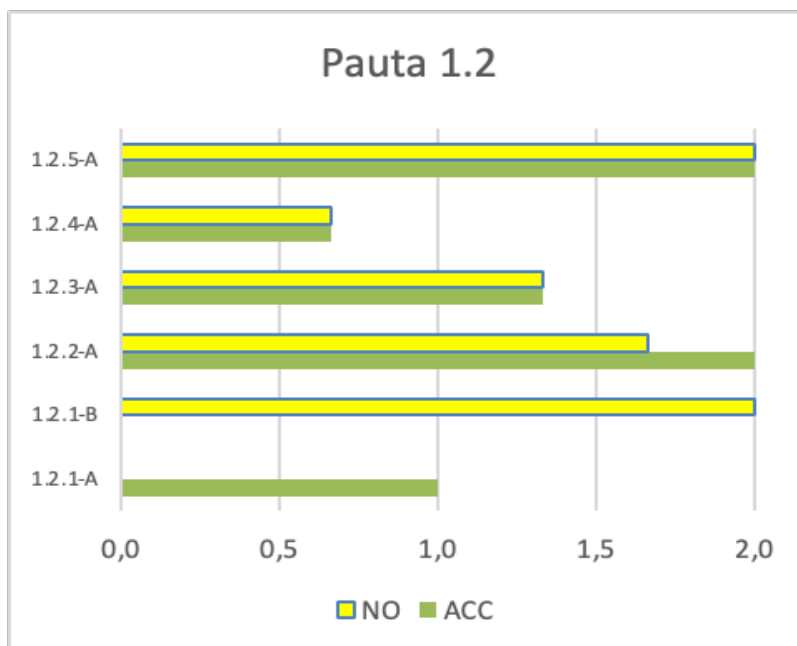


Gráfico 2: Principio 1 Perceptible. Pauta 1.2 Medios tempodependientes

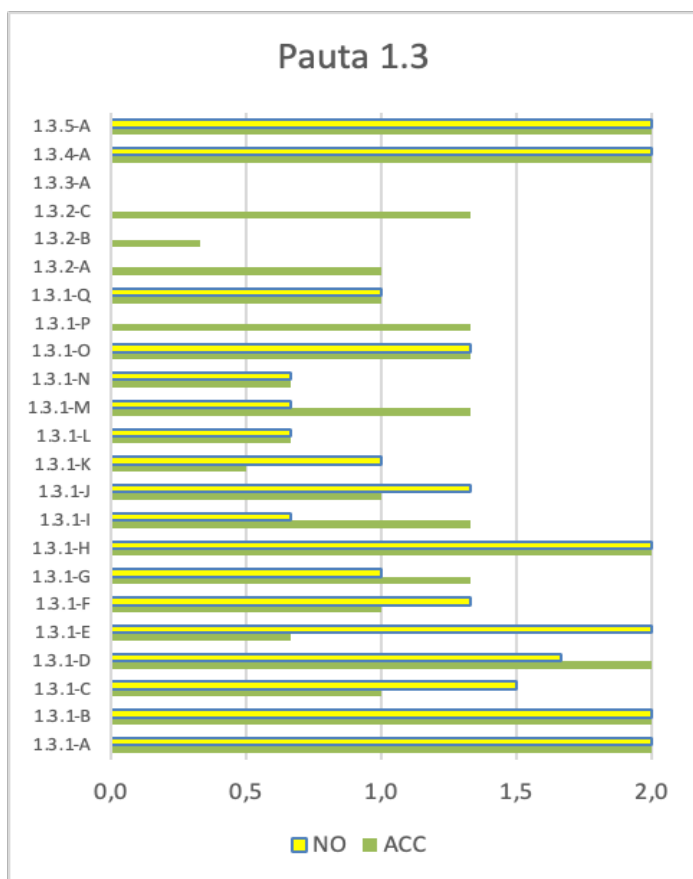


Gráfico 3: Principio 1 Perceptible. Pauta 1.3 Adaptable

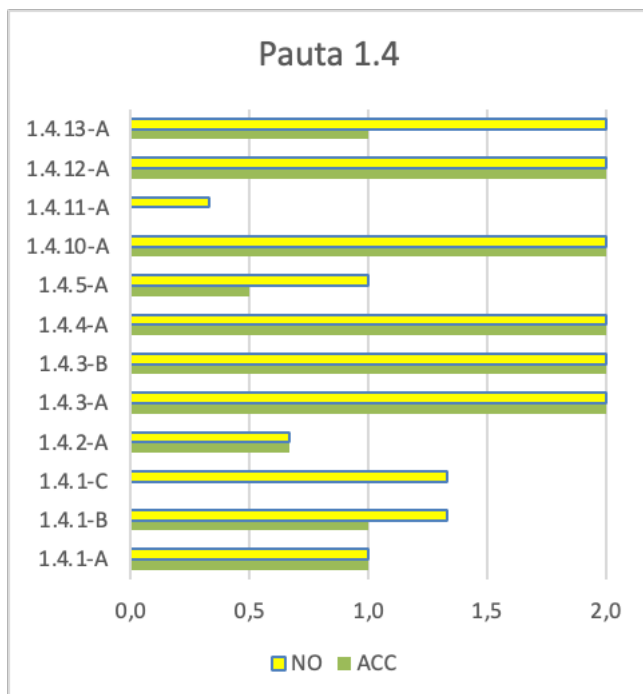


Gráfico 4: Principio 1 Perceptible. Pauta 1.4. Distingible

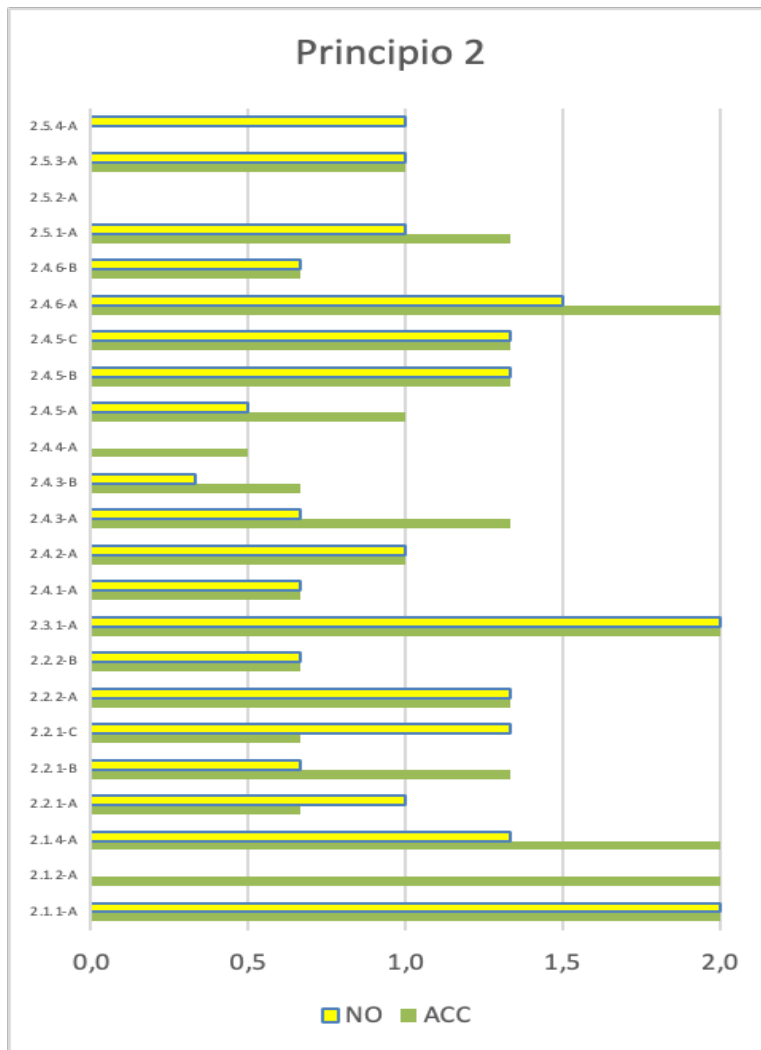


Gráfico 5: Principio 2 Operable.

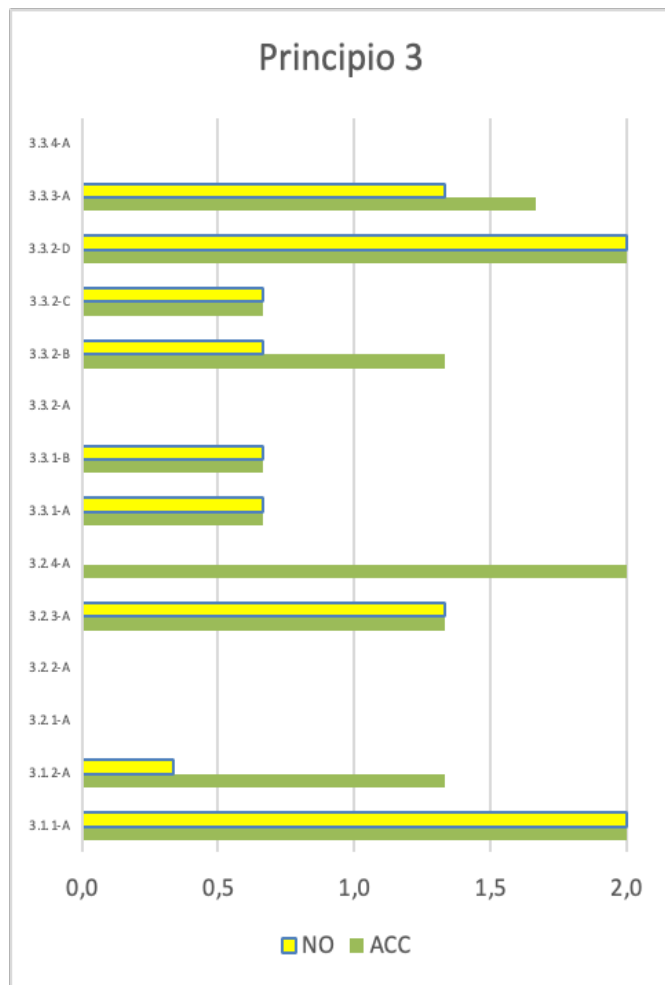


Gráfico 6: Principio 3. Comprensible

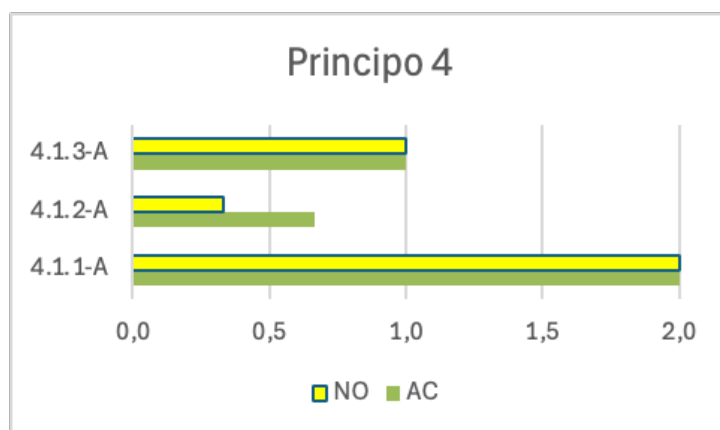


Gráfico 7: Principio 4. Robusto

Primeras experiencias para el aprendizaje de ciberseguridad a través del uso de robots sociales^{*}

Raquel Lacuesta^{1,†}, Daniel M. Pérez^{1,†}, Eva Hervás[†], Guillermo Azuara[†] and Jesús Gallardo[†]

Escuela Universitaria Politécnica de Teruel (Universidad de Zaragoza), C/Atarazana, n°2, 44003 Teruel, España

Resumen

La integración de robots sociales en el entorno educativo ha despertado un gran interés en los últimos años. Actualmente, existen dos enfoques principales en su uso, como tutores o como compañeros de aprendizaje. Mientras que la mayoría de los estudios se han centrado en su aplicación con niños en edad escolar, en este trabajo exploramos su potencial en el ámbito universitario, además nos centramos en un ámbito menos explorado con robots, la enseñanza de la ciberseguridad. En este estudio se presenta una experiencia para el aprendizaje en entornos universitarios de conceptos básicos de seguridad donde el uso del robot añade interactividad y movimiento frente a la interacción con computadores y/o tabletas. En dicha experiencia el robot interactúa con los usuarios a través de la realización de preguntas básicas sobre conceptos de seguridad. Los usuarios responden de forma táctil y/o por voz obteniendo retroalimentación instantánea por parte del robot.

Palabras clave

Implicación, aprendizaje de ciberseguridad, robots sociales, gamificación¹

1. Introducción

El uso de nuevas tecnologías en la educación es un campo en constante evolución. Recientemente, los robots sociales han emergido como una herramienta prometedora en el ámbito académico, ofreciendo oportunidades significativas para mejorar la retención del conocimiento. Gracias a su capacidad para proporcionar una experiencia gamificada e interactiva [2], estos robots facilitan la participación activa de los alumnos, rompiendo la barrera entre lo digital y lo físico y favoreciendo un aprendizaje más dinámico.

Si bien la mayoría de los estudios se han centrado en la integración de robots en entornos escolares, su aplicación en el ámbito universitario sigue siendo un tema poco explorado. No obstante, algunas investigaciones [5] sugieren una percepción positiva por parte de los estudiantes universitarios respecto al uso de estos robots en las aulas, aunque con reservas respecto a su autonomía para calificar o tomar decisiones académicas.

Aunque existen diversos estudios sobre la implementación de robots sociales en la enseñanza de diferentes disciplinas, como es el caso de la enseñanza de idiomas [7], hasta la fecha no se han encontrado investigaciones que analicen específicamente su impacto en la enseñanza de ciberseguridad. Por ello, el presente trabajo explora cómo la interacción con robots sociales puede contribuir a mejorar el aprendizaje y la retención de conceptos en esta área específica del conocimiento en el contexto de estudios universitarios.

A continuación, en la sección 2, se lleva a cabo una revisión de la literatura sobre los beneficios de integrar robots en el ámbito educativo y su impacto en la implicación de los estudiantes, presentados como trabajos previos. Posteriormente, la sección 3 detalla el diseño y la implementación de la aplicación utilizada en el robot. La sección 4 aborda las pruebas realizadas y

^{*}Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

¹Corresponding author

[†]These authors contributed equally

✉ lacuesta@unizar.es (R. Lacuesta); danielmiguel.perez@unizar.es (D. M. Pérez); eva.hervas@unizar.es (E. Hervás); guillermo.azura@unizar.es (G. Azura); jesus.gallardo@unizar.es (J. Gallardo)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

los resultados obtenidos. Finalmente, el artículo concluye con las principales conclusiones derivadas del estudio y se proponen futuras líneas de investigación.

2. Trabajos previos

El uso de robots sociales en el ámbito académico ha ganado relevancia en los últimos años, con una revisión temprana destacando su potencial como herramientas educativas versátiles en distintos niveles de enseñanza [8]. Estos robots pueden desempeñar varios roles; si bien algunos estudios los posicionan como co-estudiantes que apoyan metodologías activas y favorecen la participación [9, 11], el presente estudio se alinea con el rol de tutores o asistentes del profesorado, un enfoque investigado en [4].

La capacidad de los robots para fomentar el engagement y la personalización del aprendizaje es un área clave de investigación. Por ejemplo, el estudio de [2] demostró que la interacción personalizada de un robot, al recordar detalles del estudiante, generaba un mayor nivel de implicación y una mejor retención de la información. Esta personalización es una ventaja significativa frente a las tecnologías de enseñanza digitales tradicionales como ordenadores o tabletas. Aunque también existen enfoques educativos efectivos que utilizan robots no sociales para enseñar sobre tecnología [6], los robots sociales ofrecen beneficios únicos en términos de implicación y personalización [10] gracias a su presencia física. Esta les permite ofrecer interacciones multimodales que combinan habla, movimiento y expresión corporal. Estudios como el de [3] han evidenciado que el uso de gestos en robots puede mejorar la memorización a largo plazo del vocabulario de un segundo idioma, lo que se traduce en un aprendizaje más efectivo debido a la mayor implicación de los estudiantes.

A pesar de los avances en la integración de robots sociales en diversas disciplinas educativas, como el aprendizaje de idiomas, y el reconocimiento de su potencial en el ámbito universitario [5], la literatura carece de investigaciones que analicen específicamente su impacto en la enseñanza de la ciberseguridad. Este vacío en la investigación es precisamente lo que este trabajo busca abordar, explorando la contribución de los robots sociales en la retención de conceptos de ciberseguridad en estudiantes universitarios.

3. Diseño e implementación de la app

3.1. Diseño de la aplicación

La aplicación planteada integra una serie de ejercicios de evaluación de conceptos de seguridad. El robot Sanbot enunciará las preguntas de forma progresiva, recogerá las respuestas del usuario mediante reconocimiento de voz o de forma táctil y proporciona retroalimentación. Esta retroalimentación se manifiesta a través de voz, cambios de color, movimientos, sonrisas, entre otras acciones, dependiendo de si la respuesta es correcta o incorrecta.

El objetivo de esta interacción, aunque sencilla, es fomentar un vínculo entre el robot y el usuario, lo que contribuye a la implicación de los estudiantes en la evaluación.

A la hora de diseñar la aplicación se ha pretendido que los cuestionarios fueran sencillos de entender y con opciones que sean legibles en la interfaz táctil del usuario (pantalla). Se observa en la Figura 1 un ejemplo de una pregunta con tres opciones. Para el diseño del cuestionario se ha contado con una profesora del área de la ciberseguridad que ha validado las preguntas y respuestas, así como su dificultad.



Figura 1: Interfaz de la aplicación desarrollada en Android Studio.

3.2. Implementación

Para esta prueba, se ha decidido utilizar el robot social Sanbot ELF debido a sus características humanoides y su compatibilidad con Android. El prototipo fue desarrollado en Android Studio utilizando el lenguaje Java.

Esta implementación busca aprovechar las ventajas de la interacción robot-alumno para la enseñanza de ciberseguridad, proporcionando una experiencia más inmersiva y efectiva en comparación con los métodos tradicionales.

Al abrir la aplicación, se presenta un menú principal donde los estudiantes pueden seleccionar el test que desean realizar. Una vez seleccionado un test, se inicia la prueba y en pantalla aparece la pregunta actual junto con varias opciones de respuesta. El robot enuncia la pregunta y sus posibles respuestas en voz alta, facilitando la interacción auditiva.

Para responder, el usuario puede hacerlo de manera táctil o activando el reconocimiento de voz. En lugar de responder “la A”, “la B”, etc. se decidió que el/la usuario tuviera que repetir la respuesta completa, para que de este modo se interioriza mejor el contenido.

El sistema recoge el audio (respuesta del estudiante) y lo pasa a un filtro que subsane errores en la respuesta. Una vez convertida a texto, comprueba la similitud entre esta y la respuesta correcta con un porcentaje de error. Una vez hecho esto el robot proporcionará retroalimentación inmediata (acierto o error).

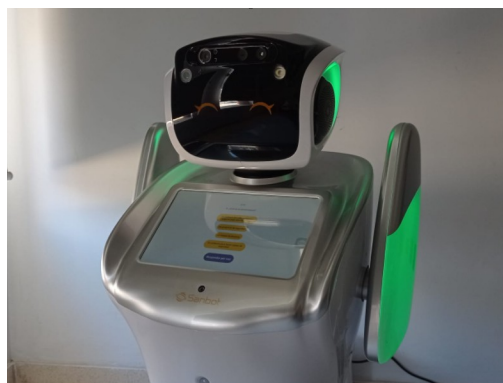


Figura 2: Sanbot en uso, ejecutando la aplicación diseñada. En estado de celebración. (2025)

- Si la respuesta es correcta, el robot cambiará su expresión facial a alegre, modificará el color de sus luces y levantará los brazos mientras felicita al usuario (Figura 2).
- Si la respuesta es incorrecta, se activará una expresión de llanto, acompañada de un mensaje de ánimo y permitirá al estudiante intentarlo nuevamente.

Una vez completadas todas las preguntas, el test finaliza y el robot informa al usuario de que ha terminado la actividad. Tras esto, se regresa al menú principal, donde se puede seleccionar otro test para continuar practicando.

4. Pruebas y valoración

Se preparó el prototipo y se presentó a estudiantes de la universidad para que resolvieran un test de preguntas sobre ciberseguridad (Tabla 1). Participaron 22 personas, los cuales cursaban segundo año del grado de Ingeniería Informática y en clase han trabajado conceptos de ciberseguridad aplicados en el diseño de interfaces de usuario. Este aspecto es relevante a la hora de analizar los datos ya que están más familiarizados con este tipo de dispositivos que la población media. La muestra estuvo compuesta mayoritariamente por hombres (60 %), mientras que el 30 % eran mujeres. La edad de los participantes estaba comprendida entre 19 y 24 años. La actividad tuvo una duración aproximada de 15 minutos por grupo, más 5 minutos adicionales dedicados a completar un cuestionario. Cada grupo Cada alumno al menos debía contestar a una pregunta del test, siendo cada test de 10 preguntas.

Se solicitó a los participantes que se organizaran en equipos de hasta cinco personas. Sin embargo, con el fin de explorar posibles diferencias en la dinámica grupal y el rendimiento en función del tamaño del grupo, se permitió que dos de los equipos inicialmente formados se unieran, conformando así un único grupo más grande, compuesto por un total de diez participantes. Esta decisión tuvo como propósito observar si el tamaño del grupo influía en la interacción de los miembros con el robot.

Las preguntas del cuestionario se resolvieron de forma conjunta por los equipos, se les pidió que todos los miembros dieran la solución a mínimo una de las preguntas (responder la solución a una de las cuestiones ya fuera de manera táctil o utilizando la voz). No se les dio ninguna dirección de cómo debían razonar las preguntas, esto fue intencionado, para así analizar el comportamiento de estos estudiantes. El número de opciones de respuesta variaba en función de la pregunta, con un rango de entre 3 y 6 alternativas.

Tabla 1

Cuestionario de ciberseguridad, realizado por los estudiantes durante la actividad.

Cuestionario de ciberseguridad
1. ¿Qué es el phishing? 2. ¿Cuál es una buena práctica para crear una contraseña segura? 3. ¿Qué es un firewall? 4. ¿Qué tipo de ataque intenta sobrecargar un servidor con tráfico excesivo? 5. ¿Cuál de estos es un método de autenticación multifactor (MFA)? 6. ¿Qué es el ransomware? 7. ¿Cuál es una señal común de un correo electrónico de phishing? 8. ¿Qué significa que una página web tenga HTTPS? 9. ¿Cuál es la mejor manera de protegerse contra malware? 10. ¿Qué es la autenticación biométrica?

4.1. Resultados de los test

Al finalizar la experiencia se pidió a los participantes que rellenaran un cuestionario de implicación, test reducido propuesto en [1], su uso en estudios anteriores relacionados con tecnologías interactivas y entornos educativos lo convierte en una herramienta coherente con los objetivos del presente trabajo. Se centra en seis áreas diferentes (preguntas 1 a 17), además de otras dos propuestas por los investigadores para conocer la diversión percibida del estudiante. Además, se les preguntó por su género y edad para el análisis posterior. Durante la experiencia los evaluadores recogieron los comentarios o emociones que presentaban y el número de aciertos y fallos.

Tabla 2

Cuestionario de implicación, repartido al finalizar la experiencia.

Responder a las preguntas marcando la respuesta que más se aproxime a tu sentimiento: (1) Totalmente en desacuerdo (2) En desacuerdo (3) Ni de acuerdo ni en desacuerdo (4) De acuerdo (5) Totalmente de acuerdo	
Atención Enfocada	1. Me perdí/dejé llevar completamente en esta experiencia
	2. Estuve tan involucrado en esta experiencia que perdí la noción del tiempo.
	3. Estaba absorto en esta experiencia.
Usabilidad Percibida	4. Me sentí frustrado mientras usaba la aplicación del robot para aprender ciberseguridad.
	5. La aplicación del robot para aprender seguridad me pareció confusa de usar.
	6. Usar la aplicación del robot para aprender seguridad fue agotador.
Atractivo Estético	7. La aplicación del robot para aprender seguridad era atractiva.
	8. La aplicación del robot para aprender seguridad era estéticamente agradable.
	9. La aplicación apeló a mis sentidos.
Durabilidad	10. Usar la aplicación del robot para aprender conceptos de seguridad valió la pena.
	11. Recomendaría la aplicación del robot para aprender conceptos de seguridad a mi familia y amigos.
	12. Seguiría usando la aplicación del robot para aprender seguridad por curiosidad.
Novedad	13. El contenido de la aplicación del robot para aprender seguridad despertó mi curiosidad.
	14. Mi experiencia fue gratificante.
Implicación Sentida	15. Realmente me sentí atraído por esta experiencia.
	16. Me sentí involucrado en esta experiencia.
	17. Me sentí interesado en esta experiencia.
Motivación	18. Aprender conceptos básicos de seguridad a través de la interacción con un robot es divertido.
	19. Creo que podría aprender seguridad a través de la interacción con el robot.

Para el análisis de los datos tenemos que tener en cuenta que las preguntas 4,5 y 6 están invertidas, por lo que un bajo resultado implica una alta usabilidad.

Respecto a la evaluación, resultados en las Figura 3 y Figura 4 cabe anotar que hubo problemas de comprensión en la primera pregunta, “Me perdí/dejé llevar completamente en esta experiencia”,

se entendió como, “me perdí haciendo el test”. Se dio el caso de que se malinterpretó la pregunta, pero varios alumnos preguntaron por el significado durante el cuestionario y finalmente no hay una gran variación entre los resultados de esta y las semejantes de su área. Como resultado podemos apreciar que esta pregunta es una con mayor dispersión en sus respuestas.

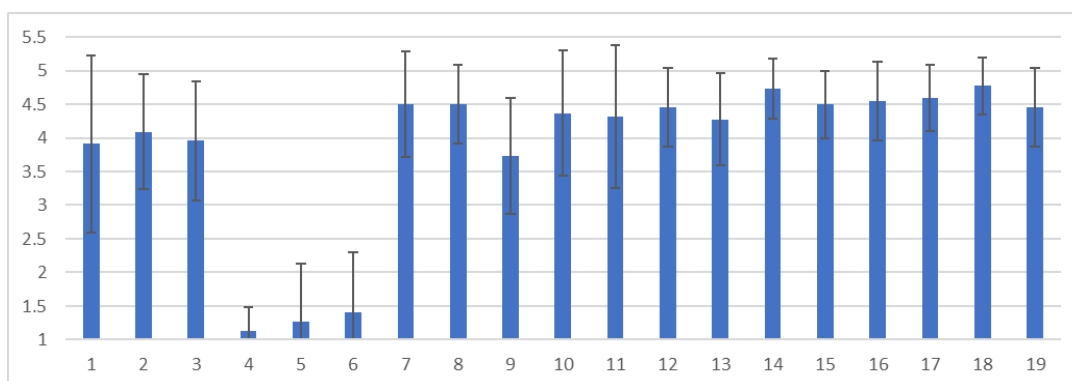


Figura 3: Resultado medio de cada pregunta.

También se aprecia en la pregunta 9 “la aplicación apeló a mis sentimientos” la media baja con respecto a las demás de su misma área y la media general, dando una media de 3’7 puntos, como se puede ver en la Figura 3. Hubo un gran porcentaje de gente que respondió de forma neutral. Se entiende que no hubo un alta “conexión emocional”.

El área de Usabilidad Percibida, tiene unas valoraciones muy positivas. En especial la Pregunta 4, “Me sentí frustrado mientras usaba la aplicación del robot para aprender ciberseguridad”, donde la mayoría de participantes respondió “Completamente en desacuerdo”. Se entiende que la actividad no causa frustración u otro tipo de sentimientos similares debido a su uso. Esto puede deberse a que hoy en día este tipo de tecnologías están ampliamente aceptadas y que la muestra seleccionada está interesada en esta área de conocimiento.

Por otro lado, una de las preguntas que destaca frente a las demás es la 18, Tabla 2, “Aprender conceptos básicos de seguridad a través de la interacción con un robot es divertido.” donde se valora el aspecto lúdico de esta experiencia, con una valoración media de 4’77. Lo que sugiere un alto nivel de satisfacción por parte de los participantes. La diversión, en este contexto, no solo se interpreta como un elemento anecdótico, sino como un indicador del potencial motivador que tiene el uso de robots sociales.

La mayoría de preguntas tienen una dispersión relativamente baja con medias altas, lo que indica un acuerdo fuerte entre los participantes. Sin embargo, las preguntas 1, 5 y 6 se podría considerar que tienen una alta dispersión. La pregunta 1 ya la hemos comentado previamente. Esta desviación en las preguntas 5 y 6 del área de usabilidad, nos dejan ver que a pesar de que la mayoría de alumnos encontraron la aplicación fácil de usar, hay un pequeño grupo al cual le fue complicado entender el funcionamiento y les resultó agotadora la experiencia.

Los grupos 1, 2 y 3 estaban formados por un máximo de 5 miembros, mientras que en el grupo 4 esta cifra se duplica. A continuación, en la Figura 4 se observa que la media en el cuestionario final no varía significativamente.

Sí se observó durante la prueba que en este último grupo los participantes se mantenían más alejados del robot y se centraron en responder bien a las preguntas individualmente sin discutirla con sus compañeros, utilizando al robot como una herramienta para realizar el test en lugar de un medio interactivo.

Por otro lado, en el grupo 1, 2 y 3, los participantes trataron las preguntas de manera conjunta, se discutía en grupo la solución y se respondió de forma incorrecta (1 vez en los grupos 2 y 3, y 3 veces en el grupo 1) intencionadamente para ver la reacción del robot. También se hicieron chistes sobre las expresiones del robot. En las respuestas se puede apreciar una media de la satisfacción notablemente más alta.

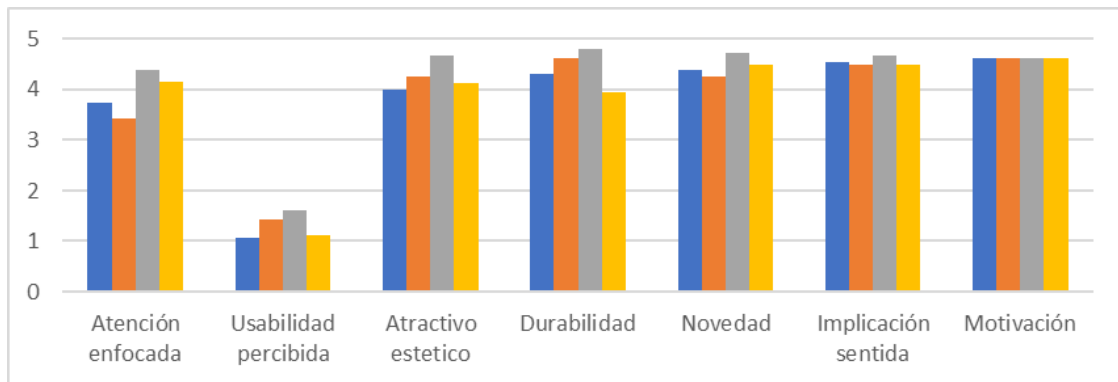


Figura 4: Media de los resultados por pregunta, test de implicación.

También queremos apuntar que la mayoría de los estudiantes tendía a responder utilizando la pantalla táctil. Esto puede relacionarse con la familiaridad de este tipo de dispositivos (smartphones, tabletas, etc.) y que únicamente el 20% de las preguntas se respondieron utilizando la respuesta por voz.

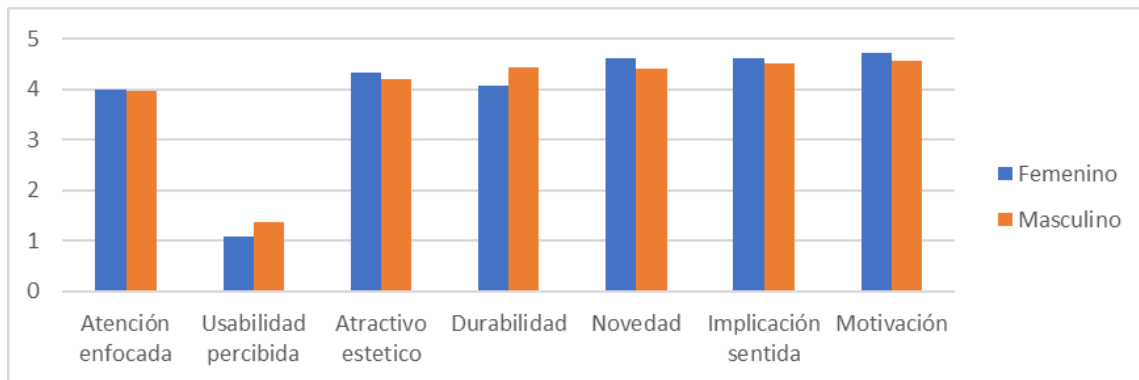


Figura 5: Media de los resultados por pregunta, test de implicación.

En el análisis de los resultados por género, se observó que las participantes femeninas otorgaron valoraciones más altas en la mayoría de las áreas evaluadas, tanto en los aspectos positivos como en los negativos. La única excepción se encontró en el área de durabilidad, que incluye ítems relacionados con la percepción de si la experiencia valió la pena y la disposición a recomendarla a otras personas, donde las valoraciones fueron ligeramente inferiores. Es importante señalar que la muestra estuvo compuesta mayoritariamente por hombres, quienes representaban el 60 % del total de participantes.

5. Discusión

Los resultados de esta experiencia sugieren que los estudiantes percibieron al robot no solo como una máquina, sino que mostraron una interacción más dinámica y, en ocasiones, lúdica. La preferencia inicial por la interacción táctil observada en los datos se complementó con una experimentación gradual de las respuestas por voz a medida que la prueba avanzaba. Los gestos y expresiones del robot contribuyeron significativamente a crear una atmósfera gamificada, evidenciada por las risas y los comentarios de asombro de los usuarios al descubrir nuevos movimientos.



Figura 6: Estudiantes realizando el test de ciberseguridad en la experiencia.

Se observó que, en grupos de tamaño reducido, donde la implicación de todos los miembros era más activa, la interacción con el robot era más fluida. Esto transformaba la experiencia, haciéndola percibida más como un juego que como una prueba teórica, lo cual podría haber influido positivamente en su implicación con la prueba.

Un comentario recurrente entre varios alumnos fue el deseo de que el robot realizara algún tipo de celebración (ej., un baile o una reacción expresiva) al finalizar la prueba. Aunque durante el test el robot acompañaba la interacción con movimientos sutiles de cabeza y brazos, así como cambios en sus expresiones faciales, su inmovilidad al concluir generaba una sensación de desconexión. Este aspecto subraya la importancia de mantener interacciones dinámicas y finales satisfactorios en el diseño de robots sociales para maximizar la experiencia del usuario.

6. Conclusiones

La revisión bibliográfica confirma el potencial de los robots sociales en entornos educativos, tanto en roles de educadores como de co-estudiantes [9]. Sin embargo, se identificó una ausencia de investigaciones que exploren su aplicación en la enseñanza de la ciberseguridad, un vacío que el presente estudio ha abordado mediante una experiencia de aprendizaje interactiva. Los resultados obtenidos evidencian el significativo potencial de este método de aprendizaje, especialmente en lo que respecta a la motivación y la implicación sentida por los estudiantes. Esta percepción positiva se ve reforzada por las valoraciones de los alumnos, quienes, como se observa en la pregunta 11 (Tabla 2), considerarían este medio útil para aprender nuevos conceptos y lo recomendarían a un amigo, destacando la experiencia como gratificante. Si bien estas buenas valoraciones podrían estar influenciadas por el efecto de novedad, su impacto a largo plazo será objeto de futuras investigaciones sobre el uso prolongado de robots en el aula.

En futuras líneas de investigación, se valorará la repercusión de este método en el aprendizaje real de los estudiantes en comparación con enfoques pedagógicos más tradicionales. Asimismo, se explorará la viabilidad de integrar este tipo de robots como complemento a las clases presenciales, sopesando la inversión de tiempo que implica la preparación de tales actividades frente al beneficio pedagógico y el valor en el aprendizaje que obtienen los participantes.

Agradecimientos

El trabajo está parcialmente financiado por el Ministerio de Ciencia e Innovación y Universidades (MCIU), la Agencia Estatal de Investigación (AEI) y la UE (FEDER) a través del contrato PID2022-136779OB-C31 y por el Gobierno de Aragón (Grupo T60_23R). Este trabajo está parcialmente financiado por EU-funded “Plan de de Recuperación, Transformación y Resiliencia” (Next Generation), Instituto Nacional de Ciberseguridad (INCIBE) (Spain), “Proyecto Estratégico de Ciberseguridad Desarrollado en Zonas Escasamente Pobladas.

Declaración de uso de IA Generativa

Durante la preparación de este trabajo, los autores utilizaron ChatGPT y Gemini con el fin de: Revisar gramática y ortografía. Después de utilizar estas herramientas/servicios, los autores revisaron y editaron el contenido según fue necesario y asumen total responsabilidad por el contenido de la publicación.

Referencias

- [1] H. L. O'Brien, P. Cairns, M. Hall, A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form, *International Journal of Human-Computer Studies* 112 (2018) 28-39. doi:10.1016/j.ijhcs.2018.01.004
- [2] R. Van Den Berghe, J. Verhagen, O. Oudgenoeg-Paz, S. Van Der Ven, P. Leseman, Social Robots for Language Learning: A Review, *Review of Educational Research* 89 (2018) 259–295.
- [3] G. Huang, R. Moore, Using social robots for language learning: are we there yet?, *Journal of China Computer-Assisted Language Learning* 3 (2023) 208-230. doi:10.1515/jccall-2023-0013
- [4] T. Belpaeme, F. Tanaka, Social robots as educators, in: *OECD digital education outlook 2021 pushing the Frontiers with artificial intelligence, blockchain and robots: pushing the Frontiers with artificial intelligence, blockchain and robots*, OECD Publishing, Paris, 2021, p. 143.
- [5] S. Serholt, W. Barendregt, Students' attitudes towards the possible future of social robots in education, in: *Workshop proceedings of RO-MAN*, 2014.
- [6] V. Girotto, C. Lozano, K. Muldner, B. Burleson, E. Walker, Lessons learned from in-school use of rtag: A robo-tangible learning environment, in: *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 919-930. doi:10.1145/2858036.2858454
- [7] F. Wang, A. Cheung, Robots' Social Behaviors for Language Learning: A Systematic Review and Meta-Analysis, *Review of Educational Research* (2024). doi:10.3102/00346543231216437
- [8] T. Kanda, M. Shimada, S. Koizumi, Children learning with a social robot, in: *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 351-358.
- [9] A. Ahtinen, A. Chowdhury, V. Ramírez Millan, C. H. Wu, G. Menon, Co-learning around social robots with school pupils and university students–Focus on data privacy considerations, in: *Proceedings of the 11th International Conference on Human-Agent Interaction*, 2023, pp. 115-123. doi:10.1145/3623809.3623816
- [10] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, Social robots for education: A review, *Science Robotics* 3 (2018). doi:10.1126/scirobotics.aat5954
- [11] S. Chandra, P. Dillenbourg, A. Paiva, Children Teach Handwriting to a Social Robot with Different Learning Competencies, *International Journal of Social Robotics* 12 (2019) 721.

Robots sociales en museos–Un enfoque centrado en el análisis de emociones*

Raquel Lacuesta, Fernando Barrado Lucía, Alejandro Guillén Florian^{1,†}, Francesc Xavier Gaya Morey, José María Buades Rubio^{2,†} and Francisco José Perales López^{3,†}

¹ Escuela Universitaria Politécnica de Teruel (Universidad de Zaragoza), C/Atarazana, n°2, 44003 Teruel, España

² Universitat de les Illes Balears, Carretera de Valldemossa, km 7.5, Nord, 07122 Palma, Illes Balears, España

³ Universitat Autònoma de Barcelona, Plaça Cívica, 08193 Bellaterra (Barcelona), España

Resumen

Actualmente, los robots sociales se presentan como una solución innovadora para guiar e interactuar con visitantes de todas las edades en entornos museísticos. Este trabajo expone un estudio centrado en el uso de un robot social humanoide en el contexto de visitas familiares a un museo. En colaboración con la Fundación Amantes de Teruel, se diseñó e implementó una aplicación lúdica para un robot interactivo con el objetivo de fomentar la participación activa y compartida entre niños y adultos. La interacción con el robot fue evaluada mediante el análisis de expresiones faciales de los miembros de la familia, como indicador del impacto emocional de las actividades. Los resultados muestran un predominio de emociones positivas durante la interacción, aunque con variaciones en función del tipo de actividad realizada. Estas conclusiones sugieren que la incorporación de robots sociales en museos no solo es viable, sino que también enriquece significativamente la experiencia familiar de los visitantes, abriendo nuevas posibilidades para la educación y el disfrute conjunto. Asimismo, se destaca la importancia de comprender en profundidad el efecto emocional que provocan las distintas actividades, con el fin de optimizar su diseño y adecuarlas a las características y necesidades del público familiar.

Palabras clave

interacción humano-robot, visitas familiares, interacción lúdica, reconocimiento facial, experiencia museística

1. Introducción

En la última década, los museos han incrementado la adopción de tecnologías interactivas para mejorar la accesibilidad y el compromiso del público [1], destacando la emergencia de los robots sociales como guías y facilitadores de experiencias. Estos robots, diseñados para interactuar naturalmente con las personas, ofrecen la posibilidad de personalizar recorridos y adaptar la experiencia del visitante.

Investigaciones previas [2, 3] han explorado el uso de robots humanoides en entornos comunitarios y educativos para fomentar interacciones intergeneracionales. Por ejemplo, estudios con preescolares, adultos mayores [2] y abuelos con nietos en juegos [3] han demostrado que los robots pueden facilitar la coordinación, el aprendizaje mutuo y mejorar la colaboración, siendo percibidos como herramientas útiles y fuentes de compañía. Estas evidencias sugieren que los robots sociales pueden actuar como puentes para la diversión y la educación, creando un espacio de participación conjunta en actividades culturales o educativas.

Basado en estos antecedentes, el presente artículo explora el uso de un robot social en un contexto museístico real: el Museo de los Amantes de Teruel. El objetivo principal es analizar cómo la interacción mediada por el robot puede ofrecer experiencias lúdicas y divertidas a grupos de

*Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

^{1†} Corresponding author.

[†] These authors contributed equally.

✉ lacuesta@unizar.es (R. Lacuesta); 801585@unizar.es (A. Guillén); josemaria.buades@uib.es (J. M. Buades);
paco.perales@uib.es (F. J. Perales)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

diferentes edades. El trabajo detalla el estado del arte, el diseño e implementación del robot, y la evaluación de la experiencia mediante el análisis de reconocimiento facial de las emociones de los usuarios.

2. Estado del arte

2.1. Robots sociales en museos

En los últimos años, los museos han integrado robots como guías y facilitadores de experiencias para mejorar la accesibilidad y el compromiso del público [1, 4]. Estos robots asumen roles de informador y facilitador social, narrando historias, respondiendo preguntas y adaptando su contenido al estado de ánimo del público. [1] reporta cómo el robot Pepper, dotado de visión artificial, personaliza recomendaciones, aumentando el involucramiento del visitante.

Sin embargo, la mayoría de estas implementaciones se han centrado en interacciones individuales o con grupos homogéneos, careciendo de un análisis profundo de la dinámica emocional en interacciones familiares o en la influencia de actividades lúdicas específicas que promuevan el aprendizaje conjunto. Aunque [2] y [3] respaldan el potencial de los robots en interacciones intergeneracionales para fomentar la participación y el aprendizaje mutuo, pocos estudios han explorado el análisis emocional detallado de estas interacciones en contextos museísticos reales, atendiendo cómo el diseño del juego influye en las respuestas afectivas.

Los robots afectivos pueden estimar emociones en tiempo real mediante expresiones faciales y señales fisiológicas [7], ajustando su comportamiento según la retroalimentación emocional del usuario. Estos sistemas se validan comparando su precisión con expertos humanos para asegurar un rendimiento fiable en entornos del mundo real [5].

A diferencia de los trabajos previos, la presente investigación ofrece un enfoque novedoso y una contribución significativa al centrarse en el potencial de los robots sociales como mediadores para fomentar la interacción y el aprendizaje compartido entre adultos y niños en museos. Nuestro estudio va más allá de la implementación, ofreciendo un análisis empírico detallado del impacto emocional de actividades educativas mediadas por el robot, utilizando reconocimiento facial para identificar cómo distintas dinámicas lúdicas influyen en las emociones de los usuarios. La bibliografía consultada no presenta estudios tan específicos en el análisis emocional del proceso de interacción humano-robot para diversas actividades y en contextos de interacción familiar en museos, lo que enriquece la comprensión de las vivencias de los usuarios y el diseño de experiencias interactivas.

3. Diseño de la aplicación

La experiencia se llevó a cabo en el Conjunto Museo de los Amantes de Teruel, un espacio cultural que narra la conocida leyenda de los Amantes.

3.1. Principios de diseño de la interacción

El diseño de las actividades interactivas para el robot Erox se guió por principios clave de la Interacción Humano-Computadora (HCI) y la educación museística, buscando optimizar la experiencia familiar y el compromiso emocional. Nos enfocamos en:

- **Colaboración intergeneracional.** Las actividades fomentaron la participación conjunta de niños y adultos, asegurando que los desafíos e interacciones fueran relevantes y accesibles para ambos grupos, incluyendo pausas y adaptación de ritmos.

- **Interactividad lúdica y afectiva.** Se priorizó el juego como vehículo de aprendizaje, integrando retroalimentación afectiva del robot (expresiones y movimientos) para una experiencia más empática y motivadora, incluso ante errores o frustraciones.
- **Claridad y consistencia.** Se diseñó una interfaz intuitiva con interacción multimodal (voz y táctil) para facilitar la familiarización del usuario. La comunicación del robot se adaptó para ser fluida, natural y humanizada.
- **Integración educativa.** Las actividades se alinearon con los objetivos del museo, reforzando conceptos clave de la leyenda a través de diversos formatos (diálogo, preguntas, ordenación de imágenes).
- **Reducción de la frustración.** Se implementaron mecanismos en el diseño del robot (respuestas afectivas, ayudas) para gestionar y reducir la aparición de emociones negativas en actividades con reto, transformándolas en oportunidades de aprendizaje y superación.

3.2. El robot Erox y sus capacidades tecnológicas

Se utilizó un robot social humanoide de tamaño mediano (aprox. 1 metro), apodado "Erox", equipado con movilidad sobre ruedas, brazos articulados, pantalla táctil frontal y un rostro digital con expresiones básicas para una comunicación humanizada. Sus sensores clave incluyen micrófonos y altavoces para reconocimiento y síntesis de voz, y una cámara frontal para interacción visual, reconocimiento facial y análisis de emociones.

El sistema de Erox integra un módulo de reconocimiento de emociones faciales (basado en [6] y [7]) capaz de categorizar en tiempo real seis emociones básicas (alegría, sorpresa, tristeza, miedo, disgusto y enojo). Para este estudio, las interacciones grabadas fueron procesadas a posteriori para extraer y organizar los datos emocionales por actividad.

Si bien estudios fundamentales [6] indican un rendimiento robusto del algoritmo subyacente en entornos controlados y bases de datos estandarizadas, reconocemos que la validación específica del módulo en ambientes no controlados y reales, como un museo, presenta desafíos por variaciones en iluminación, ángulos de cámara y expresiones sutiles. La interpretación de estas métricas en interacción real requiere cautela y es un área de investigación futura, sugiriendo validación cruzada con métodos como la auto-observación o codificación externa para afinar su precisión en poblaciones diversas y situaciones dinámicas.

3.3. Descripción de las actividades interactivas

Las actividades interactivas, diseñadas por los responsables del Museo de los Amantes, se centraron en la leyenda local para reforzar la experiencia educativa y promover la interacción familiar con el robot, generando emociones positivas. Se integraron en Erox diversas actividades lúdicas para involucrar simultáneamente a niños y adultos (ver Figura 1).

- **Actividad 1: “Diálogo con Erox”** (Figura 2). Permite a los usuarios familiarizarse con Erox a través de preguntas sobre la leyenda o peticiones de música. Erox responde por voz, con expresiones faciales y movimientos, humanizando la interacción, que también es táctil. Esta actividad busca tanto la diversión como el aprendizaje de interacción con el robot.



Figura 1: Robot Sanbot 'Erox' en el Mausoleo de los Amantes de Teruel.



Figura 2: Actividad del diálogo con el robot sobre los Amantes de Teruel.

- **Actividad 2: “Explicación ampliada”** (Figura 3). Las familias acceden a contenidos multimedia sobre la leyenda (videos, audios) mediante interfaz táctil o comandos de voz, complementando las explicaciones orales y facilitando la comprensión de la historia de forma envolvente.

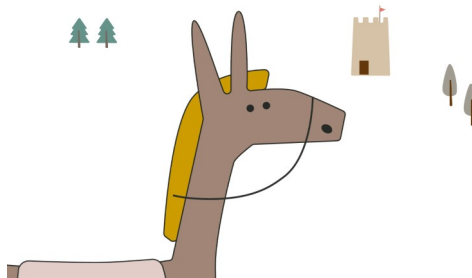


Figura 3: Uno de los videos que proyecta Erox en su pantalla para comprender la historia.

Tras la visita, Erox propone tres juegos interactivos para consolidar el aprendizaje:

- **Juego 1: “Juego del ahorcado”** (Figura 4(a)). Juego educativo para adivinar palabras de la leyenda. El robot ofrece retroalimentación afectiva (movimiento, expresiones faciales) ante aciertos o fallos, reforzando el aprendizaje a través del reto y la diversión.

- **Juego 2 “Juego de preguntas”** (Figura 4(b)). Consiste en preguntas de opción múltiple sobre la visita, evaluando la comprensión y memoria del visitante.
- **Juego 3 “Ordena imágenes de la historia”** (Figura 5). Reta al usuario a secuenciar imágenes clave de la leyenda.

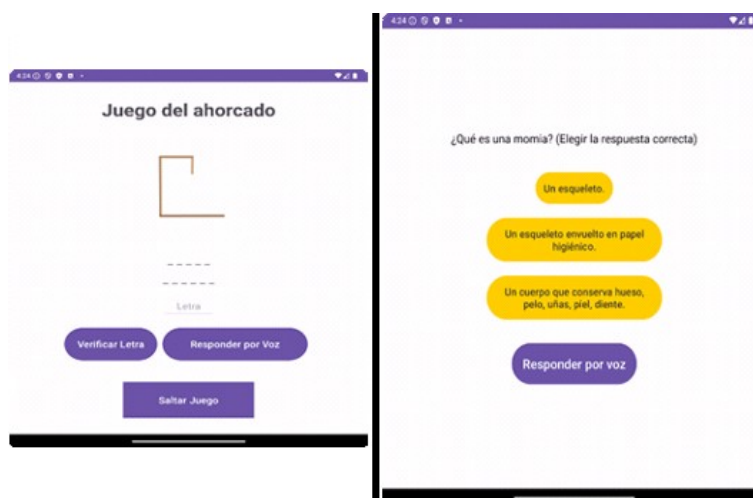


Figura 4: Interfaces de juegos interactivos. (a) Interfaz del juego “Juego del ahorcado” y (b) Interfaz del juego “Juego de preguntas”.



Figura 5: Interfaz del juego “Ordena las Imágenes según la Historia”.

3.4. Participantes y procedimiento

Para evaluar la intervención, se realizaron sesiones piloto con un total de 18 participantes, principalmente niños (8-14 años) acompañados por sus padres. De estos, 9 participantes dieron consentimiento para el estudio de emociones. Los participantes, siendo en su mayoría residentes locales de Teruel, no requirieron una visita guiada previa y pudieron interactuar directamente con el robot Erox en las actividades propuestas.

Al inicio de cada sesión, un investigador explicó el objetivo de la actividad (mejorar la experiencia museística con nuevas tecnologías). Las familias iniciaron con la "Actividad de Diálogo" y luego procedieron con los juegos, interactuando con Erox mediante voz o pantalla táctil.

El análisis de las expresiones de los participantes se realizó post-interacción usando un módulo de reconocimiento facial integrado en el robot, que detecta seis emociones básicas (alegría, sorpresa, tristeza, miedo, enfado, disgusto). Los videos grabados para este fin fueron eliminados del

sistema tras el procesamiento. La Figura 6 muestra ejemplos de usuarios interactuando con Erox durante las actividades.



Figura 6: Familias interactuando con el robot.

3.5. Consideraciones éticas y privacidad de datos

Todas las interacciones fueron grabadas en video para su análisis posterior. Se obtuvo consentimiento informado explícito y por escrito de todos los participantes mayores de edad y de los tutores legales de los menores, explicando el propósito de la grabación y el uso del reconocimiento facial. Se garantizó la voluntariedad de la participación y el derecho a retirarse en cualquier momento. Para garantizar la privacidad y la protección de datos sensibles, se implementaron los siguientes protocolos rigurosos:

- **Protección de la identidad.** Los videos fueron desidentificados inmediatamente después de su recopilación. Esto implicó eliminar cualquier identificador personal directo y asignar un código único a cada participante.
- **Almacenamiento seguro.** Todos los datos (videos brutos y procesados) se almacenaron en servidores seguros, cifrados y con acceso restringido, siguiendo las directrices del RGPD y la normativa local sobre datos personales y biométricos.
- **Acceso restringido.** Solo el equipo de investigación autorizado tuvo acceso a los datos, y este acceso estuvo limitado a las tareas necesarias para el análisis del estudio.
- **Eliminación de datos.** Todos los videos grabados fueron eliminados de forma segura e irreversible una vez completado el análisis y la extracción de datos pertinentes. Solo se retuvieron los datos cuantitativos anonimizados (ej., porcentajes de emociones) para análisis estadístico y gráficos, utilizando protocolos estandarizados para minimizar riesgos y maximizar la protección de la privacidad.

4. Resultados

4.1. Datos iniciales y análisis emocional

En la evaluación participaron 14 personas (9 para el análisis emocional). El 57,1% eran menores de 18 años, el resto adultos (menos del 10% mayores de 65 años), con un 64,3% hombres y 35,7% mujeres.

El análisis de emociones se realizó a posteriori utilizando un módulo de reconocimiento facial [6] que categoriza seis emociones básicas: felicidad, tristeza, miedo, sorpresa, disgusto y enojo. Los datos se procesaron de interacciones grabadas, segmentándose por actividad (diálogo y juegos).

4.2. Resultados emocionales por actividad

Los resultados emocionales variaron según la actividad, reflejando sus dinámicas específicas:

- **Diálogo con Erox** (Figura 7(a)). La felicidad predominó (47%), seguida de sorpresa (12%). Las emociones negativas (tristeza 10%, miedo 9%, disgusto 8%, enojo 14%) fueron mínimas. Esto sugiere una conexión emocional positiva inicial, favorecida por un diseño sin retos y la guía del personal.
- **Juego del ahorcado** (Figura 7(b)). La felicidad se mantuvo principal (37%), con aumento de sorpresa (18%), tristeza (14%), enojo (13%), disgusto (8%) y miedo (10%). La introducción del error generó momentos de frustración, pero el enfoque afectivo del robot y la sorpresa (interés) mantuvieron la experiencia positiva.

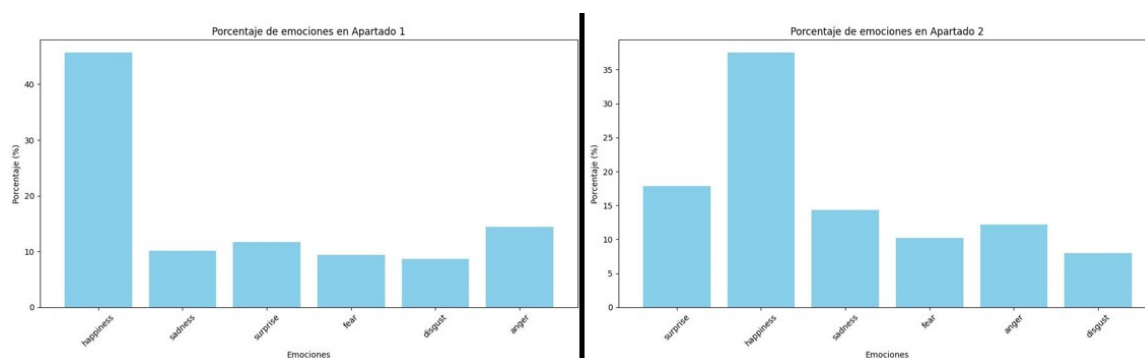


Figura 7: Gráfico de barras que detalla la distribución porcentual de diversas emociones en (a) el juego 'Diálogo' y (b) juego 'Ahorcado'.

- **Juego de preguntas** (Figura 8(a)). La felicidad predominó (29%), con notable aumento del enojo (26%) y sorpresa (18%). Tristeza (12%), disgusto (8%) y miedo (7%) fueron menores. La mayor exigencia cognitiva explica el incremento de emociones negativas por errores, aunque la retroalimentación del robot mantuvo la experiencia desafiante pero positiva.
- **Juego Ordena la historia** (Figura 8(b)). Felicidad (29%) y sorpresa (20%) prevalecieron, pero con mayor equidad en disgusto (12%), enojo (15%), tristeza (11%) y miedo (13%). La alta carga cognitiva y lógica secuencial explica la distribución emocional dispersa (necesidad de apoyo textual). No obstante, el reto fue motivador y el feedback afectivo del robot mantuvo la experiencia general positiva.

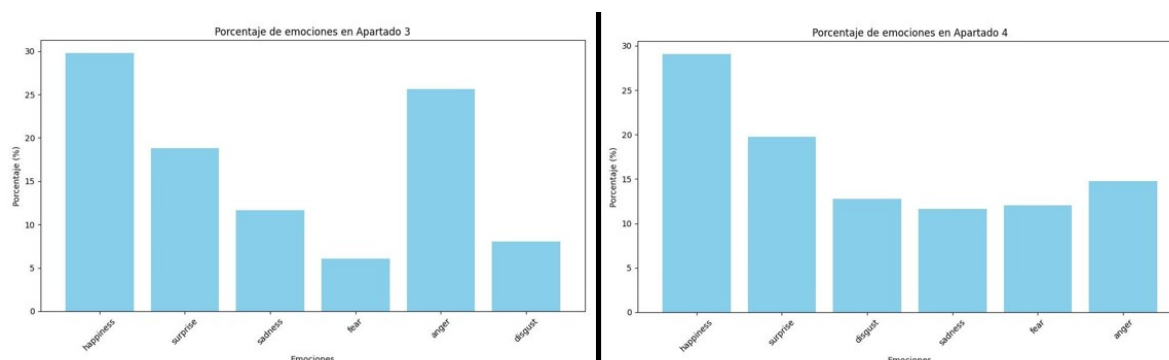


Figura 8: Gráfico de barras que detalla la distribución porcentual de diversas emociones en (a) el juego 'Juego de preguntas' y (b) juego 'Ordenar historia'.

4.3. Resultados globales

La Figura 9 muestra el promedio de emociones globales: felicidad (34%) y sorpresa (17%) fueron las más frecuentes, seguidas de enojo (17%), tristeza (12%), miedo (10%) y disgusto (10%). La prevalencia de felicidad valida el diseño para familias, y la emergencia de emociones negativas en momentos de evaluación es normal y pedagógicamente beneficiosa, activando reflexión y superación.

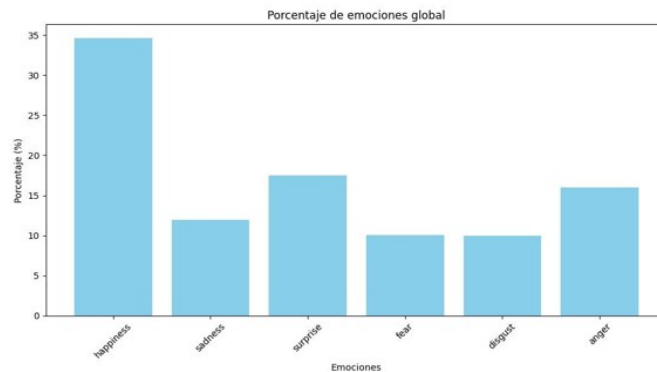


Figura 9: Gráfico de barras que ilustra la distribución porcentual global de diferentes emociones.

5. Conclusiones y trabajo futuro

Este estudio, realizado en colaboración con la Fundación Amantes de Teruel, resalta el significativo potencial de los robots sociales como mediadores innovadores en entornos educativos y culturales. Demuestra cómo el diseño de actividades puede influir profundamente en las emociones de los usuarios, fomentando el aprendizaje y el disfrute compartido entre niños y adultos en museos. Los resultados evidencian la motivación de los pequeños y la participación activa de los adultos, con un efecto emocional diferenciado según el reto propuesto por la actividad (lúdica, cognitiva, de memoria o interacción social).

La integración de un módulo de reconocimiento facial permitió una aproximación robusta a las vivencias emocionales de los usuarios durante las actividades lúdicas, identificando matices cruciales que otros métodos podrían pasar por alto. Esta tecnología ofrece una perspectiva valiosa para un diseño más adaptado y enriquecedor de futuras experiencias.

5.1. Limitaciones del estudio y sesgos

A pesar de los hallazgos, este estudio presenta limitaciones metodológicas. La muestra reducida (N=14; N=9 para análisis emocional) limita la generalización, y los resultados deben interpretarse como tendencias preliminares. El reclutamiento de participantes familiarizados con el museo pudo introducir sesgos. Aunque el módulo de reconocimiento emocional ha sido validado [6], su precisión en entornos reales y con distintos rangos de edad (especialmente en niños) requiere validación adicional, y las diferencias en expresiones faciales podrían afectar el rendimiento del algoritmo.

Este estudio demuestra la viabilidad de integrar un robot social como 'Erox' en museos para facilitar interacciones lúdicas y educativas intergeneracionales, evidenciando un impacto emocional positivo (felicidad, sorpresa). Aporta al campo de la HCI y la museología mediante el diseño de interacciones afectivas y experiencias culturales innovadoras, y subraya el valor pedagógico de las emociones (positivas y de desafío) en experiencias museísticas significativas.

Futuros estudios buscarán ampliar la muestra y su diversidad geográfica para mejorar la robustez y generalización de los resultados. Se validará más rigurosamente el sistema de reconocimiento facial, incluyendo comparaciones con codificación manual o autoevaluaciones. Además, se profundizará en el análisis del valor educativo de la emoción compleja, incorporando variables como aceptación, usabilidad y jugabilidad, con el fin de establecer un marco metodológico replicable para el diseño de experiencias culturales con robótica social.

Declaración de uso de IA Generativa

Durante la preparación de este trabajo, el(los) autor(es) utilizó(aron) ChatGPT y Gemini con el fin de: Revisar gramática y ortografía. Después de utilizar esta(s) herramienta(s)/servicio(s), el(los) autor(es) revisó(aron) y editó(aron) el contenido según fue necesario y asume(n) total responsabilidad por el contenido de la publicación.

Referencias

- [1] G. Castellano, B. De Carolis, N. Macchiarulo, and G. Vessio (2020). Pepper4Museum: Towards a Human-like Museum Guide. En *Proceedings of AVI2CH 2020: Workshop on Advanced Visual Interfaces and Interactions in Cultural Heritage*.
- [2] S. Joshi and S. Šabanović (2019). Robots for inter-generational interactions: implications for nonfamilial community settings. En *Proc. of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 478-486.
- [3] A. K. Aravamuthan, D. Šćepanović, F. Poveda, K. S. Chathuranga, and M. Obaid (2022). Exploring the role of social robots in intergenerational gameplay. En *RESNA 2022 Conference*.
- [4] Demir, Öykü; Vatan, Ahmet. Robotisation in travel and tourism: Tourist guides' perspectives on robot guides. *Tourism & Management Studies*, 2024, vol. 20, no 2, p. 13-23.
- [5] Using a Social Robot to Evaluate Facial Expressions in the Wild. *Sensors* (Basel, Switzerland), 20. <https://doi.org/10.3390/s20236716>.
- [6] Gaya Morey, F. X., Manresa-Yee, C., & Buades Rubio, J. M. (2024, June). An AI-Powered Computer Vision Module for Social Interactive Agents. In *Proceedings of the XXIV International Conference on Human Computer Interaction* (pp. 1-5).
- [7] Val-Calvo, M., Álvarez-Sánchez, J., Ferrández-Vicente, J., & Fernández, E. (2020). Affective Robot Story-Telling Human-Robot Interaction: Exploratory Real-Time Emotion Estimation Analysis Using Facial Expressions and Physiological Signals. *IEEE Access*, 8, 134051-134066. <https://doi.org/10.1109/ACCESS.2020.3007109>.

SanbotPPT: Presentador configurable para un robot social^{*}

SanbotPPT: Configurable presenter for a social robot

Loreto Matinero^{1,†}, Sandra Baldassarri^{1,2,*} y Eva Cerezo^{1,2,†}

¹ Departamento de Informática e Ingeniería de Sistemas, Universidad de Zaragoza, Zaragoza, España

² Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, Zaragoza, España

Resumen

En los últimos años, los robots sociales han mejorado considerablemente su capacidad de comunicación e interacción, por lo que se han ido introduciendo en gran diversidad de ámbitos, entre los que destaca su uso como presentadores, ya sea como guías en visitas, o como presentadores de trabajos científicos o educativos. En este artículo se presenta el desarrollo de una aplicación llamada SanbotPPT diseñada para permitir a usuarios sin conocimientos técnicos generar y reproducir presentaciones en el robot social Sanbot Elf, incorporando interacciones naturales para que el robot pueda simular el comportamiento no verbal de un presentador humano a través de un módulo reactivo. SanbotPPT permite crear presentaciones a través de una interfaz intuitiva que facilita la inclusión del contenido (texto, imágenes, vídeos,...) así como las diferentes funcionalidades del robot (voz, movimiento, expresiones faciales,...). Además, el módulo reactivo permite que el robot realice movimientos naturales aleatorios, detecte ruido para solicitar silencio, localizase el sonido para dirigir la atención, detecte personas para dar la bienvenida o reaccionar a su presencia, y realice reconocimiento facial para saludar o responder a expresiones. Se ha realizado una evaluación inicial utilizando la metodología de "pensar en voz alta", cuestionarios SUS y entrevistas para comprobar la facilidad de uso de la aplicación SanbotPPT. Los resultados han demostrado una muy buena experiencia por parte de los usuarios para la configuración y reproducción de presentaciones, aunque es necesario realizar una evaluación más exhaustiva considerando todas las acciones del comportamiento reactivo del robot.

Abstract

In recent years, social robots have considerably improved their communication and interaction capabilities, being introduced in a wide variety of fields, including their use as presenters, either as tour guides or as lecturers of scientific papers or acting as teachers in educational environments. This paper presents the development of an application called SanbotPPT designed to allow non-technical users to generate and reproduce presentations on the social robot Sanbot Elf, incorporating natural interactions so that the robot can simulate the non-verbal behavior of a human presenter through a reactive module. SanbotPPT allows the creation of presentations through an intuitive interface that facilitates the inclusion of content (text, images, videos,...) as well as the different functionalities of the robot (voice, movement, facial expressions,...). In addition, the reactive module allows the robot to perform random natural movements, detect noise to request silence, localize sound to direct attention, detect people to welcome or react to their presence, and perform facial recognition to greet or respond to expressions. An initial evaluation has been conducted using think-aloud methodology, SUS questionnaires and interviews to test the usability of the SanbotPPT application. The results have shown a very good user experience for the generation and play of presentations, although a more comprehensive evaluation considering all actions of the robot's reactive behavior is needed.

Palabras clave

Robot social, robot presentador, comportamiento natural, generador de presentaciones

^{*}Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

^{1*} Corresponding author.

[†] These authors contributed equally.

✉ lmatinero@unizar.es (L. Matinero); sandra@unizar.es (S. Baldassarri); ecerezo@unizar.es (E. Cerezo)

 0009-0004-8101-1023 (L. Matinero); 0000-0002-9315-6391 (S. Baldassarri); 0000-0003-4424-0770 (E. Cerezo)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Introducción

Los robots sociales han ido cobrando real importancia en los últimos años debido a su capacidad para interactuar y comunicarse con las personas [1, 2], incorporándose en multitud de áreas de aplicación, que van desde el entretenimiento, como acompañantes en juegos o guías en visitas y museos [3, 4], hasta en educación [5] y salud [6], como asistentes para menores [7], personas con discapacidad o personas mayores [8], o incluso en servicios de atención directa a usuarios [9].

Estos robots representan una categoría particular de robots diseñados para comunicarse e interactuar con los seres humanos de manera socialmente aceptable y natural, involucrando comportamientos sociales como escuchar, hablar y expresar emociones [10, 11]. En la mayor parte de los casos, estos robots disponen de un cuerpo físico, con extremidades (imitando brazos o piernas humanos) y una pantalla que representa la cabeza o cara del robot. Esta apariencia antropomórfica y la capacidad de actuar físicamente es de gran importancia para la percepción de los seres humanos [1, 10], pero lo más importante es su capacidad para mantener relaciones interactivas, comunicativas y en algunos casos incluso hasta afectivas [12]. Estas capacidades de comunicación han dado lugar a que este tipo de robots se hayan utilizado en ocasiones para reemplazar a los seres humanos en tareas repetitivas, como suelen ser mostrar los mismos contenidos a diferentes grupos de visitantes o estudiantes. En estos casos los robots actúan como presentadores, ya sea de trabajos, lugares o entornos [13], o incluso se utilizan como profesores virtuales [14].

En los últimos años se han llevado a cabo diversas investigaciones en las que se utilizan robots directamente como presentadores, en algunos casos con dos robots [3], o incluso como entrenadores para ayudar a un orador humano a mejorar sus presentaciones [15, 16], como copresentadores para reducir la ansiedad de hablar en público [17], o como miembros de equipos de trabajo [18].

En algunos trabajos se ha explorado la percepción de la audiencia al sustituir a un conferenciante que realiza una presentación estilo PowerPoint por un robot. Esta aproximación se ha efectuado en exhibiciones donde era necesario repetir constantemente la misma presentación para diferentes grupos de visitantes. La evaluación se realizó con 78 participantes llegando a la conclusión de que las presentaciones con robots son efectivas en aceptación y comprensión, siempre que se realicen los comportamientos no verbales correctos [19]. En otros casos se ha analizado una conferencia dada por una persona, para luego reproducirla en el robot Sota reemplazando aquellos comportamientos no verbales insuficientes e inadecuados por otros más coherentes con la finalidad y el contenido de la comunicación [20] analizando la percepción por parte de la audiencia [14]. De forma similar, en el estudio [19] graba presentaciones realizadas por una persona, para generar un escenario de presentación según los gestos y comportamientos no verbales detectados, diagnosticando si el comportamiento es adecuado y modificándolo si fuera necesario.

En el estudio realizado por [21] utilizan un robot social Pepper para que realice una presentación imitando el comportamiento de un orador humano experto (gestos, tiempos, movimientos corporales, etc) y analizan luego la diferencia en la percepción de la audiencia ante el robot y el ser humano llegando a la conclusión que el robot atrae más la atención, pero siempre que los oyentes estén interesados en el tema que presenta. Estos autores consideran que, en algunos casos, tratar de imitar de forma muy realista el comportamiento humano por parte del robot no es suficiente, o incluso puede llegar a ser contraproducente [22]. Por ello, en su siguiente trabajo, en lugar de utilizar un robot antropomórfico como Pepper, utilizan un robot de juguete no antropomórfico, el robot Anki Vector, para realizar la presentación. Al igual que en el caso anterior, extraen información sobre la postura y la expresión facial del vídeo de un orador humano, pero en

lugar de imitar fielmente al humano, reorientan esta información hacia el estilo del robot, obteniendo mejores resultados. También se utiliza un robot Pepper para enriquecer presentaciones científicas actuando como co-presentador junto a un científico e interactuando en el turno de preguntas [17]. De este modo se ha logrado reducir considerablemente la ansiedad de los oradores al hablar en público.

A pesar de los buenos resultados obtenidos en la mayor parte de estos trabajos en cuanto a aceptación por parte de la audiencia, hay que indicar que en todos ellos los gestos y comportamientos del robot estaban pre-grabados. Exceptuando en el último caso [17] en donde se ha utilizado la técnica de Wizard of Oz [23] para ejecutar las acciones pregrabadas del robot, en el resto de casos se ha realizado una programación ad-hoc del comportamiento del robot, resultando una tarea bastante ardua y que requiere de conocimientos técnicos.

Es necesario señalar que en todos los trabajos mencionados previamente los robots actúan como presentadores, pero no se ofrece ninguna herramienta de autor o aplicación que permita a los usuarios a preparar las presentaciones a mostrar, exceptuando los sistemas para generar presentaciones con scripts [19] o lenguaje de marcas (PML [24], MPML [25]) que permite controlar la presentación del robot, especificando el texto a decir por el robot y los elementos multimedia adecuados, así como los gestos que debe realizar el robot en cada momento. La gestión de estos lenguajes no es compleja, pero también requieren un mínimo de conocimientos técnicos.

Por otra parte, en todos los estudios se destaca la importancia que tiene el comportamiento similar al humano por parte del robot, pero en general todos los comportamientos suelen estar pre-grabados, y en ningún caso suelen gestionar ningún tipo de reacción en tiempo real por parte del robot durante la presentación. Aunque hay trabajos en los cuales robots sociales reaccionan en tiempo real de forma similar a los seres humanos durante la interacción con los usuarios [10, 11, 26, 27, 28], estos comportamientos no se han incorporado a las presentaciones realizadas por robots, si bien hay antecedentes en el ámbito de los personajes virtuales [29, 30, 31, 32].

La investigación sobre la que se centra este artículo, se basa en la utilización del robot social Sanbot Elf para su uso como presentador dotado con ciertos comportamientos naturales. El sistema permite que usuarios sin ningún tipo de conocimiento técnico puedan configurar presentaciones para ser reproducidas y ejecutadas en un robot. La aplicación de configuración de presentaciones, no solo permite montar una presentación mediante imágenes y textos, sino que además permite a los usuarios hacer uso de las funcionalidades internas del robot, pudiendo así los usuarios generar presentaciones más dinámicas y fáciles de adaptar a distintos contextos de presentación, como podrían ser visitas o charlas, entre otros. Además, como bien se menciona en los trabajos comentados, cabe destacar la importancia de que el robot actúe de forma natural imitando el comportamiento humano mediante gestos y comportamientos no verbales, por lo que se ha considerado imprescindible dotar al robot de interacciones naturales durante las presentaciones, y mejorar de esta forma la calidad de las presentaciones.

El resto del artículo se organiza del siguiente modo: en la Sección 2 se presenta una descripción general de la aplicación SanbotPPT; la Sección 3 expone el diseño y desarrollo de la aplicación, detallando cada uno de los módulos que la componen; la Sección 4 explica la evaluación inicial que se ha llevado a cabo y sus resultados; y finalmente la Sección 5 recoge las conclusiones y el trabajo a futuro.

2. Descripción general de la aplicación SanbotPPT

SanbotPPT es una aplicación que se ha diseñado y desarrollado para que usuarios sin conocimientos técnicos puedan diseñar, configurar, personalizar y reproducir presentaciones para un robot social. Desde la aplicación el usuario puede introducir textos, imágenes, vídeos y

movimientos, expresiones y colores del robot para armar una presentación. Además, para mejorar la calidad, dinamismo y naturalidad de las presentaciones, la aplicación SanbotPPT se ha enriquecido con interacción natural, de manera que el robot pueda imitar el comportamiento que podría tener un presentador humano mediante ciertas reacciones y gestos no verbales.

Dentro de la aplicación, los usuarios pueden crear una presentación asignando un título y una lista de acciones organizadas según bloques de acciones, que el robot ejecutará posteriormente. Mediante la herramienta SanbotPPT el usuario puede modificar cualquier aspecto de la presentación mediante la reordenación o eliminación de bloques de acciones y acciones. También se permite a los usuarios reproducir tanto la presentación, como los bloques de acciones y las acciones individualmente antes de guardarlas. La visualización de la presentación completa puede realizarse tanto desde la pantalla táctil del robot como proyectada en la pared trasera al robot si el usuario decide activar esta opción. Por último, la aplicación, mediante un módulo reactivo, permite configurar qué tipo de reacciones se aplicarán durante la reproducción de cada presentación. En el siguiente apartado, se presenta en detalle el diseño y la arquitectura de la aplicación SanbotPPT, así como el funcionamiento de cada uno de sus componentes.

La aplicación desarrollada en este trabajo ha sido diseñada para el robot social Sanbot Elf, que se muestra en la Figura 1. Este robot dispone de expresión facial y brazos, lo que le aportan ciertas características humanas. Mide 90,2 cm de altura, pesa 19 kg y cuenta con una base motorizada y un sistema de ruedas omnidireccionales que le permiten moverse con facilidad en diferentes entornos. Asimismo, dispone de una pantalla táctil de 10.1 pulgadas situada en el pecho, que sirve como una interfaz visual para la interacción directa con los usuarios. Además, está equipado con múltiples sensores, los cuales permiten al robot reconocer rostros, voces y contacto humano, además de detectar obstáculos y personas, ayudando así a la prevención de posibles colisiones o caídas del robot.



Figura 1: Robot Sanbot Elf.

Todas las funcionalidades que ofrece la aplicación se han implementado utilizando los módulos que proporciona el SDK del robot Sanbot Elf, QihanOpenSDK. Estos módulos son:

- Módulo de reconocimiento de voz, el cual permite la síntesis de voz y la detección de voz.

- Módulos de movimiento (movimiento de cabeza, brazos y ruedas) que permiten al robot mover sus distintas partes del cuerpo según un ángulo, una velocidad, o el parámetro que requiera la acción.
- Módulo de control de hardware, el cual controla el funcionamiento de los LEDs, además de recoger y tratar la información que proporcionan los distintos sensores de contacto e infrarrojos.
- Módulo de control de sistema, que permite obtener el ID del dispositivo, controlar el cambio de las expresiones faciales, obtener del nivel y estado de la batería, y monitorizar alertas de seguridad.
- Módulo de gestor de multimedia permite capturar y procesar los datos de la cámara HD del robot, tanto del flujo de vídeo como de audio.
- Módulo de control del proyector incluye funcionalidades de encendido y apagado del proyector (situado en la parte trasera de la cabeza del robot) además de otros ajustes sobre el mismo.

3. Diseño y desarrollo de la aplicación SanbotPPT

En la Figura 2 se presenta la arquitectura de la aplicación, compuesta por los siguientes módulos: interfaz de usuario, configurador de presentaciones, gestor de datos, ejecutor de acciones, reproductor de presentaciones y módulo reactivo. A continuación, se describe cada uno de ellos en detalle.

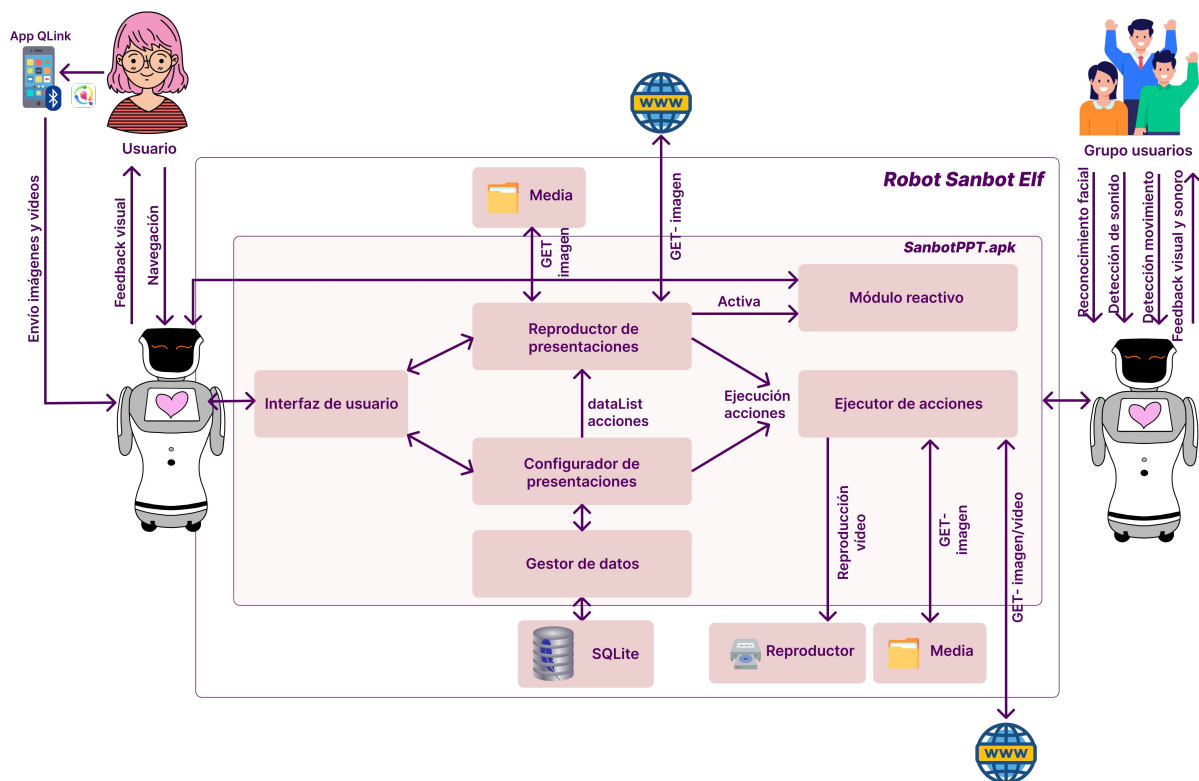


Figura 2: Arquitectura del sistema.

3.1. Interfaz de usuario

La interfaz de usuario cobra un papel importante dentro de la aplicación, ya que el proceso general de interacción que sigue el usuario se realiza utilizando la pantalla táctil del robot. La interfaz

permite al usuario navegar por la aplicación y acceder a las distintas funcionalidades. En la Figura 3 se muestra la pantalla principal de la aplicación, en la que aparece un listado de presentaciones ya generadas y configuradas, sobre las cuales el usuario puede realizar tres funcionalidades: reproducir, editar o eliminar una presentación. Además, mediante el botón disponible de la esquina inferior derecha, el usuario puede crear una nueva presentación.

Cuando el usuario clicca en el icono de reproducir, se le pregunta si desea proyectar la presentación, en caso de que acepte se activa el proyector, lo que permite visualizar la presentación no solo en la pantalla del robot, sino también proyectada en la pared trasera. Al crear o editar una presentación, el usuario accede a la pantalla de configuración que se observa en la Figura 4. En esta pantalla el usuario debe asignar un título a su presentación y puede crear, modificar, reproducir o eliminar uno o más bloques de acciones.



Figura 3: Listado de presentaciones (izquierda) y edición de una presentación (derecha).

Para cada bloque de acciones el usuario puede asignar un nombre al bloque y añadir y configurar las diferentes acciones que quiere que realice el robot en esa parte de la presentación (ver Figura 5). Estas acciones son las que aparecen listadas en la Figura 6: síntesis de voz, movimiento de brazos, movimiento de cabeza, movimiento de ruedas, encender LEDs, cambio de expresión facial, insertar imagen, insertar vídeo y pregunta de tipo verdadero o falso y será el módulo ejecutor de acciones el que realice su ejecución. Estas acciones se explicarán con más detalle en el apartado 3.4: Ejecutor de acciones.



Figura 4: Edición de un bloque de acciones (izquierda) y lista de acciones (derecha).

Finalmente, en la Figura 7 se presenta la pantalla de reproducción, a la que se accede cuando se desea reproducir una presentación completa o un bloque de acciones. Durante la reproducción el

usuario puede realizar tres acciones distintas: pausar, finalizar o activar subtítulos en la presentación.



Figura 5: Reproducción.

Este módulo de la interfaz de usuario recibe como entrada las interacciones táctiles del usuario a través de la pantalla del robot. Estas interacciones incluyen el uso de botones e iconos para ejecutar funciones como crear, editar o eliminar presentaciones, así como la introducción de textos para el nombramiento de presentaciones o bloques de acciones, como la selección del tipo de acción y los parámetros que se requieran para la configuración de las acciones. Además de las funciones que se pueden ejecutar durante la reproducción mediante el uso de botones (pausar, finalizar o activar subtítulos).

Como salidas el módulo muestra al usuario una respuesta visual inmediata, actualizando las pantallas correspondientes a la configuración y reproducción de las presentaciones. Asimismo, el módulo genera salidas en forma de instrucciones enviadas a otros componentes del sistema, como el configurador de presentaciones para guardar o modificar las configuraciones, o el módulo de reproducción para iniciar o detener la ejecución de una presentación.

3.2. Configurador de presentaciones

El configurador de presentaciones es el módulo que se encarga de gestionar la configuración de las presentaciones mediante la interacción con el resto de los elementos del sistema.

Este módulo recibe como entrada la información que le envía el módulo de interfaz de usuario. En el caso de creación o modificación de una presentación, este módulo recibe el nombre de esa presentación además de la información correspondiente a los bloques de acciones que contiene. Respecto a los bloques de acciones, el módulo recibe el nombre correspondiente del bloque y la información detallada de la configuración de cada una de las acciones que forman dicho bloque, consistente en el tipo de acción y los parámetros requeridos para su ejecución. Además, el módulo también recibe información procedente del gestor de datos, el cual le envía la información correspondiente a las presentaciones configuradas y almacenadas en la base de datos.

En cuanto a las salidas, este módulo envía al gestor de datos la información que recibe desde la interfaz de usuario, lo que permite actualizar el contenido de la presentación. A su vez, utiliza la información obtenida del gestor de datos para enviar a la interfaz de usuario la información necesaria para actualizar la visualización de la presentación. Asimismo, el configurador de presentaciones es el que se encarga de enviar al módulo reproductor de presentaciones la lista de acciones que se deben ejecutar durante la reproducción, y también transmite información sobre la

configuración de las acciones al ejecutor de acciones, para su previsualización directa sin necesidad de activar el módulo de reproducción.

3.3. Gestor de datos

El gestor de datos es el módulo encargado de establecer la relación entre la aplicación y la base de datos de almacenamiento local (SQLite). Este se encarga de gestionar las llamadas a la base de datos, y crea una capa superior la cual permite acceder a los datos disponibles y modificarlos o eliminarlos según desee el usuario.

Este módulo recibe como entrada los datos enviados por el configurador de presentaciones. Estos datos incluyen información sobre las nuevas presentaciones creadas, modificaciones realizadas en presentaciones ya existentes, bloques de acciones que deben almacenarse o modificarse, y la configuración de las acciones de cada bloque.

En el caso de las salidas del módulo, el gestor de datos proporciona al configurador de presentaciones los datos que este solicita desde la base de datos.

3.4. Ejecutor de acciones

Este módulo contiene la lógica para ejecutar cada una de las acciones que se ofrecen a los usuarios. Estas acciones son: síntesis de voz, cambios de expresión facial, encender LEDs, movimientos de brazos, movimientos de cabeza, movimientos de ruedas, insertar imagen, insertar vídeo y preguntar verdadero o falso.

Al ejecutar la acción de síntesis de voz, el robot reproduce la cadena de texto introducida por el usuario. Acciones como cambio de expresión facial y encender LEDs, se ejecutan indicando parámetros específicos como la expresión a mostrar o el color y parte del cuerpo a iluminar.

Para reducir la complejidad en la configuración de las acciones de movimientos de brazos, cabeza y ruedas, las cuales requieren completar parámetros como ángulos o velocidades, se ofrecen al usuario movimientos comunes ya pre-creados, como “Asentir”, “Negar”, o “Mirar a la derecha”.

Respecto a las acciones de insertar imagen, insertar vídeo y preguntar verdadero o falso, estas independientes del SDK del robot, permiten añadir recursos visuales a las presentaciones. Estos recursos si no se encuentran dentro del almacenamiento interno del robot, el usuario puede en todo momento transmitirlos mediante *Bluetooth* o gracias a la aplicación *Qlink* (aplicación para *smartphones* que permite comunicarse con el robot Sanbot Elf). Tanto para la inserción de vídeos como de imágenes, los usuarios pueden insertar estos recursos desde el almacenamiento interno o mediante una URL. En el caso de los recursos de almacenamiento interno, las imágenes se cargan desde la carpeta Media, y para los vídeos se redirige al reproductor interno del robot. Cuando estos recursos son externos el acceso es mediante una petición GET a la URL correspondiente. Acerca de la acción preguntar verdadero o falso, el robot dicta la pregunta configurada y espera una respuesta del usuario, evaluando si es correcta o no.

3.5. Reproductor de presentaciones

El reproductor de presentaciones se encarga de ejecutar de forma secuencial las acciones configuradas para una presentación, permitiendo así que el robot actúe como presentador. Este módulo incorpora varias funcionalidades que permiten al usuario controlar el flujo de la presentación a través de la interfaz. Estas funcionalidades son: finalizar, pausar, reanudar y activar subtítulos. Además de ejecutar las acciones configuradas, el reproductor también se encarga de activar el módulo reactivo, añadiendo así comportamientos naturales al robot durante la reproducción de las presentaciones.

El reproductor recibe principalmente dos tipos de entradas, por un lado, la lista de acciones desde el configurador de presentaciones ya sea al reproducir una presentación completa o un bloque de acciones. Y, por otro lado, el módulo recibe entradas directas desde la interfaz cuando los usuarios pulsan los botones de pausar, reanudar, finalizar o activar subtítulos.

Con respecto a las salidas, el módulo envía hacia el ejecutor de acciones los parámetros previamente definidos, para su correcta ejecución. Desde el reproductor también se envían a la interfaz de usuario las respuestas a acciones de pausar, reanudar, finalizar y activación o desactivación de subtítulos, y los cambios de imágenes de fondo de pantalla según el recurso que tenga asignado cada acción (la llamada a estos recursos se puede realizar desde almacenamiento local Media o desde Internet). Además, este módulo de reproducción es el que realiza la llamada al módulo reactivo durante el proceso de reproducción. Sin embargo, si el proyector está activo durante la reproducción de las presentaciones, el robot no tendrá habilitada la posibilidad de realizar movimientos de cabeza y torso, aunque se llamen desde el módulo reactivo o hayan sido configurados por el usuario.

3.6. Módulo reactivo

El módulo reactivo pretende mejorar la figura de presentador incluyendo reacciones naturales que simulen la actitud humana de un presentador. Durante la ejecución de una presentación oral, la realización de comportamientos no verbales y la retroalimentación adecuada [33] juegan un papel muy importante.

Este módulo reactivo, una vez que se activa por el módulo de reproducción de presentaciones, recibe como entrada diversos estímulos del entorno que le rodea: la percepción y movimientos de los usuarios que están visualizando la presentación, y la voz y sonido capturados del entorno. Las respuestas a estos estímulos generan las salidas de este módulo reactivo, que consisten principalmente en que el robot ejecute acciones de movimiento o de habla acordes con la situación percibida.

Para esto, se han incorporado cinco aspectos (ver Figura 8): la ejecución de movimientos naturales, la detección de ruido, la localización de sonido, la detección de personas y el reconocimiento facial. Estos pueden ser activados o no desde la configuración de la presentación, y se explican en detalle en los siguientes subapartados. En algunos casos como la localización de sonido, detección de personas y reconocimiento facial, se generan respuestas muy concretas que podrían interferir en el flujo de una presentación con varias personas en el público, por lo que se recomienda su uso en presentaciones preparadas para la visualización de una a cinco personas.

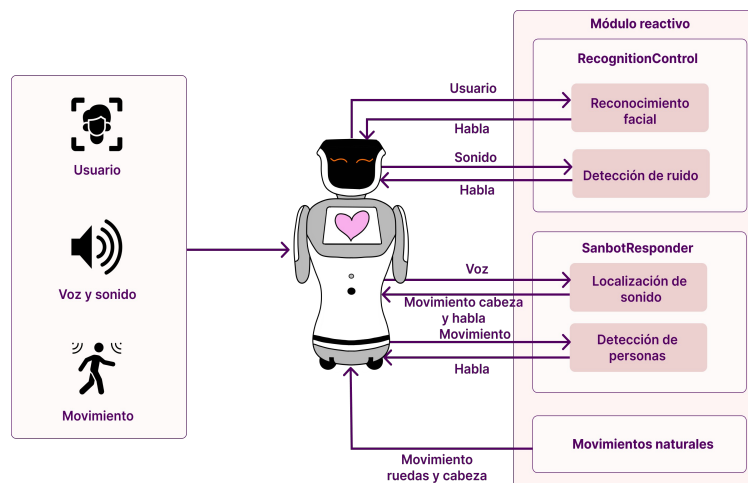


Figura 6: Módulo reactivo. Estímulos detectados (izquierda) y reacciones del robot (derecha).

3.6.1. Movimientos naturales

En muchas ocasiones, los presentadores no son conscientes de su propio lenguaje corporal, lo que los lleva a realizar gestos no intencionados pero que proporcionan naturalidad, evitando que el comportamiento sea robótico, rígido y totalmente determinista. Este tipo de movimientos forman parte del comportamiento humano, por lo que es interesante incluirlos en las reacciones del robot para mejorar el dinamismo de la presentación y humanizar su actuación durante la presentación.

Teniendo esto en cuenta y contemplando la movilidad de la que dispone Sanbot Elf (movimiento de brazos en el eje vertical 270°, movimiento de cabeza eje vertical y horizontal 180°, movimiento de ruedas avanzar/retroceder y girar 360°) se han programado los siguientes movimientos: giro de cabeza, giro de torso y desplazamiento. Estos movimientos son los únicos que no dependen de ningún estímulo externo, ya que están programados para que vayan surgiendo de forma aleatoria durante la presentación. Además, se ha tenido en cuenta que se realicen a una velocidad y rango de movimiento reducidos para evitar de esta forma la brusquedad de la acción.

3.6.2. Detección de ruido

Teniendo en cuenta los tipos de comportamientos que podrían ser interesantes para mejorar las respuestas reactivas que se ofrecen a los usuarios en el contexto de llevar a cabo una presentación de forma natural, se ha programado que el robot reaccione al ruido ambiente.

El robot Sanbot Elf cuenta con un módulo de gestor de multimedia, el cual permite capturar y procesar los datos de la cámara HD del robot, tanto del flujo de vídeo como de audio. Para incorporar este comportamiento reactivo al módulo, se ha decidido utilizar el flujo de audio para calcular los decibelios presentes en el entorno, y producir una respuesta acorde a estos. Este flujo de audio se captura mediante un array de bytes, el cual se procesa para obtener los decibelios (dB) que representan la intensidad del sonido en el ambiente. Una vez calculados estos decibelios se realizaron diversas pruebas de sonido para confirmar el rango de sonido de reacción adecuado, y producir una respuesta acorde al mismo.

La reacción a estos decibelios se ha decidido fijar en mayor a 75 dB, y siempre y cuando la duración de este valor sea por lo menos durante 1000ms, para así poder descartar otro tipo de sonidos como chirridos o golpes. Cuando estas dos condiciones se cumplen, el robot reacciona quejándose del volumen y pidiendo silencio.

3.6.3. Localización de sonido

En relación con la localización del sonido, el SDK del robot incorpora un módulo de control de hardware el cual permite localizar la fuente de sonido detectada en el entorno. Concretamente, el robot captura un sonido cualquiera, y el módulo de control de hardware calcula el ángulo respecto al robot donde se ha detectado el sonido.

Tras esta detección, se produce una reacción programada desde el módulo reactivo, que en este caso consiste en que el robot realice un movimiento de cabeza hacia la localización de la fuente de sonido y reproduzca una frase aleatoria según la posición de la detección. En el caso de que la detección se realice por la parte frontal, algunas de las frases propuestas son: “¿Qué fue eso?”, “Creo que oí algo por aquí” o “¡Escuche algo por esta zona!”. Si la detección se realiza por la parte trasera del robot, las frases son las siguientes: “¡No me asustes así!”, “¡Oh! Hay algo detrás” o “Parece que escuche algo por detrás”.

3.6.4. Detección de personas

Respecto a la detección de personas, el robot cuenta con dos sensores infrarrojos, uno en la parte frontal de su cuerpo y otro en la parte posterior, que permiten detectar fácilmente si una persona está pasando por delante o por detrás del robot.

La información capturada por estos sensores se ha gestionado también mediante el módulo de control de hardware del SDK del robot. En este caso, el movimiento es captado a través de los sensores infrarrojos, y desde el módulo reactivo se ha programado una reacción. En el caso de que el robot detecte el movimiento de una persona por delante, este reproducirá una frase aleatoria de las siguientes: “¡Hola! Bienvenido”, “¡Hola! Toma asiento y disfruta de la presentación” o “Adelante, ponte cómodo”. Si la detección de la persona se realiza por la parte trasera, el robot girará su cuerpo 180° hacia la derecha, reproducirá una frase entre las siguientes: “¡Me asustaste! No te había visto”, “¡Ah! Debería tener ojos en la espalda” o “Hola, parece que me estabas espiando”, y finalmente volverá a girar 180° hacia la izquierda para volver a su posición inicial.

La detección de personas no solo se puede realizar mediante los sensores infrarrojos, sino también mediante la cámara HD que incorpora el robot en la parte frontal de su cabeza. En este caso, el módulo gestor de multimedia del SDK procesará la entrada de vídeo capturada por la cámara HD. Esta entrada de vídeo se captura y se envía a un servidor local, donde se encuentra desplegado un módulo de visión basado en inteligencia artificial desarrollado para agentes sociales interactivos [34]. Este módulo procesa las imágenes recibidas y envía una respuesta al módulo reactivo, permitiendo así adaptar el comportamiento del robot. En este caso, el módulo de visión cuando detecta a una persona devuelve al módulo reactivo las coordenadas delimitadoras de la persona detectada. Desde el módulo reactivo cuando estas coordenadas son recibidas se ordena al robot reaccionar reproduciendo la frase “Hola ¿Hay alguien ahí?”.

Estas respuestas que ofrece el robot, se realizan de forma controlada y no continua. Para evitar que estas intervenciones del robot resulten molestas a los usuarios, cada vez que el movimiento es detectado, se verifica si han transcurrido al menos cinco minutos desde la última intervención, y solo en ese caso el robot emite una nueva respuesta.

3.6.5. Reconocimiento facial

El reconocimiento facial permite al robot realizar detecciones faciales (localización de una cara en la imagen capturada), reconocimiento de expresiones faciales, reconocimiento de edad y género, y reconocimiento facial (reconocimiento de una persona concreta). Se ha utilizado el módulo de gestor de multimedia, que procesará la entrada de vídeo capturada por la cámara HD del robot. Al igual que en el caso anterior de detección de personas, el reconocimiento facial utiliza el módulo de visión desarrollado para agentes sociales interactivos [34]. Para estas funciones el módulo reactivo recibe del módulo de visión información concreta sobre las coordenadas del cuadro delimitador del rostro (detección facial), una de las seis expresiones de Ekman [35] (reconocimiento de expresiones faciales), la edad y el género *male* o *female* (reconocimiento de edad y género), y el nombre de la persona reconocida según un conjunto predefinido de usuarios, lo que requiere que exista al menos una imagen del sujeto a reconocer dentro del conjunto (reconocimiento facial).

En todos los casos el robot reacciona mediante la función de habla, incluyendo en sus frases la información detectada. Para la detección facial el robot únicamente saluda al usuario, y en el reconocimiento de expresiones faciales el robot dice una frase acorde a la emoción detectada, por ejemplo, si detecta un rostro de enfado el robot dice la frase: “¡Eh! No te enfades conmigo, yo no tengo la culpa”.

4. Evaluación inicial y resultados

Para comprobar la usabilidad y la experiencia de usuario al utilizar el configurador y reproductor de presentaciones se ha realizado una evaluación inicial, aplicando tres métodos distintos: pensar en voz alta o *thinking aloud* [36], cuestionarios de usabilidad SUS (*System Usability Scale*) [37] y entrevistas.

Esta evaluación se ha llevado a cabo con tres usuarios de 22, 23 y 25 años (1 mujer y 2 hombres), todos ellos estudiantes de ingeniería. Previamente a la sesión, se les había pedido a los usuarios que pensaran qué presentación querían configurar utilizando la aplicación SanbotPPT y que trajeran los recursos audiovisuales necesarios.

La sesión de evaluación con cada uno de los usuarios tuvo una duración de aproximadamente 2 horas, se llevó a cabo en el laboratorio del grupo de investigación, siguiendo la siguiente estructura:

- Presentación de la aplicación. Se explica al usuario en qué consiste la aplicación y las nociones básicas de funcionamiento, y se le proporciona un manual de uso por si le surgen dudas durante la realización de las tareas propuestas.
- Asignación y ejecución de tareas. A cada usuario se le presenta una lista de tareas a realizar enfocadas en la creación y configuración de presentaciones. La lista de tareas incorpora todas las opciones disponibles para generar presentaciones, y los usuarios deben completarlas con el material traído.
- Durante la realización de las tareas se ha aplicado el método *thinking aloud* que consiste en que los participantes verbalicen libremente cualquier pensamiento o apreciación acerca de la interacción con la aplicación. Este método permite obtener datos cualitativos acerca de la percepción de la facilidad de uso de la aplicación desarrollada.
- Completar el test SUS. Tras completar la lista de tareas mencionadas previamente, se pidió a los usuarios completar el test SUS, que es un cuestionario estandarizado que permite medir la percepción de la usabilidad de un sistema, obteniendo así una medida cuantitativa de la usabilidad percibida tras utilizar la aplicación.
- Realización de entrevista. Finalmente, se realiza una entrevista semiestructurada con los usuarios, permitiendo profundizar más en algunos aspectos detectados durante la realización de las tareas, y obtener así más información cualitativa acerca de la experiencia de cada usuario.

Tras la realización de las sesiones se obtuvieron los resultados que se exponen a continuación.

En relación a la fase de realización de las tareas siguiendo el método de *thinking aloud*, los usuarios comentaron varios aspectos positivos, describiendo la aplicación como intuitiva y fácil de usar, y destacando la cantidad de opciones de configuración que la misma ofrece y enriquecen el sistema. También salieron a la luz algunos aspectos negativos, como la falta de retroalimentación en la adición de una acción, la dificultad de introducir textos en la acción de síntesis de voz debido al incómodo manejo del teclado, la dificultad de introducir URLs, y la incomodidad de tener que añadir un soporte visual a cada acción en vez de a un bloque de acciones completo.

Los resultados obtenidos de la realización del cuestionario SUS por parte de los usuarios se recogen en la Tabla 1. De media han valorado la aplicación con una puntuación de 91,7 sobre una puntuación total de 100, lo cual indica que la satisfacción de los usuarios con la aplicación es muy alta y que valoran muy bien la usabilidad del sistema.

Tabla 1

Resultados cuestionario SUS.

Usuario	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	SUS
Usuario1	5	1	5	1	5	1	4	1	5	1	97.0
Usuario2	4	1	5	4	4	1	5	2	5	1	85.0
Usuario3	5	1	5	1	5	2	4	1	4	1	92.5
Total:											91.7

Respecto a la entrevista, los usuarios destacaron que la aplicación era intuitiva, fácil de usar y visualmente atractiva. Algunos mencionaron que les gustó la variedad de acciones configurables y la inclusión de subtítulos la hacían una aplicación mucho más accesible, sin embargo, sugirieron mejorar la retroalimentación al añadir nuevas acciones y la dificultad al introducir textos a través del teclado.

Tras la realización de esta primera evaluación se decidieron llevar a cabo algunas mejoras, como la incorporación de retroalimentación al añadir nuevas acciones, y mejorar la introducción de textos mediante una opción de dictado por voz, facilitando así la configuración de acciones de síntesis de voz. Sin embargo, aún existen algunos aspectos por perfeccionar, como facilitar la introducción de URLs en la aplicación para la adición de recursos visuales a las presentaciones, y cambiar el modo de asignación de recursos, permitiendo asociarlos a bloques de acciones en lugar de acciones individuales, para evitar de esta forma la agregación múltiple de un mismo soporte visual en una presentación.

5. Conclusiones

En este trabajo se ha presentado el diseño y desarrollo de SanbotPPT, una novedosa aplicación configurable para el robot social Sanbot Elf, que permite a los usuarios la configuración de presentaciones y la reproducción de estas mismas, en las que el robot actúa como presentador y reacciona de forma natural a ciertos estímulos del entorno. Al configurar las presentaciones se pueden incluir textos a visualizar, imágenes y vídeos a mostrar a través de la pantalla frontal del robot o proyección, así como diferentes acciones a llevar a cabo por parte del robot (movimientos, encendido de LEDs, texto a decir, etc.) y posibles reacciones a diferentes estímulos durante la presentación (detección de ruido, sonido, caras y personas).

El módulo de configuración y reproducción de presentaciones ha sido evaluado de forma preliminar con 3 usuarios, lo que ha permitido verificar su usabilidad, así como realizar varias de las mejoras sugeridas. Sin embargo, queda pendiente realizar una evaluación más exhaustiva de la aplicación, con una mayor cantidad de usuarios, de diferentes edades y perfiles, preferentemente personas sin conocimientos técnicos de ningún tipo, además de emplear otros cuestionarios para evaluar la usabilidad de robots como los que se recomiendan en el artículo [38].

Por otra parte, también queda pendiente realizar una evaluación sistemática del módulo reactivo, que contemple la aceptación por parte de los usuarios del comportamiento reactivo del robot, considerando todas las reacciones posibles, y prestando especial atención a aquellas que pueden generar demasiadas interrupciones en la presentación si se realiza ante audiencias con gran número de personas.

Como trabajo a futuro, también se pretende perfeccionar los problemas detectados tras esta evaluación inicial, entre los cuales se incluye mejorar la introducción de URLs en la aplicación y

cambiar el modo de asignación de recursos. Otras líneas de trabajo que se prevé abordar en el futuro consisten en el desarrollo de una base de datos externa que permita el almacenamiento de los usuarios y sus presentaciones, permitiendo así el acceso únicamente a sus presentaciones, y además pudiendo acceder a ellas desde otros robots o dispositivos; así como incorporar ciertas mejoras de accesibilidad, como la aplicación del uso de los subtítulos en todas las acciones de habla que ejecuta el robot.

Agradecimientos

Este trabajo está parcialmente financiado por el Ministerio de Ciencia e Innovación y Universidades (MCIU), la Agencia Estatal de Investigación (AEI) y la UE (FEDER) a través del contrato PID2022-136779OB-C31, por MCIN/ AEI/10.13039/501100011033/ y por la Unión Europea NextGenerationEU/PRTR a través del proyecto TED2021-130374B-C22 y por el Gobierno de Aragón (Grupo T60_23R).

Declaración sobre la IA Generativa

Durante la preparación de este trabajo, los autores utilizaron GPT-4 para: Revisar la gramática y la ortografía. Después de utilizar estas herramientas/servicios, el autor revisó y editó el contenido según fue necesario y asume plena responsabilidad por el contenido de la publicación.

Referencias

- [1] C. Breazeal, K. Dautenhahn, T. Kanda, *Social Robotics*, Springer International Publishing, Cham, 2016, pp. 1935–1972. doi:10.1007/978-3-319-32552-1_72.
- [2] A. Henschel, G. Laban, E. S. Cross, What makes a robot social? a review of social robots from science fiction to a home or hospital near you, *Current Robotics Reports* 2 (2021) 9–19. doi:10.1007/s43154-020-00035-0.
- [3] A.-M. Velentza, D. Heinke, J. Wyatt, Museum robot guides or conventional audio guides? an experimental study, *Advanced Robotics* 34 (2020) 1571–1580. doi:10.1080/01691864.2020.1854113.
- [4] N. Gasteiger, M. Hellou, H. S. Ahn, Deploying social robots in museum settings: A quasi-systematic review exploring purpose and acceptability, *International Journal of Advanced Robotic Systems* 18 (2021) 17298814211066740. doi:10.1177/17298814211066740.
- [5] G. Barros, B. Motta, V. Teixeira, A. Gravatá, S. Silva Júnior, L. de Sá, M. Amaral, L. Almeida, Learning interactions: Robotics supporting the classroom, in: C. Stephanidis, M. Antona, S. Ntoa (Eds.), *HCI International 2021 - Posters*, Springer International Publishing, Cham, 2021, pp. 3–10.
- [6] L. Ragno, A. Borboni, F. Vannetti, C. Amici, N. Cusano, Application of social robots in healthcare: Review on characteristics, requirements, technical solutions, *Sensors* 23 (2023). doi:10.3390/s23156820
- [7] A. Triantafyllidis, A. Alexiadis, K. Votis, D. Tzovaras, Social robot interventions for child healthcare: A systematic review of the literature, *Computer Methods and Programs in Biomedicine Up-date* 3 (2023) 100108. doi: 10.1016/j.cmpbup.2023.100108.

- [8] R. A. Søråa, G. Tøndel, M. W. Kharas, J. A. Serrano, What do older adults want from social robots? a qualitative research approach to human-robot interaction (hri) studies, *International journal of social robotics* 15 (2023) 411–424. doi:10.1007/s12369- 022- 00914- w.
- [9] M. U. Khan, Z. E. and, A systematic review of social robots in shopping environments, *International Journal of Human-Computer Interaction* 0 (2024) 1–22. doi:10.1080/10447318.2024.2426740.
- [10] F. Melo, P. Moreno, Socially reactive navigation models for mobile robots, in: *2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2022, pp. 91–97. doi:10.1109/ICARSC55462.2022.9784789.
- [11] L. J. Wood, A. Zaraki, B. Robins, K. Dautenhahn, Developing kaspar: a humanoid robot for children with autism, *International Journal of Social Robotics* 13 (2021) 491–508. doi:10.1007/s12369- 019- 00563- 6.
- [12] R. Stock-Homburg, Survey of emotions in human–robot interactions: Perspectives from robotic psychology on 20 years of research, *International Journal of Social Robotics* 14 (2022) 389–411.
- [13] J. C. de Mello, G. P. N. Secci, P. H. Ribeiro, Robotics and ai in museums – the future of the present, *Robotic Systems and Applications* 4 (2024) 44–58. doi:10.21595/ras.2024.24355.
- [14] T. Ishino, M. Goto, A. Kashiara, Robot lecture for enhancing presentation in lecture, *Research and Practice in Technology Enhanced Learning* 17 (2022) 1–22. doi:10.1186/s41039-021-00176-6.
- [15] H. Trinh, R. Asadi, D. Edge, T. Bickmore, Robocop: A robotic coach for oral presentations, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1 (2017). doi:10.1145/3090092.
- [16] M. Goto, T. Ishino, K. Inazawa, N. Matsumura, T. Nunobiki, A. Kashiara, Authoring robot presentation for promoting reflection on presentation scenario, in: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 660–661. doi:10.1109/HRI.2019.8673278.
- [17] Y. Terzioğlu, P. Murali, E. Kimani, T. Bickmore, Sharing the spotlight: Co-presenting with a humanoid robot, in: *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022, pp. 551–560. doi:10.1109/HRI53351.2022.9889400.
- [18] B. Abendschein, C. Edwards, A. Edwards, V. Rijhwani, J. Stahl, Human-robot teaming configurations: A study of interpersonal communication perceptions and affective learning in higher education, *Journal of Communication Pedagogy* 4 (2021) 123–132. URL: <https://search.informit.org/doi/10.3316/INFORMIT.105941407010443>.
- [19] M. Goto, M. Yokoyama, Y. Matsuura, Impression evaluation of presentation by a communication robot in an actual exhibition, in: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, HRI '20*, Association for Computing Machinery, New York, NY, USA, 2020, p. 218–220. doi:10.1145/3371382.3378264.
- [20] T. Ishino, M. Goto, A. Kashiara, A robot for reconstructing presentation behavior in lecture, in: *Proceedings of the 6th International Conference on Human-Agent Interaction, HAI '18*, Association for Computing Machinery, New York, NY, USA, 2018, p. 67–75. doi:10.1145/3284432.3284460.

- [21] M.-L. Bourguet, M. Xu, S. Zhang, J. Urakami, G. Venture, The impact of a social robot public speaker on audience attention, in: Proceedings of the 8th International Conference on Human-Agent Interaction, HAI '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 60–68. doi:10.1145/3406499.3415073.
- [22] G. Venture, B. Muraccioli, M.-L. Bourguet, J. Urakami, Can robots be good public speakers?, in: Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3490149.3505576.
- [23] L. D. Riek, Wizard of oz studies in hri: a systematic review and new reporting guidelines, *J. Hum.-Robot Interact.* 1 (2012) 119–136. doi:10.5898/JHRI.1.1.Riek.
- [24] P. Mohan, M. Hosein, A system for giving presentations with the nao robot, in: 2022 IEEE Games, Entertainment, Media Conference (GEM), 2022, pp. 1–6. doi:10.1109/GEM56474.2022.10017306.
- [25] Y. Nozawa, H. Dohi, H. Iba, M. Ishizuka, Humanoid robot presentation controlled by multimodal presentation markup language mpml, in: RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), 2004, pp. 153–158. doi:10.1109/ROMAN.2004.1374747.
- [26] A. Edwards, C. Edwards, D. Westerman, P. R. Spence, Initial expectations, interactions, and beyond with social robots, *Computers in Human Behavior* 90 (2019) 308–314. URL: <https://www.sciencedirect.com/science/article/pii/S0747563218304175>. doi:10.1016/j.chb.2018.08.042.
- [27] M. M. Reimann, F. A. Kunneman, C. Oertel, K. V. Hindriks, A survey on dialogue management in human-robot interaction, *ACM Transactions on Human-Robot Interaction* 13 (2024). URL: <https://doi.org/10.1145/3648605>. Doi:10.1145/3648605.
- [28] E. Ang, A. Bejleri, B. Tantisira, A. Van de Velde, Considerations for the future of social robots and human-robot interactions, 2024. URL: <https://www.oxjournal.org/the-future-of-social-robots-and-human-robot-interactions/>.
- [29] F. Serón, S. Baldassarri, E. Cerezo, Maxineppt: using 3d virtual characters for natural interaction, in: Proc. 2nd International Workshop on Ubiquitous Computing & Ambient Intelligence, Citeseer, 2006, pp. 241–250.
- [30] A. Nijholt, Towards the automatic generation of virtual presenter agents, in: Informing Science+IT Education Conference, InSITE 2006, Informing Science Institute, 2006, pp. 97–110.
- [31] T. Bickmore, E. Kimani, A. Shamekhi, P. Murali, D. Parmar, H. Trinh, Virtual agents as supporting media for scientific presentations, *Journal on Multimodal User Interfaces* 15 (2021) 131–146. doi:10.1007/s12193-020-00350-y.
- [32] L. Cui, J. Liu, Virtual human: A comprehensive survey on academic and applications, *IEEE Access* 11 (2023) 123830–123845. doi:10.1109/ACCESS.2023.3329573.
- [33] A. Axelsson, G. Skantze, Multimodal user feedback during adaptive robot-human presentations, *Frontiers in Computer Science Volume 3 - 2021* (2022). URL: <https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2021.741148>. doi:10.3389/fcomp.2021.741148.

-
- [34] F. X. Gaya Morey, C. Manresa-Yee, J. M. Buades Rubio, An ai-powered computer vision module for social interactive agents, in: Proceedings of the XXIV International Conference on Human Computer Interaction, Interacción '24, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3657242.3658601.
- [35] P. Ekman, An argument for basic emotions, *Cognition and Emotion* 6 (1992) 169–200. doi:10.1080/02699939208411068.
- [36] J. Nielsen, Thinking aloud: The 1 usability tool, 2012. URL: <https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>.
- [37] J. Brooke, et al., Sus-a quick and dirty usability scale, *Usability evaluation in industry* 189 (1996) 4–7.
- [38] K. Moulai, R. Moulai, K. Bahaadinbeigy, The most used questionnaires for evaluating the usability of robots and smart wearables: A scoping review, *DIGITAL HEALTH* 10 (2024) 20552076241237384. doi:10.1177/20552076241237384.

Modelado de audiodescripción prosódica basado en narrativa semántica

Alexander Rozo-Torres, Cristian Plazas and Christian Quintero

Grupo de Investigación Multimedia (GIM), Facultad de Ingeniería, Universidad Militar Nueva Granada, Carrera 11 No. 101-80, Bogotá, Colombia

Resumen

Los sistemas de audiodescripción actuales presentan limitaciones en la generación de contenidos precisos y expresivos. Este trabajo propone un esquema de modelado de audiodescripción prosódica basado en narrativa semántica para mejorar la accesibilidad audiovisual. El enfoque integra mecanismos de atención para refinar las descripciones y técnicas de modelado emocional. Se recolectaron y analizaron descripciones de 10 escenas seleccionadas, estableciendo una comparación entre la descripción original y la generada mediante narrativa semántica. Posteriormente, se modelaron características prosódicas (intensidad, tono, ritmo y duración de pausas) para lograr una descripción oral expresiva. Además, se analizaron 16 interpretaciones emocionales utilizando métricas acústicas extraídas con Parselmouth y OpenSMILE. Los resultados demuestran que los modelos de síntesis de voz emocional logran una mayor congruencia con perfiles prosódicos emocionales, especialmente en emociones de alta activación. El esquema propuesto mejora tanto la precisión descriptiva como la expresividad emocional, ofreciendo una experiencia auditiva más inmersiva para personas con discapacidad visual.

Keywords

audiodescripción, modelado prosódico, narrativa semántica, síntesis de voz emocional, accesibilidad audiovisual

1. Introducción

El acceso equitativo al contenido audiovisual sigue siendo un desafío crucial para las personas con discapacidad visual, ya que la oferta de descripciones precisas y adaptadas resulta aún escasa y, con frecuencia, poco sensible a las complejidades narrativas de los vídeos. En los últimos años, la audiodescripción automática basada en modelos generativos ha surgido como una solución prometedora: los avances conjuntos de la visión por computadora y el procesamiento de lenguaje natural permiten generar descripciones textuales que representan el contenido visual de manera cada vez más natural e inclusiva [1, 2]. Sin embargo, estos sistemas, diseñados inicialmente para imágenes estáticas, presentan limitaciones cuando se enfrentan a la riqueza temporal y semántica de los vídeos, así como a la necesidad de transmitir emociones a través de la voz sintetizada.

La AudioDescripción (AD), definida como una técnica de accesibilidad que interpreta los elementos visuales de una escena sin diálogos en descripciones verbales, permite a las personas con discapacidad visual comprender y disfrutar de contenidos audiovisuales [3]. En este contexto, la prosodia (entendida como un conjunto de características rítmicas y melódicas del habla que modulan cómo se dice algo, más allá del contenido literal de las palabras) desempeña un papel fundamental, ya que no solo mejora la inteligibilidad del mensaje, sino que también aporta intencionalidad, emoción y significado [4, 5]. Por ejemplo, la entonación puede enfatizar momentos de tensión o calma, mientras que el ritmo y las pausas pueden marcar transiciones narrativas o destacar detalles importantes, complementando así el contenido lingüístico y enriqueciendo la experiencia auditiva [6].

En el ámbito de la AD automática, la narrativa semántica se refiere a la capacidad de generar descripciones textuales que no solo enumeran objetos o acciones, sino que construyen un relato coherente, contextualizado y adaptado a la dinámica temporal de la escena. Este enfoque va más allá de la simple identificación de elementos visuales, integrando información sobre relaciones espaciales,

Interacción 2025: XXV Congreso Internacional de Interacción Persona-Ordenador, 3-5 de septiembre de 2025, Valladolid, España

✉ est.omar.rozo@unimilitar.edu.co (A. Rozo-Torres); est.cristian.plazas@unimilitar.edu.co (C. Plazas);

christian.quintero@unimilitar.edu.co (C. Quintero)

ORCID 0000-0001-9343-2264 (A. Rozo-Torres); 0009-0008-2633-9607 (C. Plazas); 0000-0001-5763-2373 (C. Quintero)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

secuencias de eventos y matices contextuales que enriquecen la comprensión global del contenido audiovisual. Mediante el uso de modelos avanzados de procesamiento de lenguaje natural y visión por computadora, la narrativa semántica permite que las descripciones sean detalladas, fluidas y alineadas con la intención comunicativa original del material audiovisual. Así, la AD no solo es informativa, sino también significativa y envolvente para el usuario, facilitando una experiencia más cercana a la percepción visual original.

Estudios recientes demuestran que la narrativa semántica es fundamental para superar la brecha entre la simple descripción y la experiencia audiovisual completa. Al aprovechar arquitecturas basadas en transformadores y redes neuronales convolucionales, es posible identificar y describir eventos sin segmentaciones manuales, logrando secuencias textuales detalladas y coherentes que integran contexto espacial y temporal [7, 8]. Además, la incorporación de fusión multimodal, que combina información visual, auditiva y textual, refuerza la coherencia de las descripciones y enriquece su contenido contextual [9]. Sin embargo, cuando estas descripciones se transforman en audio mediante modelos de Texto a voz (TTS, por sus siglas en inglés) convencionales, la limitada expresividad prosódica genera relatos monótonos, reduciendo la inmersión del oyente y la transmisión de matices emocionales [10, 11].

Los sistemas actuales de síntesis de voz presentan limitaciones significativas en la generación de una prosodia natural y expresiva, lo que impacta negativamente en la inmersión del oyente y en la percepción de calidad del contenido [10]. En el caso específico de la AD, esta falta de variabilidad prosódica puede derivar en una experiencia monótona, disminuyendo la transmisión de matices emocionales y afectando la comprensión y el disfrute del usuario [12, 13].

En la comunicación humana, la prosodia no solo mejora la inteligibilidad, sino que comunica intención y emoción [4, 5]. Para la AD, una prosodia expresiva puede enfatizar momentos de tensión o sosiego, señalar transiciones narrativas y destacar detalles críticos, enriqueciendo la experiencia de las personas con discapacidad visual [14]. No obstante, los modelos TTS tradicionales carecen de la variabilidad necesaria para este fin. Frente a esta limitación, los modelos de síntesis de voz emocional (ETTS, por sus siglas en inglés) han tratado de plasmar patrones prosódicos humanos a partir de audios de referencia, prompts textuales o representaciones latentes, logrando resultados más naturales y adaptativos [15, 16].

Desde una perspectiva teórica, modelos emocionales como el psicoevolutivo de Plutchik y el circumplejo de Russell han sido ampliamente utilizados para mapear la transmisión de emociones en el habla [17]. Estos modelos ofrecen marcos conceptuales sólidos para comprender cómo las emociones se estructuran y se expresan, proporcionando herramientas valiosas para el diseño de sistemas de ETTS [18]. Sin embargo, la implementación de estos modelos sigue siendo un desafío debido a la complejidad de adaptar la prosodia a distintos estados emocionales y contextos narrativos [19].

Por otro lado, estudios de neurociencia cognitiva indican que las personas con ceguera (especialmente adquirida) manifiestan una sensibilidad potenciada hacia los matices emocionales de la voz [14]. Investigaciones sobre autenticidad afectiva y procesamiento auditivo en entornos ruidosos corroboran que cualquier artificialidad prosódica mal calibrada se percibe con facilidad, afectando la inmersión [16, 20]. Por consiguiente, la síntesis destinada a AD debe equilibrar naturalidad acústica y claridad emocional, adaptándose a las expectativas cognitivas de esta población.

En este contexto, el presente trabajo propone un esquema unificado que integra, por un lado, mecanismos de atención para refinar la narrativa semántica y, por otro, modelado prosódico emocional para narrarla de forma expresiva. Este enfoque busca no solo favorecer descripciones detalladas y coherentes con el contenido visual, sino también ofrecer una experiencia auditiva emocional. Al combinar una narrativa semántica y características prosódicas, se busca ofrecer una experiencia auditiva más envolvente, ajustada a las necesidades de los usuarios y capaz de reforzar la inclusión digital. Las siguientes secciones describen el marco teórico que sustenta la propuesta, la metodología implementada, los experimentos realizados y los resultados obtenidos, para finalmente trazar líneas de investigación futura que consoliden una AD emocional y accesible.

2. Revisión de literatura

El campo de la audiodescripción automática (AD) ha experimentado un crecimiento sostenido gracias a los avances combinados de la visión por computadora y el procesamiento de lenguaje natural (PLN). Sin embargo, garantizar descripciones que sean simultáneamente ricas en contenido semántico y prosódicamente expresivas sigue siendo un reto multidisciplinar. A continuación, se revisan de forma integrada los ejes conceptuales que sustentan la presente investigación: narrativa semántica, mecanismos de atención y fusión multimodal, modelado prosódico-emocional, percepción auditiva en personas con discapacidad visual y automatización del control prosódico.

La narrativa semántica —concepto que en inglés se refiere a *Dense Video Captioning*— busca generar descripciones detalladas y coherentes de los múltiples eventos que conviven en un mismo vídeo. Shen et al. [3] demuestran que la combinación de características visuales y auditivas incrementa la riqueza contextual de las descripciones, mientras que Zhu et al. [9] presentan DenseLSTM, una arquitectura de redes recurrentes densamente conectadas que retiene contexto a largo plazo y mejora la precisión al describir secuencias de acciones. Por su parte, AutoAD-Zero [21] introduce una estrategia *zero-shot* que prescinde de entrenamiento específico: primero produce una descripción pormenorizada y, después, la sintetiza, facilitando la escalabilidad a dominios con pocos datos. Finalmente, la fusión multimodal planteada en MemCap [22] integra señales visuales, textuales y auditivas para obtener narrativas estilizadas y coherentes. Estas aportaciones consolidan un marco robusto para la generación descriptiva; sin embargo, ninguna aborda explícitamente la dimensión prosódica, lo que abre una oportunidad clara para mejorar la experiencia de los usuarios.

Paralelamente, los mecanismos de atención se han convertido en una pieza angular para filtrar dinámica y selectivamente la información más relevante dentro de un flujo audiovisual. Modelos que combinan CNN y LSTM con atención jerárquica han mostrado reducciones notables de ruido descriptivo y mejoras en coherencia temporal [23, 24]. Esta atención se complementa con estrategias de fusión multimodal que alinean representaciones visuales, auditivas y textuales en un espacio semántico común, reforzando la densidad informativa de la descripción y permitiendo transiciones narrativas más naturales [25]. De este modo, el resultado textual no solo es exacto en términos de contenido, sino también adaptado pragmáticamente a la situación comunicativa.

A pesar de que los modelos actuales han logrado avances importantes en términos de precisión semántica, el control prosódico sigue siendo un aspecto subexplorado en la síntesis de voz. La literatura sobre Text-to-Speech emocional ha propuesto diversas estrategias para abordar esta limitación. Cai et al. [26] introducen un enfoque de aprendizaje semisupervisado que no requiere anotaciones emocionales explícitas, pero es capaz de generar habla con expresividad emocional a partir de corpus expresivos. Cornille et al. [27], por su parte, implementan un sistema de control multinivel (tanto global como local) sobre la prosodia dentro del modelo FastSpeech, lo que permite ajustes detallados incluso a nivel fonémico. En una línea similar, Choi y Hahn [28] utilizan módulos de atención seq2seq para modular la duración de las emociones, mientras que Kaneko y Kameoka [29] proponen CycleGAN-VC para realizar conversiones de voz no paralelas, evitando la sobresuavización típica de los modelos estadísticos tradicionales.

De forma complementaria, Liu et al. [13] incorporan funciones de pérdida perceptual a nivel de enunciado completo para mejorar la expresividad emocional. Asimismo, Wu et al. [30] y Zhang et al. [31] exploran mecanismos de control latente a través de Global Style Tokens (GST) y Variational Autoencoders (VAE), los cuales permiten transferir emociones con una cantidad limitada de datos. Si bien, estos desarrollos demuestran que el control prosódico es técnicamente viable y prometedor; sin embargo, su aplicación al ámbito de la AD automática aún se encuentra en una etapa emergente.

Estudios de neurociencia cognitiva indican que las personas con ceguera (especialmente adquirida) manifiestan una sensibilidad potenciada hacia los matices emocionales de la voz [14]. Investigaciones sobre autenticidad afectiva [16] y procesamiento auditivo en entornos ruidosos [20] corroboran que cualquier artificialidad prosódica mal calibrada se percibe con facilidad, afectando la inmersión. Por consiguiente, la síntesis destinada a AD debe equilibrar naturalidad acústica y claridad emocional, adaptándose a las expectativas cognitivas de esta población.

Por lo anterior, la automatización del control prosódico se perfila como una frontera emergente. Sistemas multimodales que combinan reconocimiento de emociones en voz y expresiones faciales [32] enlazan el análisis emocional con la generación TTS, ofreciendo ajustes en tiempo real. Enfoques como EmotionGesture [12] o Sound to Expression [33] ilustran cómo las señales acústicas pueden guiar la generación de expresiones no verbales, prefigurando narrativas accesibles que sincronicen voz y gesto. Además, arquitecturas basadas en difusión latente, como DiffProsody [19], y modelos multilingües como METTS [34] evidencian la viabilidad de transferir emociones entre dominios y lenguas con mínima supervisión.

La propuesta destaca por su capacidad de articular mecanismos de atención con un modelado prosódico emocional, todo dentro de una narrativa semántica coherente. Al priorizar dinámicamente los elementos visuales más relevantes de cada escena, el sistema genera descripciones que no solo son precisas y contextualizadas, sino que también se expresan con una prosodia adaptada a la emoción predominante. Esta combinación permite que la audiodescripción transmita tanto el contenido informativo como los matices afectivos del material audiovisual, logrando una experiencia auditiva más rica, inmersiva y alineada con las necesidades de las personas con discapacidad visual. La integración de atención, narrativa y prosodia emocional abre nuevas posibilidades para la accesibilidad, superando la monotonía y limitaciones de los enfoques convencionales.

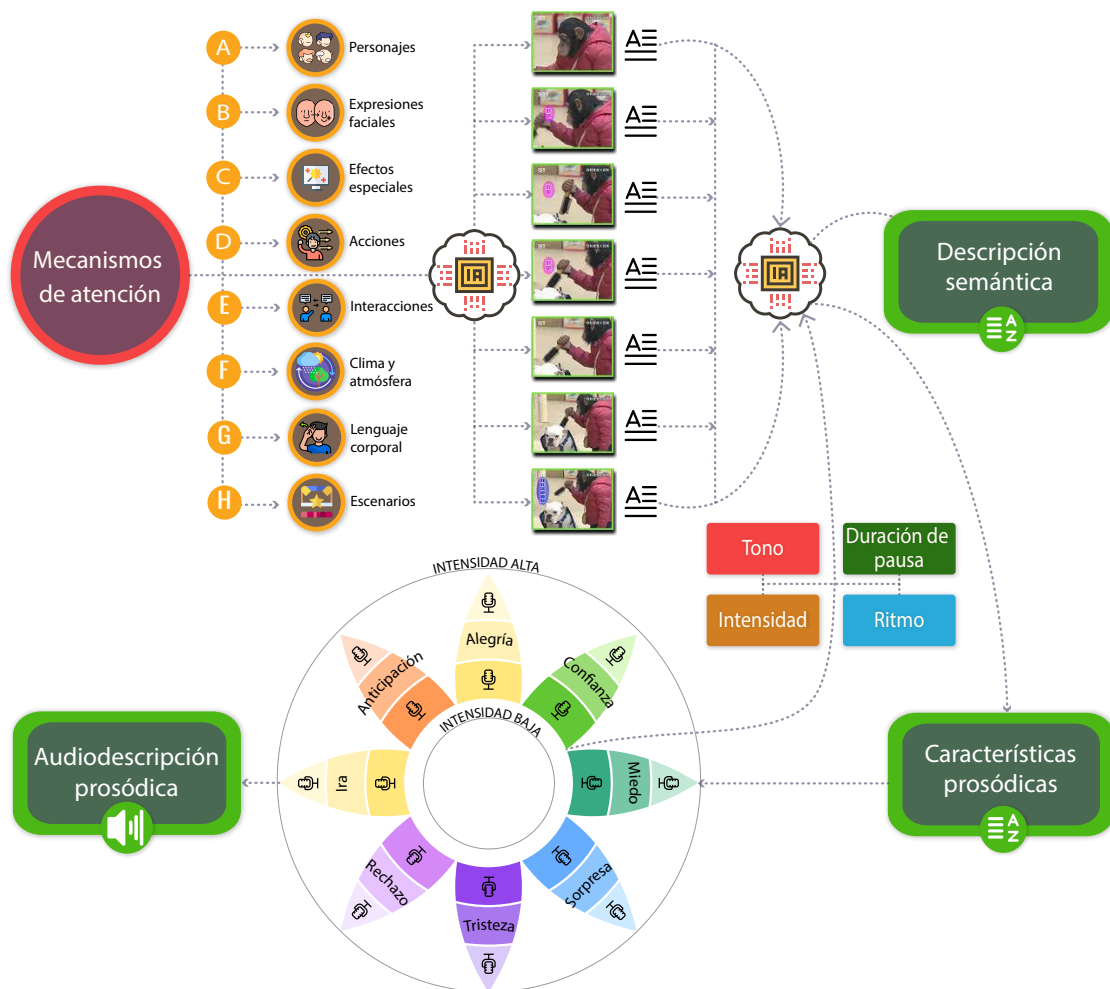


Figura 1: Esquema general del sistema propuesto para Modelado de Audiodescripción Prosódica Basado en Narrativa Semántica, estructurada en mecanismos de atención sobre elementos clave de la escena, genera una descripción semántica y ajusta las características prosódicas de la síntesis de voz en función de un esquema emocional, siguiendo el modelo de Plutchik [35].

3. Esquema propuesto

Se propone un esquema que integra técnicas de análisis de imágenes, extracción semántica contextual y caracterización prosódica, con el objetivo de optimizar la generación automática de AD emocionalmente expresivas. Este enfoque está orientado a mejorar la accesibilidad digital y enriquecer la experiencia narrativa de personas con discapacidad visual. La Figura 1 ilustra la arquitectura general del Modelado de Audiodescripción Prosódica Basado en Narrativa Semántica, estructurada en varias etapas interrelacionadas.

En la primera etapa, se aplican **Mecanismos de Atención** sobre distintos elementos de la escena audiovisual, tales como personajes, expresiones faciales, efectos especiales, acciones, interacciones, clima y atmósfera, lenguaje corporal y escenarios. Estos mecanismos permiten identificar y priorizar la información más relevante para la comprensión de la escena, evitando la sobrecarga descriptiva y asegurando la pertinencia contextual.

Posteriormente, se genera una **Descripción Semántica** detallada, que no solo enumera objetos o acciones, sino que construye un relato coherente y contextualizado, adaptado a la dinámica temporal de la escena. Esta narrativa semántica se apoya en modelos de procesamiento de lenguaje natural y visión por computadora, garantizando la alineación entre el contenido visual y la descripción verbal.

El sistema integra la dimensión emocional mediante la identificación de la emoción predominante en la escena, siguiendo el modelo de Plutchik. Para ello, se consideran 16 interpretaciones emocionales, cada una asociada a una emoción principal del modelo. A su vez, cada emoción se vincula a un perfil prosódico específico, definido por parámetros como *intensidad*, *entonación*, *ritmo* y *duración de pausas*.

En la última etapa, la síntesis de voz ajusta sus **Características Prosódicas** para reflejar la emoción detectada, generando una AD prosódica que enriquece la experiencia auditiva y facilita la inmersión del usuario. Este flujo modular garantiza que la AD no solo sea informativa y precisa, sino también expresiva y emocionalmente congruente con el contenido audiovisual.

Adicionalmente, el sistema se alinea con el modelo de Plutchik para establecer una correspondencia sistemática entre emociones básicas y características acústicas. Por ejemplo, emociones como *alegría* o *ira* tienden a presentar un tono más alto, mayor intensidad y menor duración de pausas, mientras que *tristeza* o *serenidad* se caracterizan por tonos descendentes y ritmos pausados. Esta correspondencia se visualiza en la Figura 2, que sintetiza el efecto de cada emoción sobre la modulación prosódica del habla sintetizada. Este esquema modular y progresivo permite no solo generar descripciones visualmente contextualizadas, sino también expresarlas oralmente con una prosodia afectiva, coherente, natural y perceptiblemente enriquecida, sentando así las bases para una AD más natural y expresiva.

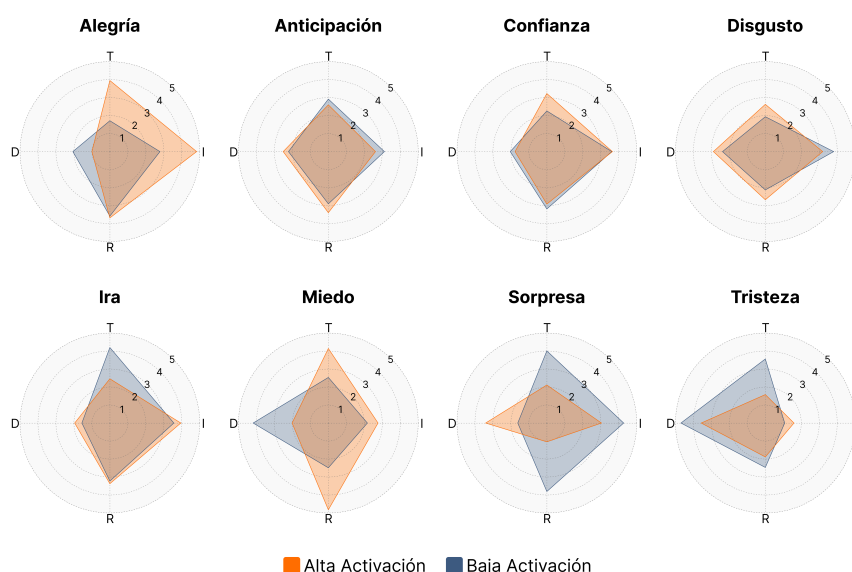


Figura 2: Correspondencia entre emociones del modelo de Plutchik y parámetros prosódicos del habla sintetizada. Cada emoción se asocia con variaciones específicas en intensidad (I), tono (T), ritmo (R) y duración de pausas (D).

Finalmente, se implementa un protocolo de evaluación que establece criterios basados en la coherencia, naturalidad y expresividad del habla generada. Para ello, se emplean desde Python las librerías Parselmouth PRAAT [36] y OpenSMILE, especializadas en el análisis de señales acústicas. Estas herramientas permiten extraer y verificar los valores de los parámetros prosódicos, comparándolos con los perfiles definidos en la matriz emocional, y así establecer contrastes entre los diferentes modelos de síntesis evaluados.

4. Metodología

Para desarrollar un sistema de audiodescripción automatizado y con una prosodia emocionalmente alineada, se diseñó un flujo de trabajo sistemático que integra técnicas de visión por computador, procesamiento de lenguaje natural y síntesis de voz. Este proceso abarca desde el análisis de contenido visual hasta la generación y evaluación acústica de la narrativa accesible. En la Figura 3, se muestra cada una de las fases implementadas: selección y segmentación de escenas, generación de descripciones semánticas, análisis y modelado prosódico emocional, comparación de modelos de descripción y síntesis, síntesis de voz y estrategias de generación, y evaluación acústica y validación.



Figura 3: Diagrama de flujo de la metodología propuesta. Se detallan las seis fases secuenciales: selección de escenas, generación de descripciones, modelado prosódico, comparación de modelos, síntesis de voz y evaluación acústica.

4.1. Selección y segmentación de escenas:

Se empleó el conjunto de datos **Microsoft Research Video Description Corpus (MSVD)** [37]. Este corpus está conformado por videos cortos, cada uno acompañado de múltiples descripciones generadas manualmente, lo que facilita la evaluación directa de la calidad narrativa y el cálculo de métricas objetivas de similitud semántica. El MSVD contiene aproximadamente 1,800 fragmentos de video y 120,000 anotaciones humanas en varios idiomas [2]. Habitualmente, los grupos de investigación dividen este conjunto en tres particiones: entrenamiento (1,200 videos), validación (100 videos) y prueba (670 videos). Todos los videos tienen el audio silenciado y se excluyen aquellos que presentan subtítulos o texto superpuesto. La duración típica de los videos varía entre 10 y 25 segundos, y la mayoría representa una sola actividad. Además, es ampliamente utilizado en aplicaciones como la asistencia a personas con discapacidad visual, interacción humano-robot, sistemas de videovigilancia y la comprensión automática de escenas, entre otras [1].

Para el experimento, se seleccionaron 10 escenas representativas del corpus MSVD 4, priorizando aquellas con alta carga emocional y variedad de elementos visuales. Los videos fueron segmentados en fotogramas clave mediante el modelo CLIP, con una segmentación adaptativa ajustada dinámicamente a la duración del video y su tasa de fotogramas por segundo (FPS). La diarización inversa permitió identificar intervalos sin diálogos para insertar las AD; cabe resaltar que los videos no contenían audio.



Figura 4: Videos utilizados para el estudio con su descripción del corpus MSVD.

4.2. Generación de descripciones semánticas:

A partir de los fotogramas extraídos, se generaron descripciones automáticas usando un modelo de inteligencia artificial generativa, estableciendo un límite de caracteres proporcional a la duración del video (30 caracteres por segundo). Se incorporaron mecanismos de atención avanzados para mejorar la coherencia y el nivel de detalle en la descripción de cada escena, priorizando los elementos más relevantes.

4.3. Análisis y modelado prosódico emocional:

Se elaboró un análisis de prosodia, ajustando el tono, ritmo e intensidad de la narración con base en las emociones predominantes de cada video. A través del procesamiento de los fotogramas clave y las descripciones generadas, se identificaron emociones principales y secundarias, asignando porcentajes de relevancia para reflejar la intencionalidad emocional de la escena. Este paso permitió que la AD no solo fuera precisa y clara, sino también coherente con la carga emocional del contenido visual, mejorando así la experiencia de accesibilidad.

Las descripciones generadas se asociaron a una emoción predominante, identificada según el modelo de Plutchik. Se construyó un corpus de referencia con 16 interpretaciones emocionales, grabadas por 3 locutores expertos en AD, a partir del cual se extrajeron parámetros prosódicos clave: intensidad, tono, ritmo y duración de pausas. Estos valores promedio por emoción se utilizaron para parametrizar la síntesis de voz, presentando la relación entre la emoción detectada y la expresión prosódica.

4.4. Comparación de modelos de descripción y síntesis:

Posteriormente, se evaluaron comparativamente ocho modelos de inteligencia artificial generativa para la descripción de videos, contrastándolos con las descripciones oficiales de referencia. El análisis abarcó 10 videos con anotaciones en español, seleccionados estratégicamente para garantizar una diversidad de escenas y contextos visuales. Los resultados de esta evaluación comparativa permitieron determinar

que GPT-4o ofrece el mejor rendimiento en la generación de descripciones. Con base en estos hallazgos, se consolidó un protocolo sistemático para la generación de las AD, abarcando desde la reutilización de narrativas previamente generadas hasta la síntesis de voz emocional y el análisis prosódico comparativo.

4.5. Síntesis de voz y estrategias de generación:

Para la etapa de síntesis de voz, se aplicó el esquema metodológico descrito previamente, utilizando dos estrategias diferenciadas:

1. **Generación mediante instrucciones textuales (prompts):** Se utilizaron instrucciones explícitas para inducir emociones en los modelos TTS que lo permitían.
2. **Generación mediante audios de referencia:** Este fue el enfoque central del estudio. Se usaron grabaciones del corpus de referencia emocional para que los modelos de clonación de voz replicaran sus patrones prosódicos.

La operativización se realizó con un conjunto de herramientas seleccionadas por su robustez, accesibilidad y compatibilidad con ajustes prosódicos. Los modelos específicos empleados para el modelado a partir de muestras de audio fueron:

- **XTTS v2:** Se utilizó la plataforma de código abierto Coqui TTS, específicamente el modelo XTTS v2. Es reconocido por su alta capacidad de clonación de voz multilingüe en modo zero-shot, lo que fue clave para reproducir fielmente el tono, ritmo e intensidad de las grabaciones de referencia y generar una voz natural y alineada emocionalmente.
- **Spanish-F5:** Se empleó el modelo Spanish-F5, una implementación para el español basada en la arquitectura F5-TTS. Este modelo fue entrenado con un corpus de 218 horas de audio que incluye una variedad de acentos latinoamericanos y españoles. Al igual que con XTTS v2, se utilizó su capacidad para replicar las características de una voz de referencia, permitiendo así generar las audiodescripciones con la riqueza prosódica y regional capturada de los locutores.

Cada uno de estos modelos fue empleado para sintetizar las AD correspondientes a los 10 videos previamente seleccionados. A fin de mantener la comparabilidad entre salidas, se conservó constante el contenido semántico en todas las muestras, y únicamente se modificaron los parámetros prosódicos conforme a las capacidades y configuraciones específicas de cada herramienta. En los modelos controlados por prompt, se utilizaron instrucciones textuales explícitas para inducir emociones específicas. Por su parte, los modelos por referencia reprodujeron patrones prosódicos previamente definidos, basados en perfiles emocionales mapeados en la fase de modelado prosódico.

Cada modelo generó una AD por video, con excepción de las herramientas basadas en audio de referencia (Coqui y Spanish-F5), las cuales utilizaron como insumo 3 audios de referencia correspondientes a los locutores previamente seleccionados para el modelado prosódico emocional. De este modo, por cada video se generaron 3 versiones de AD utilizando estas herramientas, asociadas a los perfiles prosódicos derivados de cada uno de los 3 locutores. En conjunto, el proceso de síntesis produjo un total de 60 archivos de audio: 30 provenientes de los modelos por referencia (2 herramientas × 10 videos × 3 locutores). Este corpus constituido por AD sintetizadas fue posteriormente sometido a un análisis comparativo de las características prosódicas, con el fin de evaluar el grado de alineación entre la prosodia generada y las emociones objetivo definidas conforme al modelo emocional de Plutchik. Este análisis permitió realizar una valoración integral de la expresividad emocional lograda por cada enfoque de síntesis en el contexto de la narración automatizada para accesibilidad.

4.6. Evaluación acústica y validación:

La evaluación de las AD sintetizadas se centró en analizar su expresividad, naturalidad y adecuación emocional a través de un estudio acústico de sus propiedades prosódicas. Con este fin, se emplearon de manera complementaria dos herramientas especializadas: Parselmouth, una interfaz de Python para

PRAAT, y OpenSMILE, un robusto entorno para la extracción de características acústicas de bajo nivel utilizado en el análisis de emociones en la voz. La utilización conjunta de estas herramientas aseguró la precisión y completitud en la extracción de los parámetros prosódicos (Tono, Intensidad, Ritmo y Duración de pausas), como se detalla en la Tabla 1.

Tabla 1

Parámetros prosódicos analizados y herramienta de extracción.

Parámetro	Descripción	Herramienta
Tono (T)	Representa la entonación de la voz; se asocia a la altura tonal y es fundamental para la percepción emocional.	Parselmouth
Duración de pausas (D)	Tiempo entre unidades fonéticas o frases; influye en el ritmo, la segmentación del discurso y la expresividad.	Parselmouth / OpenSMILE
Intensidad (I)	Nivel de energía acústica de la señal vocal; afecta la percepción de énfasis, fuerza o suavidad emocional.	OpenSMILE
Ritmo (R)	Velocidad relativa del habla, basada en la duración de sílabas y pausas; indica dinamismo o lentitud del discurso.	Parselmouth / OpenSMILE

Esta aproximación metodológica no solo permitió una caracterización cuantitativa y reproducible de la prosodia, facilitando la comparación entre los distintos modelos de síntesis utilizados. A partir de estos parámetros, se analizó el grado de correspondencia entre las características acústicas producidas por cada sistema TTS y los perfiles prosódicos definidos previamente según el modelo emocional de Plutchik. Además, sentó las bases para una evaluación sistemática del alineamiento emocional y la eficacia expresiva de las narraciones automatizadas desde una perspectiva multimodal y comparativa.

5. Resultados

La evaluación de los modelos de AD prosódica se realizó sobre el conjunto de 10 videos seleccionados del corpus MSVD, generando un total de 40 archivos de audio sintetizados mediante las distintas estrategias y herramientas de TTS descritas en la metodología. El análisis se articuló en torno a tres dimensiones principales: coherencia semántica, expresividad prosódica y alineación emocional.

Coherencia semántica: las métricas de similitud (como BLEU, METEOR y ROUGE) indicaron que el modelo GPT-4o obtuvo el puntaje más alto al comparar las descripciones generadas con las anotaciones de referencia en español. Las descripciones de GPT-4o fueron consistentemente evaluadas como más precisas, detalladas y contextualmente adecuadas.

Expresividad prosódica: el análisis acústico con OpenSMILE y Parselmouth reveló diferencias notables entre los enfoques de síntesis. Específicamente, los modelos basados en audio de referencia (XTTS v2 y Spanish-F5) lograron una mayor variabilidad promedio en tono, intensidad y ritmo, aproximándose a los perfiles definidos en el corpus de referencia emocional. En contraste, los modelos controlados por prompts textuales, aunque permitieron un control explícito de la emoción, mostraron una modulación prosódica menos marcada.

Alineación emocional: la correspondencia entre la emoción objetivo (según el modelo de Plutchik) y la prosodia generada fue consistentemente más alta en los sistemas por referencia. Dicha tendencia fue especialmente pronunciada en emociones de alta activación como alegría, ira o sorpresa. Las emociones de baja activación, como tristeza o serenidad, también fueron reproducidas con mayor fidelidad en estos modelos, evidenciando pausas más largas, tonos descendentes y menor intensidad. La Tabla 2 resume los valores promedio de los parámetros prosódicos obtenidos para cada emoción.

Tabla 2

Valores promedio de parámetros prosódicos por emoción y modelo de síntesis.

Emoción	Tono (Hz)	Intensidad (dB)	Ritmo (sílabas/s)	Duración de pausas (ms)
Alegría	220	75	5.2	180
Tristeza	160	62	3.1	350
Ira	240	78	5.5	150
Serenidad	170	65	3.4	320
Sorpresa	230	76	5.0	200

5.1. Asignación de la prosodia

Para la descripción de los videos se utilizó un enfoque basado en inteligencia artificial generativa, donde se analizaron los fotogramas clave de cada video junto con la descripción generada por nuestro modelo. En la Tabla 3 se muestra el análisis, donde se determinaron las características prosódicas más adecuadas para la narración, considerando emociones principales, intensidad, tono, duración de pausas y ritmo.

Tabla 3

Análisis de emociones, tono e intensidad en la descripción de videos.

VideoID	Emoción Principal	Intensidad	Tono	Duración	Ritmo
zTn-nNj5Bng	Confianza	Baja	Medio, cálido, poca variabilidad	Moderada	Regular, fluido, sin cambios bruscos
zpgW7m7_LZw	Alegría	Baja	Medio-bajo, estable, sin grandes variaciones	Larga, pausas relajadas	Lento, regular, pausado
ZdIG8fjGjlo	Confianza	Alta	Alto, a veces ligeramente ascendente	Moderada	Enfático, con cierto realce, pero no brusco
za-9mBZyNfQ	Alegría	Alta	Alto, con variaciones amplias	Moderada-corta, con exclamaciones	Rápido, energético, dinámico
Z3XarHZ3HpY	Anticipación	Alta en alerta	Medio-alto, a veces tenso	Moderada, con pausas para observar	Ágil, algo acelerado, atento a cambios
z0Si1XxMibg	Alegría	Baja	Medio-bajo, estable, sin grandes variaciones	Larga, pausas relajadas	Lento, regular, pausado
YerslyzsOpc	Tristeza	Baja	Bajo, monótono o levemente descendente	Larga, con pausas reflexivas	Lento, reflexivo, casi meditativo
YEmS8ZQlhTA	Alegría	Alta	Alto, con variaciones amplias	Moderada-corta, con exclamaciones	Rápido, energético, dinámico
yAD_TS5L2d4	Anticipación	Baja	Medio-alto, ligeramente ascendente	Variable	Fluido, ágil, mostrando atención sostenida
XZ0Ch_CIBfo	Alegría	Alta	Alto, con variaciones amplias	Moderada-corta, con exclamaciones	Rápido, energético, dinámico

Se analizaron 10 videos del conjunto de datos MSVD, con el objetivo de evaluar cómo diferentes modelos generan descripciones automáticas de los eventos visuales y cómo estas se alinean con las expectativas de los usuarios. Cada modelo fue probado en su capacidad para describir con precisión los personajes, objetos y eventos, detallando las características relevantes de la escena, y generando narrativas claras y coherentes.

En tal sentido, se determinó la capacidad expresiva de los modelos de síntesis de voz en la generación de AD emocional mediante 4 dimensiones prosódicas fundamentales: intensidad (I), tono (T), ritmo (R) y duración de pausas (D).

Las características prosódicas de cada AD fueron extraídas utilizando OpenSMILE, lo que permitió una caracterización robusta de la señal prosódica. A partir de los valores obtenidos, se realizó una comparación cuantitativa con los perfiles ideales definidos para cada emoción según el modelo de Plutchik, considerando variantes de alta y baja intensidad emocional, permitiendo construir un corpus homogéneo y equilibrado de 60 ADs. Es de anotar que se realizó un promedio de los datos obtenidos de las interpretaciones prosódicas de los locutores, esto en correspondencia a los criterios para evaluación prosódica sobre el cual se evaluó la congruencia entre las emociones objetivo y la prosodia sintetizada.

La Figura 5 muestra una comparación visual, donde se representa el ajuste entre los valores prosódicos generados por los modelos XTTS v2 y Spanish-F5 y los valores de referencia esperados para las emociones predominantes en cada video.

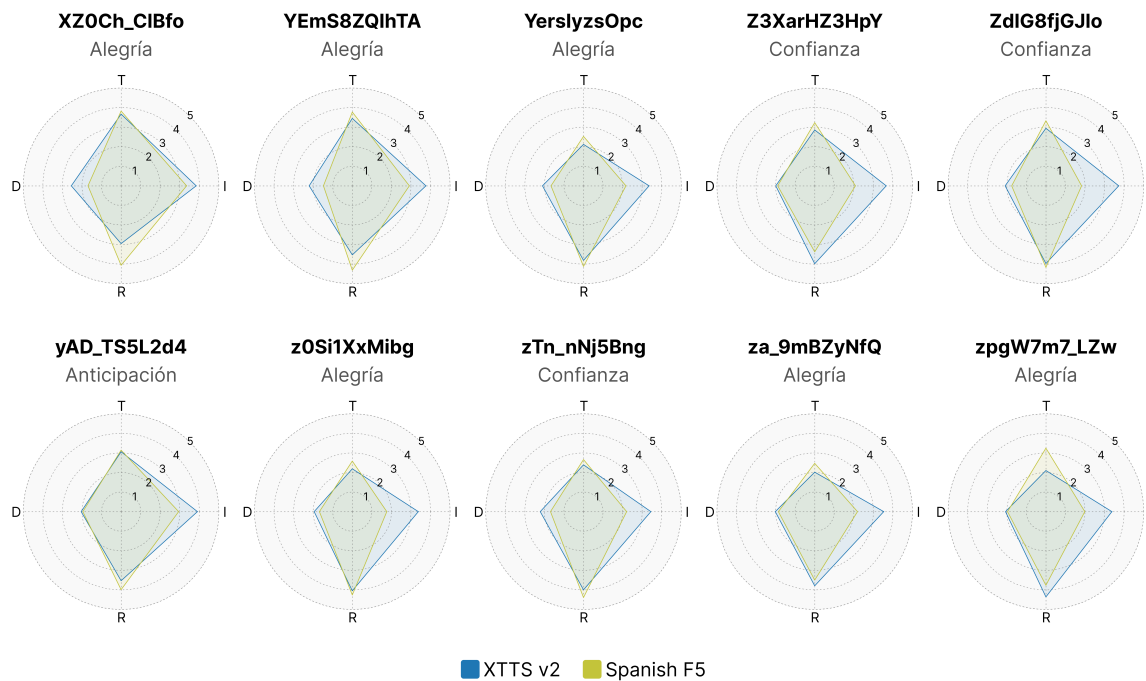


Figura 5: Comparación de valores prosódicos obtenidos por modelos de síntesis de voz emocional, evaluados con OpenSMILE. Cada gráfico corresponde a un video con su emoción predominante, según el modelo de Plutchik y mediante las dimensiones prosódicas: intensidad (I), tono (T), ritmo (R) y duración de pausas (D).

6. Discusión

Los resultados obtenidos en este trabajo validan la hipótesis central de que la integración de una narrativa semántica con un modelado prosódico-emocional mejora de manera significativa la calidad y expresividad de la AD automática. La capacidad de ajustar dinámicamente la prosodia en función de la emoción predominante en la escena es un paso crucial para ofrecer una experiencia auditiva más inmersiva y, en última instancia, más accesible para las personas con discapacidad visual.

El proceso de evaluación se basó en tres criterios principales: precisión, nivel de detalle, y claridad y comprensión. La precisión midió la capacidad del modelo para identificar correctamente personajes, objetos y eventos dentro del video, incluyendo la identificación clara de los sujetos, sus características distintivas, las acciones que realizan y la contextualización adecuada del entorno y los escenarios. El nivel de detalle evaluó la inclusión de información relevante sin redundancias, mientras que la claridad y comprensión verificó la coherencia narrativa de las descripciones generadas (4).

Un hallazgo clave es la aparente superioridad de los modelos por referencia (XTTS v2 y Spanish-F5) en la alineación emocional y la expresividad prosódica. El uso de perfiles acústicos derivados de locutores humanos parece ser un método más eficaz para transferir los matices del habla emocional a la voz sintetizada, especialmente en emociones de alta activación, donde la variabilidad prosódica es un marcador clave. Por el contrario, aunque los modelos basados en prompts ofrecen una mayor flexibilidad y un control más explícito, su modulación emocional tiende a ser menos matizada y natural.

Estos resultados refuerzan la importancia de contar con un corpus de referencia emocionalmente etiquetado y de emplear estrategias de síntesis que permitan la transferencia precisa de perfiles prosódicos. Asimismo, la metodología propuesta demuestra que es posible mantener la coherencia semántica y la adecuación contextual de las descripciones, sin sacrificar la expresividad emocional, lo que contribuye a una experiencia auditiva más inmersiva y accesible para personas con discapacidad visual.

No obstante, persisten desafíos en la reproducción de emociones sutiles y en la estabilidad de la síntesis en escenas complejas o con cambios emocionales rápidos. La variabilidad entre locutores y la dependencia de la calidad de los audios de referencia también representan retos para la generalización del sistema. Además, la evaluación automática de la expresividad prosódica, aunque robusta, podría

complementarse con estudios perceptuales con usuarios finales para validar la experiencia subjetiva de accesibilidad e inmersión.

En síntesis, estos hallazgos abren nuevas líneas de investigación orientadas a perfeccionar los mecanismos de control prosódico, ampliar la cobertura emocional de los corpus de referencia y desarrollar métricas perceptuales que reflejen mejor la experiencia del usuario. La combinación de análisis acústico cuantitativo con pruebas estadísticas y estudios perceptuales con usuarios finales será fundamental para avanzar hacia una AD automatizada que sea verdaderamente natural, expresiva y centrada en las necesidades de la comunidad con discapacidad visual.

7. Limitaciones

A pesar de los resultados prometedores, este estudio presenta ciertas limitaciones que deben ser consideradas para contextualizar los hallazgos y guiar futuras investigaciones. Una de las principales limitaciones se deriva del corpus utilizado, Microsoft Research Video Description Corpus (MSVD). Los videos de este corpus tienen duraciones cortas, típicamente entre 10 y 25 segundos, y en su mayoría representan actividades individuales y acotadas. Esta característica, si bien es útil para análisis controlados, no refleja la complejidad narrativa de contenidos audiovisuales más extensos como películas o documentales. Esta característica del conjunto de datos puede haber influido en la efectividad del modelo prosódico, especialmente en escenarios que requieren transiciones emocionales más complejas o narrativas de mayor duración.

Como consecuencia directa de la corta duración de los videos, las descripciones generadas para el análisis también fueron breves y concisas. Esto dificultó una comparación más profunda y un análisis detallado de matices narrativos complejos en los resultados.

El estudio se llevó a cabo exclusivamente en español. Esto restringe la aplicabilidad directa de los hallazgos a otros idiomas, que poseen características prosódicas y estructuras gramaticales distintas. Adicionalmente, no se consideraron de manera sistemática las variaciones dialectales y regionales dentro del propio español, las cuales podrían influir en la percepción de la naturalidad y la emoción.

Estas limitaciones abren líneas claras para investigaciones futuras, orientadas a la evaluación del sistema en contextos narrativos más extensos y variados, así como a la adaptación y validación en otros idiomas y variantes dialectales.

8. Conclusiones

El presente trabajo ha propuesto un esquema innovador que integra mecanismos de atención para la generación de narrativas semánticas y características prosódicas en la AD automática, demostrando avances significativos en la calidad y expresividad de las descripciones generadas. Los resultados confirman que el enfoque propuesto supera a modelos alternativos en términos de precisión semántica, riqueza descriptiva y claridad narrativa, ofreciendo una solución más completa para mejorar la accesibilidad audiovisual.

La incorporación de mecanismos avanzados de atención ha sido crucial para generar descripciones enriquecidas con contexto, permitiendo identificar y describir con mayor precisión los elementos relevantes de cada escena. Esta capacidad para distinguir la importancia de objetos, acciones y eventos dentro del contenido visual representa un avance significativo respecto a enfoques convencionales, que tienden a generar descripciones genéricas o incompletas.

El modelado prosódico emocional implementado ha permitido superar una de las limitaciones más persistentes en la AD automática: la monotonía expresiva. Al adaptar dinámicamente parámetros como intensidad, tono, duración y ritmo según la carga emocional de cada escena, este enfoque logra narraciones más naturales y emocionalmente alineadas con el contenido visual. Esta riqueza prosódica mejora la experiencia auditiva y facilita la comprensión y el disfrute del material audiovisual por parte de personas con discapacidad visual.

La integración de características prosódicas en la AD representa un cambio importante en el campo, yendo más allá del enfoque tradicional centrado solo en la precisión semántica. Los resultados demuestran que esta aproximación dual, que considera tanto el “*qué*” como el “*cómo*” de la narración, enriquece significativamente la experiencia narrativa, permitiendo transmitir no solo el contenido visual, sino también su carga emocional.

9. Trabajo futuro

Para trabajos futuros, se identifican varias áreas prometedoras. La adaptabilidad prosódica en escenas cortas requiere mayor atención, ya que la variabilidad emocional en estos segmentos demanda ajustes más dinámicos en parámetros como tono y ritmo. Implementar estrategias avanzadas de fusión multimodal podría mejorar la precisión descriptiva en escenarios complejos o con múltiples eventos simultáneos.

Otra línea de investigación importante es el ajuste dinámico del número de caracteres por segundo en las descripciones. Si bien se estableció un límite de 30 caracteres por segundo, no se evaluó sistemáticamente su impacto en la comprensión y sincronización de las narraciones. Este aspecto es fundamental para garantizar que las AD automáticas sean precisas, expresivas y temporalmente adecuadas al ritmo visual del contenido.

Además, se destaca la necesidad de incorporar procesos de evaluación perceptual con usuarios finales, especialmente personas con discapacidad visual, con el objetivo de validar la calidad emocional y la naturalidad de las AD generadas. Solo mediante esta retroalimentación empírica será posible ajustar los sistemas de síntesis a las preferencias y necesidades reales de los usuarios, favoreciendo el diseño de experiencias auditivas verdaderamente inclusivas.

Finalmente, este trabajo sienta las bases para explorar modelos adaptativos que integren prosodia emocional, narrativa contextual y aprendizaje continuo. Esta perspectiva podría derivar en sistemas TTS más inteligentes y personalizables, capaces de ajustar automáticamente su estilo narrativo y emocional según el tipo de contenido, la audiencia o el entorno de aplicación. Así, el enfoque propuesto no solo avanza el estado del arte en AD automática, sino que también abre nuevas rutas para la democratización del acceso a contenidos audiovisuales a través de tecnologías emocionalmente conscientes.

Agradecimientos

Producto derivado del proyecto INV-ING-4048 financiado por la Universidad Militar Nueva Granada - Vigencia 2024. Los autores expresan su agradecimiento a la Universidad Militar Nueva Granada por el apoyo brindado. Asimismo, se reconoce la vinculación como profesor de tiempo completo del autor Christian Quintero a la Universidad Militar Nueva Granada, lo que contribuyó significativamente a la realización de esta investigación.

Declaración de uso de IA Generativa

Durante la preparación de este trabajo, los autores emplearon herramientas de IA Generativa. Específicamente, se utilizó el modelo GPT-4o para la generación y evaluación comparativa de las descripciones semánticas de los videos. Asimismo, los modelos de síntesis de voz XTTS v2 y Spanish-F5, basados en transferencia de características de voz, se usaron para generar las audiodescripciones prosódicas a partir de audios de referencia grabados por locutores como parte de este estudio. Tras utilizar estas herramientas, los autores revisaron y editaron el contenido según fue necesario y asumen la total responsabilidad por el contenido de la publicación.

Referencias

- [1] D. Sharma, C. Dhiman, D. Kumar, Evolution of visual data captioning methods, datasets, and evaluation metrics: A comprehensive survey, *Expert Systems with Applications* 221 (2023) 119773. doi:10.1016/j.eswa.2023.119773.
- [2] S. Islam, A. Dash, A. Seum, A. H. Raj, T. Hossain, F. M. Shah, Exploring video captioning techniques: A comprehensive survey on deep learning methods, *SN Computer Science* 2 (2021) 120. doi:10.1007/s42979-021-00487-x.
- [3] M. Shen, G. Huang, Y. Wu, S. Song, S. Zhou, L. Li, Z. Yu, W. Wang, J. Bu, Making accessible movies easily: An intelligent tool for authoring and integrating audio descriptions to movies, in: *Proceedings of the 21st International Web for All Conference*, ACM, 2024, pp. 160–164. doi:10.1145/3677846.3677855.
- [4] B. Iavarone, M. S. Morelli, D. Brunato, S. Ghiasi, E. P. Scilingo, N. Vanello, F. Dell’Orletta, A. Greco, The linguistic structure of an emotional text influences the sympathetic activity and the speech prosody of the reader, *Biomedical Signal Processing and Control* 89 (2024). doi:10.1016/j.bspc.2023.105776.
- [5] Z. Gu, K. He, Affective Prompt-Tuning-Based Language Model for Semantic-Based Emotional Text Generation, *INTERNATIONAL JOURNAL ON SEMANTIC WEB AND INFORMATION SYSTEMS* 20 (2024). doi:10.4018/IJSWIS.339187.
- [6] C. Nussbaum, M. Poehlmann, H. Kreysa, S. R. Schweinberger, Perceived naturalness of emotional voice morphs, *COGNITION & EMOTION* 37 (2023) 731–747. doi:10.1080/02699931.2023.2200920.
- [7] L. Zhou, Y. Zhou, J. J. Corso, R. Socher, C. Xiong, End-to-end dense video captioning with masked transformer (2018).
- [8] Y. Li, T. Yao, Y. Pan, H. Chao, T. Mei, Jointly localizing and describing events for dense video captioning, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [9] Y. Zhu, S. Jiang, Attention-based densely connected lstm for video captioning, in: *Proceedings of the 27th ACM International Conference on Multimedia*, ACM, 2019, pp. 802–810. doi:10.1145/3343031.3350932.
- [10] T. M. Taha, Z. B. Messaoud, M. Frikha, Convolutional Neural Network Architectures for Gender, Emotional Detection from Speech and Speaker Diarization, *International Journal of Interactive Mobile Technologies* 18 (2024) 88–103. doi:10.3991/ijim.v18i03.43013.
- [11] K. Cusworth, G. Paulik, N. Thomas, D. Preece, G. Campitelli, D. C. Mathersul, Emotion processes in voice-hearers: Understanding differences in emotional reactivity, emotion regulation and alexithymia, *Psychology and Psychotherapy: Theory, Research and Practice* 97 (2024) 706 – 721. doi:10.1111/papt.12554.
- [12] X. Qi, C. Liu, L. Li, J. Hou, H. Xin, X. Yu, EmotionGesture: Audio-Driven Diverse Emotional Co-Speech 3D Gesture Generation, *IEEE Transactions on Multimedia* 26 (2024) 10420 – 10430. doi:10.1109/TMM.2024.3407692.
- [13] M. Liu, C. Wang, J. Hu, Older adults’ intention to use voice assistants: Usability and emotional needs, *Heliyon* 9 (2023). doi:10.1016/j.heliyon.2023.e21932.
- [14] B. M. Ben-David, D. R. Chebat, M. Icht, “Love looks not with the eyes”: supranormal processing of emotional speech in individuals with late-blindness versus preserved processing in individuals with congenital-blindness, *Cognition and Emotion* (2024). doi:10.1080/02699931.2024.2357656.
- [15] T. Bott, F. Lux, N. T. Vu, Controlling Emotion in Text-to-Speech with Natural Language Prompts (2024). arXiv:2406.06406.
- [16] J. Sarzedas, C. F. Lima, M. S. Roberto, S. K. Scott, A. P. Pinheiro, T. Conde, Blindness influences emotional authenticity perception in voices: Behavioral and ERP evidence, *Cortex* 172 (2024) 254–270. doi:10.1016/j.cortex.2023.11.005.
- [17] Z. Zhou, Y. Shen, X. Chen, D. Wang, GERP: A Personality-Based Emotional Response Generation Model, *Applied Sciences (Switzerland)* 13 (2023) 1–15. doi:10.3390/app13085109.

- [18] X. Zheng, Y. Du, X. Qin, CoMaSa: Context Multi-aware Self-attention for emotional response generation, *NEUROCOMPUTING* 611 (2025). doi:10.1016/j.neucom.2024.128692.
- [19] H.-S. Oh, S.-H. Lee, S.-W. Lee, DiffProsody: Diffusion-Based Latent Prosody Generation for Expressive Speech Synthesis With Prosody Conditional Adversarial Training, *IEEE-ACM TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING* 32 (2024) 2654–2666. doi:10.1109/TASLP.2024.3395994.
- [20] Y.-L. Liu, Y.-X. Zhang, Y. Wang, Y. Yang, Evidence for early encoding of speech in blind people, *Brain and Language* 259 (2024) 105504. doi:https://doi.org/10.1016/j.bandl.2024.105504.
- [21] J. Xie, T. Han, M. Bain, A. Nagrani, G. Varol, W. Xie, A. Zisserman, Autoad-zero: A training-free framework for zero-shot audio description (2024).
- [22] W. Zhao, X. Wu, X. Zhang, Memcap: Memorizing style knowledge for image captioning, *Proceedings of the AAAI Conference on Artificial Intelligence* 34 (2020) 12984–12992. doi:10.1609/aaai.v34i07.6998.
- [23] F. Huang, S. Zeng, J. Ke, S. Lei, J. Wang, A video description model with improved attention mechanism, *Journal of Physics: Conference Series* 2384 (2022) 012015. doi:10.1088/1742-6596/2384/1/012015.
- [24] T. C. Smits, S. L'yi, A. P. Mar, N. Gehlenborg, Altgosling: Automatic generation of text descriptions for accessible genomics data visualization (2024). doi:10.1093/bioinformatics/btae670/7900296.
- [25] D. Li, Y. Zhao, R. Cui, L. Zhao, Research on image text generation based on word2vec visual vocabulary attention, in: *2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, IEEE, 2021, pp. 344–348. doi:10.1109/ACCTCS52002.2021.00075.
- [26] X. Cai, D. Dai, Z. Wu, X. Li, J. Li, H. Meng, Emotion controllable speech synthesis using emotion-unlabeled dataset with the assistance of cross-domain speech emotion recognition (2020). doi:https://doi.org/10.48550/arXiv.2010.13350.
- [27] T. Cornille, F. Wang, J. Bekker, Interactive multi-level prosody control for expressive speech synthesis, in: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 8312–8316. doi:10.1109/ICASSP43922.2022.9746654.
- [28] H. Choi, M. Hahn, Sequence-to-sequence emotional voice conversion with strength control, *IEEE Access* 9 (2021) 42674–42687. doi:10.1109/ACCESS.2021.3065460.
- [29] T. Kaneko, H. Kameoka, CycleGAN-VC: Non-parallel voice conversion using cycle-consistent adversarial networks, in: *2018 26th European Signal Processing Conference (EUSIPCO)*, IEEE, 2018, pp. 2100–2104. doi:10.23919/EUSIPCO.2018.8553236.
- [30] P. fei Wu, Z. hua Ling, L. juan Liu, Y. Jiang, H. chuan Wu, L. rong Dai, End-to-end emotional speech synthesis using style tokens and semi-supervised training (2019). doi:https://doi.org/10.48550/arXiv.1906.10859.
- [31] Y.-J. Zhang, S. Pan, L. He, Z.-H. Ling, Learning latent representations for style control and transfer in end-to-end speech synthesis (2018).
- [32] C. P. Loizou, An automated integrated speech and face image analysis system for the identification of human emotions, *Speech Communication* 130 (2021) 15–26. doi:https://doi.org/10.1016/j.specom.2021.04.001.
- [33] W. Liu, S. Zhang, L. Zhou, N. Luo, Q. Chen, Sound to expression: Using emotional sound to guide facial expression editing, *Journal of King Saud University - Computer and Information Sciences* 36 (2024) 101998. doi:https://doi.org/10.1016/j.jksuci.2024.101998.
- [34] X. Zhu, Y. Lei, T. Li, Y. Zhang, H. Zhou, H. Lu, L. Xie, METTS: Multilingual Emotional Text-to-Speech by Cross-speaker and Cross-lingual Emotion Transfer (2023). arXiv:2307.15951.
- [35] R. Plutchik, The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, *American Scientist* 89 (2001) 344–350.
- [36] P. Boersma, V. Van Heuven, Speak and unspeak with praat, *Glott International* 5 (2001) 341–347.
- [37] A. Yang, A. Nagrani, P. H. Seo, A. Miech, J. Pont-Tuset, I. Laptev, J. Sivic, C. Schmid, Vid2seq: Large-scale pretraining of a visual language model for dense video captioning (2023).

A. Prompt de Control Emocional para la Modulación Prosódica

El siguiente conjunto de instrucciones fue diseñado para guiar la síntesis de voz emocional mediante la definición de perfiles prosódicos específicos por emoción. Cada entrada describe los parámetros de intensidad, tono, duración y ritmo asociados a una emoción principal y su complementaria, permitiendo así una generación de habla sintética emocionalmente coherente.

- *Emoción principal: Alegría; Emoción complementaria: Serenidad; Intensidad: Baja y suave; Tono: Medio-bajo, estable, sin grandes variaciones; Duración: Larga, pausas relajadas; Ritmo: Lento, regular, pausado.*
- *Emoción principal: Alegría; Emoción complementaria: Éxtasis; Intensidad: Muy alta y entusiasta; Tono: Alto, con variaciones amplias (picos frecuentes); Duración: Moderada-corta, con exclamaciones; Ritmo: Rápido, enérgico, dinámico.*
- *Emoción principal: Confianza; Emoción complementaria: Aceptación; Intensidad: Baja y estable; Tono: Medio, cálido, poca variabilidad; Duración: Moderada; Ritmo: Regular, fluido, sin cambios bruscos.*
- *Emoción principal: Confianza; Emoción complementaria: Admiración; Intensidad: Alta y expresiva; Tono: Medio-alto, a veces ligeramente ascendente; Duración: Moderada; Ritmo: Enfático, con cierto realce, pero no brusco.*
- *Emoción principal: Miedo; Emoción complementaria: Aprensión; Intensidad: Baja, con tensión; Tono: Ligeramente ascendente, posible temblor o quiebre; Duración: Moderada, con pausas vacilantes; Ritmo: Irregular, inseguro, a veces acelerado.*
- *Emoción principal: Miedo; Emoción complementaria: Terror; Intensidad: Muy alta y potente; Tono: Muy alto, con quiebres marcados; Duración: Corta, entrecortada; Ritmo: Muy acelerado o brusco, pausas de jadeo.*
- *Emoción principal: Sorpresa; Emoción complementaria: Distracción; Intensidad: Muy baja e inconstante; Tono: Variable, a veces errático; Duración: Variable (puede ser corta o larga); Ritmo: Inconsistente, con cambios inesperados.*
- *Emoción principal: Sorpresa; Emoción complementaria: Asombro; Intensidad: Alta y enfocada en la exclamación; Tono: Alto y ascendente al inicio, luego variable; Duración: Corta o moderada, con pausas breves; Ritmo: Brusco e irregular, marcando sorpresa.*
- *Emoción principal: Tristeza; Emoción complementaria: Pensamiento profundo; Intensidad: Baja y tenue; Tono: Medio-bajo, monótono o levemente descendente; Duración: Larga, con pausas reflexivas; Ritmo: Lento, reflexivo, casi meditativo.*
- *Emoción principal: Tristeza; Emoción complementaria: Duelo; Intensidad: Muy alta y sufrimiento; Tono: Bajo, descendente, monótono; Duración: Larga, con pausas prolongadas; Ritmo: Muy lento, pesado, con silencios marcados.*
- *Emoción principal: Disgusto; Emoción complementaria: Aburrimiento; Intensidad: Muy baja y monótona; Tono: Bajo, con poca variación; Duración: Moderada-larga, habla sostenida; Ritmo: Lento, plano, sin cambios de intensidad.*
- *Emoción principal: Disgusto; Emoción complementaria: Aversión; Intensidad: Alta con desagrado; Tono: Medio a alto, seco o cortante; Duración: Moderada, a veces con pausas bruscas; Ritmo: Cortante, agresivo, enfatizando el rechazo.*
- *Emoción principal: Ira; Emoción complementaria: Molestia; Intensidad: Baja, a veces variable; Tono: Medio, con pequeños picos de tensión; Duración: Moderada, frases más cortas; Ritmo: Un tanto acelerado, interrumpido, con quiebres.*
- *Emoción principal: Ira; Emoción complementaria: Furia; Intensidad: Muy alta y dominante; Tono: Alto, agresivo, con cambios bruscos; Duración: Moderada o corta, intercalada con exclamaciones; Ritmo: Rápido, explosivo, con picos intensos.*
- *Emoción principal: Anticipación; Emoción complementaria: Interés; Intensidad: Baja y enfocada; Tono: Medio-alto, ligeramente ascendente; Duración: Variable (depende de la interacción); Ritmo: Fluido, ágil, mostrando atención sostenida.*
- *Emoción principal: Anticipación; Emoción complementaria: Vigilancia; Intensidad: Alta en alerta; Tono: Medio-alto, a veces tenso; Duración: Moderada, con pausas para observar; Ritmo: Ágil, algo acelerado, atento a cambios."*

Conferencias invitadas



HCI en tiempos de IA: ética, diseño y regulación

Isabelle Hupont

Ministerio de Ciencia, Innovación y Universidades

Resumen

La inteligencia artificial (IA) ya no es una tecnología reservada a perfiles técnicos: hoy está integrada en aplicaciones de uso cotidiano y en sectores de alto impacto como la educación, la sanidad o la administración pública, siendo accesible para cualquier persona. Este cambio plantea nuevos retos para el diseño de la interacción persona-ordenador (HCI), que debe dar respuesta a cuestiones como la transparencia, la supervisión humana, la explicabilidad o la prevención de sesgos.

Frente a la reciente crítica a la opacidad de ciertos sistemas de IA, el diseño centrado en las personas se consolida como una herramienta clave para que estas tecnologías sean comprensibles, controlables y adaptadas a la diversidad de quienes las utilizan. La nueva regulación europea, en particular el Reglamento de Inteligencia Artificial (AI Act), refuerza esta visión: aunque exige ciertas garantías sobre la robustez de los algoritmos, sitúa el foco regulatorio en el uso previsto del sistema y en los riesgos que ese uso puede generar para los derechos fundamentales. Esto convierte a la HCI en un elemento central para garantizar un despliegue responsable de la IA.

En esta charla exploraremos cómo integrar los principios éticos, la supervisión humana y la atención a la diversidad desde las fases iniciales del diseño, tanto en sistemas de IA clásicos como generativos. Abordaremos también cómo estos sistemas pueden amplificar estereotipos y narrativas sesgadas, pero también cómo, mediante un diseño consciente, podemos fomentar representaciones más justas e inclusivas. Por último, compartiré mi experiencia como asesora científica en políticas públicas, desde la Comisión Europea hasta el Ministerio de Ciencia de España, para reflexionar sobre el papel estratégico que la investigación en HCI puede desempeñar en la construcción de un ecosistema digital centrado en las personas y alineado con los valores democráticos.

Interfaces cerebro-máquina: hacia una nueva era en la interacción persona-ordenador

Sergio López Bernal

Departamento de Ingeniería de la Información y las Comunicaciones (DIIC), Universidad de Murcia

Resumen

En esta ponencia se abordará el papel de las interfaces cerebro-máquina (BCIs) en la evolución de la interacción humano-computador, con especial atención a su potencial en escenarios emergentes como el metaverso. A partir de una introducción sobre qué son las BCIs, sus tipos y funcionamiento, se explorarán las posibilidades que ofrecen para establecer una conexión más directa y natural entre el cerebro humano y los entornos digitales. Se presentarán ejemplos actuales de aplicación, como la detección de emociones o diversos estados cognitivos, así como los contextos donde ya se están utilizando. Posteriormente, se analizará cómo el uso de BCIs podría transformar profundamente la experiencia en el metaverso, si bien aún se requieren avances tecnológicos significativos para su adopción generalizada. La charla también examinará las principales tendencias, los progresos que están haciendo estas interfaces más accesibles, y los desafíos que plantean en términos de privacidad, seguridad, inclusión y ética. Todo ello con el objetivo de ofrecer una visión clara, crítica y realista del potencial de las BCIs como próxima frontera en la interacción entre personas y sistemas digitales.

*Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain



Demostraciones



AffeX: a real-time emotion monitoring and visualization tool

Juan Camilo Méndez Flórez¹, Nelly Condori-Fernández^{1,2,*}, Alejandro Catala^{1,2},
Salvador Alonso Figueroa Vásquez¹ and Jose M. Alonso-Moral^{1,2}

¹*Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain*

²*Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain*

Abstract

Non-invasive, real-time emotion detection enables continuous and discreet monitoring of affective states, supporting system adaptability while safeguarding user privacy, an essential feature in clinical, educational, and research contexts. This paper presents AffeX, an intuitive tool for the real-time visualization and analysis of facial emotions, designed with a focus on privacy and data transparency. It offers customizable visual feedback, allowing the exploration of emotional responses without compromising identity. Unlike commercial solutions, AffeX ensures full control over data handling, enabling further analysis and multimodal integration, making it especially suitable for privacy-aware and adaptable research scenarios.

Keywords

human-computer interaction, real-time visualization, emotion monitoring, face anonymization

1. Introduction

Emotions play a key role in cognition, behavior, and interaction, influencing processes such as attention, memory, and decision making. Their automatic detection—whether through discrete categories or continuous models like valence and arousal—has become increasingly relevant for applications in healthcare, education, and human-computer interaction [1, 2].

Recent works have explored real-time emotional dashboards, combining facial data with other signals [3], or applying them to specific domains such as virtual classrooms [4] and educational video analysis [5]. In parallel, commercial solutions like FaceReader¹ or Kopernica² have emerged, though they often involve closed licenses, proprietary models, or the need for specific hardware.

This paper presents *AffeX*, an intuitive research tool for real-time analysis and visualization of emotions, with a primary focus on protecting user privacy through anonymization and facilitating further analyses in research contexts. This tool does not aim at competing with, nor substitute, existing commercial tools for emotion recognition, which indeed could provide highly specialized and well fine-tuned solutions in some niches, but complement them to enable real-time and research-oriented tasks in academic contexts. In this sense, the tool may be expanded in the future to be fed with multimodal signals coming from commercial emotion recognition engines instead of using the freely available models by DeepFace [6].

The remainder of this demo paper is organized as follows: Section 2 describes the architecture and modules. Section 3 addresses privacy, ethics, and limitations. Section 4 summarizes the main contributions and outlines future work.

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Corresponding author.

✉ juancamilo.mendez.florez@usc.es (J. C. M. Flórez); n.condori.fernandez@usc.es (N. Condori-Fernández);
alejandro.catala@usc.es (A. Catala); salvador.figueroa.vasquez@usc.es (S. A. F. Vásquez); josemaria.alonso.moral@usc.es
(J. M. Alonso-Moral)

ORCID 0009-0006-7282-0835 (J. C. M. Flórez); 0000-0002-1044-3871 (N. Condori-Fernández); 0000-0002-3677-672X (A. Catala);
0009-0004-7670-2997 (S. A. F. Vásquez); 0000-0003-3673-421X (J. M. Alonso-Moral)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹Noldus Information Technology, *FaceReader*, 2025. <https://www.noldus.com/facereader>

²Kopernica AI, *Kopernica Platform*, 2025. <https://www.kopernica.ai/?lang=es>

2. System design and architecture

2.1. Modules of the system

The AffeX system, depicted in Figure 1-a, is composed of a set of modules that form a structured processing workflow, from the initial configuration to the visualization and storage of the user's emotional data, through the integration of computer vision technologies, deep learning, and web-based visualization tools. The modules are:

- **Device configuration:** The workflow begins with the administrator logging in and accessing an interface to configure the devices required for the experimental session. This includes selecting cameras, microphones, and external sensors. This configuration defines the consent form that will later be presented to the user participant, adapting it to the enabled devices.
- **User selection:** Once the environment is configured, the participant user accesses the system, accepts the customized terms of the informed consent, and activates the monitoring process. The system employs MediaPipe for fast and accurate real-time facial detection through its BlazeFace model [7]. If a single face remains visible for three seconds, it is automatically selected; otherwise, the administrator can manually select the face to be monitored.
- **Facial anonymization:** Once the participant user's face is selected, the system uses the face_recognition³ library, based on deep neural networks (dlib), which converts faces into 128-dimensional feature vectors. This allows the user to be represented anonymously, without storing images, ensuring privacy throughout the entire process.
- **Emotion detection:** The system uses DeepFace to detect, in real time, the seven basic emotions (happiness, sadness, anger, fear, surprise, disgust, and neutrality), generating emotional weights that feed into interactive visualizations. These data are stored as lists in the repository, along with their corresponding valence, arousal, and dominance (VAD) values, derived from the emotional weights and updated at a frequency of 1 Hz.
- **Event marker:** During the session, the administrator can insert events that provide context to the collected data, improving data quality and enriching subsequent analysis.
- **Sensor data aggregation:** This module collects and synchronizes data from external sensors integrated into AffeX. When additional sensors (such as GSR) are activated, their records—stored in the proprietary repository—are retrieved and integrated with facial emotion data, enabling a multimodal emotional representation.

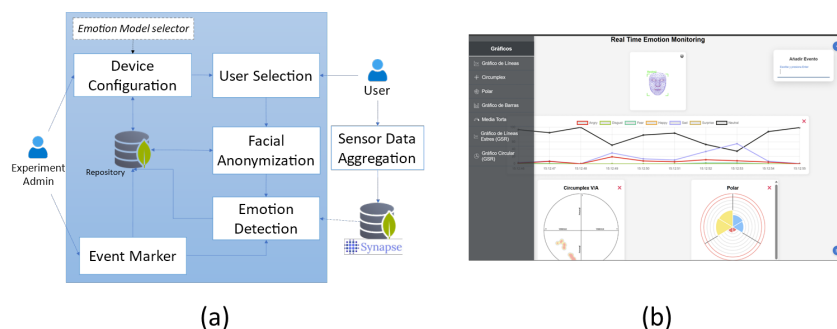


Figure 1: (a) General system architecture of AffeX. (b) UI of the application showing the graphs.

All emotional and physiological data collected by AffeX are securely stored in a proprietary repository under development, called SynapXAI, which enables data retrieval and reuse for future research. Stored data is anonymized through facial feature vectors, ensuring both user privacy and data traceability.

³face_recognition, *Documentation*, 2025. Available at: <https://face-recognition.readthedocs.io/en/latest/>

2.2. Visualization and user options

The user interface (see Figure 1-b) allows for intuitive real-time monitoring of emotions, offering visualization and configuration options tailored to the needs of either the participant or the administrator. AffeX generates dynamic visualizations using Chart.js, integrated into a customizable interface built with GridStack. Detected emotions are displayed in real time through several graphs:

- **Line chart:** Depicts temporal evolution of the seven detected emotions. (*Categorical, Temporal*).
- **Circumplex VA:** A two-dimensional heatmap that positions emotions according to valence (positive/negative) and arousal (high/low), based on Russell's model. (*Dimensional, Non-temporal*).
- **Polar VAD chart:** A radial diagram that shows average emotions in the VAD three-dimensional space, updated every 10 seconds. (*Dimensional, Temporal*).
- **Bar chart:** Summarizes emotional averages every minute. (*Categorical, Temporal*).
- **GSR graphs (if enabled):** Display stress levels from 1 to 5 in linear and circular formats, complementing the emotional analysis from electrodermal activity (EDA) sensors. (*Categorical, Temporal/Non-temporal*).

Visualizations can be reorganized, resized, or removed according to the user's preferences. It can be chosen to view either the original video or the facial mesh overlay as generated by MediaPipe.

Table 1

Performance of Models Used in AffeX (see model indicators at [8, 9])

System Module	Model	Role	Metric
Facial Detection	MediaPipe (BlazeFace)	Real-time face detection	Accuracy: 99.9% Recall: 98.4%
Facial Encoding	face_recognition (dlib)	Face anonymization and encoding	Accuracy: 99.38%
Emotion Analysis	DeepFace	Emotional analysis	Accuracy: 87.25% Recall: 80.2%

3. Privacy, ethics and limitations

The tool was designed based on principles of ethics and privacy by design, avoiding invasive practices and ensuring that the user retains control over their visualization at all times. No images or raw visual data are stored; instead, faces are converted into numerical representations that capture their features without retaining directly identifiable information or raw data. Informed consent is requested at the beginning of the usage flow and is adapted to the configuration previously established by the experiment administrator, who defines the active devices and monitoring parameters. The interface, designed for educational, clinical, or experimental settings, offers a configurable and intuitive experience that allows users to decide what information to visualize and how to interact with the charts in real time, without compromising privacy. The entire design is aligned with the principles of the General Data Protection Regulation (GDPR) and AI act, including data minimization, and transparency in processing.

This demonstrator presents some limitations derived from its technical design. First, face detection: the BlazeFace model from MediaPipe is optimized for detecting faces at short distances and loses accuracy when users are far from the camera, limiting its use in environments with variable changing positions. By design, AffeX focuses on monitoring a single user at a time, aligning with typical experimental settings in which the emotional state of one subject is the primary target of analysis. This intentional constraint reinforces data minimization principles and helps ensure privacy and clarity in data interpretation. However, we acknowledge that some research scenarios may require multi-user analysis. In such cases, future versions of AffeX could be extended to support this functionality.

Although AffeX currently integrates a single emotion recognition model, this choice was guided by its performance, as shown in Table 1. Accuracy and recall served as key selection criteria to ensure reliable real-time results. The system's modular architecture is designed to support additional models in future versions.

4. Conclusion

This work presented an interactive tool for real-time facial emotion detection and visualization, integrating technologies such as MediaPipe, face_recognition, and DeepFace, with a user-centered focus on privacy and usability. The proposed modular architecture enabled the efficient integration of computer vision, deep learning, and web visualization components. To facilitate compliance with ethical and legal principles the tool works without storing images. Additionally, the interface offers flexibility and user control, which is especially valuable in educational, clinical, and experimental contexts. Overall, we expect AffeX to facilitate experimenters to run lab experiments. Future work will focus on expanding the range of detectable emotions, integrating additional physiological sensors, and enabling the detection of multiple users simultaneously, which would be useful for typical research settings involving dyads.

Acknowledgements

All authors acknowledge the support of the Galician Ministry of Culture, Education, Professional Training and University (grants ED431G2023/04 and ED431C2022/19) and the European Regional Development Fund (ERDF). This work is also supported by the Interreg VI-A Spain-Portugal Program (POCTEP) 2021-2027 with grant 0144_TRANSFIRESAUDE_1_E, and the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033/) with grant PID2021-123152OB-C21.

Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly for Grammar and spelling checks.

References

- [1] P. Ekman, An argument for basic emotions, *Cogn. Emot.* 6 (1992) 169–200. doi:10.1080/02699939208411068.
- [2] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, X. Yang, A review of emotion recognition using physiological signals, *Sensors* 18 (2018). doi:10.3390/s18072074.
- [3] H. Jo, J. Lee, H. W. Park, M. Kim, Y. Kim, W. H. Lee, Developing an integrated dashboard to analyze multimodal data for user experience evaluation, in: *2023 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 2023, pp. 1–4.
- [4] A. Sassi, W. Jaafar, S. Cherif, J. B. Abderrazak, H. Yanikomeroğlu, Video traffic analysis for real-time emotion recognition and visualization in online learning, *IEEE Access* 11 (2023) 99376–99386.
- [5] Emodash: A dashboard supporting retrospective awareness of emotions in online learning, *International Journal of Human-Computer Studies* 139 (2020) 102411.
- [6] S. Serengil, A. Ozpinar, A benchmark of facial recognition pipelines and co-usability performances of modules, *Journal of Information Technologies* 17 (2024) 95–107. doi:10.17671/gazibtd.1399077.
- [7] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, M. Grundmann, Blazeface: Sub-millisecond neural face detection on mobile gpus, *arXiv* (2020). URL: <https://arxiv.org/pdf/1907.05047>.
- [8] B. et al., Mediapipe blazeface model card, 2021. URL: [https://storage.googleapis.com/mediapipe-assets/MediaPipe%20BlazeFace%20Model%20Card%20\(Short%20Range\).pdf](https://storage.googleapis.com/mediapipe-assets/MediaPipe%20BlazeFace%20Model%20Card%20(Short%20Range).pdf).
- [9] V. et al., Human emotion detection using deepface ai, *Eng. Proc.* 59 (2023). doi:10.3390/engproc2023059037, see Table 3 for accuracy metrics of the proposed emotion detection method.

TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados

María D. Lozano^{1,*}, Luis Orozco¹, José A. Gallud¹, Víctor M. R. Penichet¹, Gabriel Sebastián-Rivera², Ricardo Tesoriero¹, Juan E. Garrido¹, Felipe Costa-Tebar¹, Lucia Polo-Martínez¹, Belen Luna-Lorente¹, Alejandro Navarro¹, Eva Cerezo³ and Sandra Baldassarri³

¹Albacete Research Institute of Informatics, Universidad de Castilla-La Mancha, 02071 - Albacete, Spain

²School of Engineering and Technology, International University of La Rioja, 26006 Logroño (Spain)

³Departamento de Informática e Ingeniería de Sistemas, Universidad de Zaragoza, 50018 - Zaragoza (Spain)

Abstract

TecnoCRA es un Entorno Virtual de Aprendizaje gamificado desarrollado en el marco del proyecto regional de la Junta de Comunidades de Castilla-La Mancha titulado “Innovaciones Tecnológicas de Soporte a la Docencia en Colegios Rurales Agrupados (INTECRA)”, activo hasta Diciembre 2025. TecnoCRA es el resultado de varias líneas de investigación del proyecto, cubriendo especialmente el diseño e implementación de modelos inteligentes de gamificación aplicados en el campo de la enseñanza, entre otros. Con la participación activa de los agentes educativos implicados, alumnos y profesores, TecnoCRA da solución a muchos de los problemas que los Colegios Rurales Agrupados tienen, especialmente la sensación de unidad y pertenencia a un único colegio, independientemente de la población donde se resida. En TecnoCRA se han diseñado e implementado soluciones adaptadas a las necesidades de profesores y alumnos teniendo en cuenta directrices de usabilidad y bajo los principios de diseño centrado en el usuario. Además, alrededor de esta plataforma se han implementado numerosos prototipos aplicando diversos mecanismos de interacción.

Keywords

Entorno Virtual de Aprendizaje, Gamificación, Aprendizaje enriquecido con tecnología (TEL)

1. Introducción

Los Colegios Rurales Agrupados (CRA) tienen la particularidad de que se encuentran distribuidos en centros de diferentes localidades separadas geográficamente unas de otras, pero constituyendo a todos los efectos un único Colegio de Educación Infantil y Primaria. En cada centro se ubican un número reducido de alumnos de diferentes cursos, compartiendo aula y profesor. Los profesores del colegio, en algunos casos atienden un único centro de una localidad, pero otros casos viajan entre las distintas localidades para impartir las clases. Dada la distancia geográfica de los distintos centros que componen el colegio, los niños de un mismo curso no tienen oportunidad de trabajar juntos ni comunicarse entre sí. Y lo mismo pasa entre los profesores de un aula en la que tienen alumnos de distintos cursos a la vez, que resulta difícil la coordinación con otros profesores de otras localidades del mismo curso. Ante esta situación, se plantea un proyecto de investigación para dar solución a esta problemática real que plantean los CRAs y cubrir muchas de las necesidades tanto de docentes como de alumnos. En esta demostración, se presenta el Entorno Virtual de Aprendizaje denominado TecnoCRA, que ha sido desarrollado en el marco del proyecto regional de la Junta de Comunidades de Castilla-La Mancha. Este entorno se ha desarrollado para cubrir de forma personalizada las necesidades específicas de los Colegios Rurales Agrupados.

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Corresponding author.

✉ maria.lozano@uclm.es (M. D. Lozano); luis.orocho@uclm.es (L. Orozco); jose.gallud@uclm.es (J. A. Gallud); victor.penichet@uclm.es (V. M. R. Penichet); gabriel.sebastian@unir.net (G. Sebastian-Rivera); ricardo.tesoriero@uclm.es (R. Tesoriero); juanenrique.garrido@uclm.es (J. E. Garrido); felipe.costa@uclm.es (F. Costa-Tebar); lucia.polo@uclm.es (L. Polo-Martínez); belen.lunalorente@uclm.es (B. Luna-Lorente); alejandro.navarro@uclm.es (A. Navarro); ecerezo@unizar (E. Cerezo); sandra@unizar (S. Baldassarri)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Descripción de TecnoCRA

TecnoCRA se ha implementado en versión de escritorio, Web y móvil. Se ofrece una Landing Page para su descarga e información general, tal y como se muestra en la Figura 1.



Figura 1: Landing page de TecnoCRA

Una vez en el sistema, se accede al hall del colegio virtual, utilizando una metáfora de pasillos y puertas (ver Figura 2), desde donde se ofrece acceso al interior del entorno principal y a la parte de Gestión, que será diferente según si el acceso lo hace un profesor/a o un alumno/a.

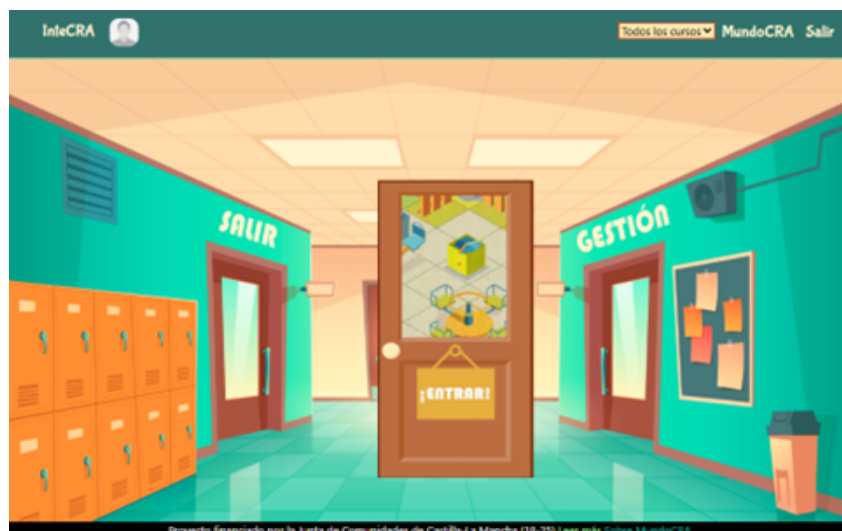


Figura 2: Acceso principal

La puerta central permite el acceso al entorno principal donde todos los usuarios conectados estarán representados por un avatar diferente, según sea profesor/a o alumno/a, y donde tienen lugar todas las interacciones entre los usuarios conectados, mediante sistemas de chat. El entorno creado recrea las instalaciones de un colegio, con zonas de recreo, zonas deportivas, aulas con pupitres y estanterías, biblioteca, aparcamiento, etc.

En la Figura 3 se muestra el mapa general del entorno principal, donde tienen lugar las actividades principales y los usuarios se mueven libremente representados por un Avatar concreto que los identifica.



Figura 3: Mapa general del entorno virtual de aprendizaje

Según el rol con el que se accede al sistema, la funcionalidad será diferente. En los siguientes apartados se resume cada una de ellas.

2.1. Funcionalidad del rol profesor

Al acceder con el rol de profesor, el usuario se muestra con un avatar identificativo y se puede mover libremente por el entorno (ver Figura 4). En las baldosas se pueden ir creando actividades de distinto tipo, según la materia y el curso al que vayan destinadas (Figura 4 – sección izquierda), que posteriormente los alumnos deberán ir realizando. La realización de las actividades se queda registrada en el sistema para el posterior seguimiento del profesor.

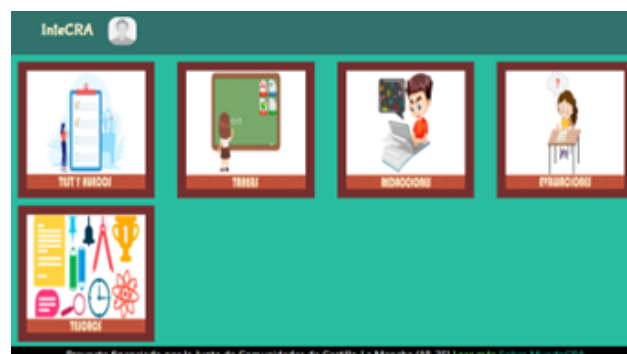


Figura 4: Acceso con rol de profesor

Al acceder a la parte de Gestión (ver Figura 2), el profesor dispone de diferentes opciones para la gestión general de su actividad, tal y como se muestra en la Figura 5a.



(a) Opciones de gestión del rol profesor

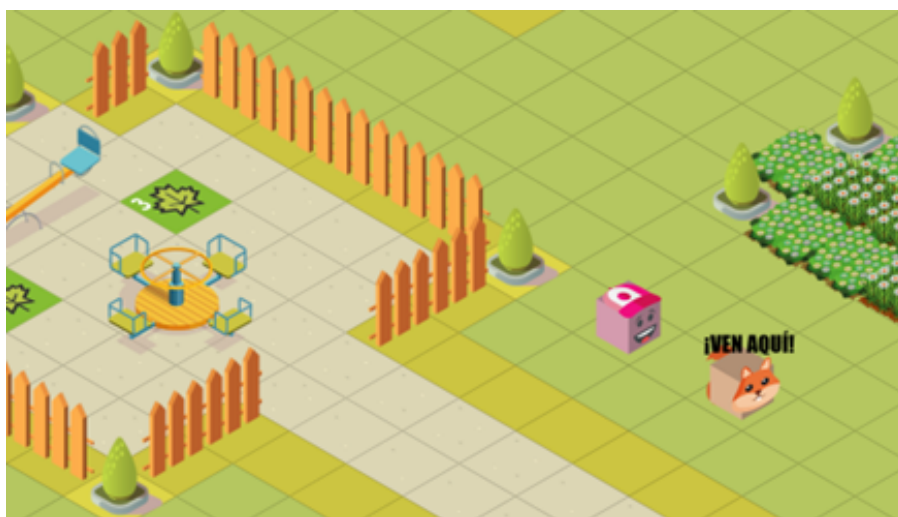


(b) Resultados del rol alumno

Figura 5: Comparativa de interfaces según el rol: profesor y alumno

2.2. Funcionalidad del rol alumno

Al acceder al Sistema con el rol de alumno (ver Figura 6), aparece un elemento de gamificación, la ardilla Pilla, que conduce al alumno a las actividades que tiene pendientes de hacer. Además, en lugar de opciones de Gestión, como al profesor, al alumno se le ofrece la consulta de Resultados de las diferentes actividades realizadas y su progreso (ver Figura 5b).

**Figura 6:** Acceso con rol de alumno

3. Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto nacional concedido por el Ministerio de Ciencia e Innovación (España) con referencia PID2022-140974OB-I00 y por el proyecto regional con referencia SBPLY-21-180501-000056, concedido por la Junta de Comunidades de Castilla-La Mancha y los Fondos Europeos de Desarrollo Regional (FEDER).

4. Declaración sobre la IA generativa

Los autores no han utilizado ninguna herramienta de inteligencia artificial generativa.

SimpLee: an artificial intelligence agent for Easy Reading in support of demographic challenge*

Margarita Cea-López¹, Alberto Jiménez-Sánchez¹, Clara Rodríguez-Marcos¹,
Javier Manso-Conde¹, Laura Grande-Pérez² and Ana-Belén Gil-González^{2,*}

¹AIR Institute, IoT Digital Innovation Hub, Av. Santiago Madrigal, 39, 37003 Salamanca, Spain

²BISITE Research Group, University of Salamanca, Edificio Multiusos I+D+i, Calle Espejo 2, 37007 Salamanca, Spain

Abstract

Spain's ageing population and rural depopulation demand innovative solutions for the digital inclusion of vulnerable groups. This study presents a comprehensive analysis of agents based on Artificial Intelligence (AI) for easy reading as a strategic tool to address these challenges. It reviews existing platforms, such as EASIER and ATECA, their integration with the national strategy against demographic challenges, and opportunities to improve the quality of life of people with reading comprehension difficulties. As a case study, it presents SimpLee, an AI-based tool developed to support easy reading within the needs associated with demographic challenges. The results demonstrate that these technologies can contribute significantly to the democratisation of access to information and construction of a more inclusive society.

Keywords

Artificial Intelligence, ease of reading, cognitive accessibility, digital inclusion, aging populations

1. Introduction

The demographic challenge facing Spain, characterised by an accelerated aging of the population and rural depopulation, requires innovative solutions to ensure the cognitive accessibility and digital inclusion of the most vulnerable groups. The Vulnerability and Employment Observatory Adecco Foundation, based on Spanish National Statistics Institute (INE) data in 2024, published that Spain has reached a historic record in population aging, with a ratio of 133 individuals aged 65 and over for every 100 under the age of 16, which represents the highest growth recorded in the entire historical series. In this context, artificial intelligence-based agents for easy reading have emerged as fundamental tools for democratising access to information and improving the quality of life of people with reading comprehension difficulties[1]. Spain faces an unprecedented demographic challenge that manifests in multiple dimensions. This demographic transformation has direct implications for the provision of public services, territorial organisations, and the cognitive accessibility needs of the population.

The SimpLee project is a functional prototype of a digital assistant with artificial intelligence (AI) to help people who have difficulty understanding complex documents, such as legal texts, regulations, or technical manuals. It focuses on simplifying information to make it clearer and more accessible, especially for older adults, people with cognitive difficulties, and citizens without training in technical subjects. To achieve this, various strategies have been investigated, ranging from traditional simplification methods (e.g. automatic summaries and language adaptation) to advanced techniques using AI models.

The system exploits the potential of large-scale language models and the explainability and reliability of classical text-simplification operations. All will be unified through an agent-tool scheme that integrates language models with classical techniques in a seamless and collaborative manner. This

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Corresponding author.

[†]These authors contributed equally.

✉ mcea@air-institute.com (M. Cea-López); ajimenez@air-institute.com (A. Jiménez-Sánchez); crodriguez@air-institute.com (C. Rodríguez-Marcos); jmanso@air-institute.com (J. Manso-Conde); lauragrande@usal.es (L. Grande-Pérez); abg@usal.es (A. Gil-González)

ORCID: 0009-0004-8071-9310 (M. Cea-López); 0009-0004-2994-8628 (L. Grande-Pérez); 0000-0001-7235-6151 (A. Gil-González)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

approach allows simplification tasks to be performed more accurately by combining the best of both worlds, to provide more understandable and accessible results for the end user.

2. Development of AI agents for Easy Reading

2.1. Easy Reading as a tool for cognitive accessibility

Easy reading[2] is a method of writing texts adapted to the needs of everyone, especially those with reading comprehension difficulties resulting from intellectual disabilities, age-related cognitive impairments, or language barriers. It is estimated that 30% of the population have reading or reading comprehension difficulties, making reading an essential tool for social inclusion easy.

The target audience for easy-to-read texts includes multiple groups, which are particularly relevant in the context of demographic challenges.

- **Elderly people:** who may suffer cognitive impairment and need more time to understand and interpret information.
- **People with intellectual disabilities:** who represent approximately 0.6
- **Migrants:** whose first language is not Spanish and face difficulties in understanding technical texts
- **People with low educational level:** who have fewer tools to deal with complex texts.

2.2. Some existing platforms

Recently, multi-agent system integration has emerged as a promising strategy in natural language processing, enabling the collaboration of specialised agents to tackle complex tasks in a decentralised and efficient manner. In this context, studies have introduced a novel framework based on multiagent systems and large-scale language models that simulate the collaboration of a team of human experts to simplify documents [3, 4, 5]. This approach explores different document communication and reconstruction strategies, achieving significant improvements in both simplification and coherence of the resulting texts based on automatic and human evaluations. The EASIER platform [6], developed by the Human Language and Accessibility Technologies (HULAT) group at Carlos III University of Madrid, represents a significant advance in the application of artificial intelligence for cognitive accessibility. This tool provides lexical simplification of Spanish texts; detects complex words; and offers synonyms, definitions, and pictograms to improve comprehension.

EASIER uses natural language processing techniques and has been validated with the participation of people with intellectual disabilities and the elderly, thus guaranteeing its effectiveness in real-world contexts. The platform collaborates with Plena Inclusión Madrid to provide easy-to-read definitions and uses resources for pictograms.

ATECA (Adaptación de TExtos en CAstellano) is a web application based on artificial intelligence developed by the Polytechnic University of Madrid in collaboration with the Spanish Centre for Cognitive Accessibility (CEACOG). [7]. This application follows the easy-to-read UNE 153101:2018 EX Standard [8] and automates the first two activities of the adaptation process: document analysis and transformation suggestions.

3. SimpLee. The proposal

SimpLee is a modular tool based on AI designed to simplify documents using large language models (LLMs), prompt engineering techniques, multi-agent systems, and Retrieval Augmented Generation (RAG). The main objective is to facilitate the understanding of complex texts, without sacrificing normative accuracy, for audiences with different levels of functional literacy, thus contributing to a more accessible and understandable public administration

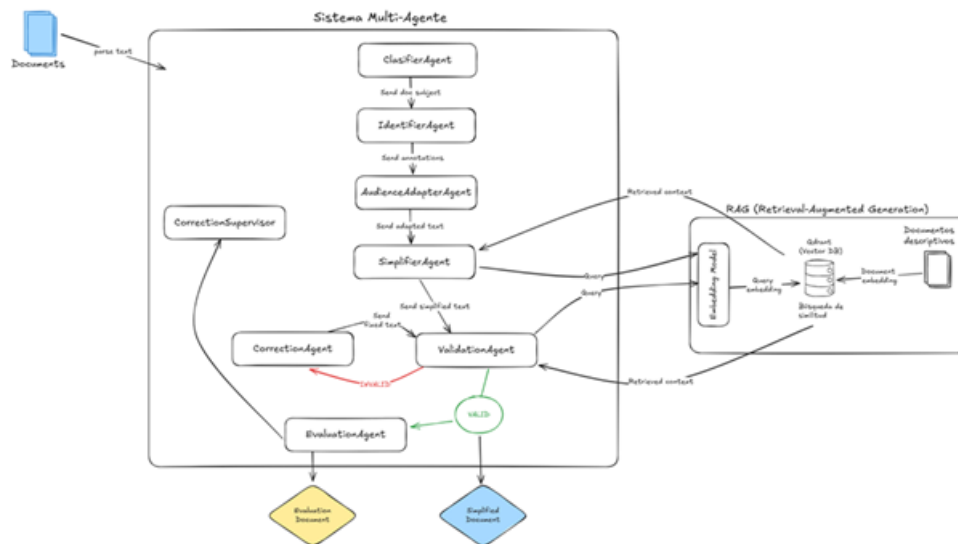


Figure 1: Simplee architecture

Aquí le presento la evaluación de la calidad del texto simplificado respecto al original según los 6 criterios:

Criterio	Puntuación /5	Comentario
Claridad del Lenguaje	4/5	El lenguaje es claro y fácil de entender, pero en algunos puntos se simplifica demasiado. Por ejemplo, la parte "1. Que por medio del presente formulo DENUNCIA." se reduce a "Qué se denuncia".
Estructura y Coherencia	4/5	La estructura es similar al original, pero se han eliminado algunos detalles importantes como la mención a la autoridad competente. La coherencia generalmente está bien, aunque en algunos puntos se pierde la lógica del texto original.
Adecuación al Público	5/5	El texto simplificado está claramente dirigido a un público no especializado y utiliza un lenguaje accesible.
Formato Visual	4/5	El formato visual es simple y fácil de leer, pero se ha perdido la estructura original en algunos puntos. Por ejemplo, la numeración de los puntos "1.", "II." y "III." no está presente en el texto simplificado.
Fidelidad al Original	3/5	El texto simplificado no sigue fielmente el original en muchos puntos. Se han eliminado detalles importantes y se ha simplificado demasiado la estructura y el lenguaje.
Comprensibilidad General	4.5/5	En general, el texto simplificado es fácil de entender y comprender, pero en algunos puntos se pierde la claridad y la coherencia del original.

Puntuación total: 26/30

Resumen final:
El texto simplificado es una buena versión del original, pero no sigue fielmente el formato y la estructura del mismo. Se han eliminado detalles importantes y se ha simplificado demasiado en algunos puntos. Sin embargo, el lenguaje es claro y fácil de entender, y está adecuadamente dirigido a un público no especializado. En general, es una buena versión para fines prácticos, pero no es una traducción literal del original.

Figure 2: Example of evaluation output

3.1. Simplee technical functionalities

The architecture described in this paper, see figure 1, represents a preliminary version corresponding to the initial phase of the project. The current definition responds to the functional and technical requirements identified at an early stage of development; therefore, it is assumed to be a base subject to evolution. It is foreseeable that, as the implementation, empirical validation, and fine-tuning of the system components deepen, the architecture will undergo substantial modifications to optimise its performance, scalability, and suitability for use. The system is coordinated by LangGraph, which orchestrates a flow of eight agents, this separation of responsibilities and agent roles allows for more accurate error handling, as well as the ability to improve the accuracy of that role through prompt engineering techniques. The multiagent system is composed of ClassifierAgent, IdentifierAgent, AudienceAdapterAgent, SimplifierAgent, ValidationAgent, CorrectionAgent, and EvaluationAgent.

This design allows for modular and traceable surgical intervention, where each agent acts only on the aspects they master, promoting a higher final quality and more precise and specific validation. Integration with RAG enables context retrieval with semantic embeddings to improve the accuracy and coherence of generated content [9]. During simplification and validation, agents perform semantic queries on a vector database (Qdrant) containing official descriptions, regulations, and glossaries. This context was dynamically injected into prompts to ensure that the answers were normatively accurate.

3.1.1. Evaluation and validation

The system includes an automatic validation loop in which simplified texts are evaluated based on their lexical and syntactic clarity, conceptual fidelity, and appropriateness to the selected level, among others, as shown in figure 2. If problems are detected, the text is forwarded to the correction agent, maintaining the process until a VALID status is reached or if a maximum number of iterations are carried out. This

avoids infinite loops and allows us to determine the metrics that do not comply with the standards to improve it. At the end, a report is generated that includes: (1) Scores per dimension assessed.(2) Comments generated by stakeholders, and (3) visual comparisons between the original and simplified versions.

4. Conclusions

SimpLee is an AI-driven tool that leverages LLM, multiagent systems, and retrieval-augmented generation to simplify complex documents while maintaining regulatory accuracy. By enhancing text accessibility for users with varying literacy levels, SimpLee supports a more inclusive public administration. Such AI solutions are essential for addressing Spain’s demographic challenges, particularly for promoting accessibility and digital inclusion for vulnerable groups.

5. Acknowledgments

This research was supported by Grant TSI-100933-2023-1 funded by the “University-Enterprise Chair Call (Cátedras ENIA 2022)” of the “Spanish Ministry for the Digital Processing and Civil Service” and by the “European Union NextGenerationEU/PRTR”.

Declaration on Generative AI

During the preparation of this work, the authors used Paperpal in performance grammar and spelling checks. After using these tool the authors reviewed and edited the content as needed and takes full responsibility for the publication’s content

References

- [1] F. Adecco, El envejecimiento se dispara en españa y alcanza el récord de 142: se contabilizan 142 personas mayores de 64 años por cada 100 menores de 16, 2024. URL: <https://fundacionadecco.org/notas-de-prensa/el-envejecimiento-se-dispara-en-espana>, nota de prensa.
- [2] R. Jonuzaj, Cognitive accessibility in hci interface design: The easy reading tool (2024). Submitted manuscript.
- [3] G. Karadzhov, I. Barr, Y. Zhang, X. Li, D. Khashabi, Collaborative document simplification using multi-agent systems, in: Proceedings of the 30th International Conference on Computational Linguistics (COLING 2025), Association for Computational Linguistics, 2025, pp. 716–729. URL: <https://aclanthology.org/2025.coling-main.60.pdf>.
- [4] P. Martínez, A. Ramos, L. Moreno, Exploring large language models to generate easy to read content, *Frontiers in Computer Science* 6 (2024) 1394705.
- [5] G. Wu, J. Qian, S. Castelo Quispe, S. Chen, J. Rulff, C. Silva, ARTiST: Automated text simplification for task guidance in augmented reality, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, ACM, 2024, pp. 1–24.
- [6] R. Alarcón, L. Moreno, P. Martínez, J. A. Macías, Easier system: Evaluating a spanish lexical simplification proposal with people with cognitive impairments, *International Journal of Human-Computer Interaction* 40 (2024) 1195–1209.
- [7] M. C. Suárez de Figueroa Baonza, P. M. Blanco, A. Blanco, I. Diab Lozano, M. García-Agudo, A. Muñoz, C. Torres, Ateca: una aplicación web para agilizar las adaptaciones a lectura fácil, *Siglo Cero* 56 (2025) 115–137.
- [8] UNE 153101:2018 EX – Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos, Norma técnica experimental 153101:2018 EX, AENOR / UNE, 2018.
- [9] C. Lyu, G. Pergola, Society of medical simplifiers, 2024. [arXiv:2410.09631](https://arxiv.org/abs/2410.09631), arXiv preprint.

Coloquio doctoral



Virtual interfaces and augmented reality for orthopedic surgery^{*}

Ánxela Pérez Costa^{1,*}, Nieves Pedreira Souto²

¹Laboratory of Mechanical Engineering, Centro de Investigación en Tecnologías Navales e Industriales (CITENI), Campus Industrial de Ferrol, Universidade da Coruña, Ferrol, Spain

²Department of Computer Science and Information Technologies, Faculty of Computer Science, CITIC-Research Center of Information and Communication Technologies, Universidade da Coruña, Biomedical Research Institute of A Coruña (INIBIC), A Coruña, Spain

Abstract

The evolution of User Interaction and Experience (UX) has advanced significantly with emerging technologies like virtual, augmented, and mixed reality. In healthcare, Augmented Reality (AR) has shown transformative potential, particularly in surgical settings, by enhancing precision, spatial awareness, and decision-making. This thesis explores interaction alternatives—such as touch, voice, and gestures—for AR devices in orthopedic surgery, focusing on intuitive and efficient interfaces that accommodate both technological and contextual constraints. Special attention is given to AR Head-Mounted Displays (HMDs) used in Total Knee Arthroplasty (TKA), which provide real-time data in the surgeon's field of view. While offering clear benefits, these systems introduce human-computer interaction challenges that must be addressed to ensure usability, reduce cognitive load, and maintain surgical safety and effectiveness.

Keywords

Augmented reality (AR), head-mounted displays (HMD), total knee arthroplasty (TKA), interaction methods, usability evaluations

1. Introduction

Throughout history, the evolution of User Interaction and Experience (UX) has been strongly influenced by technological advancements. In recent times, the field of UX in Virtual Interfaces has experienced rapid growth and development, driven by technologies such as virtual reality (VR), augmented reality (AR), and mixed reality (MR) [1, 2]. AR has found various applications in the healthcare sector, recently being used to assist surgeons in visualizing real-time data and medical images during surgical procedures [3]. This has improved accuracy and decision-making, highlighting the transformative potential of immersive technologies in enhancing healthcare. These technologies not only enhance precision and efficiency but also improve the experience for both healthcare professionals and patients. The objective of this thesis is to explore alternatives for interacting (touch, voice, gestures, etc.) with virtual reality devices in orthopedic surgery, aiming to design intuitive and effective user interfaces. Both the technological limitations of current market products and the restrictions associated with specific usage conditions will be considered. A solid understanding of the factors influencing the user experience in immersive environments will enable the development of effective interfaces, adaptation to this new technology, and prevention of errors that could have serious consequences, especially in the medical sector. The use of Augmented Reality (AR) Head-Mounted Displays (HMDs) in orthopedic surgery, more specifically in Total Knee Arthroplasty (TKA) is reshaping the surgical landscape. These devices provide surgeons with immediate, three-dimensional visual data directly within their line of sight, enhancing both precision and spatial awareness during procedures [4]. By projecting essential information onto the surgical field, AR HMDs support real-time decision-making without requiring surgeons to glance away at external screens. Despite these benefits, incorporating AR HMDs into

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

^{*}Corresponding author.

✉ anxela.perez@udc.es (Á. P. Costa); nieves.pedreira@udc.es (N. P. Souto)

🆔 0009-0004-6156-4763 (Á. P. Costa); 0000-0002-8122-0886 (N. P. Souto)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the surgical workflow introduces a new set of challenges—particularly in terms of human-computer interaction (HCI). Surgeons must engage with digital content under high-pressure conditions where intuitive, efficient, and user-friendly interfaces are critical [5]. If interaction methods are poorly designed or too complex, they risk increasing cognitive load, interrupting procedural flow, or contributing to mistakes [6, 7, 8]. Moreover, the surgical context places specific demands on interaction techniques, requiring them to be valid, safe, and ergonomically suitable [9]. As such, optimizing the interface between the surgeon and AR systems is key to leveraging this technology effectively, while safeguarding performance and patient outcomes.

2. Background

In our research, we began with a commercially available computer-assisted orthopedic surgery system designed for use during total knee arthroplasty (TKA). This system utilizes smart glasses that allow the surgeon to view a virtual model superimposed on the patient's anatomy, enabling accurate alignment of surgical instruments in accordance with the preoperative plan. The performance evaluations provided by the manufacturer served as a starting point for our investigation, offering a benchmark to assess current capabilities and identify opportunities for enhancement. This study focuses on how surgeons interact with an AR-based system during TKA, aiming to uncover both the benefits and the challenges of integrating AR into routine surgical workflows.

3. Objectives

The primary objectives of these studies are to investigate and identify effective interaction methods for user interfaces and to assess their suitability for integration into orthopedic surgery applications. This evaluation focuses on key factors such as interaction time, user comfort, and perceived workload. Additionally, the study aims to improve the recognition and integration of surgical tools that interact with the AR device, enhancing the overall user experience while ensuring accurate data input and system reliability.

4. Methodology

4.1. Definition of a research plan

To address surgeons' concerns regarding the integration of Augmented Reality (AR) head-mounted displays (HMDs) in Total Knee Arthroplasty (TKA), this research begins by leveraging both quantitative and qualitative data provided by a commercial AR solution for orthopedic surgery. This initial phase aims to understand user needs, identify friction points, and uncover opportunities to improve the AR-assisted surgical workflow. Insights derived from this analysis form the foundation for a broader research strategy, supported by an extensive literature review. The overarching goal is to optimize the AR experience for surgeons by enhancing system usability, improving interface design, and reducing the likelihood of user error in critical medical procedures.

4.1.1. Evaluation of interaction methods

The second phase focuses on evaluating various interaction methods for AR HMDs, considering the specific demands of the surgical environment. Five interaction techniques were tested: head orientation, holographic touch, hand gestures, gaze fixation, and a gaze-gesture bimodal approach. Each method was assessed through both objective metrics—task completion time and error rate—and subjective measures such as perceived workload and user comfort. While prior studies have explored AR interaction methods, few have systematically tested these techniques under conditions that simulate real surgical settings. This research addresses that gap by reproducing realistic constraints, allowing for a more accurate

evaluation of usability and ergonomics. Adapting existing human-computer interaction evaluation frameworks, the study provides a comprehensive comparison to determine which interaction methods offer the best balance of efficiency, intuitiveness, and comfort for surgical applications.

4.1.2. Enhancement of surgical tool recognition and integration (future work)

A recurring issue observed in the studies was the inconsistent recognition of surgical tools by the AR system, often influenced by the user's height or positioning. Given the need for seamless and rapid interaction during surgery, enhancing the accuracy and reliability of tool recognition is essential. This component of the research will explore ways to refine recognition algorithms and interaction design to reduce variability, improve ergonomics, and minimize cognitive load. Effective tool integration will ultimately contribute to a smoother surgical workflow and reduce the potential for interaction-related errors.

4.1.3. Computer simulation of pathologies and treatment prediction (future work)

The final component of this methodology examines the integration of advanced computer simulations into the AR surgical interface. These simulations are capable of modeling specific pathologies and predicting treatment outcomes, offering valuable insights to support intraoperative decision-making. By embedding these predictive tools into the AR environment, surgeons can access enhanced clinical intelligence in real time. This study will focus particularly on the usability and interactivity of these simulations when presented through AR HMDs, ensuring they are both informative and practical within the fast-paced surgical context.

5. Results

5.1. Definition of a research plan

The resulting research plan was structured into four sequential phases. The first phase, which spanned 12 months, was grounded in a series of evaluations of a commercial AR interface for Total Knee Arthroplasty (TKA), previously conducted and provided by an industry partner. These evaluations were instrumental in identifying usability challenges and areas for potential enhancement, serving as the foundation for shaping the subsequent research direction. This phase also included a systematic literature review with two primary objectives: (1) to identify key factors affecting both surgical performance in orthopedic procedures and the usability of AR systems, thereby establishing a framework for evaluation; and (2) to determine feasible, hands-free interaction methods compatible with AR headsets, ensuring suitability within sterile surgical environments. The second phase, completed over a 10-month period, involved a comprehensive assessment of the selected interaction techniques. This phase was structured into three sub-stages: an initial feasibility analysis of the proposed methods, iterative rapid testing to refine interaction design based on early user feedback, and a final validation stage involving orthopedic surgeons to ensure clinical applicability and realism in a simulated surgical setting. The third phase, currently in preparation, will focus on improving the recognition and integration of surgical tools that serve as biometric inputs to the AR system. Various tool geometries and recognition strategies will be explored to enhance ergonomic visibility, increase recognition robustness across different angles, and maintain or improve the fidelity of the transmitted data. Finally, the fourth phase will involve the integration of advanced biomechanical simulations into the AR interface. These simulations will be visualized through the headset to support surgical planning and intraoperative decision-making by predicting treatment outcomes based on patient-specific anatomical models.

5.2. Evaluation of interaction methods

Statistical analysis using repeated-measures ANOVA ($\alpha = 0.05$) revealed significant differences in task completion time across the five interaction methods ($p < 0.001$). Gaze fixation yielded the fastest

completion times but was associated with a small number of errors. In contrast, gestures and head orientation selection (HOS) were completely error-free, with gestures requiring slightly more time and HOS achieving mid-range performance in terms of speed. Holographic touch was the slowest method and showed moderate error rates, while the bimodal gaze + gesture method offered a balanced trade-off between speed and usability, albeit with a higher occurrence of near-errors.

Perceived workload, measured using the NASA-TLX scale, also varied notably between methods. HOS and gestures registered the lowest overall workload scores (means = 16.08 and 16.50, respectively), despite HOS showing elevated physical demand. In contrast, the gaze + gesture combination recorded the highest workload (mean = 26.25), driven primarily by increased mental demand and frustration levels. Subjective preference rankings, validated through high internal consistency (Cronbach's α = 0.903), showed gaze fixation as the most favored method, with the highest mean scores for both preference (8.3) and comfort (8.0). It was closely followed by HOS (8.2 preference, 6.7 comfort) and gestures (7.7 preference, 7.3 comfort). Holographic touch ranked lowest in both categories. When participants selected their top three interaction methods, 90% included gaze fixation and gestures, with gaze fixation being most frequently selected as both the preferred (30%) and most comfortable method (40%), highlighting its strong balance between efficiency and user comfort.

Acknowledgments

Ánxela Pérez Costa would like to acknowledge the support of the Galician Government and the Ferrol Industrial Campus by means of her predoctoral research contract 2023/CP/209.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] D. Casterlvecchi, Low-cost headsets boost virtual reality's lab appeal, *Nature* 533 (2016) 153–154. doi:10.1038/nature17671.
- [2] C. Ebert, Looking into the Future, Technical Report, IEEE Computer Society, 2015.
- [3] R. Q. Mao, L. Lan, J. Kay, R. Lohre, O. R. Ayeni, D. P. Goel, D. de SA, Immersive virtual reality for surgical training: A systematic review, 2021. doi:10.1016/j.jss.2021.06.045.
- [4] S. Pokhrel, A. Alsadoon, P. W. Prasad, M. Paul, A novel augmented reality (ar) scheme for knee replacement surgery by considering cutting error accuracy, *International Journal of Medical Robotics and Computer Assisted Surgery* 15 (2019). doi:10.1002/rcs.1958.
- [5] S. P. Canton, C. N. Austin, F. Steuer, S. Dadi, N. Sharma, N. M. Kass, D. Fogg, E. Clayton, O. Cunningham, D. Scott, D. LaBaze, E. G. Andrews, J. T. Biehl, M. C. V. Hogan, Feasibility and usability of augmented reality technology in the orthopaedic operating room, *Current Reviews in Musculoskeletal Medicine* 17 (2024) 117–128. doi:10.1007/s12178-024-09888-w.
- [6] J. Buchner, K. Buntins, M. Kerres, The impact of augmented reality on cognitive load and performance: A systematic review, 2022. doi:10.1111/jcal.12617.
- [7] T. Schneider, T. Cetin, S. Uppenkamp, D. Weyhe, T. Muender, A. V. Reinschluessel, D. Salzmann, V. Uslar, Measuring bound attention during complex liver surgery planning: Feasibility study, *JMIR Formative Research* 9 (2025). doi:10.2196/62740.
- [8] E. J. Brown, K. Fujimoto, B. Blumenkopf, A. S. Kim, K. L. Kontson, H. L. Benz, Usability assessments for augmented reality head-mounted displays in open surgery and interventional procedures: A systematic review, 2023. doi:10.3390/mti7050049.
- [9] M. Birlo, P. J. E. Edwards, M. Clarkson, D. Stoyanov, Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review (2022). URL: <http://arxiv.org/abs/2202.04141><http://dx.doi.org/10.1016/j.media.2022.102361>. doi:10.1016/j.media.2022.102361.

Diseño de un modelo de empatía en diálogos de agentes sociales interactivos para la mejora de la aceptación tecnológica

Alberto Hitos-García^{1,*}, Francisco Luis Gutiérrrez-Vela^{1,†} y Patricia Paderewski-Rodríguez^{1,†}

¹Universidad de Granada, Dpto. de Lenguajes y Sistemas Informáticos - ETSIT, Granada, España

Resumen

La empatía es un componente clave en las interacciones humanas que podría mejorar significativamente la experiencia y aceptación de los agentes sociales interactivos (SIA). Esta tesis propone el diseño de un modelo de empatía aplicado a los diálogos de los SIA, con el objetivo de fomentar su aceptación tecnológica y potenciar su impacto en la integración social. A partir de una revisión del estado del arte, se definirá un modelo teórico de empatía y se desarrollará una guía de diseño de diálogos empáticos para agentes conversacionales. Además, se diseñarán herramientas específicas de evaluación y se crearán experiencias lúdicas sociales dirigidas especialmente a colectivos vulnerables, como los adultos mayores, en las que los SIA actúan como facilitadores de la interacción.

Palabras clave

empatía, diálogo, agentes sociales interactivos, aceptación tecnológica

1. Introducción y antecedentes

La **empatía** es, según Davis [1], una respuesta afectiva que surge cuando un individuo percibe e imita los estados emocionales de otra persona, facilitando así una comprensión más profunda de sus experiencias. Este proceso involucra un esfuerzo activo por parte del observador para entrar en la perspectiva del otro, utilizando capacidades cognitivas que permiten no solo compartir las emociones, sino también comprender los pensamientos y motivaciones subyacentes.

Con el auge de los **agentes sociales interactivos (SIA – Socially Interactive Agent)** como los asistentes virtuales durante la última década, junto con el reciente avance de la inteligencia artificial generativa basada en modelos de lenguaje de gran escala (LLM – Large Language Models) y la progresiva incorporación de robots sociales en el mercado, se abre un nuevo debate sobre si estos sistemas deben exhibir una empatía artificial que podría enriquecer las interacciones con los usuarios y fomentar una mayor aceptación social de estas tecnologías emergentes.

Sin embargo, implementar la **empatía humano-agente** no es una tarea fácil ya que es una cualidad puramente humana. No obstante, investigadores como Park y Whang [2] han identificado, mediante la revisión de diversos estudios, una serie de pasos generales que permiten a un agente simular la empatía:

1. **Reconocimiento de señales emocionales:** el agente detecta señales empáticas del humano que pueden ser tanto verbales como no verbales. Por ejemplo, la expresión facial, el tono de voz o el lenguaje corporal pueden ayudar a identificar el estado de ánimo de la persona.
2. **Procesamiento interno:** una vez captadas una o varias señales empáticas, el agente comienza un procesamiento interno utilizando mecanismos tanto afectivos (imitación motora) como cognitivos (condicionamiento clásico, toma de perspectiva, etc.).
3. **Generación de la respuesta empática:** al finalizar el procesamiento interno, se genera una respuesta empática en el agente. Esta respuesta puede ser tanto paralela (es decir, reflejando la

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Doctorando.

†Directores de tesis.

✉ albertohitos@ugr.es (A. Hitos-García); fgutierr@ugr.es (F. L. Gutiérrrez-Vela); patricia@ugr.es (P. Paderewski-Rodríguez)

🆔 0009-0009-3136-5022 (A. Hitos-García); 0000-0001-6629-7597 (F. L. Gutiérrrez-Vela); 0000-0001-6626-9633

(P. Paderewski-Rodríguez)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

misma emoción que el humano) como reactiva (respondiendo de manera diferente pero apropiada, con simpatía o preocupación).

4. **Expresión de la respuesta empática:** por último, el agente decide si debe comunicar la respuesta empática generada en el paso anterior, modulándola en función de diversos parámetros como las características del humano (edad, relación previa, situación) o el contexto (un entorno de trabajo, una conversación informal, etc.). En el caso de comunicar la respuesta empática, el agente lo podrá hacer a través de múltiples canales mediante expresiones faciales, gestos, palabras o incluso acciones físicas.

Para **medir la empatía percibida** en interacciones humano-agente existen diversos instrumentos y escalas que adaptan métricas previamente utilizadas en el estudio de la empatía en relaciones interpersonales. Uno de los primeros es el propuesto por Tisseron et al. en 2015, denominado “Questionnaire on Empathy and Auto-empathy” (QEAE) [3]. El cuestionario está compuesto por 16 ítems para evaluar 4 dimensiones: Auto-empatía, Empatía Directa, Empatía Recíproca y Empatía Intersubjetiva; en 3 tipos de interacciones: entre humanos, entre humanos y robots, y entre humanos y avatares.

Por otro lado, al ser los SIA una tecnología novedosa, es muy importante evaluar la predisposición de los usuarios a utilizarla en su vida diaria. Para **medir el nivel de aceptación tecnológica** existen varios modelos, siendo el más utilizado globalmente el “Technology Acceptance Model” (TAM) [4].

En la actualidad, están surgiendo multitud de estudios para evaluar el potencial de los SIA como una herramienta para **reducir el aislamiento social** y la soledad no deseada, siendo los primeros resultados muy prometedores [5]. Sin embargo, todos los estudios coinciden en que es necesaria una mayor investigación en este campo por ser una tecnología novedosa, con mucho potencial pero con sus riesgos e implicaciones éticas.

2. Hipótesis y justificación

La tesis partirá de dos hipótesis que servirán de guía durante todo el desarrollo y que permitirán explorar el potencial de la empatía aplicada a los SIA, tanto para mejorar su aceptación tecnológica como para promover las relaciones interpersonales.

Hipótesis 1: La empatía aplicada en los diálogos de los SIA mejora su aceptación tecnológica y la efectividad de la interacción humano-agente.

Para validar esta hipótesis será necesario construir un **modelo teórico** para definir las dimensiones, parámetros y variables de la empatía. Este modelo servirá como base para desarrollar una **guía de diseño de diálogos** que permita a los desarrolladores construir agentes empáticos de manera simple, eficaz y ética.

A partir de esta guía, se diseñarán **diálogos empáticos** que serán evaluados con usuarios reales. El objetivo será analizar si los usuarios son capaces de percibir la empatía en las interacciones y si los agentes que utilizan estos diálogos son mejor aceptados que aquellos que no exhiben comportamientos empáticos.

Si bien actualmente existen herramientas para medir tanto la empatía percibida [3] como la aceptación tecnológica [4] en sistemas interactivos, la precisión de estas herramientas depende en gran medida del contexto de uso y del tipo de agente utilizado. Por ello, es importante desarrollar **herramientas propias de evaluación** que estén adaptadas al contexto del proyecto y que deberán ser validadas tanto por expertos como empíricamente.

Hipótesis 2: El uso de los SIA empáticos en entornos sociales mejora la integración social y reduce la soledad no deseada.

La validación de esta hipótesis implicará desarrollar y evaluar **experiencias lúdicas sociales** que promuevan la interacción interpersonal en un entorno social, utilizando un SIA empático como guía o elemento dinamizador de la experiencia. Durante estas experiencias, el SIA debe garantizar que

todos los usuarios participen de manera equitativa para que nadie se sienta excluido y debe fomentar la creación de vínculos interpersonales mediante el compañerismo (colaboración, trabajo en equipo, etc.), el conocerse mejor (compartir gustos personales, anécdotas, etc.) o la reducción de la ansiedad social (estructurar las interacciones para que las personas más tímidas se sientan cómodas al participar, uso del refuerzo positivo, uso del humor para relajar el ambiente, etc.).

Sin embargo, validar esta hipótesis es una tarea compleja porque la mejora en la integración social no depende únicamente de la presencia del SIA, sino también de factores individuales y contextuales, como la predisposición de los participantes a socializar, la calidad de las interacciones y el entorno en el que se desarrollan. Por ello, será fundamental diseñar un **modelo de evaluación propio** para analizar el impacto del SIA en la creación y fortalecimiento de vínculos interpersonales, combinando métodos cuantitativos (encuestas de percepción social, análisis de redes de interacción, etc.) y cualitativos (observación de dinámicas grupales, entrevistas con los participantes, etc.).

3. Objetivos

3.1. Objetivo general

Analizar el impacto de la empatía en los diálogos durante la interacción humano-agente para mejorar la aceptación tecnológica de los SIA y explorar su potencial para fomentar la integración social y reducir la soledad no deseada en entornos comunitarios, a través del diseño, desarrollo y evaluación de experiencias lúdicas sociales.

3.2. Objetivos específicos

1. Desarrollar un **modelo teórico de empatía** para agentes conversacionales, definiendo las dimensiones, parámetros y variables clave para su implementación en los SIA.
En primer lugar, se hará una revisión de la literatura existente con el fin de conocer el estado del arte de la empatía aplicada en los diferentes tipos de SIA, así como los modelos computacionales utilizados actualmente. Basado en la revisión, se desarrollará un modelo teórico que deberá incluir también propiedades importantes de los agentes interactivos como son la proactividad, la persuasión, la socialización y la confianza, además de otras transversales como son la ética, el género, la personalización, etc.
2. Crear una **guía de diseño de diálogos empáticos** para desarrolladores, que incluya principios éticos y directrices prácticas para la implementación de respuestas empáticas en los SIA.
Esta guía incluirá también directrices y recomendaciones a la hora de usar LLM para la generación inteligente de diálogos. Para ello se recomendarán *prompt* específicos para los diferentes aspectos del diálogo como son la detección y expresión de emociones, generación de diálogos basados en la personalidad, análisis del contexto, etc.
3. Desarrollar **herramientas de evaluación específicas** para medir la efectividad de la empatía en los diálogos de los SIA, tanto para mejorar su aceptación tecnológica como para fomentar las relaciones interpersonales de los usuarios.
Para ello, se emplearán principalmente cuestionarios que permitan recopilar información tanto cualitativa como cuantitativa.
4. Diseñar y desarrollar **experiencias lúdicas sociales** en las que un SIA empático actúe como facilitador de la interacción interpersonal en entornos sociales.
Estas experiencias se apoyarán en el uso de técnicas de codiseño aplicadas a colectivos específicos como los adultos mayores, con el objetivo de fomentar comportamientos prosociales en los SIA.
5. **Validar el modelo teórico y las herramientas de evaluación** mediante la revisión con expertos y la realización de pruebas empíricas con usuarios reales.
Esto incluirá análisis cualitativos y cuantitativos para ajustar los modelos según las respuestas y el *feedback* recibido.

4. Resultados

Actualmente, la tesis se encuentra en su primer año y está en una fase inicial de revisión de la literatura y diseño de pequeñas experiencias preliminares con un grupo reducido de usuarios. Sin embargo, ya se han presentado algunos artículos en diversos congresos:

1. **Enhancing Elderly Care with Empathetic Technologies: Insights from Human-Computer Interaction.**
11 JIHC 25 - 11 Jornadas Iberoamericanas de Interacción Humano - Computadora 2025. Zacatecas, México.
2. **Empathy and Human-Computer Interaction in Social Interactive Agents.**
CISTI 25 - 20th Iberian Conference on Information Systems and Technologies. Lisboa, Portugal.
3. **Comparison of instruments for assessing perceived empathy in Human-Agent Interactions.**
INTERACCIÓN 25 - XXV Congreso Internacional de Interacción Persona-Ordenador. Valladolid, España.
4. **Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales.**
INTERACCIÓN 25 - XXV Congreso Internacional de Interacción Persona-Ordenador. Valladolid, España.

5. Agradecimientos

Esta tesis está cofinanciada por el proyecto PLEISAR-Social, Ref. PID2022-136779OB-C33, con fondos del Ministerio de Ciencia e Innovación del Gobierno de España (MCIN/AEI/ 10.13039/501100011033, UE) y por el proyecto C-ING-179-UGR23 financiado por la Consejería de Universidades, Investigación e Innovación y por el Programa FEDER Andalucía 2021-2027.

Declaración sobre IA generativa

Los autores declaran que no han usado ninguna herramienta de IA generativa.

Referencias

- [1] M. H. Davis, *Empathy: A social psychological approach*, 1st ed., Routledge, New York, NY, USA, 1996.
- [2] S. Park, M. Whang, *Empathy in human–robot interaction: Designing for social robots*, *International journal of environmental research and public health* 19 (2022) 1889. doi:10.3390/ijerph19031889.
- [3] S. Tisseron, F. Tordo, R. Baddoura, *Testing empathy with robots: a model in four dimensions and sixteen items*, *International Journal of Social Robotics* 7 (2015) 97–102. doi:10.1007/s12369-014-0268-5.
- [4] F. D. Davis, *Perceived usefulness, perceived ease of use, and user acceptance of information technology*, *MIS quarterly* 13 (1989) 319–340. doi:10.2307/249008.
- [5] R. A. Marzali, C. Franceschetti, A. Dinculescu, A. Nistorescu, D. M. Kristály, A. A. Moşoi, R. Broekx, M. Marin, C. Vizitiu, S.-A. Moraru, et al., *Reducing loneliness and social isolation of older adults through voice assistants: Literature review and bibliometric analysis*, *Journal of medical Internet research* 26 (2024). doi:10.2196/50534.

Presentaciones de grupos o proyectos de investigación



EXPLainable Artificial INtelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME)*

Cristina Manresa-Yee^{1,*}, Antoni Jaume-i-Capó¹, Adel Ghazel², Isabelle Hupont³, Luis A. Leiva⁴ and I. Scott MacKenzie⁵

¹Universitat de les Illes Balears, Crta. Valldemossa km. 7.5, 07122 Palma, Spain

²Univ Rouen Normandie, ESIGELEC, Normandie Univ, IRSEEM UR 4353, F-76000 Rouen, France

³Spanish Ministry of Science, Innovation and Universities, Spain

⁴Université du Luxembourg, Maison du Nombre 6, avenue de la Fonte, L-4364 ESCH, Luxembourg

⁵York University, 4700 Keele St, Toronto, M3J 1P3, Canada

Abstract

This work describes the research directions of the four-year national research project EXPLAINME, *EXPLainable Artificial INtelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI* funded by MICIU/AEI and by ERDF/EU. The project aims at designing and developing methods and metrics for human-centered explainability in AI systems, considering ethical aspects. The multidisciplinary research team brings together expertise from human-computer interaction, artificial intelligence, law, ethics, and psychology to address the technical, legal, and human factors in creating transparent and trustworthy AI.

Keywords

explainable AI, human-centered XAI, human-computer interaction

1. Introduction

This work presents the main research directions of the four-year national research project EXPLAINME, *EXPLainable Artificial INtelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI* (2024-2028) funded by MICIU/AEI and by ERDF, EU.

Artificial Intelligence (AI) has proven effective in many areas, but outcomes are hard for people to understand, leading to confusion and reduced trust. To address this, eXplainable AI (XAI) aims to enhance the transparency and interpretability of AI systems—particularly in applications like medical imaging, affective state recognition, and human action recognition. XAI not only helps users comprehend AI behavior but also supports responsible development by identifying potential sources of bias in data and algorithms.

The project EXPLAINME is a continuation of the national research project EXPLAINING (EXPLainable Artificial INtelligence for health and wellbeING, PID2019-104829RA-I00, 2020-2024) funded by MICIU/AEI and by ERDF/EU. During this project, the team focused on developing and analyzing XAI techniques applied to various domains [1], including facial expression recognition [2, 3, 4, 5, 6], human action recognition [7, 8], and medical imaging [9, 10, 11]. Results showed that although some XAI methods reveal how black-box models work, they are often hard for non-experts to understand. As a result, more research is needed to develop human-centered XAI (HCXAI) that focuses on user experience and perceptions, especially as XAI systems become more common in real-world applications.

On the one hand, recent studies reveal that XAI systems often underuse established human-centered methods from Human-Computer Interaction (HCI) [12]. Applying these approaches can guide user-

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Corresponding author.

✉ cristina.manresa@uib.es (C. Manresa-Yee); antoni.jaume@uib.es (A. Jaume-i-Capó); Adel.Ghazel@esigelec.fr (A. Ghazel); isabelle.hupont@ciencia.gob.es (I. Hupont); luis.leiva@uni.lu (L. A. Leiva); mack@yorku.ca (I. S. MacKenzie)

ORCID 0000-0002-8482-7552 (C. Manresa-Yee); 0000-0003-3312-5347 (A. Jaume-i-Capó); 0000-0001-5286-2110 (A. Ghazel); 0000-0002-9811-9397 (I. Hupont); 0000-0002-5011-1847 (L. A. Leiva); 0000-0002-4593-4101 (I. S. MacKenzie)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

driven technical choices, identify pitfalls in existing methods, and provide conceptual frameworks for human-compatible XAI [13].

On the other hand, current research highlights major gaps in XAI evaluation, including a lack of user studies [12], standardized procedures for measuring, quantifying and comparing the explainability of AI systems [14], and sufficient frameworks or metrics to assess explainability [15, 16]. Two challenges must be addressed: (1) evaluating explainability to analyze whether similar responses have similar explainability, and (2) evaluating explanations to determine if users understand and trust AI system's outputs.

Advancing these research lines can help address open challenges to AI adoption, including building trust, ensuring ethical practices like fairness and bias reduction, promoting human-centered design, and meeting regulatory requirements such as the EU's GDPR and AI Act, which emphasizes the right to explanation in automated decision-making.

2. Research hypothesis and aims

Considering the knowledge acquired in the EXPLAINING project, EXPLAINME poses the next research hypothesis:

"HCXAI design and evaluation provides more trustworthy, human-centric, and ethical AI. The development of XAI models and interfaces and XAI evaluation methods and metrics following human-centered guidelines and ethical considerations, will allow gaining further insights or evidence from the AI system to improve decision making, be aware of the system's limitations, increase the trust towards AI and verify and ensure compliance with applicable laws and regulations, ethical and legal principles and values. Further, it will help in avoiding biases based on gender, age, ethnicity, or other characteristics."

Based on this hypothesis, we define the general objective as follows: **To design and develop methods and metrics for human-centered explainability in AI systems.**

The focus is as follows:

- Designing and developing human-centered methods for XAI, including explanations, evaluation strategies, and user interfaces. This objective focuses on creating XAI methods grounded in human-centered principles that account for users, their context, and their explainability needs, while explicitly incorporating trustworthiness and ethical considerations.
- Defining methodologies for the design and evaluation of XAI from a human-centered perspective. This objective aims to create comprehensive methodologies that integrate human-centered design and evaluation principles into the development of XAI, to ensure trustworthiness and ethical values.

3. Method

3.1. Team

The team comprises 17 researchers. In addition to Computer Science experts in HCI and AI, the research team includes specialists in law, ethics, and psychology. Legal and ethical guidance ensures the development of trustworthy AI, while insights from cognitive and social sciences help deepen understanding of how people interact with and interpret explanations.

3.2. Methodology and work packages

The project comprises two phases: research and validation (see Fig. 1) supported by an overarching project management work package.

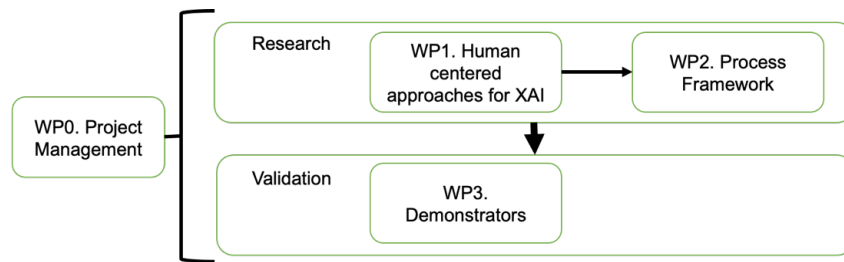


Figure 1: Workpackages of the project.

In the research phase, the team will analyze existing XAI techniques and their evaluation methods. Empirical research will examine defined research questions with both quantitative and qualitative data. This phase includes theoretical work on XAI methods and metrics (e.g., formalizations and algorithms), implementation of AI models, and adaptation of explainability techniques. It also integrates a Human-Centered Design approach to ensure the interactive systems are usable, effective, and aligned to user needs, thereby enhancing trust, satisfaction, and accessibility.

In the validation phase, empirical methods will assess user interaction with the developed systems. Research questions will be addressed through user studies, usability evaluations, and established instruments from HCI and psychology—such as trust and technology adoption questionnaires—to evaluate user attitudes and behaviours.

Specifically, WP1 aims to develop and evaluate XAI for visual challenges (both images and video) considering the most appropriate human-centered methods from the HCI community. The aim of WP2 is to advance a framework to ensure the accomplishment of legal, social and ethical issues to achieve Responsible AI systems. Finally, tasks in WP3 will demonstrate the applicability of results of previous WPs. We will develop two demonstrations: XAI for medical diagnostic support and XAI for elderly.

4. Expected impact

The research aims to advance knowledge in XAI systems and human-centered AI solutions. Expected outcomes include a process framework addressing legal, social, and ethical concerns, analysis of human-centered methods in XAI design, new metrics for XAI system evaluation, and the development of explainability models for image analysis, affective state recognition, human action recognition, and addressing bias issues. We expect to advance HCXAI in video and image analysis. Our contributions will impact AI adoption, design, transparency, and ethical aspects.

Acknowledgments

Project PID2023-149079OB-I00 (EXPLainable Artificial Intelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME) funded by MICIU/AEI/10.13039/ 501100011033/ and by ERDF/EU.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] M. Miró-Nicolau, A. J. i Capó, G. Moyà-Alcover, Assessing fidelity in XAI post-hoc techniques: A comparative study with ground truth explanations datasets, *Artificial Intelligence* 335 (2024) 104179. doi:<https://doi.org/10.1016/j.artint.2024.104179>.

- [2] F. X. Gaya-Morey, S. Ramis-Guarinos, C. Manresa-Yee, et al., Unveiling the human-like similarities of automatic facial expression recognition: An empirical exploration through explainable AI, *Multimedia Tools and Applications* 83 (2024) 85725–85753. doi:10.1007/s11042-024-20090-5.
- [3] C. Manresa-Yee, S. Ramis, J. M. Buades, Analysis of gender differences in facial expression recognition based on deep learning using explainable artificial intelligence, *International Journal of Interactive Multimedia and Artificial Intelligence* 9 (2024) 18–27. doi:10.9781/ijimai.2023.04.003.
- [4] G. del Castillo Torres, M. F. Roig-Maimó, M. Mascaró-Oliver, E. Amengual-Alcover, R. Mas-Sansó, Understanding how CNNs recognize facial expressions: A case study with LIME and CEM, *Sensors* 23 (2023). doi:10.3390/s23010131.
- [5] C. Manresa-Yee, S. Ramis, F. X. Gaya-Morey, J. M. Buades, Impact of explanations for trustworthy and transparent artificial intelligence, in: *Proc. of the XXIII International Conference on Human Computer Interaction, Interacción '23*, ACM, New York, 2024. doi:10.1145/3612783.3612798.
- [6] S. Ramis Guarinos, C. Manresa Yee, J. M. Buades Rubio, F. X. Gaya-Morey, Explainable facial expression recognition for people with intellectual disabilities, in: *Proc. of the XXIII International Conference on Human Computer Interaction, Interacción '23*, ACM, New York, 2024. doi:10.1145/3612783.3612789.
- [7] F. X. Gaya-Morey, C. Manresa-Yee, J. M. Buades-Rubio, Deep learning for computer vision based activity recognition and fall detection of the elderly: A systematic review, *Applied Intelligence* 54 (2024) 8982–9007. doi:10.1007/s10489-024-05645-1.
- [8] F. X. Gaya-Morey, C. Manresa-Yee, J. M. Buades-Rubio, Explainable activity recognition for the elderly, in: *Proc. of the XXIII International Conference on Human Computer Interaction, Interacción '23*, ACM, New York, 2024. doi:10.1145/3612783.3612790.
- [9] N. Petrović, G. Moyà-Alcover, A. J. i Capó, M. González-Hidalgo, Sick-cell disease diagnosis support selecting the most appropriate machine learning method: Towards a general and interpretable approach for cell morphology analysis from microscopy images, *Computers in Biology and Medicine* 126 (2020) 104027. doi:10.1016/j.combiomed.2020.104027.
- [10] M. Miró-Nicolau, G. Moyà-Alcover, A. Jaume-i Capó, Evaluating explainable artificial intelligence for x-ray image analysis, *Applied Sciences* 12 (2022). doi:10.3390/app12094459.
- [11] P. D. Marrero-Fernandez, J. M. Buades-Rubio, A. Jaume-i Capó, T. Ing Ren, An approach for selecting the most explanatory features for facial expression recognition, *Applied Sciences* 12 (2022). doi:10.3390/app12115637.
- [12] Y. Rong, T. Leemann, T.-T. Nguyen, L. Fiedler, P. Qian, V. Unhelkar, T. Seidel, G. Kasneci, E. Kasneci, Towards human-centered explainable AI: A survey of user studies for model explanations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46 (2024) 2104–2122. doi:10.1109/TPAMI.2023.3331846.
- [13] Q. V. Liao, K. R. Varshney, Human-centered explainable AI (XAI): From algorithms to user experiences, 2022. arXiv:2110.10790.
- [14] N. Burkart, M. F. Huber, A survey on the explainability of supervised machine learning, *J. Artif. Int. Res.* 70 (2021) 245–317. doi:10.1613/jair.1.12228.
- [15] L. Floridi, J. Cows, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, AI4People—An ethical framework for a Good AI society: opportunities, risks, principles, and recommendations, *Minds and Machines* 28 (2018) 689–707. doi:10.1007/s11023-018-9482-5.
- [16] R. R. Hoffman, S. T. Mueller, G. Klein, J. Litman, Metrics for explainable AI: challenges and prospects, 2019. arXiv:1812.04608.

Artículos relevantes ya publicados



Resumen del artículo: un experimento empírico de un método de elicitación de requisitos de usabilidad para diseñar GUI basadas en entrevistas*

Yeshica Isela Ormeño^{1,*†}, José Ignacio Panach^{2,†} and Oscar Pastor^{3,†}

¹ Universidad Nacional de San Antonio Abad del Cusco, Av. De la Cultura 733 Cusco, Perú

² Universitat de València, Avenida de la Universidad s/n, Valencia, Burjassot 46100, España

³ Universitat Politècnica de València, Camino de Vera s/n, Valencia, Valencia 46022, España

Resumen

Este estudio enfrenta el reto de elicitar requisitos de usabilidad cuando los analistas no son expertos. Se propone y se evalúa UREM, un método estructurado basado en entrevistas guiadas con árboles de decisión que se comparó con entrevistas no estructuradas en un experimento con 48 participantes organizados en parejas, midiendo variables como efectividad, eficiencia y satisfacción. UREM demostró mejorar significativamente la identificación de requisitos de usabilidad y la aplicación de guías de diseño, sin afectar negativamente la eficiencia ni la satisfacción de analistas y usuarios. Se concluye que UREM mejora la calidad del proceso de elicitación de requisitos.

Palabras Clave

Usabilidad, experimento, guías, interfaces usuario

1. Introducción

La usabilidad es un atributo de calidad en el desarrollo de software, determinante para el éxito o fracaso de un sistema. Las normas ISO 9241-11 e ISO/IEC 25010 [1] definen la usabilidad en términos de efectividad, eficiencia y satisfacción del usuario en un contexto específico. En la actualidad, donde se utilizan nuevas formas de interacción con los sistemas, es fundamental diseñar interfaces altamente usables que reduzcan problemas y riesgos asociados a la experiencia del usuario. Sin embargo, en la práctica del desarrollo de software, los requisitos de usabilidad suelen ser ignorados en las primeras fases del ciclo de vida, especialmente durante la elicitación de requisitos. Esto incrementa los costos y afecta la calidad del producto final. A pesar de que existen diversas técnicas de evaluación de usabilidad, como entrevistas y prototipado, su eficacia depende de una correcta estructuración y del conocimiento del analista, quien frecuentemente no es experto en usabilidad.

Con el objetivo de apoyar a los analistas, se propone un método llamado UREM (Usability Requirements Elicitation Method) [2], el cual se basa en entrevistas estructuradas guiadas por árboles de decisión que contienen preguntas y respuestas derivadas de guías de usabilidad. Este método permite identificar diferentes alternativas de diseño y sugiere, con base en guías existentes, cuáles optimizan la usabilidad [3].

En general, los métodos existentes suelen basarse en entrevistas y cuestionarios, pero carecen de recomendaciones explícitas que guíen al usuario final en decisiones de diseño, y la mayoría son entrevistas no estructuradas que dependen en gran medida de la experiencia del entrevistador. UREM se propone como una alternativa estructurada y orientada a la usabilidad para cubrir esta carencia.

*Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

Resumen del artículo: An empirical experiment of a usability requirements elicitation method to design GUIs based on interviews, Information and Software Technology, 164 (2023), 107324. doi: <https://doi.org/10.1016/j.infsof.2023.107324>

^{1*} Corresponding author.

[†] These authors contributed equally.

✉ yeshica.ormeno@unsaac.edu.pe (Y. Ormeño); joigpana@uv.es (I. Panach); opastor@dsic.upv.es (O. Pastor)

 0000-0002-5497-6928 (Y. Ormeño); 0000-0002-7043-6227 (I. Panach); 0000-0002-1320-8471 (O. Pastor)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Antecedentes

Se realizó una revisión sobre métodos para la elicitación de requisitos de usabilidad, en base a propuestas que incluyen modelos, metodologías y validaciones empíricas. La revisión fue realizada en repositorios como Scopus e IEEEExplore, seleccionándose 15 trabajos relevantes, agrupados en cuatro categorías: i) métodos aplicados a contextos específicos, Gunduz y Pathan[4] identificaron problemas de usabilidad en apps móviles de aerolíneas. Fahey [5] emplearon entrevistas Delphi y modelado de procesos en un hospital. Otros estudios trataron temas como autenticación biométrica y desarrollo dirigido por comportamiento usando historias de usuario, ii) contextos generales, De Carvalho et al. utilizaron FRAM y BPMN para descubrir requisitos en sistemas complejos. Nhavoto [6] desarrolló un sistema de mensajería médica para seguimiento de pacientes con VIH/TB. Elias [7] propuso un sistema semiautomático con ontologías y agentes en entornos educativos. Yuan y Zhang [7] crearon un modelo ontológico para personalizar requisitos mediante diálogo con usuarios. Abad [8]. compararon prototipado en papel silencioso y “en voz alta”, destacando su utilidad para captar requisitos no funcionales, iii) uso de guías de usabilidad, Márquez y Taramasco [9] diseñaron un marco para generar recomendaciones en sistemas clínicos. Abdallah [10] propusieron QXSBD, un método cuantitativo basado en escenarios que redujo defectos de usabilidad, y iv) En validaciones empíricas, Vitiello [11] aplicó un enfoque centrado en empoderamiento del usuario. Tanikawa [12] diseñó formularios para alinear necesidades y diseño. Abad evaluó la técnica Wizard-of-Oz. Peruzzini y Germani [12] desarrollaron plataformas asistidas para adultos mayores.

En conclusión, se ha observado que muchos métodos carecen de estructura guiada o de recomendaciones sistemáticas. El método UREM, busca cubrir esa carencia con entrevistas estructuradas apoyadas por guías de usabilidad.

3. Definición del experimento y planeación

En esta sección se describe el diseño del experimento según Juristo and Moreno [13]. El objetivo del experimento es comparar dos métodos para la elicitación de requisitos de usabilidad en el diseño de interfaces gráficas (GUI): entrevistas no estructuradas (control) versus entrevistas estructuradas usando UREM. Se busca evaluar su efectividad, eficiencia y nivel de satisfacción desde la perspectiva tanto del analista como del usuario final.

Las cinco preguntas de investigación son:

- Efectividad para elicitar requisitos, RQ1r = ¿La eficacia del analista para obtener los requisitos de usabilidad es afectada por el método de elicitación?
- Efectividad en la aplicación de guías de usabilidad, RQ1g = ¿La eficacia del analista para aplicar las guías de usabilidad es afectada por el método de elicitación?
- Eficiencia, RQ2r = ¿La eficiencia del analista es afectada por el método de elicitación de los requisitos de usabilidad?
- Satisfacción del analista, RQ3a = ¿La satisfacción del analista es afectada por el método de elicitación de los requisitos de usabilidad?
- Satisfacción del usuario, RQ3e = ¿La satisfacción del usuario final es afectada por el método de elicitación de los requisitos de usabilidad?

Se utilizó un diseño intra-sujetos, donde se distribuyeron los 48 participantes en dos grupos cada grupo G1 y G2 que estaban conformados por parejas con el rol de analista y usuario final. Cada grupo y por ende cada participante usó ambos métodos. Se definió dos problemas para determinar los requisitos de

usabilidad: uno para un centro de salud (P1) y otro para la banca móvil (P2). Ambos problemas se enmarcan en el contexto de las aplicaciones móviles. Se miden indicadores de efectividad, eficiencia y satisfacción (se utilizó cuestionarios MAM y CSUQ). Los sujetos experimentales fueron estudiantes de ingeniería de software de la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC). El uso de estudiantes está justificado por la naturaleza del experimento, centrado en comparar métodos bajo condiciones similares. Todos los instrumentos utilizados: cuestionarios demográficos, descripción de los objetos experimentales, cuestionarios de satisfacción, hojas de cálculo para ejecutar el experimento están disponibles en Zenodo [14]. UREM se apoya en la herramienta desarrollada (<http://hci.dsic.upv.es/urem>) y el árbol de decisiones que guía al usuario final a través de las alternativas de diseño, recomendando aquellas que optimizan la usabilidad. Se aplican pruebas estadísticas (Mann-Whitney y GLM), además de análisis de tamaño de efecto (eta parcial cuadrado) y potencia estadística.

4. Resultados del experimento

Los resultados mostraron que UREM fue significativamente más efectivo que las entrevistas no estructuradas en dos aspectos: mayor cantidad de requisitos de usabilidad correctamente elicitados (RQ1r) y mejor aplicación de guías de usabilidad (RQ1g). Ambas diferencias fueron estadísticamente significativas ($p < 0.001$), con efectos fuertes. No se encontraron diferencias significativas en eficiencia (RQ2r), lo que indica que UREM no reduce la productividad a pesar de su estructura. Tampoco hubo diferencias relevantes en la satisfacción del analista (RQ3a) ni del usuario final (RQ3e), aunque este último valoró levemente mejor las recomendaciones del método UREM. En conjunto, UREM mejora notablemente la calidad del proceso de elicitación sin comprometer el tiempo ni la experiencia subjetiva de los involucrados. Esto lo posiciona como una herramienta poderosa y práctica en contextos donde no se tiene expertos en usabilidad.

5. Discusión y Conclusiones

Los hallazgos permiten afirmar que UREM aporta beneficios claros al proceso de diseño de interfaces gráficas. Mejora la cobertura de requisitos de usabilidad, fomenta la aplicación de buenas prácticas y permite a analistas sin experiencia aplicar recomendaciones expertas de manera guiada. Además, se mantiene competitivo en tiempo y satisfacción, lo que lo hace viable en contextos reales. En comparación con la literatura existente, UREM destaca por ser un método estructurado reutilizable y fundamentado en guías, algo ausente en la mayoría de propuestas previas. Entre las limitaciones se encuentra el uso de estudiantes como sujetos experimentales y el hecho de que el árbol de decisiones fue diseñado por un experto. Como trabajo futuro, se propone validar UREM en entornos industriales, extender el árbol a otros dominios y vincularlo con herramientas de prototipado visual.

En conclusión, UREM representa una contribución metodológica importante para la ingeniería de requisitos de usabilidad, especialmente en etapas tempranas del desarrollo de software.

6. Agradecimientos

Este trabajo se desarrolló con el apoyo de la Universidad Nacional de San Antonio Abad del Cusco - Perú, con la ayuda de la Generalidad Valenciana con TENTACLE (CIAICO/2023/089), y con la ayuda del Ministerio Español de Ciencia e Innovación cofinanciado con FEDER en el proyecto SREC (PID2021-123824OB-I00).

Declaración de uso de IA Generativa

Los autores no han empleado ninguna herramienta de IA Generativa.

7. Referencias

- [1] I. O. f. Standardization, ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs): Part 11: Guidance on usability. 1998.
- [2] Y. Ormeño, J. I. Panach, O. Pastor, A Proposal to Elicit Usability Requirements within a Model-Driven Development Environment, *International Journal of Information System Modeling and Design*, 5 (2014), 1-21. doi:10.4018/ijismd.2014100101.
- [3] Y. I. Ormeño, J. I. Panach, O. Pastor, An empirical experiment of a usability requirements elicitation method to design GUIs based on interviews, *Information and Software Technology*, 164 (2023), doi.org/10.1016/j.infsof.2023.107324.
- [4] F. Gunduz, A. S. K. Pathan, Usability improvements for touch-screen mobile flight booking application: A case study, in: *Proceedings of the ACSAT International Conference on Advanced Computer Science Applications and Technologies*, 2012, pp. 49-54. doi: 10.1109/ACSAT.2012.9.
- [5] P. Fahey, C. Harney, S. Kesavan, A. McMahon, L. McQuaid, B. Kane, Human computer interaction issues in eliciting user requirements for an Electronic Patient Record with multiple users, in: *Proceedings of the 24th International Symposium on Computer-Based Medical Systems, CBMS 2011*, pp. 1-6. doi:10.1109/CBMS.2011.5999130.
- [6] J. A. Nhavoto, Å. Grönlund, W. P. Chaquilla, SMSaúde: Design, development, and implementation of a remote/mobile patient management system to improve retention in OF care for HIV/aids and tuberculosis patients, *JMIR mHealth and uHealth*, 3 (2015). doi:10.2196/mhealth.3854.
- [7] E. Elias et al. Towards an ontology-based system to improve usability in collaborative learning environments, volume 7315 of *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, Springer, Berlin, Heidelberg, 2012. doi:10.1007/978-3-642-30950-2_39.
- [8] Z. S. H. Abad, S. D. V. Sims, A. Cheema, M. B. Nasir, P. Harisinghani, Learn More, Pay Less! Lessons Learned from Applying the Wizard-of-Oz Technique for Exploring Mobile App Requirements, in: *IEEE 25th International Requirements Engineering Conference Workshops (REW)*, 2017, pp. 132-138. doi:10.1109/REW.2017.71.
- [9] G. Márquez, C. Taramasco, Using Dissemination and Implementation Strategies to Evaluate Requirement Elicitation Guidelines: A Case Study in a Bed Management System, *IEEE Access*, 8 (2020), 145787-145802. doi:10.1109/ACCESS.2020.3015144.
- [10] A. Abdallah, R. Hassan, M. A. Azim, Quantified extreme scenario based design approach, in: *Proceedings of the ACM Symposium on Applied Computing, SAC'13*, ACM Press, New York, NY, 2013, pp. 1117-1122. doi:10.1145/2480362.2480572.
- [11] G. Vitiello, R. Francese, M. Sebillio, G. Tortora, M. Tucci, UX-requirements for patient's empowerment - The case of multiple pharmacological treatments: A case study of it support to chronic disease management, in: *Proceedings of the IEEE 25th International Requirements Engineering Conference Workshops, REW'17*, 2017, pp. 139-145. doi: 10.1109/REW.2017.67
- [12] Y. Tanikawa, R. Okubo, S. Fukuzumi, Process support method for improved user experience, *NEC Technical Journal*, 8 (2014), 28-32.
- [13] N. Juristo, A. M. Moreno, *Basics of software engineering experimentation*. Springer Science & Business Media, 2013.
- [14] Y. I. Ormeño, J. I. Panach, Ó. Pastor. Experimental material of the article "An Empirical Experiment of a Usability Requirements Elicitation Method based on Interviews", 2023. URL: <https://doi.org/10.5281/zenodo.7646554>.

Concurso TFG/TFM



Heurísticas de evaluación de usabilidad para plataformas de comercio electrónico

Alejandra Yicel Vargas Muñoz^{1,*†}, César A Collazos^{2,†} and Antoni Granollers i Saltiveri^{3,†}

¹ Universidad del Cauca, Cauca, Colombia, 190001

² Universidad del Cauca, Cauca, Colombia, 190001

³ Universitat de Lleida, Escola Politècnica Superior, INSPIRES research institute, Lleida, Catalunya-España

Resumen

Las plataformas de comercio electrónico han transformado las interacciones digitales para la compra de bienes y servicios, convirtiendo la usabilidad en un factor clave para la satisfacción del usuario, la eficiencia de las transacciones y la fidelización de los clientes. Sin embargo, las metodologías tradicionales de evaluación de la usabilidad, como las heurísticas de Nielsen, no abordan completamente los desafíos específicos del comercio electrónico, como la navegación por catálogos extensos, los procesos de pago y la confianza del usuario. Este estudio propone un conjunto de heurísticas personalizadas, diseñadas específicamente para evaluar la usabilidad en plataformas de comercio electrónico. A través de una revisión sistemática de la literatura y la validación por parte de expertos, las heurísticas propuestas buscan mejorar las evaluaciones de usabilidad en este ámbito. Los resultados demuestran su eficacia para identificar problemas de usabilidad con mayor precisión que los métodos tradicionales.

Palabras clave

Heurísticas de usabilidad, evaluación de usabilidad en comercio electrónico, interacción humano-computadora, experiencia de usuario, evaluación heurística, pruebas de usabilidad, plataformas de comercio digital.

1. Introducción

El comercio electrónico ha transformado la manera en que los usuarios interactúan con los sistemas digitales para la adquisición de bienes y servicios. La usabilidad de estas plataformas juega un papel crucial en la satisfacción del usuario, la eficiencia de las transacciones y la retención de clientes. Sin embargo, las metodologías tradicionales de evaluación de usabilidad, como las heurísticas de Nielsen, no abordan completamente las particularidades del comercio electrónico, como la navegación por catálogos extensos, los procesos de pago y la confianza en la plataforma.

Este trabajo presenta un conjunto de heurísticas diseñadas específicamente para evaluar la usabilidad de plataformas de comercio electrónico. La investigación parte de una revisión sistemática de la literatura existente y culmina en la validación de las heurísticas mediante un proceso experimental con expertos en usabilidad.

2. Objetivos

El objetivo general de esta investigación es diseñar, desarrollar y validar un conjunto de heurísticas específicas para evaluar la usabilidad en plataformas de comercio electrónico, con el fin de mejorar la experiencia del usuario y la eficiencia de las interacciones en estos entornos digitales. Los objetivos específicos son:

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain.

*Corresponding author.

† These authors contributed equally.

✉ alejandravargas@unicauca.edu.co (A. Vargas); ccollazo@unicauca.edu.co (C. Collazos); toni.granollers@udl.cat (T. Granollers)

 0009-0007-0678-6832 (A. Vargas); 0000-0002-7099-8131 (C. Collazos); 0000-0001-9189-7308 (T. Granollers)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2.1 Realizar una revisión de la literatura

Sobre heurísticas de usabilidad existentes y sus aplicaciones en plataformas de comercio electrónico.

2.2 Desarrollar un conjunto de heurísticas específicas

Basadas en el análisis de literatura y necesidades identificadas en el contexto del comercio electrónico.

2.3 Validar las heurísticas propuestas

Mediante un proceso experimental con expertos en usabilidad, evaluando su aplicabilidad y efectividad.

3. Resultados

La investigación se desarrolló en tres fases principales: (i) revisión sistemática de la literatura, (ii) diseño del conjunto de heurísticas y (iii) validación experimental.

3.1 Revisión de la literatura

Se analizaron estudios previos sobre evaluación heurística en plataformas de comercio electrónico, identificando limitaciones en las heurísticas tradicionales y la necesidad de nuevas categorías específicas.

3.2 Desarrollo del conjunto de heurísticas

El proceso de desarrollo de las heurísticas siguió una metodología estructurada que incluyó varias etapas clave:

3.2.1 Análisis comparativo de heurísticas existentes

Se realizó una revisión de múltiples enfoques heurísticos utilizados en evaluación de usabilidad, identificando sus limitaciones en el contexto del comercio electrónico. Se analizaron conjuntos heurísticos ampliamente utilizados, como los de Nielsen, Tognazzini y Granollers, para determinar los vacíos en la evaluación de plataformas comerciales digitales.

3.2.2 Identificación de dimensiones clave

A partir del análisis comparativo, se definieron dimensiones esenciales para la usabilidad en comercio electrónico, tales como facilidad de navegación, claridad en la presentación de información de productos, confianza y seguridad en las transacciones, eficiencia del proceso de compra y capacidad de recuperación ante errores. Estas dimensiones guiaron la formulación de las nuevas heurísticas.

3.2.3 Definición inicial del conjunto heurístico

Con base en las dimensiones clave identificadas, se desarrolló un primer conjunto de heurísticas específicas para comercio electrónico. Se aseguraron principios de aplicabilidad universal, con un lenguaje claro y una estructura que facilitara su uso en evaluaciones prácticas.

3.2.4 Proceso iterativo de refinamiento

El conjunto inicial de heurísticas fue sometido a evaluación por parte de expertos en usabilidad y HCI. Se llevaron a cabo sesiones de validación en las que los expertos aplicaron las heurísticas a plataformas reales y proporcionaron retroalimentación sobre su claridad, aplicabilidad y efectividad. En cada iteración, se ajustaron definiciones, se reformularon criterios y se eliminaron redundancias.

3.2.5 Diseño de instrumentos de evaluación

Para facilitar la aplicación de las heurísticas, se desarrollaron herramientas como cuestionarios estructurados y guías detalladas de evaluación. Estas herramientas estandarizaron el proceso de análisis y aseguraron una evaluación consistente en diferentes plataformas de comercio electrónico.

3.3 Validación experimental

El conjunto de heurísticas se validó mediante un estudio con expertos en usabilidad. Se aplicaron las heurísticas a una plataforma real, midiendo su eficacia para detectar problemas de usabilidad. Los resultados demostraron que las heurísticas propuestas permiten identificar problemas específicos de estas plataformas con mayor precisión que los métodos tradicionales.

4. Declaración de uso de inteligencia artificial

En el desarrollo de este artículo se utilizó inteligencia artificial (IA) como herramienta de apoyo exclusivamente para la redacción preliminar de textos, revisión gramatical y generación de ideas. Todas las decisiones conceptuales, metodológicas y analíticas fueron tomadas por los autores, quienes también realizaron la validación final del contenido. Las herramientas de IA no sustituyeron el trabajo académico ni científico de los autores, y su uso se realizó en conformidad con los principios de integridad y transparencia científica.

Entrando en Oz: simulación virtual y evaluación de robots sociales a través de técnicas de Mago de Oz

Guillermo Cubero^{1,*}, Laura Villa^{1,†} and Ramón Hervás^{1,†}

¹Universidad de Castilla-La Mancha, UCLM, Ciudad Real, España

Resumen

El trabajo aborda el desafío de evaluar robots sociales asistenciales, específicamente SHARA3 para adultos mayores, mediante el desarrollo de una infraestructura web que permite simulación y evaluación remota. La solución implementa un enfoque de Wizard of Oz con tres componentes: una aplicación web con simulación 3D del robot, un servidor central para gestionar comunicaciones, y una aplicación de control para que operadores supervisen el comportamiento del robot en tiempo real. Utilizando tecnologías como React, Three.js y servicios cognitivos en la nube, el sistema fue evaluado positivamente por expertos internacionales, demostrando su potencial para mejorar la adaptabilidad robótica, crear interacciones realistas y aplicarse en diversos escenarios de investigación, estableciendo una base sólida para el desarrollo futuro de sistemas de evaluación de robots sociales asistenciales.

Palabras clave

Robots asistentes sociales, metodología Wizard of Oz, sistema conversacional, interacción humano-robot

1. Introducción

El envejecimiento global de la población ha incrementado la demanda de soluciones tecnológicas para el cuidado de adultos mayores. Los Robots Sociales Asistentes (SAR) han emergido como herramientas prometedoras para brindar apoyo emocional y funcional[1]. Sin embargo, su desarrollo y evaluación enfrentan desafíos significativos[2]:

- **Complejidad logística:** Transportar robots físicos y equipos de profesionales a entornos reales para pruebas con usuarios.
- **Costes elevados:** Equipos de evaluadores y adaptaciones hardware/software para cada prueba.
- **Evaluación multidimensional:** Necesidad de analizar no solo funcionalidades técnicas, sino también aspectos sociales y emocionales.

Este trabajo se centra en el robot SHARA3, diseñado para interactuar con adultos mayores[3, 4]. Para abordar estos desafíos, se propone una infraestructura web basada en la metodología Wizard of Oz (WoZ), que permite simular y evaluar interacciones humano-robot de forma remota, reduciendo costes y facilitando la colaboración entre equipos de investigación[5].

2. Objetivos

2.1. Objetivo general

Desarrollar un sistema de simulación virtual y evaluación remota del robot SHARA3 mediante un enfoque semi-automático basado en WoZ, que permita:

- **Evaluación remota:** Eliminar la necesidad de pruebas físicas con el robot.

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Corresponding author.

†These authors contributed equally.

✉ Guillermo.Cubero@uclm.es (G. Cubero); Laura.Villa@uclm.es (L. Villa); Ramon.HLucas@uclm.es (R. Hervás)

🌐 <https://github.com/GuillermoCuberoCharco/> (G. Cubero); <https://github.com/Laura-VFA/> (L. Villa)

🆔 0009-0000-9392-1658 (G. Cubero); 0000-0001-9928-8945 (L. Villa); 0000-0001-9924-5443 (R. Hervás)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- **Desarrollo colaborativo:** Facilitar la participación de múltiples equipos.
- **Prototipado rápido:** Permitir ajustes ágiles en el comportamiento del robot.
- **Control flexible:** Combinar autonomía del robot con supervisión humana.

2.2. Objetivos específicos

- Simular en 3D el robot SHARA3 con animaciones emocionales y funcionalidades clave (ej. detección de voz y "wakeface").
- Implementar un servidor central para gestionar comunicaciones y servicios cognitivos (IBM Watson, Google Cloud).
- Desarrollar una interfaz WoZ para operadores, con modos automático y semi-automático.
- Replicar las interacciones del robot físico en un entorno virtual.

3. Resultados y conclusiones

El sistema se estructura en tres componentes principales, como se ilustra en la figura 1. En este entorno interactúan dos tipos de usuarios: el usuario final, quien dialoga directamente con la simulación del robot social, y el operador, responsable de supervisar y gestionar la calidad de la interacción.

El flujo de interacción se desarrolla de la siguiente manera: inicialmente, el usuario interactúa con el robot social a través de la plataforma web denominada *Tierra de Oz*, donde se ejecuta la simulación. Esta aplicación captura diversos datos de la interacción natural del usuario, incluyendo la transcripción de voz, la detección facial y la grabación de vídeo.

Posteriormente, todos estos datos son transmitidos al servidor, el *Camino de baldosas amarillas*, donde se procesan mediante los mismos mecanismos y servicios en la nube que utilizaría el robot social real. Una vez generadas las respuestas por el servidor, estas no se envían directamente al usuario, sino que primero son remitidas a la *Ciudad Esmeralda*, la interfaz donde el operador puede evaluarlas, aprobarlas tal como fueron generadas o modificarlas según considere necesario. De este modo, la continuidad de la interacción requiere siempre la autorización del operador.

Cabe destacar que la interfaz del operador incorpora diversas herramientas diseñadas para agilizar el proceso de respuesta, minimizando así el tiempo de espera del usuario y evitando que perciba la intervención humana. Adicionalmente, el operador dispone de funcionalidades avanzadas: puede enviar respuestas por iniciativa propia sin necesidad de una entrada previa del usuario, y tiene la opción de activar un modo autónomo en la simulación. En este último caso, el robot responderá automáticamente, aunque el registro completo de la interacción permanecerá visible en la aplicación del operador, permitiéndole intervenir cuando lo estime oportuno.

3.1. Tierra de Oz (aplicación web)

Tecnologías: React, Three.js, Face-api.js.

Funcionalidades:

- Modelo 3D interactivo de SHARA3 con animaciones emocionales (ej. felicidad, tristeza).
- Detección facial ("wakeface") y de voz para iniciar interacciones con el usuario.
- Comunicación bidireccional con el operador WoZ a través de un chat escrito y de voz.

3.2. Camino de baldosas amarillas (servidor central)

Tecnologías: Node.js, Socket.IO, WebSocket.

Funcionalidades:

- Gestión de mensajes y streaming de vídeo en tiempo real.
- Integración con servicios cognitivos (IBM Watson para diálogos y análisis emocional, Google Cloud para síntesis de voz y traducción).

3.3. Ciudad Esmeralda (interfaz WoZ)

Tecnologías: Python, PyQt6.

Funcionalidades:

- Control en tiempo real de las interacciones, permitiendo al operador intercalar entre un comportamiento autónomo de la simulación a una completamente manual.
- Visualización del usuario y del robot simulado, junto con todos los datos pertenecientes a la interacción en proceso.
- Respuestas predefinidas para agilizar la intervención del operador y reducir la personificación del mismo sobre el robot.

Enfoque iterativo: El desarrollo siguió un proceso incremental con 11 iteraciones, desde prototipos básicos hasta un sistema integrado y evaluado por expertos[6].

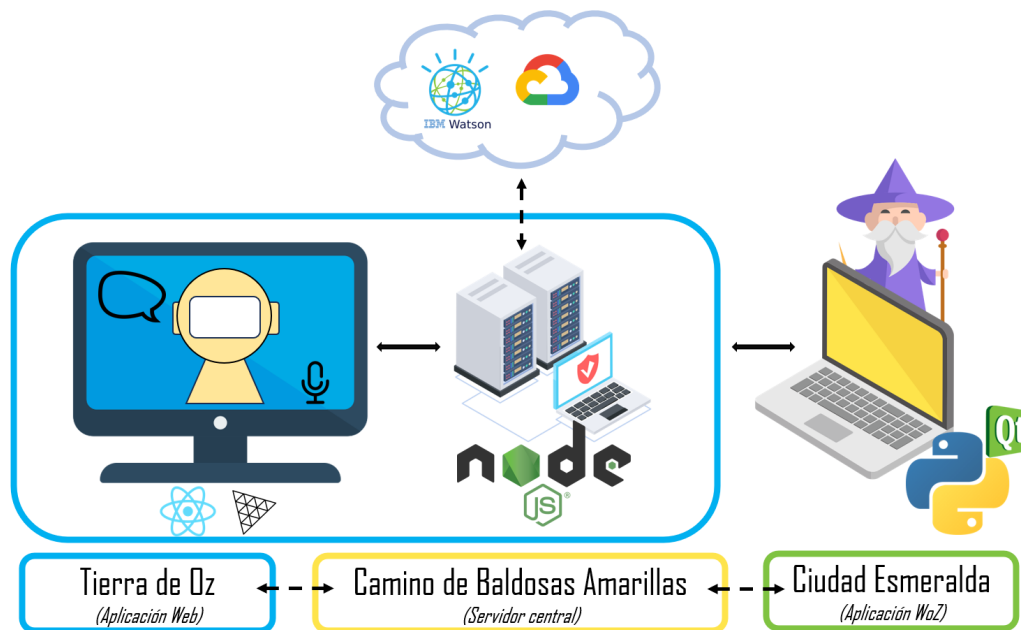


Figura 1: Arquitectura general del sistema.

3.4. Principales logros

- **Simulación realista:** El modelo 3D replica las expresiones emocionales y funcionalidades de SHARA3, validado por un panel internacional de expertos en HRI, junto con otros componentes del sistema.
- **Reducción de costes:** Elimina la necesidad de despliegues físicos, permitiendo pruebas remotas.
- **Flexibilidad:** El modo semi-automático permite evaluar tanto comportamientos autónomos como intervenciones humanas.
- **Evaluación positiva:** Los expertos destacaron la utilidad del sistema para refinar interacciones y su potencial como herramienta colaborativa.

Este proyecto de investigación ha dado lugar a una contribución científica relevante mediante la publicación de un artículo en el congreso nacional UCAmI (*Ubiquitous Computing and Ambient Intelligence*). El trabajo, titulado *SHARA in the Land of Oz: A platform for the rapid development of human-robot-interactions using social robot simulator and Wizard of Oz*[7], permitió presentar a la comunidad científica una de las primeras implementaciones funcionales del sistema desarrollado.

3.5. Trabajo futuro

- Incorporar LLMs locales para reducir dependencia de servicios externos y latencias producto de las respuestas de los mismos.
- Ampliar el sistema como marco genérico para otros robots asistenciales y aumentar la capacidad de edición y control del operador sobre la simulación.
- Mejorar el análisis de lenguaje corporal y protocolos de comunicación social.

3.6. Conclusiones

Este trabajo establece una alternativa en la evaluación de robots sociales mediante simulación virtual, ofreciendo una solución escalable y práctica. Personalmente, destacó la importancia de la computación afectiva y la robótica social en el apoyo a poblaciones vulnerables, abriendo líneas de investigación futuras en interacciones humano-robot más naturales y accesibles.

Declaración de uso de la IA

(utilizando la taxonomía de actividades en ceur-ws.org/genai-tax.html):

Durante la preparación de este trabajo, el/los autor(es) utilizó/utilizaron X-GPT-4 y Gramby con el fin de: Revisión de gramática y ortografía. Después de usar esta(s) herramienta(s)/servicio(s), el/los autor(es) revisó/revisaron y editó/editaron el contenido según fue necesario y asume(n) la responsabilidad completa del contenido de la publicación.

Referencias

- [1] B. Sawik, S. Tobis, E. Baum, A. Suwalska, S. Kropińska, K. Stachnik, E. Pérez-Bernabeu, M. Cildoz, A. Agustin, K. Wiczorowska-Tobis, Robots for elderly care: Review, multi-criteria optimization model and qualitative case study, *Healthcare* 11 (2023) 1286.
- [2] D. Cruz-Sandoval, A. Morales-Tellez, E. B. Sandoval, J. Favela, A social robot as therapy facilitator in interventions to deal with dementia-related behavioral symptoms, in: *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 161–169.
- [3] L. Villa, R. Hervás, D. Cruz-Sandoval, J. Favela, Design and evaluation of proactive behavior in conversational assistants: approach with the eva companion robot, in: *International Conference on Ubiquitous Computing and Ambient Intelligence*, Springer, 2022, pp. 234–245.
- [4] L. Villa, R. Hervás, C. C. Dobrescu, D. Cruz-Sandoval, J. Favela, Incorporating affective proactive behavior to a social companion robot for community dwelling older adults, in: *International Conference on Human-Computer Interaction*, Springer, 2022, pp. 568–575.
- [5] S. Honig, T. Oron-Gilad, Understanding and resolving failures in human-robot interaction: Literature review and model development, *Frontiers in Psychology* 9 (2020) 861.
- [6] A. J. González, O. a Objeto, *Ingeniería de software: Metodologías*, Visitado el 8 (2007). URL: <https://slideplayer.es/slide/9892803/>.
- [7] G. Cubero, L. Villa, J. Favela, S. Ochoa, C. Díaz-Fernández, R. Hervás, Shara in the land of oz: A platform for the rapid development of human-robot-interactions using social robot simulator and wizard of oz, in: *Proceedings of 16th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAmI 2024)*, Springer, 2024.

Análisis y modelado del comportamiento de estudiantes en cursos en línea desde el enfoque de la Analítica del Aprendizaje Multimodal

Alvaro Becerra¹, Ruth Cobos¹

¹Grupo GHIA, Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, España

Resumen

En los últimos años, la enseñanza en línea ha crecido de manera exponencial. Sin embargo, la naturaleza a distancia de este tipo de aprendizaje hace difícil saber si los estudiantes están atentos o si están sufriendo distracciones. Por ello, ha surgido el campo de las Analíticas del Aprendizaje Multimodal cuyo objetivo es obtener una visión holística del proceso del aprendizaje utilizando una gran cantidad de datos. Para entender cómo afectan las distracciones, y en especial el uso de móvil, en estudiantes en aprendizaje en línea, el año pasado como parte de mi TFG se realizó una captura de datos biométricos y multimodales en una sesión de aprendizaje con 120 estudiantes divididos en tres grupos en función del uso que podían hacer del móvil. El objetivo de este TFM es estudiar si dichas distracciones provocan cambios en las distintas variables biométricas capturadas mediante biosensores. Para ello, en primer lugar se ha realizado un análisis estadístico para determinar qué estudiantes han aprendido más, cuáles sentían más ansiedad y las diferencias en variables fisiológicas como atención, pulso o las ondas cerebrales entre grupos, sexos y actividades y cómo estas varían antes, durante y después del uso de móvil. En segundo lugar, se ha modelado el comportamiento de uso de móvil mediante varios modelos de aprendizaje automático para detectar si un estudiante ha usado o no el móvil en un periodo determinado. Finalmente, también se comentará el trabajo que se está realizando para poder proporcionar *feedback* personalizado a los estudiantes.

Palabras clave:

Analíticas del Aprendizaje Multimodal, Aprendizaje en línea, Biometría, Detección de comportamiento, Machine Learning, Uso de móvil

1. Introducción

El aprendizaje en línea y los MOOCs (*Massive Open Online Courses*) han crecido significativamente, especialmente desde la pandemia del COVID-19 [1]. Sin embargo, su naturaleza a distancia dificulta saber si los estudiantes están realmente atentos o sufren distracciones [2], lo que ha impulsado el desarrollo de campos como las Analíticas del Aprendizaje Multimodal (*Multimodal Learning Analytics* o MMLA) [3], que es un subcampo de las Analíticas del Aprendizaje [4], que se caracteriza por el uso de sensores como eye-tracker, bandas EEG o webcams para capturar datos biométricos de estudiantes mientras aprenden para obtener una visión global y holística del proceso del aprendizaje [5].

En la Universidad Autónoma de Madrid, en los últimos años el grupo GHIA [6] y el grupo BiDA-Lab [7] han estado colaborando en el avance de este campo con investigaciones que buscan, por ejemplo, predecir la atención de un estudiante a través de los datos de una webcam [8]. Durante la realización de mi TFG, se trabajó en la monitorización de estudiantes mientras realizaban una sesión de aprendizaje en un MOOC con el objetivo de entender como afectaban las interrupciones debidas al uso del móvil.

El objetivo de este TFM se centra en analizar datos multimodales de 120 estudiantes que fueron monitorizados en mi TFG y que gracias al sistema M2LADS [9, 10], están etiquetados en función de la actividad que estaba realizando el estudiante (viendo vídeo, leyendo material, etc.) durante una sesión de aprendizaje en un MOOC de la UAM. Los estudiantes de estas monitorizaciones fueron divididos en tres grupos dependiendo del uso que podían hacer del móvil. Varios estudios se han centrado en ver si las interrupciones o distracciones por el uso del móvil afectan al aprendizaje [11]. La motivación de este TFM es estudiar si dichas distracciones se ven reflejadas en las distintas variables biométricas.

Interacción '25: XXV Congreso Internacional de Persona-Ordenador, Septiembre 03-05, 2025, Valladolid, España

✉ alvaro.becerra@uam.es (A. Becerra); ruth.cobos@uam.es (R. Cobos)

id 0009-0003-7793-2682 (A. Becerra); 0000-0002-3411-3009 (R. Cobos)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Objetivos

Los objetivos principales del TFM son, en primer lugar estudiar el estado del arte de MMLA y de estudios sobre uso de móvil. En segundo lugar, realizar un análisis estadístico del rendimiento y de las variables biométricas atención, meditación, ondas cerebrales (ondas alfa, beta, gamma, delta y theta) y pulso en toda la sesión de aprendizaje y antes durante y después del uso de móvil para descubrir las posibles diferencias entre grupos y por sexo. En tercer lugar, entrenar diversos modelos de aprendizaje automático para poder detectar si un estudiante estaba o no usando el móvil en un período determinado. Finalmente, también se resumen las investigaciones iniciales que se han llevado a cabo para poder proporcionar feedback personalizado a los estudiantes para que puedan mejorar su aprendizaje.

3. Contexto: datos multimodales

Con el objetivo de entender cómo afecta el uso de móvil al rendimiento y a las distintas señales biométricas de estudiantes en cursos en línea en 2023 se realizaron una serie de monitorizaciones a 120 estudiantes de la UAM (59 mujeres) contando con el respaldo del Comité de Ética de la universidad. Estos estudiantes debían realizar una sesión de aprendizaje de unos 30 minutos dentro de un MOOC. Además, se definieron tres grupos en función del uso de móvil (Figura 1):

- **Grupo 1 (Uso de móvil):** Este grupo tiene el móvil encima de la mesa, con las notificaciones activadas y con volumen. Los estudiantes de este grupo tenían que usar el móvil cada vez que recibieran una notificación y además, recibieron dos mensajes de los investigadores a los que tuvieron que responder obligatoriamente. Asimismo, entre las dos primeras lecturas se les pidió que consultasen su móvil de manera libre (por ejemplo, correo o notificaciones).
- **Grupo 2 (Móvil visible, pero prohibido):** Este grupo tiene el móvil encima de la mesa boca abajo, con las notificaciones activadas y con volumen. Sin embargo, durante la prueba no podían usarlo en ningún momento. Durante la sesión, recibieron mensajes de los investigadores en dos momentos. Además, podían recibir notificaciones de sus contactos (y tampoco podían leerlas).
- **Grupo 3 (Uso de móvil prohibido):** A los participantes de este grupo se les retiró el móvil silenciado de forma que ni podían usarlo durante la sesión ni lo tenían a la vista.

Antes de la sesión, los estudiantes realizaron un pretest sobre el contenido del curso. A continuación, fueron monitorizados con los sensores de la plataforma edBB [12] y M2LADS, en una sesión de aprendizaje (Figura 1) sobre HTML de un MOOC con varios tipos de materiales para aprender y preguntas para comprobar si han adquirido conocimientos (estas preguntas se corresponden al posttest).

4. Resultados análisis estadísticos

Mediante test estadísticos como ANOVA, test *t* de Student o test de Welch se realizaron diversos análisis para buscar diferencias significativas. En primer lugar, se analizó el rendimiento de los estudiantes en función de su grupo y su sexo. Todos los grupos tenían unas notas similares en el pretest (diferencia no significativa) y no se puede concluir que ninguno de ellos haya obtenido mejores resultados en el posttest. Esto difiere de los resultados encontrados en clases presenciales [11, 13]. Nuestra hipótesis es que como los estudiantes tenían acceso a todos los materiales del curso, después de una distracción podían retomar por ejemplo la lectura y, por lo tanto, el uso del móvil no es tan perjudicial como en educación presencial. Analizando el tamaño del efecto, hemos visto que el grupo que más ha aprendido ha sido el grupo 2. Por sexo, las mujeres partían con menos conocimientos iniciales que los hombres y esta diferencia se traslada al posttest, donde los hombres tienen mejor media. Respecto al aprendizaje, las mujeres son las que más aprenden (al partir de menos conocimientos pueden aprender mucho más). Si comparamos por sexo y por grupo, cabe remarcar que los hombres del grupo 3 no aprenden nada (la diferencia entre pretest y posttest no es significativa). Esta diferencia no puede ser explicada por la demografía, lo que sugiere que el hecho de no tener el móvil afectó a su rendimiento. En el caso

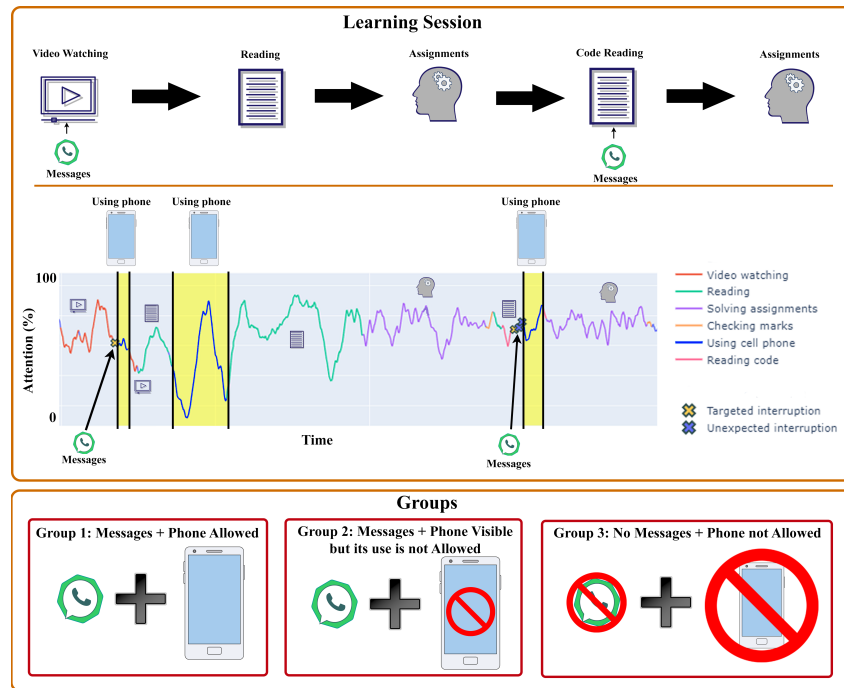


Figura 1: Diagrama de los grupos del experimento y del desarrollo de la sesión de aprendizaje

de las mujeres, es el grupo 1 el que no presenta evidencia estadística para afirmar que han aprendido (mientras que el grupo 2 y el grupo 3 sí que han aprendido). Nuestra hipótesis es que como las mujeres recibieron de media más mensajes que los hombres, les afectó más el tener que usar el móvil.

A continuación, se analizaron diferencias significativas entre las medias de las variables fisiológicas entre grupo y sexo. Las variables atención e intensidad de ondas alfa son las únicas que presentan diferencias significativas al comparar los grupos en la sesión completa. En particular, el grupo 2 es el que más atención media presenta y una menor intensidad de ondas alfa. Como el grupo 2 era el que más había aprendido, nuestra hipótesis es que el mayor aprendizaje se ha debido a una mayor concentración (reflejada en mayores niveles de atención y menores niveles de ondas alfa [14]). Si comparamos por actividades realizadas, podemos ver que los estudiantes tienen mayor atención en lectura y ejercicios que en vídeo. En el grupo 2 esta diferencia es especialmente grande. Por sexo, hemos visto que las mujeres tienen mayor pulso como ya se ha visto en [15]. También hemos observado mayor nivel de meditación en los hombres. Nuestra hipótesis es que quizás los hombres sentían más calma mental porque partían de más conocimiento previo del tema del curso. Respecto a las ondas cerebrales, las mujeres presentaban más intensidad de ondas beta y gamma. Ambas ondas están relacionadas con concentración y funciones cognitivas como la memoria y la atención selectiva [16, 17], explicando el mayor aprendizaje de las mujeres en la sesión.

Finalmente, para entender cómo afecta el uso de móvil en los estudiantes, comparamos los cambios en las medias de las distintas variables biométricas 15 segundos antes de usar el móvil, durante el uso del móvil y 15 segundos después de usarlo. Los resultados muestran que el pulso aumenta durante y después de usar el móvil, especialmente en los hombres. Nuestra hipótesis es que el incremento de las pulsaciones viene dado por el nerviosismo de mirar y responder los mensajes y la posibilidad de recibir más. Asimismo, la meditación es menor 15 después de usar el móvil que durante el uso del móvil, especialmente en los hombres sugiriendo que esta disminución pueda ser fruto de una preocupación por recibir otro mensaje. Después de usar el móvil la intensidad de las ondas beta y gamma se incrementa con respecto a antes y durante el uso del móvil en ambos sexos, indicando que los estudiantes tienen que concentrarse más para volver a enfocarse en el curso y sugiriendo que el uso del móvil provoca después de su uso un incremento en la carga cognitiva. Por último, la atención es inferior mientras usan el móvil, especialmente en las mujeres. Usar el móvil no parece requerir tanta atención como aprender.

5. Resultados modelos de aprendizaje automático y feedback

Con el objetivo de detectar si un estudiante ha usado el móvil o no, se desarrollaron varios modelos de aprendizaje automáticos unimodales. El objetivo de la predicción es dada una ventana temporal, determinar si ese estudiante ha usado el móvil. Para ello, hemos usado datos de estudiantes del grupo 1 (estudiantes que han sufrido distracciones) y datos de estudiantes del grupo 3 (estudiantes que no han sufrido ninguna distracción). Para obtener los datos de entrenamiento del modelo para cada una de las variables se han seleccionado ventanas de 40 segundos. En el caso del grupo 1, estas ventanas contendrán 20 segundos antes del uso de móvil y 20 segundos en los que habrá uso de móvil. Para los estudiantes del grupo 3 se han seleccionado ventanas también de 40 segundos en las actividades en las que ocurrían las distracciones en el grupo 1. A continuación, esas ventanas se dividen por la mitad, esto es, se obtienen dos ventanas de 20 segundos cada una. Para cada una de esas ventanas se obtienen características globales “hechas a mano” (*handcrafted features*) usadas en el estado del arte [18, 19, 20], que luego se juntan en un solo vector, que es normalizado.

Para cada una de las variables (atención, pulso, meditación, ondas alfa, ondas beta, ondas gamma, ondas delta y ondas theta) se han probado los algoritmos random forest, SVM con núcleo lineal y SVM con núcleo gaussiano con todas las características, 40 de las 65 características seleccionadas con selectKBest y PCA con el número de componentes necesarias para explicar el 95 % de la varianza. Además, se han probado con los datos sin suavizado y con los datos suavizados con ventanas de tamaño 5, 10, 15, 20, 25 y 30. Los modelos que mejor rendimiento obtienen han sido los correspondientes a las variables pulso, ondas beta y ondas gamma con una exactitud de 70 %. Para esas variables se había detectado cambios significativos por el uso del móvil en los análisis estadísticos. Los modelos con peor rendimiento son los de las variables de meditación y de las ondas theta (exactitud del 61 %).

Una vez detectado el uso de móvil es necesario informar o dar feedback a los estudiantes para que puedan mejorar en su aprendizaje. Es por ello, que también se está trabajando en formas de crear un feedback personalizado y adaptado para los estudiantes. Aprovechando los últimos avances en inteligencia artificial generativa con desarrollos como ChatGPT [21], se ha decidido trabajar en el diseño de una herramienta llamada GePeTo (acrónimo en inglés de “**G**enerative **A**I-based **P**ersonalized **G**uidance **T**ool”) que busca poder generar feedback personalizado [22].

6. Conclusiones

En este TFM se ha realizado un análisis de los datos multimodales y biométricos obtenidos en una monitorización con varios biodispositivos de 120 estudiantes mientras aprendían en un MOOC. El objetivo era comprender cómo afectaba el uso de móvil a los estudiantes y por ello se los dividió en tres grupos en función de si podían usar el móvil o no. Los resultados obtenidos han permitido entender de una manera más profunda como afecta el uso de móvil a distintas variables fisiológicas.

Antes de defender este TFM, se publicaron resultados iniciales en tres artículos de conferencia: [23] en el track organizado por el Simposio Internacional de Informática Educativa (SIIE) celebrado en el VII Congreso Español de Informática (CEDI) [24], [22] presentó GePeTo en el congreso internacional EDUCON (*IEEE Global Engineering Education Conference*) [25] y [10] presentó el sistema M2LADS en el XXIII Congreso Internacional de Interacción Persona-Ordenador [26] del año 2023 organizado por AIPO [27]. Después de defender mi TFM, continué con el desarrollo de los modelos de aprendizaje automático, combinando todos ellos y añadiendo nuevas variables biométricas como la pose, lo que ha dado lugar a un artículo de conferencia [28] enviado al EC-TEL25 [29]. También se ha enviado un artículo de revista aceptado [30] en *Scientific Data* [31] (IF: 5.8, Q1) y otro artículo con una revisión sistemática del estado del arte de MMLA [32] a la revista *Behaviour & Information Technology* [33] (IF: 2.9, Q2).

Este TFM obtuvo una calificación de 10 y Mención Honorífica.

Declaración de uso de IA

Los autores declaran que no se ha utilizado ninguna herramienta de inteligencia artificial generativa, como ChatGPT, Gemini o similares, para redactar, traducir o generar contenido textual o visual de este trabajo.

Reconocimientos

Support by projects: *Cátedra ENIA UAM - VERIDAS en IA Responsable* (Next Generation EU, PRTR TSI-100927-2023-2), *HumanCAIC* (TED2021-131787B-I00, MICINN), *SNOLA* (RED2022-134284-T), *e-Madrid-CM* (S2018/TCS-4307). Thanks to the members of the research groups *GHIA* and *BiDA-LAB*.

Referencias

- [1] A. N. Sharin, E-learning during covid-19: a review of literature, *Jurnal Pengajian Media Malaysia* 23 (2021) 15–28.
- [2] R. N. Blasiman, D. Larabee, D. Fabry, Distracted students: A comparison of multiple types of distractions on learning in online lectures., *Scholarship of Teaching and Learning in Psychology* 4 (2018) 222.
- [3] M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, R. Hammad, *The Multimodal Learning Analytics Handbook*, Springer Nature, 2022.
- [4] C. Lang, G. Siemens, A. Wise, D. Gasevic, *Handbook of learning analytics*, SOLAR, Society for Learning Analytics and Research New York, 2017.
- [5] K. Sharma, M. Giannakos, Multimodal data capabilities for learning: What can multimodal data tell us about learning?, *British Journal of Educational Technology* 51 (2020) 1450–1484.
- [6] Grupo GHIA, <https://vghia.ii.uam.es/>, Accedido mayo 2025.
- [7] Grupo BiDA-LAB, <https://bidalab.eps.uam.es/>, Accedido mayo 2025.
- [8] R. Daza, L. F. Gomez, A. Morales, J. Fierrez, R. Tolosana, R. Cobos, J. Ortega-Garcia, MATT: Multimodal Attention Level Estimation for e-learning Platforms, in: *Proc. AAAI Workshop on Artificial Intelligence for Education*, 2023.
- [9] Á. Becerra, R. Daza, R. Cobos, A. Morales, M. Cukurova, J. Fierrez, M2lads: A system for generating multimodal learning analytics dashboards, in: *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, IEEE, 2023, pp. 1564–1569.
- [10] Á. Becerra, R. Daza, R. Cobos, A. Morales, J. Fierrez, User experience study using a system for generating multimodal learning analytics dashboards, in: *Proceedings of the XXIII International Conference on Human Computer Interaction*, 2023, pp. 1–2.
- [11] J. S. Mendoza, B. C. Pody, S. Lee, M. Kim, I. M. McDonough, The effect of cellphones on attention and learning: The influences of time, distraction, and nomophobia, *Computers in Human Behavior* 86 (2018) 52–60.
- [12] R. Daza, A. Morales, R. Tolosana, L. F. Gomez, J. Fierrez, J. Ortega-Garcia, edBB-Demo: Biometrics and Behavior Analysis for Online Educational Platforms, in: *Proc. AAAI Conference on Artificial Intelligence (Demonstration)*, 2023.
- [13] S. Lee, M. W. Kim, I. M. McDonough, J. S. Mendoza, M. S. Kim, The effects of cell phone use and emotion-regulation style on college students' learning, *Applied Cognitive Psychology* 31 (2017) 360–366.
- [14] F.-R. Lin, C.-M. Kao, Mental effort detection using eeg data in e-learning contexts, *Computers & Education* 122 (2018) 63–79.
- [15] K. Umetani, D. H. Singer, R. McCraty, M. Atkinson, Twenty-four hour time domain heart rate variability and heart rate: relations to age and gender over nine decades, *Journal of the American College of Cardiology* 31 (1998) 593–601.

- [16] M. Rangaswamy, B. Porjesz, D. B. Chorlian, K. Wang, K. A. Jones, L. O. Bauer, J. Rohrbaugh, S. J. O'Connor, S. Kuperman, T. Reich, et al., Beta power in the eeg of alcoholics, *Biological psychiatry* 52 (2002) 831–842.
- [17] C. S. Herrmann, A. Mecklinger, Gamma activity in human eeg is related to highspeed memory comparisons during object selective attention, *Visual Cognition* 8 (2001) 593–608.
- [18] J. Fierrez-Aguilar, L. Nanni, J. Lopez-Penalba, J. Ortega-Garcia, D. Maltoni, An on-line signature verification system based on fusion of local and global information, in: *Audio-and Video-Based Biometric Person Authentication: 5th International Conference, AVBPA 2005, Hilton Rye Town, NY, USA, July 20-22, 2005. Proceedings 5*, Springer, 2005, pp. 523–532.
- [19] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Feature-based dynamic signature verification under forensic scenarios, in: *3rd International Workshop on Biometrics and Forensics (IWBF 2015)*, IEEE, 2015, pp. 1–6.
- [20] Z. Mohamed, M. El Halaby, T. Said, D. Shawky, A. Badawi, Characterizing focused attention and working memory using eeg, *Sensors* 18 (2018) 3743.
- [21] ChatGPT, <https://chat.openai.com>, Accedido mayo 2025.
- [22] A. Becerra, Z. Mohseni, J. Sanz, R. Cobos, A generative ai-based personalized guidance tool for enhancing the feedback to mooc learners, in: *2024 IEEE Global Engineering Education Conference (EDUCON)*, May 8-11, 2024, Kos, Greece, 2024.
- [23] Á. Becerra, J. Irigoyen, R. Daza, R. Cobos, A. Morales, J. Fierrez, M. Cukurova, Biometrics and behavioral modelling for detecting distractions in online learning, *CEDI 2024*. (In press), 2024.
- [24] CEDI 2024, <https://www.congresocedi.es/>, Accedido mayo 2025.
- [25] EDUCON24, <https://2024.ieee-educon.org/>, Accedido mayo 2025.
- [26] XXIII Congreso Internacional de Interacción Persona-Ordenador, <https://interaccion2023.udl.cat/>, Accedido mayo 2025.
- [27] AIPO, <https://aipo.es/>, Accedido mayo 2025.
- [28] A. Becerra, R. Daza, R. Cobos, A. Morales, J. Fierrez, M. Cukurova, Ai-based multimodal biometrics for detecting smartphone distractions: Application to online learning, in: *EC-TEL 2025 (Sended)*, Springer, 2025.
- [29] Twentieth European Conference on Technology Enhanced Learning ECTEL 2025, <https://ea-tel.eu/ectel2025>, Accedido mayo 2025.
- [30] R. Daza, A. Becerra, R. Cobos, J. Fierrez, A. Morales, Improve: Impact of mobile phones on remote online virtual education, *arXiv preprint arXiv:2412.14195*. Accepted in *Scientific Data* (2024).
- [31] *Scientific Data*, <https://www.nature.com/sdata/>, Accedido mayo 2025.
- [32] A. Becerra, R. Cobos, C. Lang, Enhancing online learning by integrating biosensors and multi-modal learning analytics for detecting and predicting student behavior: A review, *Behaviour & Information Technology (Major Revision)* (2025).
- [33] *Behaviour & Information Technology*, <https://www.tandfonline.com/journals/tbit20>, Accedido mayo 2025.

PlanTEA-WM: migración y mejora de una aplicación de planificación para personas con TEA

José Lara^{1,*†}, Ana I. Molina^{1,†} and Carmen Lacave^{1,†}

¹Universidad de Castilla-La Mancha, Avenida Camilo José Cela, s/n, 13071 Ciudad Real, España.

Resumen

PlanTEA-WM es una plataforma *web* que permite la planificación visual de rutinas para personas con Trastorno del Espectro Autista (TEA). Surge como evolución de la *app* PlanTEA para Android, incorporando mejoras clave tras un proceso de diseño colaborativo con asociaciones expertas en TEA (AUTRADE, FACLM y FESPAU). Entre sus funcionalidades, destacan la gestión multiusuario con roles diferenciados, el acceso desde distintos dispositivos, una vista de calendario, el buscador de pictogramas conectado con la base de datos de ARASAAC, la traducción de texto a pictogramas y la generación de rutinas mediante Inteligencia Artificial Generativa (IAG). El desarrollo ha estado guiado por principios de accesibilidad y usabilidad, aplicando heurísticas específicas para personas con TEA, pruebas de operabilidad y validaciones de contraste cromático, siguiendo las pautas WCAG. Las validaciones realizadas con profesionales y asociaciones expertas en TEA han permitido ajustar la herramienta a las necesidades reales del entorno clínico y familiar, dando lugar a una solución práctica, accesible y centrada en el usuario.

Palabras clave

Tecnologías asistivas, Trastorno del Espectro Autista (TEA), Planificación y anticipación, Sistemas Aumentativos y Alternativos de Comunicación (SAAC), Pictogramas, Inteligencia Artificial Generativa (IAG), Diseño Centrado en el Usuario, Usabilidad, Accesibilidad

1. Introducción

Este trabajo describe el desarrollo de **PlanTEA-WM**, la versión *web* de la aplicación *PlanTEA* [1], una *app* para dispositivos Android de soporte a la planificación y anticipación de actividades para personas con Trastorno del Espectro Autista (TEA). El TEA es un trastorno del neurodesarrollo con una fuerte base genética que se manifiesta desde una edad temprana y se caracteriza principalmente por alteraciones en la adquisición de habilidades en múltiples dominios, rutinas rígidas, intereses repetitivos y déficits persistentes en la socialización [2].

En el contexto de los Sistemas Aumentativos y Alternativos de Comunicación (SAAC), los pictogramas constituyen un recurso visual clave para representar rutinas, anticipar acciones y fomentar la autonomía [3]. Además, su uso ha demostrado ser eficaz para reducir la incertidumbre y la ansiedad que pueden generar los cambios inesperados o la falta de claridad en las indicaciones [4].

PlanTEA fue concebida como una herramienta para ayudar a familiares de personas con TEA, cuidadores, terapeutas y profesionales del ámbito clínico en la creación y seguimiento de secuencias visuales de actividades representadas mediante pictogramas, que facilitan la estructuración, planificación y anticipación de rutinas. Aunque su versión inicial ha tenido una valoración positiva entre las asociaciones colaboradoras, una evaluación preliminar permitió identificar diferentes limitaciones: depende de una base de datos local, lo cual obliga a usar la aplicación en un único dispositivo, y carece de capacidades multiusuario que permitan la colaboración entre distintos agentes implicados en el cuidado o el apoyo de una persona con TEA. Estas carencias limitan su adopción en entornos que exigen gestionar las rutinas de uno o varios individuos con TEA entre distintas personas.

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain.

*Autor de correspondencia.

†Estos autores han contribuido por igual a este trabajo.

✉ Jose.Lara3@alu.uclm.es (J. Lara); AnaIsabel.Molina@uclm.es (A. I. Molina); Carmen.Lacave@uclm.es (C. Lacave)

🌐 <https://jose02.github.io/> (J. Lara); <https://blog.uclm.es/anaisabelmolina/> (A. I. Molina); <https://blog.uclm.es/carmenlacave/> (C. Lacave)

🆔 0009-0001-9879-5782 (J. Lara); 0000-0002-3449-2539 (A. I. Molina); 0000-0003-2770-8482 (C. Lacave)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Para abordar estas limitaciones, se ha contado con la participación activa de asociaciones del ámbito del TEA a nivel regional y nacional, que han participado en el análisis de requisitos y en distintas validaciones basadas en su experiencia directa. Entre ellas se encuentran **AUTRADE** (*Asociación Regional de Afectados de Autismo y Otros Trastornos del Desarrollo*), **FACLM** (*Federación Autismo Castilla-La Mancha*) y **FESPAU** (*Federación Española de Autismo*). A través de reuniones periódicas y jornadas oficiales con estas entidades, se ha refinado la interfaz y priorizado aquellas funcionalidades esenciales, orientando el desarrollo hacia una mayor accesibilidad y usabilidad, con especial atención a la anticipación de actividades, la colaboración multiusuario y multirrol y la facilidad de uso para personas con TEA.

Como resultado, se tiene una solución escalable y colaborativa, *PlanTEA-WM*, que proporciona:

- **Almacenamiento centralizado** en una base de datos relacional (PostgreSQL), posibilitando el acceso desde cualquier navegador y dispositivo con conexión a Internet.
- **Gestión multirrol**: Se definen los roles de *planificador* (por ejemplo, familiares o terapeutas) y usuario *planificado* (persona con TEA). El primero es el rol encargado de la creación y gestión de las planificaciones, mientras que el rol de *planificado* es el que tiene acceso a una interfaz simplificada, en forma de reproductor, que permite al usuario con TEA visualizar y hacer el seguimiento de sus rutinas de forma secuencial.
- **Gestión multiusuario**: Se permite que un mismo usuario *planificado* pueda ser supervisado por varios usuarios *planificadores*. Además, un *planificador* puede gestionar a varios *planificados*.
- **Funciones avanzadas** de búsqueda de pictogramas (API de ARASAAC¹), traducción de texto a pictogramas y recomendación de rutinas mediante Inteligencia Artificial Generativa (IAG).
- **Interfaz accesible**: Se ha aplicado un Diseño Centrado en el Usuario con la colaboración de expertos en TEA, priorizando la sencillez y la coherencia en la navegación.

El desarrollo de *PlanTEA-WM* ha coincidido en el tiempo con la formalización de diferentes contratos de investigación en el seno del **grupo CHICO**² (*Computer Human Interaction and Collaboration*) de la UCLM (Universidad de Castilla-La Mancha), en el contexto de los proyectos de investigación **APTEA** (Ref. TED2021-131956B-I00) y **PlanTEAAF** (Ref. 2022-GRIN-34175).

2. Objetivos

El objetivo general del Trabajo Final de Grado (TFG) es desarrollar una plataforma *web* robusta y escalable (*PlanTEA-WM*) que **extienda las funcionalidades de la aplicación *PlanTEA***, superando sus limitaciones y permitiendo la planificación de rutinas con pictogramas y la colaboración multiusuario. De este proyecto se derivan los siguientes objetivos parciales:

1. **Desarrollar un sistema multirrol y multiusuario**, que gestione adecuadamente el rol *planificador* (encargado de crear y gestionar rutinas) y el rol *planificado* (personas con TEA que siguen y visualizan las actividades marcadas en dichas rutinas o planificaciones). Además, un *planificador* podrá gestionar a varios *planificados*, y un usuario *planificado* podrá, a su vez, ser supervisado por diferentes *planificadores*.
2. **Ofrecer una vista de calendario**, que permita asociar cada planificación a un día y hora concretos, programar eventos y marcarlos como “visibles” para el usuario con TEA *planificado*.
3. **Integrar un buscador de pictogramas**, enlazado con la API de ARASAAC, que permita ampliar el catálogo de pictogramas disponibles y simplificar la creación de secuencias.
4. **Incorporar un traductor de texto a pictogramas**, que automatice la conversión de oraciones o instrucciones a secuencias visuales fáciles de comprender.

¹ARASAAC (*Centro Aragonés para la Comunicación Aumentativa y Alternativa*) es un proyecto que proporciona pictogramas y otros recursos gráficos para apoyar la comunicación de personas con dificultades con el lenguaje oral. Ofrece una API que permite acceder a su catálogo de pictogramas, facilitando así su integración en aplicaciones y plataformas externas. Más información en <https://arasaac.org>.

²<https://blog.uclm.es/grupochico/>

5. **Implantar una funcionalidad de IAG** en el traductor en la que, dada una instrucción textual (*prompt*), genere automáticamente rutinas, para posteriormente traducirlas a secuencias de pictogramas.
6. **Implementar un sistema de invitaciones**, para que múltiples profesionales o familiares colaboren en la gestión de un mismo usuario con TEA.
7. **Exportar e importar planificaciones** en formatos de amplio uso (XML, PDF), favoreciendo así la interoperabilidad y la portabilidad de las rutinas.
8. **Asegurar la accesibilidad y usabilidad**, teniendo en cuenta las Pautas de Accesibilidad para el Contenido Web (WCAG) y las recomendaciones de las entidades colaboradoras, garantizando un contraste cromático adecuado y un diseño centrado en la persona con TEA [5].

3. Resultados

Los resultados de *PlanTEA-WM* se han estructurado en tres bloques principales: la **reingeniería del sistema** y sus funcionalidades clave, la validación con asociaciones de TEA y la aplicación de **heurísticas de accesibilidad y usabilidad**.

3.1. Migración y funcionalidades clave

En las siguientes secciones se presentan las principales mejoras funcionales y técnicas introducidas en *PlanTEA-WM*, incluyendo el rediseño de la arquitectura, nuevas herramientas para las planificaciones y mecanismos que favorecen la colaboración entre distintos usuarios.

3.1.1. Arquitectura multiusuario y calendario

Se ha desarrollado un *backend* con NodeJs + Express, almacenando la información en PostgreSQL. Para desarrollar la Interfaz de Usuario (IU) se ha hecho uso de Flutter Web, que permite el acceso multiplataforma [6].

En la IU, se ha incluido un **dashboard** de gestión de *planificados* y sus eventos, y una **vista de calendario**, en la que cada *planificador* podrá crear eventos y asociarlos a una planificación (secuencia de pictogramas). El *planificador* podrá marcar uno de esos eventos como “visible” para la persona con TEA, que podrá consultarlo y seguir la planificación asociada al evento activo.

3.1.2. Buscador y traductor de pictogramas

Para hacer uso del **buscador**, el usuario introducirá una palabra clave. A continuación, la aplicación se conectará con la API de ARASAAC, obteniendo todas las coincidencias en formato de pictograma para dicho término, que podrán añadirse a la planificación.

Por su parte, el **traductor** cuenta con un modo “**normal**”, que divide el texto introducido en palabras, las lematiza (por ejemplo, “*jugando*” → “*jugar*”), para mejorar la tasa de coincidencias en ARASAAC, y genera una secuencia de pictogramas. En el modo **IA** (Inteligencia Artificial), integrado con IAG, el usuario introduce un *prompt* como “*Explica cómo lavarse las manos en pasos simples*” y el sistema (mediante HuggingFace) genera una respuesta simplificada, que posteriormente se transforma en una secuencia de pictogramas.

3.1.3. Compartición de usuarios

El *planificador* puede invitar a otro planificador (profesional, cuidador, familiar) a gestionar a un *planificado* (usuario con TEA) concreto, enviándose de forma automática un enlace por correo electrónico (mediante Brevo). Si el destinatario acepta, se añade automáticamente la asociación en la base de datos. Este mecanismo facilita la colaboración multiusuario, permitiendo la gestión conjunta de una misma persona con TEA desde diferentes perfiles de *planificador*.

3.1.4. Importación y exportación de planificaciones

Se permite al usuario *planificador* guardar o cargar las rutinas en formato **XML** o **PDF**. De esta forma, las planificaciones podrán reutilizarse o compartirse en otros contextos sin necesidad de acceder directamente al *backend*.

3.2. Validación con asociaciones colaboradoras

La participación activa de asociaciones expertas en TEA, como **AUTRADE**, **FACLM** y **FESPAU**, ha sido esencial para asegurar la pertinencia y la utilidad de *PlanTEA-WM*. Dicha colaboración se organizó en varias reuniones y jornadas específicas:

- **Reuniones en Ciudad Real y en Madrid:** en ellas, se revisaron prototipos y se recogió el *feedback* de terapeutas y expertas en TEA sobre la usabilidad y funcionalidades del sistema (adecuación del calendario, la necesidad de compartir usuarios, etc).
- **Jornadas de trabajo:** se efectuaron diferentes demostraciones de la funcionalidad de la IAG para proponer secuencias de rutinas cotidianas. Se destaca la reunión que se tuvo con Pedro Ugarte, presidente de la Confederación Autismo España, en el laboratorio de usabilidad del grupo CHICO, quien valoró muy positivamente la integración de la IAG en el sistema y el ahorro de tiempo que supondría para los profesionales.

Entre otros aspectos, las diferentes asociaciones destacaron la importancia de la interfaz simplificada y los colores cuidadosamente elegidos, fundamentales para las personas con TEA.

3.3. Aplicación de heurísticas de accesibilidad y chequeo cromático

En cuanto a la **accesibilidad** y **usabilidad** se han aplicado criterios basados en las WCAG (1.4.3, 1.4.6, etc.) [7], así como guías específicas centradas en el diseño de interfaces accesibles para personas con TEA, que recogen recomendaciones sobre estructura visual, claridad, consistencia y reducción de carga cognitiva [5, 8]:

1. **Heurísticas de accesibilidad**, como el uso de mensajes de error claros y textos descriptivos en elementos interactivos y un diseño centrado en el usuario con TEA: secuencias de pictogramas ordenadas, panel de control simplificado y consistencia en la ubicación de menús.
2. **Pruebas de operabilidad:** navegación por teclado, estructura de menú coherente y formularios con instrucciones breves. Todo ello reduce la sobrecarga cognitiva y facilita la previsión de acciones en usuarios con TEA.
3. **Chequeo cromático:** se comprobó la paleta de colores utilizada, mediante herramientas de validación, como WhoCanUse, asegurando así un contraste adecuado (AA e incluso AAA) en los textos y botones.

4. Conclusión

PlanTEA-WM resuelve las limitaciones de la versión para dispositivos móviles (*PlanTEA*), habilitando la gestión colaborativa, la anticipación de rutinas mediante un calendario y la traducción de texto a pictogramas (incluyendo IAG). Las validaciones con asociaciones de TEA confirman que este enfoque aporta un valor real, no solo en entornos clínicos, sino también en el ámbito familiar. Además, la aplicación de heurísticas de usabilidad y accesibilidad y la verificación del contraste cromático han garantizado un diseño usable y coherente.

Entre las líneas de **trabajo futuro** se incluyen la introducción de un cuaderno completo de comunicación, la integración con servicios de notificaciones y calendarios externos y la realización de una evaluación formal de la aplicación en contextos clínicos y educativos, para medir su impacto en la autonomía de las personas con TEA. Con todo ello, *PlanTEA-WM* se presenta como una solución versátil y abierta a la evolución, diseñada específicamente para facilitar la planificación y la comunicación visual de las rutinas de usuarios con TEA.

Declaración sobre el uso de IA Generativa

Durante la preparación de este trabajo, los autores utilizaron herramientas de apoyo basadas en inteligencia artificial para realizar revisiones gramaticales y de estilo. Todo el contenido fue revisado y editado por los propios autores, quienes asumen la plena responsabilidad del trabajo presentado.

Referencias

- [1] P. Hernández, A. I. Molina, C. Lacave, C. Rusu, A. Toledano-González, PlanTEA: Supporting planning and anticipation for children with ASD attending medical appointments, *Applied Sciences* 12 (2022). doi:10.3390/app12105237.
- [2] C. Lord, M. Elsabbagh, G. Baird, J. Veenstra-Vanderweele, Autism spectrum disorder, *The Lancet* 392 (2018) 508–520. doi:10.1016/S0140-6736(18)31129-2.
- [3] P. Morales-Hidalgo, J. Roigé-Castellví, C. Hernández-Martínez, N. Voltas, J. Canals, Prevalence and characteristics of autism spectrum disorder among Spanish school-age children, *Journal of Autism and Developmental Disorders* 48 (2018) 3176–3190. doi:10.1007/s10803-018-3581-2.
- [4] H. Neimy, B. Fossett, *Augmentative and Alternative Communication (AAC) Systems*, Springer International Publishing, Cham, 2022, pp. 375–401. doi:10.1007/978-3-030-96478-8_20.
- [5] K. Valencia, C. Rusu, F. Botella, User experience factors for people with autism spectrum disorder, *Applied Sciences* 11 (2021). doi:10.3390/app112110469.
- [6] R. T. Fielding, *Architectural Styles and the Design of Network-based Software Architectures*, Ph.d. dissertation, University of California, Irvine, Irvine, CA, 2000. URL: <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>.
- [7] World Wide Web Consortium (W3C), Web Content Accessibility Guidelines (WCAG) 2.1, <https://www.w3.org/TR/WCAG21/>, 2018. Accessed: May 6, 2025.
- [8] Y. P. C. Aguiar, E. Galy, A. Godde, M. Trémaud, C. Tardif, AutismGuide: a usability guidelines to design software solutions for users with autism spectrum disorder, *Behaviour & Information Technology* 41 (2022) 1132–1150. doi:10.1080/0144929X.2020.1856927.

Índice de autores

- Albiol-Pérez, Sergio . . Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales, [14](#)
- Alonso-Moral, Jose M. . . AffeX: a real-time emotion monitoring and visualization tool, [164](#)
- Arribas Mateo, Adrian . . . Desarrollo de un módulo conversacional basado en IA para la interacción emocional con robots sociales, [35](#)
- Azuara, Guillermo . Primeras experiencias para el aprendizaje de ciberseguridad a través del uso de robots sociales, [109](#)
- Baldassarri, Sandra Desarrollo de un módulo conversacional basado en IA para la interacción emocional con robots sociales, [35](#)
- Baldassarri, Sandra SanbotPPT: presentador configurable para un robot social, [127](#)
- Baldassarri, Sandra . . . TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Baldassarri, Sandra . . “Robots in the wild”: grupos intergeneracionales y robots sociales en espacios públicos, [49](#)
- Barrado Lucía, Fernando Robots sociales en museos – un enfoque centrado en el análisis de emociones, [118](#)
- Bautista, Susana Evaluación de la comprensión de las pautas WCAG 2.1: análisis de ejemplos prácticos en tres universidades españolas, [80](#)
- Bautista, Susana . Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales, [14](#)
- Becerra, Alvaro . . Análisis y modelado del comportamiento de estudiantes en cursos en línea desde el enfoque de la Analítica del Aprendizaje Multimodal, [203](#)
- Buades Rubio, José María . Robots sociales en museos – un enfoque centrado en el análisis de emociones, [118](#)
- Catala, Alejandro AffeX: a real-time emotion monitoring and visualization tool, [164](#)
- Catala, Alejandro Evaluación de detectores de estrés durante la interacción con juegos serios que estimulan actividades cognitivas, [64](#)
- Cea-López, Margarita SimpLee: an artificial intelligence agent for Easy Reading in support of demographic challenge, [172](#)
- Cerezo, Eva Desarrollo de un módulo conversacional basado en IA para

- la interacción emocional con robots sociales, [35](#)
- Cerezo, Eva SanbotPPT: presentador configurable para un robot social, [127](#)
- Cerezo, Eva TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Cerezo, Eva . . “Robots in the wild”: grupos intergeneracionales y robots sociales en espacios públicos, [49](#)
- Cobos, Ruth Análisis y modelado del comportamiento de estudiantes en cursos en línea desde el enfoque de la Analítica del Aprendizaje Multimodal, [203](#)
- Collazos, César A. Heurísticas de evaluación de usabilidad para plataformas de comercio electrónico, [196](#)
- Condori-Fernández, Nelly AffeX: a real-time emotion monitoring and visualization tool, [164](#)
- Condori-Fernández, Nelly . . Evaluación de detectores de estrés durante la interacción con juegos serios que estimulan actividades cognitivas, [64](#)
- Costa, Ánxela Pérez Virtual interfaces and augmented reality for orthopedic surgery, [177](#)
- Costa-Tebar, Felipe . . . TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Cubero, Guillermo Entrando en Oz: simulación virtual y evaluación de robots sociales a través de técnicas de Mago de Oz, [199](#)
- Figueroa Vásquez, Salvador Alonso AffeX: a real-time emotion monitoring and visualization tool, [164](#)
- Gallardo, Jesús Primeras experiencias para el aprendizaje de ciberseguridad a través del uso de robots sociales, [109](#)
- Gallud, Jose A. TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Garrido, Juan E. TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Gaya Morey, Francesc Xavier Robots sociales en museos – un enfoque centrado en el análisis de emociones, [118](#)
- Ghazel, Adel EXPLainable Artificial INtelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME), [186](#)
- Gil-González, Ana-Belén SimpLee: an artificial intelligence agent for Easy Reading in support of demographic challenge, [172](#)
- Grande-Pérez, Laura SimpLee: an artificial intelligence agent for Easy Reading in support of demographic challenge, [172](#)
- Granollers i Saltiveri, Antoni Heurísticas de evaluación de usabilidad para plataformas de comercio electrónico, [196](#)
- Guillén Florian, Alejandro Robots sociales en museos – un enfoque centrado en el análisis de emociones, [118](#)
- Gutiérrez-Vela, Francisco Luis Diseño de un

- modelo de empatía en diálogos de agentes sociales interactivos para la mejora de la aceptación tecnológica, [181](#)
- Gutiérrez-Vela, Francisco Luis . . . Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales, [14](#)
- Hervás, Eva Primeras experiencias para el aprendizaje de ciberseguridad a través del uso de robots sociales, [109](#)
- Hervás, Ramón Entrando en Oz: simulación virtual y evaluación de robots sociales a través de técnicas de Mago de Oz, [199](#)
- Hitos-García, Alberto Diseño de un modelo de empatía en diálogos de agentes sociales interactivos para la mejora de la aceptación tecnológica, [181](#)
- Hitos-García, Alberto . Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales, [14](#)
- Hupont, Isabelle . . . EXPLainable Artificial Intelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME), [186](#)
- Hupont, Isabelle . . . HCI en tiempos de IA: ética, diseño y regulación, [161](#)
- Jaume-i-Capó, Antoni EXPLainable Artificial Intelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME), [186](#)
- Jiménez Sánchez, Alberto SimpLee: an artificial intelligence agent for Easy Reading in support of demographic challenge, [172](#)
- Lacave, Carmen PlanTEA-WM: migración y mejora de una aplicación de planificación para personas con TEA, [209](#)
- Lacuesta, Raquel . . . Primeras experiencias para el aprendizaje de ciberseguridad a través del uso de robots sociales, [109](#)
- Lacuesta, Raquel Robots sociales en museos – un enfoque centrado en el análisis de emociones, [118](#)
- Lacuesta, Raquel “Robots in the wild”: grupos intergeneracionales y robots sociales en espacios públicos, [49](#)
- Lara, José PlanTEA-WM: migración y mejora de una aplicación de planificación para personas con TEA, [209](#)
- Leiva, Luis A. EXPLainable Artificial Intelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME), [186](#)
- Li, Shuoshuo Automatización del protocolo Thinking Aloud en evaluaciones de usabilidad, [23](#)
- López Bernal, Sergio Interfaces cerebro-máquina: hacia una nueva era en la interacción persona-ordenador, [162](#)
- Lozano, María D. TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Luna-Lorente, Belen . TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)

- Macías, José A. Automatización del protocolo Thinking Aloud en evaluaciones de usabilidad, [23](#)
- MacKenzie, I. Scott EXPLainable Artificial Intelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME), [186](#)
- Manresa-Yee, Cristina EXPLainable Artificial Intelligence Methods and Evaluation for trustworthy, human-centric, and ethical AI (EXPLAINME), [186](#)
- Manso-Conde, Javier SimpLee: an artificial intelligence agent for Easy Reading in support of demographic challenge, [172](#)
- Martínez-Limeres, Leticia . . Evaluación de detectores de estrés durante la interacción con juegos serios que estimulan actividades cognitivas, [64](#)
- Matinero, Loreto . SanbotPPT: presentador configurable para un robot social, [127](#)
- Matinero, Loreto “Robots in the wild”: grupos intergeneracionales y robots sociales en espacios públicos, [49](#)
- Medina-Medina, Nuria Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales, [14](#)
- Méndez Flórez, Juan Camilo AffeX: a real-time emotion monitoring and visualization tool, [164](#)
- Mendoza, Liany . Desarrollo de un módulo conversacional basado en IA para la interacción emocional con robots sociales, [35](#)
- Molina, Ana I. PlanTEA-WM: migración y mejora de una aplicación de planificación para personas con TEA, [209](#)
- Navarro, Alejandro . . TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Ormeño, Yeshica Isela Resumen del artículo: un experimento empírico de un método de elicitación de requisitos de usabilidad para diseñar GUI basadas en entrevistas, [191](#)
- Orozco, Luis TecnoCRA: entorno virtual de aprendizaje gamificado de soporte a la docencia en colegios rurales agrupados, [168](#)
- Paderewski-Rodríguez, Patricia . Diseño de un modelo de empatía en diálogos de agentes sociales interactivos para la mejora de la aceptación tecnológica, [181](#)
- Paderewski-Rodríguez, Patricia . . Guías de usabilidad para mejorar la interacción de adultos mayores con Asistentes Virtuales, [14](#)
- Panach, José Ignacio Resumen del artículo: un experimento empírico de un método de elicitación de requisitos de usabilidad para diseñar GUI basadas en entrevistas, [191](#)
- Pascual-Almenara, Afra . . Evaluación de la comprensión de las pautas WCAG 2.1: análisis de ejemplos prácticos en tres universidades españolas, [80](#)
- Pastor, Oscar Resumen del artículo: un experimento empírico de un método de elicitación de requisitos

- de usabilidad para diseñar GUI
basadas en entrevistas, [191](#)
- Pedrero-Esteban, Alberto Evaluación de la
comprensión de las pautas WCAG
2.1: análisis de ejemplos prácticos
en tres universidades españolas, [80](#)
- Penichet, Víctor M. R. TecnoCRA: entorno
virtual de aprendizaje gamificado
de soporte a la docencia en
colegios rurales agrupados, [168](#)
- Perales López, Francisco José Robots
sociales en museos – un enfoque
centrado en el análisis de
emociones, [118](#)
- Pérez, Daniel M. . . . Primeras experiencias
para el aprendizaje de
ciberseguridad a través del uso de
robots sociales, [109](#)
- Plazas, Cristian Modelado de
audiodescripción prosódica basado
en narrativa semántica, [144](#)
- Polo-Martinez, Lucía . TecnoCRA: entorno
virtual de aprendizaje gamificado
de soporte a la docencia en
colegios rurales agrupados, [168](#)
- Quintero, Christian Modelado de
audiodescripción prosódica basado
en narrativa semántica, [144](#)
- Rodríguez Marcos, Clara SimpLee: an
artificial intelligence agent for Easy
Reading in support of demographic
challenge, [172](#)
- Rozo-Torres, Alexander Modelado de
audiodescripción prosódica basado
en narrativa semántica, [144](#)
- Sebastián Rivera, Gabriel TecnoCRA:
entorno virtual de aprendizaje
gamificado de soporte a la docencia
en colegios rurales agrupados, [168](#)
- Souto, Nieves Pedreira . . Virtual interfaces
and augmented reality for
orthopedic surgery, [177](#)
- Tesoriero, Ricardo TecnoCRA: entorno
virtual de aprendizaje gamificado
de soporte a la docencia en
colegios rurales agrupados, [168](#)
- Vargas Muñoz, Alejandra Yicel Heurísticas
de evaluación de usabilidad para
plataformas de comercio
electrónico, [196](#)
- Villa, Laura . . Entrando en Oz: simulación
virtual y evaluación de robots
sociales a través de técnicas de
Mago de Oz, [199](#)