

Interacción natural y emocional con robots sociales

Natural and emotional interaction with social robots

Liany Mendoza

Departamento de
Informática e Ingeniería
de Sistemas, Universidad
de Zaragoza, Zaragoza,
España
lmendoza@unizar.es

Eva Cerezo

Departamento de
Informática e
Ingeniería de Sistemas,
Universidad de
Zaragoza, Zaragoza,
España
ecerezo@unizar.es

Loreto Matinero

Departamento de
Informática e Ingeniería
de Sistemas,
Universidad de
Zaragoza, Zaragoza,
España
lmatinero@unizar.es

Adrián Arribas

Departamento de
Informática e Ingeniería
de Sistemas,
Universidad de
Zaragoza, Zaragoza,
España
795593@unizar.es

Sandra Baldassarri

Departamento de
Informática e Ingeniería
de Sistemas, Universidad
de Zaragoza, Zaragoza,
España
sandra@unizar.es

Recibido: 15.11.2025 | Aceptado: 01.12.2025

Palabras Clave

Robots sociales
Computación afectiva
LLM
Interacción natural
Interacción emocional

Resumen

En los últimos años, se ha demostrado la gran cantidad de beneficios que los robots sociales pueden ofrecer a las personas, tanto como herramientas de aprendizaje como en funciones de acompañamiento social. Para que estas interacciones sean efectivas, es necesario que los robots sean capaces de conversar de forma inteligente y actúen de forma natural y emocionalmente coherente. En este trabajo se presenta un sistema orientado a mejorar la interacción humano-robot mediante la integración de herramientas de inteligencia artificial basada en modelos de lenguaje (LLM), interfaces gráficas (visual y táctil), movimientos, luces y voz. El sistema, implementado en Android sobre el robot Sanbot Elf, se compone de tres módulos principales: el Módulo Conversacional, el Módulo Emocional y el Módulo Reactivo, diseñados para lograr un diálogo fluido, expresar estados emocionales y aportar naturalidad al comportamiento del robot. Se realizó una evaluación inicial del sistema conversacional con 18 usuarios, cuyos resultados fueron muy positivos, y dieron lugar a una serie de mejoras para optimizar el desempeño y la experiencia de interacción.

Keywords

Social robots
Affective computing
LLM
Natural interaction
Emotional interaction

Abstract

In recent years, social robots have been shown to offer a wide range of benefits to people, both as learning tools and in social companionship roles. For these interactions to be effective, robots need to be able to converse intelligently and act in a natural and emotionally coherent manner. This paper presents a system aimed at improving human-robot interaction through the integration of artificial intelligence tools based on language models (LLM), graphical interfaces (visual and tactile), movements, lights, and voice. The system, implemented in Android on the Sanbot Elf robot, consists of three main modules: the Conversational Module, the Emotional Module, and the Reactive Module, designed to achieve fluid dialogue, express emotional states, and bring naturalness to the robot's behavior. An initial evaluation of the conversational system was carried out with 18 users, the results of which were very positive and led to a series of improvements to optimize performance and interaction experience.

1. Introducción

Los robots sociales están adquiriendo una presencia cada vez más significativa en la sociedad actual. Más allá del uso de robots para trabajo y asistencia en tareas industriales que requieren alta precisión y exactitud, los robots sociales están comenzando a integrarse en la vida cotidiana como

asistentes, ayudantes, tutores e incluso compañeros (Van y Reed, 2010). Aunque las definiciones de los robots sociales son heterogéneas (Henschel, Laban y Cross, 2021), en el trabajo de Sarica, Brondi y Fortunati (2020) se identificaron ciertos rasgos comunes: los robots sociales son agentes corpóreos que poseen distintos grados de autonomía y participan en interacciones sociales con humanos,

comunicándose, cooperando y tomando decisiones. Este componente social es fundamental para que los robots proporcionen apoyo físico y emocional, manteniendo interacciones con las personas a largo plazo (Churamani, Kalkan y Gunes, 2020).

A pesar de que los robots no igualan las habilidades humanas en la interacción social (Cross, Hortensius y Wykowska, 2019), diversos estudios destacan su potencial cuando se introducen de manera contextualizada y ética (Wullenkord y Eyssel, 2020), mejorando la calidad de vida y promoviendo el bienestar (Yang et al., 2020). Una base de pruebas cada vez mayor documenta cómo los robots sociales podrían funcionar como herramientas autónomas de apoyo en intervenciones cognitivas (Alnajjar et al., 2019), rehabilitación física y fisioterapia (Assad-Uz-Zaman et al., 2019; Chen, García-Vergara y Howard, 2018; Mohebbi, 2020), intervenciones psicosociales y mejoras en el bienestar (Robinson, Cottier y Kavanagh, 2019; Scoglio et al., 2019), así como en el ámbito educativo y del entretenimiento (Kyprianou et al., 2023; Calvo-Barajas, Perugia y Castellano, 2020). En los últimos años, los robots sociales diseñados para acompañar a las personas se han vuelto más comunes, especialmente en contextos de atención a mayores o niños siendo de gran importancia para su aceptación, que estos robots sociales sean capaces de mantener conversaciones naturales y empáticas, y tengan en cuenta aspectos emocionales. Este tipo de sistemas ha crecido gracias a la integración de modelos de IA y del procesamiento del lenguaje natural (Spezialetti, Placidi y Rossi, 2020). La naturalidad conversacional es un elemento esencial para lograr interacciones significativas entre humanos y robots (Romat et al., 2016). Las técnicas modernas de IA, especialmente los modelos de lenguaje de gran tamaño (LLM), permiten mantener conversaciones más fluidas y contextualmente coherentes, considerando emociones, tonos y contextos del interlocutor (Laban y Cross, 2024). Asimismo, la expresividad corporal y la comunicación no verbal son fundamentales para la credibilidad y eficacia de la interacción (Breazeal, 2003). En este trabajo se propone el desarrollo de un sistema conversacional basado en un modelo de lenguaje de gran tamaño (LLM), que aprovecha las capacidades de interacción multimodal de un robot social (Sanbot Elf) que soporta la interacción natural, emocional y personalizada humano-robot. Además, se ha considerado importante que el robot actúe de forma natural imitando el comportamiento humano mediante gestos y comportamientos no verbales.

A continuación, en el apartado 2, se realiza el estado del arte donde se abordan las principales características que un robot debe tener para lograr una buena interacción entre el

humano y el robot. En el apartado 3 se describe el robot, las consideraciones de diseño e implementación del sistema conversacional; mientras que el apartado 4 se ofrece un ejemplo de conversación entre una persona y el robot. Luego, en el apartado 5, se describe la evaluación llevada a cabo con usuarios y las mejoras realizadas en base a los resultados obtenidos. Por último, se finaliza con las conclusiones de la investigación y el trabajo futuro.

2. Estado del arte

Para lograr una conversación natural y emocional entre humanos y robots es necesario comprender cómo un robot puede desarrollar comportamientos sociables. Para ello, se ha realizado una revisión de la literatura, con el propósito de identificar y analizar las características de comunicación y sociabilidad que deben integrarse en el diseño y funcionamiento de los robots sociales que permitan una interacción emocional y natural cuando conversan con las personas. 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 (Wang et al., 2024). En este sentido, se están introduciendo como una herramienta fundamental de soporte a la conversación humano-robot (Kim, Lee y Mutlu, 2024; Cherakara et al., 2023) debido a su gran potencial en el desarrollo de aplicaciones conversacionales capaces de expresar y reconocer emociones humanas, ampliando las fronteras de la interacción humano-robot (Jiang et al., 2025). En cuanto a su uso en robots sociales, Rawal et al. (2024) proponen un agente multimodal que integra información visual y textual para generar respuestas empáticas, combinando las expresiones faciales del usuario utilizando el modelo LLaMA2 para que el sistema reconozca y refleje estados emocionales, generando interacciones naturales y afectivas. Recientemente, Pinto-Bernal, Biondina y Belpaeme (2025) demuestran que la incorporación de LLMs en robots sociales posibilita interacciones más fluidas, coherentes y culturalmente adaptadas, al incluir mecanismos de memoria conversacional, ajuste de estilo y adaptación multilingüe. Ambos trabajos evidencian que estos modelos no solo mejoran la coherencia lingüística y contextual de los diálogos, sino que también permiten diseñar sistemas conversacionales emocionalmente inteligentes, capaces de establecer vínculos sociales más cercanos y significativos con los usuarios.

Sin embargo, no es suficiente con que el robot sea capaz de sostener una conversación verbal, sino que también debe emplear mecanismos de comunicación no verbal, tales como las emociones y las expresiones faciales (Bonarini, 2020). En

diferentes contextos, como el educativo se ha observado (Kyprianou et al., 2023) que los estudiantes esperan que los robots 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, 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 (Calvo-Barajas, Perugia y Castellano, 2020) 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 (2015), 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 (Paiva, 2017) constituye un aspecto clave en la interacción entre humanos y robots (Brave, Nass y Hutchinson, 2005; Leite et al., 2013).

Para lograr estos objetivos, es fundamental el estudio e incorporación de componentes emocionales (Gou et al., 2014), ya que los robots que manifiestan emociones mediante gestos o expresiones faciales 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 (Xu et al., 2015), lo que, a su vez, contribuye a que sea visto como confiable y empático. Diversos estudios demuestran, además, que 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 (Lee et al., 2012; Dziergwa et al., 2018). Por tanto, aunque es esencial mejorar la funcionalidad de los robots sociales, también es importante que el robot sea capaz de comunicarse de una forma que se considere socialmente aceptable (De Graaf, Allouch y Van Dijk, 2015). Por otra parte, es importante también que el robot tenga un comportamiento similar al de un humano (Melo y Moreno, 2022), por lo que debe ser capaz de reaccionar en tiempo real a estímulos del entorno durante una conversación (Reimann et al., 2024). Esta combinación de elementos verbales y no verbales (Babel et al., 2021) 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, señales verbales y no verbales, y

comportamientos naturales o reactivos para lograr una comunicación real entre humanos y robots.

Los estudios mencionados previamente abordan, de forma separada, diferentes aspectos. Por ejemplo, algunos trabajos se centran en la comunicación y el potencial de los modelos de lenguaje de gran escala (LLM) (Kim, Lee y Mutlu, 2024), incluso incluyendo expresiones faciales (Cherakara et al., 2023; Rawal et al., 2024). Otros trabajos se focalizan en estudiar la interacción, la aceptación y la confianza que generan los robots cuando son capaces de expresar emociones (Bonarini, 2020; Calvo-Barajas, Perugia y Castellano, 2020; Kyprianou et al., 2023). Y, en otros trabajos, se centran en que el robot presente comportamientos y reacciones similares a las humanas (Babel et al., 2021; Melo y Moreno, 2022; Reimann et al., 2024). Sin embargo, estos estudios, a diferencia del que se presenta en este trabajo, no integran en un único sistema la conversación fluida basada en IA, la expresión emocional a través del movimiento, el cuerpo y las expresiones faciales, la detección y reconocimiento de estímulos externos y las reacciones y comportamientos físicos naturales.

3. Descripción del robot y consideraciones de diseño del sistema conversacional

En este apartado se analizan los componentes esenciales que permiten desarrollar una interacción conversacional, emocional y natural entre el ser humano y un robot social. Para ello, se presenta en primer lugar el robot Sanbot Elf y sus características esenciales, como plataforma de interacción. Dado el importante papel que desempeña la inteligencia artificial generativa, y en particular los modelos de lenguaje de gran escala (LLMs), se realiza un análisis comparativo de distintos modelos LLM con el fin de decidir el modelo más óptimo para nuestro desarrollo. En tercer lugar, se aborda la definición y generación de reacciones naturales en Sanbot Elf como elemento clave para una interacción más natural y amigable.

3.1 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). 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 QihanOpenSDK, el cual fue utilizado para el desarrollo de la herramienta conversacional que se presenta en la arquitectura.

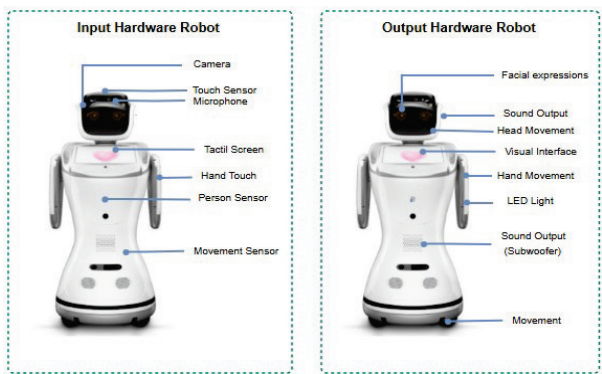


Figura 1: Componentes físicos del robot Sanbot Elf.



Figura 2: Expresiones faciales del robot Sanbot Elf

3.2 Determinación de la herramienta de IA para conversaciones fluidas

Para lograr una conversación fluida y natural entre el robot y las personas es necesario generar respuestas coherentes y contextualizadas, y luego reproducir las respuestas mediante voces artificiales más cercanas a la voz humana. Con este propósito, se llevó a cabo una comparativa de diferentes

modelos disponibles en mercado, evaluando sus capacidades: procesamiento lingüístico y síntesis de voz natural. En primer lugar, se decidieron descartar herramientas como: Speechify, Murf AI, Wondercraft, ElevenLabs o Lovo, las cuales sí permitían la reproducción de texto usando TTS (Text to Speech), pero no eran LLMs (Large Language Model). En la Tabla 1 se presentan otras herramientas LLM analizadas.

Tabla 1: Comparación de herramientas de Inteligencia Artificial

Herramienta	TTS	Observaciones
LLaMA	Sí	Requiere un buen equipo o servidor adecuado.
Google Cloud	Sí	Plataforma con muchas API confusas.
OpenAI	Sí	Precios variables según el modelo LLM utilizado.

Para la selección definitiva se descartó LLaMA, por no contar con los recursos necesarios para su instalación y Google Cloud debido a su confusa biblioteca con numerosas API, quedando la herramienta de OpenAI como la más óptima para esta investigación. Esta herramienta ofrece una muy alta calidad y consistencia en cuanto a sus modelos de diálogo y razonamiento, además de ser fácil de usar, estar bien documentada y no necesitar infraestructura propia. Una vez seleccionada **OpenAI**, fue necesario determinar qué modelo de LLM resultaba más adecuado para los objetivos del trabajo. Para ello, se realizó una comparativa de las ventajas y desventajas de los diferentes modelos disponibles, entre ellos GPT-3.5, GPT-4, y GPT-4o mini, considerando factores como precisión, velocidad de respuesta, capacidad de razonamiento, consumo de recursos y costo. Finalmente, se decidió seleccionar **GPT-4o mini**, un modelo optimizado que mantiene el rendimiento y la calidad de razonamiento del modelo GPT-4o (versión multimodal de GPT-4), pero con un menor consumo computacional y una velocidad de inferencia significativamente mayor. Este modelo tiene un menor coste y un buen rendimiento, lo que lo hace especialmente apropiado para entornos donde se requiere procesamiento rápido y respuestas coherentes. Además, GPT-4o mini es compatible con tareas multimodales (texto, imagen, audio y video) y su entrenamiento se ha realizado con datos multilingües lo que permite realizar con calidad, diálogos prolongados. Además, OpenAI cuenta con funciones Text to Speech, con la posibilidad de realizar la petición enviando un prompt personalizado, ofreciendo al cliente dos posibles modelos: TTS y TTS-HD. Ya que se buscaba la menor latencia posible y la diferencia de calidad es apenas apreciable, se seleccionó el modelo TTS estándar.

3.3 Diseño de reacciones naturales en Sanbot Elf

Para mejorar el dinamismo de la conversación y humanizar la actuación del robot, es importante incluir movimientos propios del comportamiento humano, como el lenguaje corporal o las reacciones espontáneas durante la conversación entre la persona y el robot. En este caso, y teniendo en cuenta las posibilidades de movilidad y reconocimiento de las que dispone el robot Sanbot Elf, se ha considera importante diseñar e incorporar las siguientes funcionalidades:

Ejecución de movimientos naturales: Esta funcionalidad permite incluir una serie de movimientos los cuales no dependan de ningún estímulo externo, y permitan al robot moverse de forma natural y aleatoria durante una conversación.

Detección de ruido: Esta funcionalidad permite al robot reaccionar reactivamente ante ruidos que se producen en el entorno, mediante una queja o pidiendo silencio.

Localización de sonido: Esta función permite al robot reaccionar frente a una fuente de sonido moviendo su cabeza hacia la posición del sonido percibido.

Detección de personas: Esta funcionalidad es la que va a permitir al robot reaccionar en el caso de que una persona pase por delante o detrás suyo. La reacción consiste en una frase aleatoria como: "¡Hola! ¿Te apetece hablar?" o "¡Hola! ¿Qué tal estas?", o en el caso de que la persona pase por detrás el robot girará su cuerpo 180° y reproducirá una frase de las siguientes: "¡Me asustaste! No te había visto", "¡Ah! Debería tener ojos en la espalda" o "Hola, parece que me estabas espiando".

Reconocimiento facial: Esta función 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), y reaccionar ante esa detección. Por ejemplo, si el robot detecta una expresión de enfado reproducirá una frase como frase: "¡Eh! No te enfades conmigo, yo no tengo la culpa".

3.4 Implementación del sistema

En la Figura 3 se muestra el 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. A continuación, se comentan los componentes principales: Interfaces de usuario, Módulo conversacional (que incluye gestor de dialogo, modulo emocional y modulo reactivo), y finalmente se presenta la Integración con otros

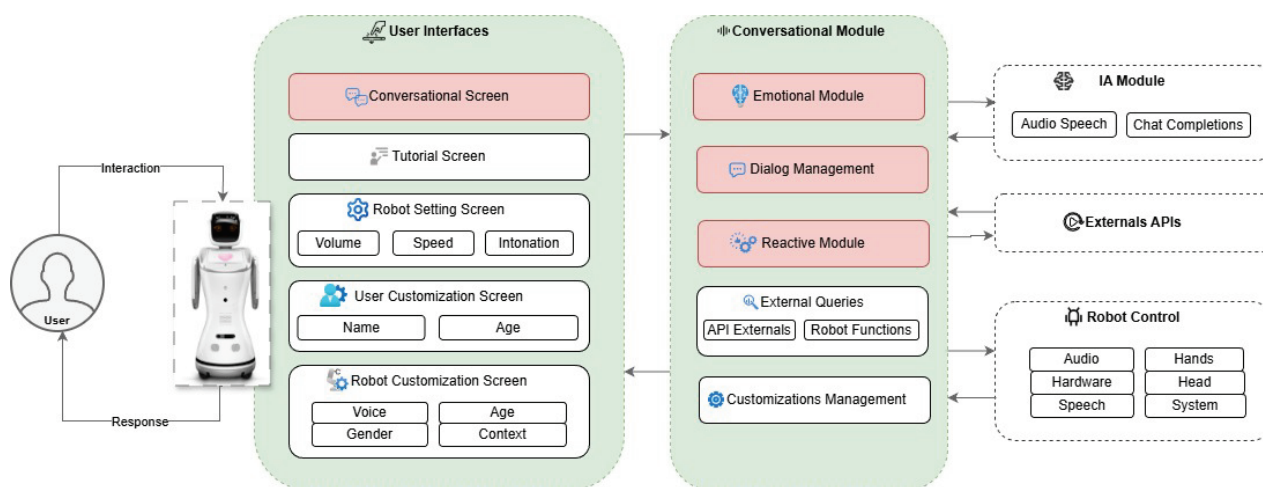


Figura 3: Sistema conversacional desarrollado (sombreados en color rojo los componentes claves de la herramienta).

componentes (módulo de IA y control del robot).

3.4.1 Interfaces de Usuario

Consta de tres pantallas principales (ver Figura 4):

- Pantalla Conversacional (*Conversational Screen*): Es la interfaz 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.

3.4.2 Módulo Conversacional

Destacan tres componentes:

- Gestor de Diálogo (*Dialogue Management*): Gestiona la interacción del chat entre el usuario y el modelo LLM (*Large Language Model*). Este componente es el encargado de crear las burbujas de diálogo para que la interfaz visual de conversación pueda mostrar el hilo de la

conversación.

- 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 LLM y del reconocimiento emocional facial, dicho módulo 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. Se utilizó las seis expresiones de Ekman (2014), lo que implicó un mapeo al catálogo de expresiones faciales disponibles en el robot. En relación con los colores LED del robot, se tomó como referencia el código de colores mencionado por E. Son (2022), referenciando la película de Disney *Inside Out*. Se consideró que este esquema cromático sería fácilmente reconocible, facilitando así la interpretación emocional del robot.
- Módulo Reactivo (*Reactive Module*): Módulo que recibe como entrada diversos estímulos del entorno que le rodea, como la percepción y movimientos de los usuarios que están conversando con el robot, o la voz o sonidos 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 las cinco funcionalidades mencionadas en el apartado anterior (ver Figura 5), y las cuales funcionan de la siguiente forma:

Reconocimiento facial: Se ha utilizado el módulo de gestor de multimedia, el cual procesa la entrada



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

de vídeo capturada por la cámara HD del robot y el módulo de visión desarrollado para agentes sociales interactivos (Ang et al., 2024). Para estas funciones el módulo reactivo recibe del módulo de visión la información concreta sobre las coordenadas del cuadro delimitador del rostro (detección facial), una de las seis expresiones de Ekman (2014) (reconocimiento de expresiones faciales), la edad y el género (reconocimiento de edad y género), y el nombre de la persona reconocida (reconocimiento de una persona) según un conjunto predefinido de usuarios. En todos los casos el robot reacciona mediante la función de habla, incluyendo en sus frases la información detectada.

Detección de ruido: Mediante el módulo camera HD que incorpora el SDK del robot, se captura y procesa el flujo de audio, y se calculan los decibelios (dB). Se considera ruido cuando se superan los 75dB y se mantienen durante más de 1000 ms (para evitar confundir con otro tipo de sonidos como chirridos o golpes). En este caso, el robot reacciona al ruido quejándose del volumen o preguntando: "¿Qué es ese ruido?".

Localización de sonido: La localización se realiza mediante el módulo de control de hardware que incorpora el SDK del robot, el cual calcula el ángulo respecto al robot donde se ha detectado el sonido. Tras esta detección el robot reacciona mediante un movimiento de cabeza hacia la fuente de sonido.

Detección de personas: Esta detección se realiza mediante los sensores infrarrojos y la cámara HD que incorpora el robot en la parte frontal de su cabeza. En este segundo caso, el módulo gestor de multimedia del SDK procesa y envía la imagen a un servidor local, donde se encuentra desplegado un módulo de visión basado en inteligencia artificial desarrollado para agentes sociales interactivos (Ang et al., 2024). Este módulo procesa las imágenes recibidas y envía una respuesta al módulo reactivo, permitiendo así que el robot reaccione, en este caso mediante la reproducción de una frase.

Ejecución de movimientos naturales: Para los simular movimientos naturales se ha tenido en cuenta la movilidad del robot (movimiento de brazos en el eje vertical 270°, movimiento de cabeza eje vertical y horizontal 180°, movimiento de ruedas avanzar/retroceder y girar 360°) y se han programado movimientos de giro de cabeza, giro

de torso y desplazamiento, los cuales se ejecutan a una velocidad y rango de movimiento reducidos para evitar la brusquedad de la acción.

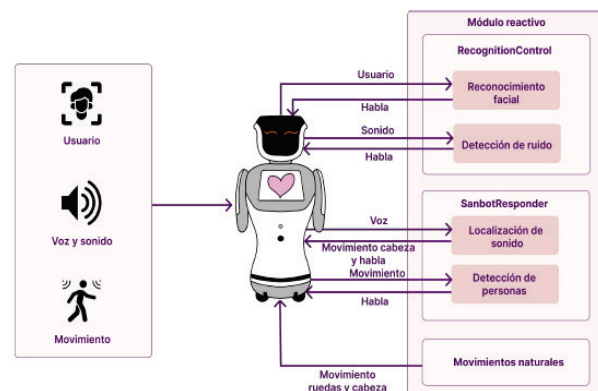


Figura 5: Módulo reactivo.

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.

3.4.3 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. Este 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. Para obtener la emoción del robot, se le solicita a *ChatCompletions* mediante el *prompt* que analice y categorice cada entrada del usuario según el modelo de Eckman 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. Asimismo, se solicita también que en la respuesta se tenga también en cuenta la edad y el nombre de la persona. 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, además dicha API recibe un *prompt* opcionalmente, que nos permitió reducir la entonación anglosajona.

- 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.

4. Conversación natural y emocional entre una persona y el robot

Una vez presentado el sistema y explicados los distintos módulos que lo componen, se muestra en este apartado un ejemplo de flujo conversacional entre una persona y el robot. El proceso comienza cuando el usuario selecciona el botón que inicia la conversación:

Respuesta del robot: ¡Hola Lucía!, ¿cómo estás hoy?

El sistema ha podido reconocer al usuario, por lo que la ha llamado por su nombre.

Entrada del usuario: "¡Hola! Estoy bien, con muchas ganas de conversar"

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

Respuesta del robot: [(1)/(1)] ¿Sobre qué te gustaría conversar? <Lucía,5>

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

1. Interpretación emocional: Los números entre corchetes, como (1)/(1), representan emociones. El primer número (1) corresponde a la emoción del usuario, que en este caso sería "Alegría". El segundo número (1) es la emoción que el robot debe transmitir en su respuesta, en este caso sería "Alegría" igualmente. 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: "¿Sobre qué te gustaría conversar?"
3. Datos del usuario: En este apartado se incluye información personal proporcionada por el usuario o por el módulo reactivo, como su nombre y edad, en este caso: <Lucía,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 alegría 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. El Módulo Reactivo detecta en una primera instancia el rostro de la persona, para comprobar si se encuentra en su base de datos, en caso de que sea así, comienza a hablar mencionando su nombre y adapta la conversación según la edad del usuario. 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.

Por otra parte, el robot, gracias al Módulo Reactivo, siempre está pendiente a sonidos o movimientos de su entorno, pudiendo darse las siguientes situaciones:

Ejecución de movimientos naturales: Mientras Lucía mantiene la conversación con el robot este ejecuta pequeños movimientos de cabeza o de torso cada pocos segundos, lo que ayuda a que Lucía perciba el robot como un objeto más animado, y no tan rígido y robótico.

Detección de ruido: Si el nivel de ruido en la sala se mantiene por encima del umbral establecido, el sistema detecta durante la conversación que hay ruido, e interviene introduciendo una frase aleatoria como "Por favor, bajar el volumen". En ese caso, la interacción sería la siguiente:

Robot: ¡Hola Lucía!, ¿cómo estás hoy?

Usuario: "¡Hola! Estoy bien, con muchas ganas de conversar"

Grupo de personas: <<Barullo>>

Robot: "Por favor, bajar el volumen".

Robot: ¿Sobre qué te gustaría conversar?

Localización de sonido: Si durante la conversación con Lucía alguien da un portazo o se produce algún sonido detectable por el sistema, mediante la localización de sonido el robot girará su cabeza hacia la fuente de sonido y dirá una frase como: "Me parece haber escuchado algo por allí" y la conversación seguirá su curso en base a la respuesta de Lucía.

Detección de personas: Si durante la conversación el sistema detecta que alguien pasa por enfrente o por detrás del robot, éste emitirá una respuesta como: "¡Me asustaste!"

No te había visto” y si la detección es por detrás, girará su torso. Por ejemplo:

Robot: ¡Hola Lucía!, ¿cómo estás hoy?

<<Una persona pasa por detrás>>

Robot: <<Gira torso>> “¡Me asustaste! No te había visto”
<<Gira torso>>

Usuario 1: "Parece que te ha visto ¿Quieres saludar?"

Usuario 2: "Hola, ¿Cómo estás?"

Respuesta del robot: "Estoy bien, pero me había asustado."

Reconocimiento facial: Durante la conversación, el sistema es capaz de reconocer expresiones faciales, edad y género, y la persona con la que interacciona. Ante este reconocimiento el sistema reacciona mediante una frase que el robot introduce en la conversación, mencionando alguna de las características detectadas. En este caso el reconocimiento facial permite que se realice la identificación del usuario Lucía:

Robot: ¡Hola Lucía!, ¿cómo estás hoy? Te veo un poco triste.

Usuario: "Tienes razón, he perdido mi pulsera favorita y me ha dado mucha rabia"

Respuesta del robot: "Vaya, no te preocupes, seguro que está donde menos te lo esperas."

5. Evaluación con usuarios

El objetivo fundamental de la primera evaluación realizada consistió en 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. He de mencionar que, en esta primera evaluación, el módulo reactivo estaba en desarrollo, por lo cual no se pudo evaluar en esta instancia. A continuación, se presenta la metodología que se llevó a cabo en la evaluación, y el análisis de los resultados obtenidos.

5.1 Metodología

En la evaluación participaron 18 usuarios en total, 7 hombres y 11 mujeres con edades entre 18 y 83 años ($M= 41.16$, $SD=22.95$). Las sesiones se llevaron a cabo en el laboratorio del grupo de investigación con una duración de aproximadamente 30 minutos por sesión, y 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; tres voces eran 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 (Kumar, 2019): 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. 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...).

Cuestionarios: Preguntas específicas sobre el funcionamiento del módulo conversacional (ver Tabla 2). Cuestionario de usabilidad con siete preguntas extraídas del SUS (Jordan, 1996) (ver Tabla 3). Cuestionario de aceptación POST VAVA-Q (Cerezo et al., 2025), generado y validado en el grupo de investigación para evaluar la aceptación tecnológica (ver Tabla 4), y que aborda cuatro dimensiones diferentes: Control Percibido, Intención de Uso, Actitud hacia el uso y Norma Subjetiva.

Tabla 2. Cuestionario de Preguntas de interacción por voz.
Respuestas del tipo Likert (1-Totalmente en desacuerdo a 5-
Totalmente de acuerdo)

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

Tabla 3. Cuestionario de Usabilidad. Respuestas del tipo Likert (1-
Totalmente en desacuerdo a 5-Totalmente de acuerdo)

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

Tabla 4. 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)

5.2 Resultados

A continuación, se muestran los resultados obtenidos de esta primera evaluación con usuarios, así como el análisis de los mismos y las mejoras a realizar. 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 (ver figura 6).

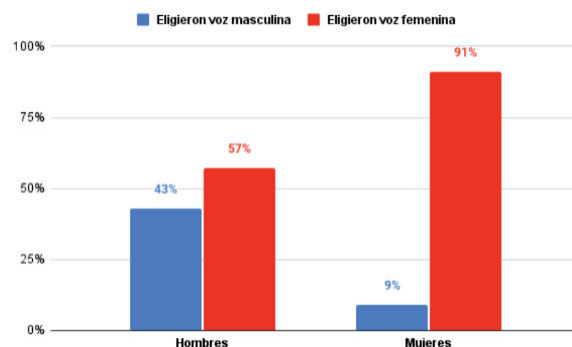


Figura 6: Resultados de la prueba de preferencia género de voz.

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. En el **análisis de género** (ver figura 7), se pudo observar 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.

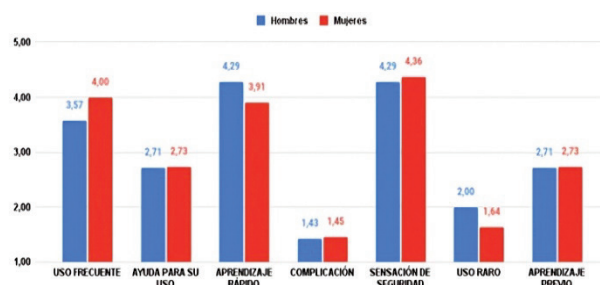


Figura 7: Resultados de la prueba de usabilidad preguntas SUS.

En cuanto al **análisis por grupo (A/B)**, 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 (ver Figura 8). 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.

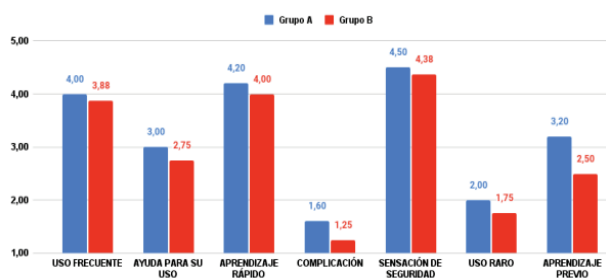


Figura 8: Resultados de la prueba de usabilidad por grupo.

También se analizó la **Usabilidad (Preguntas interacción por voz)**, donde el resultado del análisis general arrojó que todos están de acuerdo o muy de acuerdo en que hubo una buena 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. El grupo B ha comprendido mejor las respuestas del robot con un valor medio de 4.80, por lo que la incomprensión por parte del robot tuvo solo una media de 1.75, probablemente debido a los problemas de pronunciación de las voces OpenAI (nativamente inglesas), mientras que el grupo A tuvo resultados un poco menos positivos, con medias de 4.80 y 2.00, respectivamente. Los participantes del grupo B opinaron que el robot entendía peor las consultas con una media de 4.75 respecto a la media del grupo A ($M=4.90$) debido a que en ese grupo no se aplicaba ninguna personalización. Por último, el grupo A ($M=4.20$) consideró las respuestas del robot más lentas que las del grupo B ($M=4.75$), seguramente debido a que la petición a OpenAI contenía más datos de tipo emocional, personalización o similares. Finalmente, se analizó la **Aceptación**, donde los resultados fueron muy satisfactorios, pudiendo observar que todos los usuarios consideraron que el sistema no era 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, solo dos votaron de forma neutral. También se pudo observar que las mujeres encuentran más útil el uso del robot ($M=4.91$) respecto a los hombres ($M=4.29$), sin embargo, los hombres opinaron que tienen mayor intención de usarlo nuevamente ($M=4.86$)

respecto a las mujeres ($M=4.55$). En general, consideraron que la prueba fue muy divertida.

Analizando los resultados en su conjunto, se puede concluir que todos los participantes calificaron el uso del robot como 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 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 difícil de usar. El segundo punto puede estar relacionado con la necesidad de pasar, por varias pantallas de personalización, tanto del robot como del usuario, por lo que habría que reconsiderarlas o rediseñarlas.

5.3 Mejoras Realizadas

Debido a los resultados antes planteados, se realizaron diversas modificaciones para mejorar el sistema conversacional desarrollado. En primer lugar, se modificó la vista inicial de configuración para que sea más sencilla e intuitiva y el usuario no necesite navegar por otras pantallas de la aplicación, seleccionando "Ajuste Automático" el usuario ya no necesita insertar su nombre y edad, ya que esta funcionalidad le redirige a la pantalla de conversación, donde el robot lo identificará mediante el reconocimiento de rostro del módulo reactivo, en caso de que no lo reconozca le preguntará su nombre y edad para la personalización de la experiencia. En segundo lugar, la generación de voz se realizó mediante el uso de *prompts* adaptativos, donde se le indicaba que debía ser hispanohablante para así obtener una entonación más natural y con menos acento anglosajón, además se situó en la pantalla inicial de configuración las cuatro voces con menor acento foráneo (ver Figura 9).



Figura 9: Pantalla de configuración modificada

Además, se ha optimizado el código para que el tiempo de procesamiento de los diferentes módulos sean mínimos e imperceptibles para el usuario. Sin embargo, el tiempo de

respuesta de los *endpoints* de OpenAI dependerá de la velocidad y latencia de la red que está configurada en el robot. Con estas mejoras implementadas se realizó una primera evaluación muy preliminar en una residencia de personas mayores. De manera empírica se pudo observar que la mayoría de las personas usuarias mantuvieron una conversación fluida y natural con el robot, y resultó interesante comprobar que algunos participantes compartían con los robots aspectos personales. En general tanto el robot como la aplicación tuvieron una buena aceptación por parte de los usuarios. También se realizó otra prueba informal durante un taller de co-creación con adultos y personas mayores. En este caso, la experiencia no fue tan satisfactoria debido al elevado nivel de ruido ambiental, que activaba con frecuencia el módulo reactivo del robot. Esta situación evidenció que, en entornos grupales o ruidosos, es necesario ajustar la sensibilidad del sistema para evitar reacciones no deseadas. En términos generales, la aceptación del robot fue muy positiva y la latencia casi imperceptible. Sin embargo, es necesario hacer un estudio más exhaustivo utilizando instrumentos cuantitativos que permitan evaluar en detalle las percepciones de los usuarios. De este modo será posible determinar qué diferencias hay entre las experiencias realizadas en el laboratorio y las realizadas en estos contextos no controlados.

6. Conclusiones y trabajo futuro

En este trabajo se ha presentado un sistema conversacional diseñado para fortalecer la interacción natural entre personas y robots, incorporando inteligencia artificial para generar diálogos fluidos acompañados de expresiones faciales, movimientos y colores que permiten reflejar las emociones del robot, y reaccionar de forma natural en base a situaciones que detecta el robot en su entorno. El sistema desarrollado fue evaluado, de forma preliminar, con 18 usuarios, evidenciando resultados positivos en términos de usabilidad y aceptación, aunque también se identificaron algunas limitaciones, como la utilización de voces no idóneas para la interacción en español, el incremento en los tiempos de procesamiento al integrar componentes emocionales y la necesidad de mantener interfaces de configuración simples frente a la complejidad de los módulos de personalización. A partir de estos resultados, se llevaron a cabo modificaciones

en el desarrollo del sistema para solucionar los problemas detectados. Sin embargo, aunque se realizaron algunas evaluaciones empíricas, aún es necesario evaluar la aplicación, de forma más exhaustiva y rigurosa. Por ello, como trabajo futuro a corto plazo, se debe tener en cuenta, reducir la latencia o fallos que la aplicación pudiera tener en relación con la conexión a internet en entornos externos al laboratorio, para esto se prevé implantar un modelo LLM en un servidor local que se encuentre en la misma red del centro en el que se realicen las pruebas y se pretende llevar a cabo una segunda fase de evaluación con el fin de analizar el impacto de estos nuevos componentes en la experiencia de interacción.

Esta fase incluirá otros instrumentos de evaluación como *GoodSpeed* (Bartneck, Croft y Kulic, 2009) o *RoSAS* (Pan, Croft, y Niemeyer, 2018) teniendo en cuenta también un mayor número y diversidad de participantes y la aplicación de protocolos estandarizados en el ámbito de la interacción humano-robot (HRI), así como el uso de instrumentos específicos para medir variables como la empatía percibida, la fluidez conversacional y la naturalidad en la comunicación. Por otra parte, también sería interesante evaluar si las expresiones faciales, los movimientos y los colores que muestra el robot para demostrar sus emociones, tienen algún impacto en la percepción por parte de los usuarios, de forma similar al análisis realizado por (Raggioli et al., 2025) para determinar si ciertas expresiones y movimientos aumentaban la tasa de reconocimiento de ciertas emociones. De este modo, se podrán validar las mejoras implementadas, así como detectar nuevas necesidades de los usuarios y profundizar en el estudio de la interacción humano-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).

Referencias

- Ang E., Bejleri A., Tantisira B, Van de Velde. A., (2024), Considerations for the future of social robots and human-robot interactions. URL: <https://www.oxjournal.org/the-future-of-social-robots-and-human-robot-interactions/>.
- Alnajjar, F., Khalid, S., Vogan, A. A., Shimoda, S., Nouchi, R., & Kawashima, R. (2019). Emerging cognitive intervention technologies to meet the needs of an aging population: a systematic review. *Frontiers in Aging Neuroscience*, 11, 291.
- Assad-Uz-Zaman, M., Rasedul Islam, M., Miah, S., & Rahman, M. H. (2019). NAO robot for cooperative rehabilitation training. *Journal of rehabilitation and assistive technologies engineering*, 6, 2055668319862151.

- Babel, F., Kraus, J., Miller, L., Kraus, M., Wagner, N., Minker, W., & Baumann, M. (2021). 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(6), 1485-1498.
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1), 71-81.
- Bonarini, A. (2020). Communication in human-robot interaction. *Current Robotics Reports*, 1(4), 279-285.
- Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International journal of human-computer studies*, 62(2), 161-178.
- Breazeal, C. (2003). Emotion and sociable humanoid robots. *International journal of human-computer studies*, 59(1-2), 119-155.
- Calvo-Barajas, N., Perugia, G., & Castellano, G. (2020, August). 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)* (pp. 165-171). IEEE.
- Cerezo, E., Lacuesta, R., Gallardo, J., & Aguelo, A. (2025). Exploring the use of voice assistants in nursing homes. *International Journal of Human-Computer Interaction*, 1-17.
- Chen, Y., Garcia-Vergara, S., & Howard, A. M. (2018). 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(8), 490-496.
- Cherakara, N., Varghese, F., Shabana, S., Nelson, N., Karukayil, A., Kulothungan, R., & Lemon, O. (2023). Furchat: An embodied conversational agent using llms, combining open and closed-domain dialogue with facial expressions. *arXiv preprint arXiv:2308.15214*.
- Churamani, N., Kalkan, S., & Gunes, H. (2020, August). Continual learning for affective robotics: Why, what and how?. In *2020 29th IEEE international conference on robot and human interactive communication (RO-MAN)* (pp. 425-431). IEEE.
- Cross, E. S., Hortensius, R., & Wykowska, A. (2019). From social brains to social robots: applying neurocognitive insights to human-robot interaction. *Philosophical Transactions of the Royal Society B*, 374(1771), 20180024.
- de Graaf, M. M., Ben Allouch, S., & Van Dijk, J. A. (2015, October). What makes robots social?: A user's perspective on characteristics for social human-robot interaction. In *International Conference on Social Robotics* (pp. 184-193). Cham: Springer International Publishing.
- Dziergwa, M., Kaczmarek, M., Kaczmarek, P., Kędzierski, J., & Wadas-Szydłowska, K. (2018). Long-term cohabitation with a social robot: a case study of the influence of human attachment patterns. *International Journal of Social Robotics*, 10(1), 163-176.
- Ekman, P. (2014). Expression and the nature of emotion. *Approaches to emotion*, 319-343.
- Gou, M. S., Vouloutsi, V., Grechuta, K., Lallée, S., & Verschure, P. F. (2014, July). Empathy in humanoid robots. In *Conference on Biomimetic and Biohybrid Systems* (pp. 423-426). Cham: Springer International Publishing.
- Henschel, A., Laban, G., & Cross, E. S. (2021). What makes a robot social? A review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports*, 2(1), 9-19.
- Jiang, Y., Shao, S., Dai, Y., & Hirota, K. (2024, July). A LLM-Based Robot Partner with Multi-modal Emotion Recognition. In *International Conference on Intelligent Robotics and Applications* (pp. 71-83). Singapore: Springer Nature Singapore.
- Jordan, P. W., Thomas, B., McClelland, I. L., & Weerdmeester, B. (Eds.). (1996). *Usability evaluation in industry*. CRC press.
- Kim, C. Y., Lee, C. P., & Mutlu, B. (2024, March). Understanding large-language model (llm)-powered human-robot interaction. In *Proceedings of the 2024 ACM/IEEE international conference on human-robot interaction* (pp. 371-380).
- Kumar, R. (2019, March). Data-driven design: Beyond a/b testing. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (pp. 1-2).
- Kyprianou, G., Karousou, A., Makris, N., Sarafis, I., Amanatiadis, A., & Chatzichristofis, S. A. (2023). Engaging learners in educational robotics: Uncovering students' expectations for an ideal robotic platform. *Electronics*, 12(13), 2865.
- Laban, G., & Cross, E. S. (2024). Sharing our Emotions with Robots: Why do we do it and how does it make us feel?. *IEEE Transactions on Affective Computing*.
- Lee, M. K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., & Savetsila, S. (2012, March). Personalization in HRI: A longitudinal field experiment. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 319-326).
- Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human-robot relations. *International journal of human-computer studies*, 71(3), 250-260.
- Melo, F., & Moreno, P. (2022, April). Socially reactive navigation models for mobile robots. In *2022 IEEE international conference on autonomous robot systems and competitions (ICARSC)* (pp. 91-97). IEEE.
- Mohebbi, A. (2020). Human-robot interaction in rehabilitation and assistance: a review. *Current Robotics Reports*, 1(3), 131-144.
- Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3), 1-40.
- Pan, M. K., Croft, E. A., & Niemeyer, G. (2018, February). Evaluating social perception of human-to-robot handovers using the robot social attributes scale (rosas). In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 443-451).

- Pinto-Bernal, M., Biondina, M., & Belpaeme, T. (2025). Designing Social Robots with LLMs for Engaging Human Interaction. *Applied Sciences*, 15(11), 6377.
- Raggioli, L., Esposito, R., Rossi, A., & Rossi, S. (2025). Exploring the Role of Robot's Movements for a Transparent Affective Communication. *IEEE Robotics and Automation Letters*.
- Rawal, N., Maharjan, R. S., Romeo, M., Bigazzi, R., Baraldi, L., Cucchiara, R., & Cangelosi, A. (2024, September). Intelligent multimodal artificial agents that talk and express emotions. In *International Workshop on Human-Friendly Robotics* (pp. 240-254). Cham: Springer Nature Switzerland.
- Reimann, M. M., Kunneman, F. A., Oertel, C., & Hindriks, K. V. (2024). A survey on dialogue management in human-robot interaction. *ACM Transactions on Human-Robot Interaction*, 13(2), 1-22.
- Robinson, N. L., Cottier, T. V., & Kavanagh, D. J. (2019). Psychosocial health interventions by social robots: systematic review of randomized controlled trials. *Journal of medical Internet research*, 21(5), e13203.
- Romat, H., Williams, M. A., Wang, X., Johnston, B., & Bard, H. (2016, March). Natural human-robot interaction using social cues. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 503-504). IEEE.
- Sarrica, M., Brondi, S., & Fortunati, L. (2020). How many facets does a "social robot" have? A review of scientific and popular definitions online. *Information Technology & People*, 33(1), 1-21.
- Scoglio, A. A., Reilly, E. D., Gorman, J. A., & Drebing, C. E. (2019). Use of social robots in mental health and well-being research: systematic review. *Journal of medical Internet research*, 21(7), e13322.
- Son, E. (2022). 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(1), 225-240.
- Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Frontiers in Robotics and AI*, 7, 532279.
- Van Oost, E., & Reed, D. (2010, June). Towards a sociological understanding of robots as companions. In *International conference on human-robot personal relationship* (pp. 11-18). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Wang, C., Hasler, S., Tanneberg, D., Ocker, F., Joubin, F., Ceravola, A., & Gienger, M. (2024, May). Lami: Large language models for multi-modal human-robot interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-10).
- Wullenkord, R., & Eyssel, F. (2020). Societal and ethical issues in HRI. *Current Robotics Reports*, 1(3), 85-96.
- Xu, J., Broekens, J., Hindriks, K., & Neerincx, M. A. (2015). Mood contagion of robot body language in human robot interaction. *Autonomous Agents and Multi-Agent Systems*, 29(6), 1216-1248.
- Yang, G. Z., J. Nelson, B., Murphy, R. R., Choset, H., Christensen, H., H. Collins, S., ... & McNutt, M. (2020). Combating COVID-19— The role of robotics in managing public health and infectious diseases. *Science robotics*, 5(40), eabb5589.