

Enfoques arquitectónicos en plataformas de evaluación *Wizard of Oz* para robots sociales: el caso de SHARA-WoZ

Architectural approaches in Wizard of Oz evaluation platforms for social robots: the case of SHARA-WoZ

Guillermo Cubero

Departamento de Tecnologías y Sistemas de
Información

Universidad de Castilla-La Mancha

Ciudad Real, España

Guillermo.Cubero@uclm.es

Laura Villa

Departamento de Tecnologías y
Sistemas de Información

Universidad de Castilla-La Mancha

Ciudad Real, España

Laura.Villa@uclm.es

Ramón Hervás

Departamento de Tecnologías y
Sistemas de Información

Universidad de Castilla-La Mancha

Ciudad Real, España

Ramon.HLucas@uclm.es

Recibido: 06.11.2025 | Aceptado: 03.12.2025

DOI: <https://doi.org/10.65234/interaccion.131>

Palabras Clave

Robots sociales asistenciales
Mago de Oz
Interacción humano-robot
Cuidado geriátrico
Inteligencia artificial
conversacional

Resumen

Los robots sociales asistenciales emergen como solución para el cuidado de personas mayores, pero su validación efectiva requiere metodologías que faciliten la participación de múltiples implicados técnicos y no técnicos. La técnica Wizard of Oz (WoZ) permite evaluar interacciones antes de su implementación autónoma completa, aunque las plataformas actuales carecen frecuentemente de infraestructuras estandarizadas y modulares. Este artículo presenta SHARA-WoZ v2.0.0, una evolución de una plataforma de evaluación para robots sociales asistenciales que aborda sistemáticamente las limitaciones identificadas en su versión anterior. Las contribuciones principales incluyen: (1) migración a una aplicación de escritorio nativa en PyQt6 que reduce la latencia de renderizado en un 42% y el uso de memoria en un 49%; (2) integración de GPT-4o-mini que elimina servicios de traducción intermedios y reduce la latencia conversacional total en un 55%; (3) sistema de generación de respuestas múltiples basadas en estados emocionales que reduce el tiempo de respuesta en modo semi-automático a 2-3 segundos; (4) rediseño modular de la interfaz que facilita la supervisión simultánea de múltiples canales de interacción; y (5) arquitectura orientada a servicios que garantiza escalabilidad y extensibilidad. Los resultados demuestran mejoras cuantificables en rendimiento, latencia y usabilidad, estableciendo una plataforma que equilibra sofisticación técnica con accesibilidad para implicados no técnicos en el desarrollo y evaluación de robots sociales para contextos geriátricos. Estos resultados pueden ser generalizables como modelo arquitectónico efectivo para la aplicación de técnicas de Mago de Oz para el contexto de los Robots Asistenciales y Sociales.

Keywords

Socially assistive robots
Wizard of Oz
Human-robot interaction
Geriatric care
Conversational artificial
intelligence

Abstract

Socially assistive robots emerge as a solution for elderly care, but their effective validation requires methodologies that facilitate the participation of multiple technical and non-technical stakeholders. The Wizard of Oz (WoZ) method enables the evaluation of interactions before full autonomous implementation, although current platforms frequently lack standardized and modular infrastructures. This article presents SHARA-WoZ v2.0.0, an evolution of an evaluation platform for socially assistive robots that systematically address limitations identified in its previous version. The main contributions include: (1) migration to a native desktop application in PyQt6 that reduces rendering latency by 42% and memory usage by 49%; (2) integration of GPT-4o-mini that eliminates intermediate translation services and reduces total conversational latency by 55%; (3) multiple response generation system based on emotional states that reduces response time in semi-automatic mode to 2-3 seconds; (4) modular interface redesign that facilitates simultaneous supervision of multiple interaction channels; and (5) service-oriented architecture that guarantees scalability and extensibility. Results demonstrate improvements in performance, latency, and usability, establishing a

platform that balances technical sophistication with accessibility for non-technical stakeholders in the development and evaluation of social robots for geriatric contexts. These results can be generalized as an effective architectural model for the application of Wizard of Oz techniques in the context of Assistive and Social Robots.

1. Introducción

El envejecimiento progresivo de la población mundial ha impulsado el desarrollo de tecnologías asistenciales orientadas al cuidado de personas mayores. Los robots sociales asistenciales (SARs) emergen como solución prometedora para abordar necesidades de acompañamiento, monitorización de salud y asistencia en actividades diarias (Abdi et al., 2018; Kachouie et al., 2014). Sin embargo, la validación efectiva de estos sistemas requiere de la participación de múltiples implicados (stakeholders), incluyendo usuarios finales, profesionales clínicos, familiares y cuidadores, más que depender exclusivamente de expertos técnicos, lo que incrementa considerablemente la complejidad de coordinación (Tobis et al., 2023). La necesidad de medir resultados sociales, emocionales y terapéuticos demanda marcos de evaluación capaces de acomodar perspectivas de múltiples grupos interesados manteniendo validez científica.

La técnica Wizard of Oz (WoZ) constituye una herramienta fundamental en el desarrollo de sistemas de interacción en general, y especialmente útil en interacción humano-robot (HRI), permitiendo evaluar respuestas de usuarios ante comportamientos de robots antes de su implementación automática completa (Riek, 2012). Esta técnica, en la cual un operador humano controla discretamente las acciones del robot de forma transparente para el usuario, facilita la recopilación de retroalimentación durante las fases tempranas del desarrollo. No obstante, las implementaciones tradicionales de WoZ en robótica social frecuentemente carecen de infraestructuras estandarizadas y modulares que faciliten la participación de implicados no técnicos en el proceso de refinamiento de funcionalidades (Hoffman, 2016; Porfirio et al., 2018).

En respuesta a los desafíos identificados en trabajos previos (Cubero et al., 2024), el presente artículo describe una evolución significativa de la plataforma de evaluación de robots asistenciales. Esta implementación inicial demostró la viabilidad del enfoque de simulación virtual para evaluación de robots sociales. Sin embargo, se detectaron limitaciones arquitectónicas para este tipo de soluciones que son abordadas en este artículo.

Aunque el trabajo presenta una infraestructura generalista, lo ejemplificamos con la versión virtual de un robot social físico existente denominado SHARA (Figura 1) (Villa et al., 2025), un robot conversacional con capacidades afectivas y proactivas, enfocado a la asistencia y acompañamiento de

personas mayores que viven solas en su hogar. SHARA cuenta con la capacidad de detectar la presencia de sus usuarios, reconocerlos e iniciar conversaciones de forma proactiva con ellos (Villa et al., 2022). Además, puede contextualizar cada una de sus respuestas gracias a recordar las conversaciones previas (Villa et al., 2022).

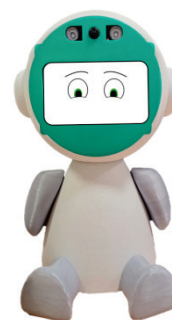


Figura 1: Robot asistencial social SHARA

El sistema presentado en este artículo proporciona una simulación virtual del robot asistencial SHARA (imitando gran parte de sus capacidades) y permite el control remoto del robot (virtual en este caso) por parte de un operador y en general realizar evaluaciones con diferentes implicados de forma más flexible.

Las contribuciones principales de este trabajo se centran en un modelo de arquitectura para sistemas basados en la técnica de WoZ para SARs y se abordan sistemáticamente las limitaciones identificadas en trabajos anteriores, materializadas mediante: (1) migración de una aplicación de escritorio desarrollada en PyQt6 que optimiza el rendimiento y reduce significativamente la latencia de comunicación mediante una arquitectura nativa; (2) integración de modelos de lenguaje de OpenAI (GPT-4o-mini¹) para generación automática de respuestas contextualizadas que mantienen coherencia conversacional y personalización, además de ser el modelo de conversación que usa el robot real; (3) implementación de un sistema de generación de respuestas múltiples basadas en emociones que proporcionan al operador alternativas adaptadas a un contexto afectivo del robot; (4) rediseño completo de la interfaz de usuario con componentes modulares que facilitan la supervisión simultánea de múltiples canales de interacción

¹<https://platform.openai.com/docs/models/gpt-4o-mini> Último acceso: 28/10/2025

(conversación, vídeo, estado de la simulación); y (5) implementación de una arquitectura de sistema orientada a servicios que garantiza escalabilidad, mantenibilidad y futura extensibilidad del sistema.

El presente artículo se estructura de la siguiente manera: 2. Estado del Arte contextualiza las contribuciones en el marco de investigación previas en robótica social asistencial, metodologías WoZ y plataformas de evaluación; 3. Materiales y Métodos describe la arquitectura del sistema mejorado, las tecnologías empleadas y las decisiones de diseño fundamentales; 4. Resultados presenta análisis comparativos detallados entre ambas versiones del sistema, incluyendo mejoras cuantificables en latencia y análisis de capacidades funcionales ampliadas; finalmente, 5. Conclusiones sintetiza los aportes del trabajo y delinean direcciones futuras de investigación en el desarrollo de plataformas WoZ para robótica social asistencial.

2. Estado del arte

El desarrollo y evaluación de robots sociales asistenciales representa un campo en constante evolución que combina múltiples dominios de investigación, incluyendo la HRI, tecnologías asistenciales y metodologías de evaluación basadas en simulación. Esta sección proporciona una revisión de los enfoques actuales en la robótica social para el cuidado de personas mayores, identificando los desafíos principales que el sistema propuesto aborda y contextualizando las mejoras implementadas respecto al trabajo previo.

2.1 Robots sociales asistenciales para el cuidado de personas mayores

Los robots sociales asistenciales han emergido como un paradigma distintivo en el cual proporcionan asistencia a través de interacción social más que física (Feil-Seifer & Mataric, 2005). Este enfoque ha ganado relevancia significativa para abordar las necesidades crecientes de una población global envejecida, con aplicaciones que abarcan desde recordatorios de medicación hasta acompañamiento social y estimulación cognitiva.

Se identifican revisiones sistemáticas de implementaciones SARs han revelado tanto beneficios potenciales significativos como limitaciones actuales en entornos de cuidado geriátrico (Zhao et al, 2023). Algunos de los factores clave que influyen en la aceptación de los robots son la importancia de interacciones personalizadas, sensibilidad cultural y la necesidad de introducción gradual de dicha tecnología en la vida de sus usuarios (Rigaud et al., 2024).

Para tratar esto factores, varios estudios recientes enfatizan en que el diseño de robots sociales efectivos requiere

comprensión profunda de principios de experiencia de usuario y metodologías de diseño centradas en él. El diseño de estos sistemas debe entender la experiencia de usuario en interacción humano máquina, identificar factores clave que influyen en la aceptación y validación del usuario, recalcando que las funcionalidades técnicas por sí solas no son suficientes. Por ello, se deben proporcionar experiencias emocionales positivas en robots sociales y mantener la confianza del usuario durante periodos de interacción extendidos (Alenjung et al., 2019). Al evaluar estos sistemas, debe prestarse especial atención al impacto de la edad en su aceptación y usabilidad (Bevilacqua et al., 2022).

Estas evaluaciones se han destacado por varias investigaciones, enfatizando la importancia de factores culturales y contextuales en la evaluación de robots asistenciales (Papadopoulos et al., 2022). Para ello, las metodologías de evaluación deben considerar diversos antecedentes de usuarios, entornos de cuidado y contextos sociales para asegurar que los resultados de estas evaluaciones sean generalizables a través de diferentes escenarios de implementación (Asi et al., 2022).

2.2 Metodologías de evaluación para robótica social

La evaluación de robots sociales para el cuidado geriátrico presenta desafíos únicos que se extienden más allá de las pruebas tradicionales de usabilidad. Los enfoques actuales frecuentemente carecen de la escalabilidad, accesibilidad y rigor metodológico necesarios para realizar evaluaciones en diversos entornos y poblaciones de usuarios (Koh et al., 2021). Teniendo en cuenta que para medir resultados sociales, emocionales y terapéuticos se requieren marcos de evaluación que permitan a varios interesados, de distintas ramas de estudios, evaluar dichos resultados manteniendo validez científica.

Por ello, los enfoques de evaluación de métodos mixtos son particularmente valiosos para capturar la naturaleza multifacética de la efectividad de robots sociales (Asl et al., 2022). Estos enfoques combinan medidas cuantitativas con conocimientos cualitativos, permitiendo a investigadores evaluar tanto resultados medibles como experiencias subjetivas de usuarios, las cuales son muy importantes para entender la aceptación de robots en entornos de cuidado geriátrico. La integración de evaluaciones cuantitativas estructuradas con técnicas de observación etnográfica aporta comprensión de cómo funcionan los robots sociales dentro de contextos de cuidado complejos.

Otra metodología de interés es la evaluación longitudinal para robots sociales en entornos de cuidado geriátrico. A diferencia de estudios de laboratorio de corto plazo, los

enfoques longitudinales rastrean la aceptación del usuario, cambios comportamentales y resultados terapéuticos durante períodos extendidos. Esta metodología es particularmente útil para poblaciones amplias, donde los efectos iniciales de novedad pueden disminuir y emergen patrones verdaderos de aceptación del robot y eficacia terapéutica (Coronado et al., 2022).

De forma similar, las metodologías de evaluación basadas en escenarios proporcionan enfoques estructurados para probar robots sociales a través de diversos contextos de interacción relevantes para el cuidado geriátrico (Koh et al., 2021). Estas metodologías implican diseñar escenarios de cuidado realistas que reflejan desafíos e interacciones comunes en entornos de cuidado geriátrico. Es importante destacar que las metodologías de evaluación participativa involucran activamente a los usuarios mayores, cuidadores y profesionales de salud en el proceso de evaluación, asegurando que los criterios de evaluación reflejen las necesidades y preferencias de los implicados (Tobis et al., 2023).

2.3 La técnica Wizard of Oz en robótica social

Entre las metodologías experimentales controladas, la técnica Wizard of Oz (WoZ) representa una metodología de investigación en interacción humano-máquina que ha sido extensivamente adaptada para estudios de interacción humano-robot (HRI) (Riek, 2012). En experimentos WoZ, los participantes interactúan con lo que ellos creen ser un sistema robótico autónomo, mientras que en realidad un operador humano controla de forma encubierta algunas o todas las acciones del robot.

Estos estudios WoZ pueden evaluar efectivamente estrategias de interacción social para robots operando bajo restricciones perceptuales (Sequeira et al., 2016). Incluso cuando los robots tienen capacidades sensoriales limitadas, los estudios WoZ cuidadosamente diseñados pueden revelar patrones de interacción exitosos para posteriormente ser incorporados en sistemas autónomos.

Gracias a estos estudios, WoZ se ha establecido como metodología relevante en investigación HRI por varias razones. En primer lugar, permite prototipados rápidos y pruebas de hipótesis de comportamientos robóticos sin la inversión sustancial de tiempo y recursos requerida para desarrollar sistemas completamente autónomos. En segundo lugar, la metodología proporciona control experimental preciso sobre variables de interacción mientras mantiene el foco en respuestas humanas naturales, lo cual es particularmente valioso cuando se estudian resultados sociales, emocionales y terapéuticos complejos. En tercer lugar, WoZ facilita enfoques de diseño centrados en el

usuario, ya que permite a investigadores probar y refinar comportamientos de robots basándose en retroalimentación auténtica de usuarios en entornos realistas antes de comprometerse con la implementación autónoma completa (Porfirio et al., 2018).

A pesar de los avances en robótica social y metodologías de evaluación, aún persisten brechas en sistemas WoZ actuales. Frecuentemente carecen de infraestructuras estandarizadas y modulares que permitirían a los implicados no técnicos participar de forma efectiva en procesos de desarrollo y evaluación de robots (Porfirio et al., 2018). Esta falta de estandarización se manifiesta en las implementaciones ad-hoc desarrolladas con poco tiempo, que suelen sufrir limitaciones como retrasos de entrada, respuestas inconsistentes del operador y falta de protocolos establecidos. Estas deficiencias pueden afectar negativamente tanto la autenticidad del comportamiento del robot como la percepción del participante.

Aunque existen esfuerzos por abordar estos problemas de estandarización, como la plataforma de código abierto WoZ4U, que proporciona atajos de teclado, cadenas de comportamiento automatizadas y acciones del operador optimizadas para reducir la carga cognitiva (Hoffman, 2016), la mayoría de las plataformas existentes se enfocan en sistemas completamente automatizados o controlados por un operador. Este enfoque binario limita su aplicabilidad en escenarios complejos de robótica social.

Para superar esta limitación, los sistemas WoZ híbridos emergen como una solución prometedora. Estos sistemas combinan control del operador con componentes autónomos, representando un avance sobre metodologías tradicionales (Hoffman, 2016). El enfoque híbrido permite evaluaciones más realistas de robots sociales semi-autónomos y facilita transiciones graduales de control del operador hacia autonomía completa, resultando práctico para desarrollar y probar comportamientos adaptativos que deben responder a situaciones sociales impredecibles.

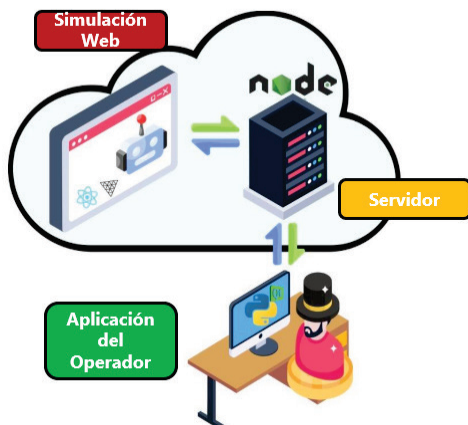


Figura 2: Arquitectura tripartita del sistema de evaluación remota

2.4 Lecciones aprendidas del Estado del Arte

Gracias a un estudio profundo de la robótica social y los métodos de evaluación más utilizados en la actualidad, podemos ver cómo deberían desarrollarse y evaluarse estos sistemas enfocados a adultos mayores.

En primer lugar, el desarrollo técnico no es lo más importante. Para que un robot sea aceptado, debe generar confianza, ser agradable y entender las limitaciones del contexto cultural de la persona mayor. Un robot útil es aquel que se adapta al usuario y a su entorno.

En segundo lugar, no es suficiente con evaluar el comportamiento y rendimiento del robot en un entorno de laboratorio controlado. Para conocer su verdadero impacto, se necesitan métodos que recojan experiencias reales y estudios que revelen el uso continuo del robot. La evaluación se debe hacer en un entorno realista.

Finalmente, la técnica WoZ híbrida resulta ser la solución ideal para poner en práctica estas lecciones ya que:

- Permite probar ideas rápidamente eliminando la necesidad de programar un robot completamente autónomo. Un operador puede simular comportamientos sociales complejos y ver cómo reaccionan los adultos mayores, refinando el diseño durante el desarrollo.
- El operador humano aporta el sentido común y la empatía para manejar situaciones impredecibles, mientras que el sistema automatizado se encarga de tareas repetitivas. Esto garantiza interacciones fluidas y realistas.
- Permite pasar de manera gradual y controlada de un prototipo controlado por un operador a un robot cada vez más autónomo, incorporando sólo los comportamientos que han sido previamente validados.

La técnica WoZ híbrida es una herramienta que permite

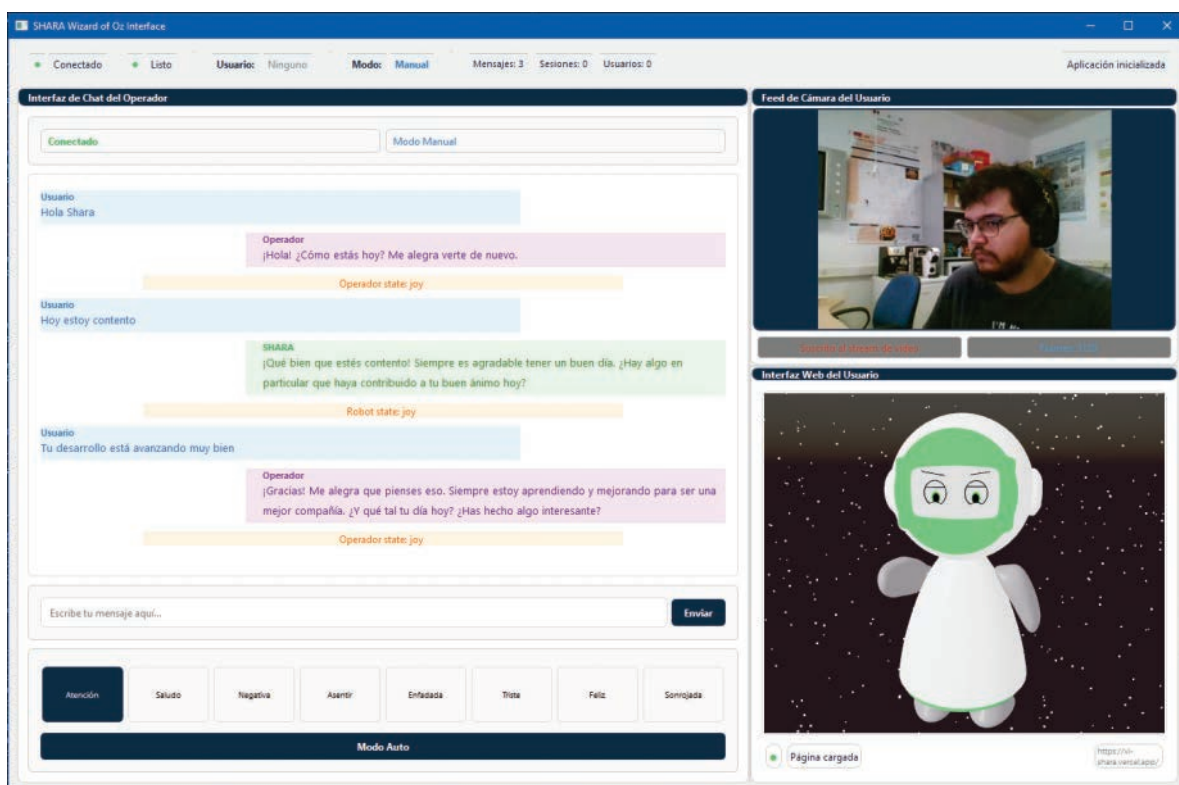


Figura 3: Interfaz del operador. La interfaz conjunta un historial de chat conversacional de la interacción, video en tiempo real del usuario, navegador con la interfaz web actualizada que ve el usuario y panel de control para una manipulación fuera del flujo conversacional.

centrarse en los usuarios a la hora de desarrollar robots sociales, puesto que con ella se pueden desarrollar y evaluar interacciones sociales significativas de una manera directa y efectiva.

3. Materiales y métodos

Esta sección describe la evolución de la arquitectura del sistema SHARA-WoZ desde su versión original (v1.0.0) hasta la versión mejorada actual (v2.0.0). Primero se presenta el sistema inicial con las limitaciones identificadas mediante evaluación por panel de expertos, que motivaron la reestructuración del sistema.

Posteriormente se detalla la arquitectura del sistema en su siguiente versión, describiendo sus componentes principales, las novedades implementadas y la justificación técnica de las decisiones de diseño que abordan las limitaciones previamente identificadas, manteniendo el mismo modelo de comunicación de componentes.

3.1 Sistema inicial (v1.0.0): Arquitectura y limitaciones identificadas

El sistema original SHARA-WoZ v1.0.0 implementaba una arquitectura web tripartita (Figura 2): una aplicación web React/Three.js para simulación 3D del robot, un servidor central Node.js para coordinación de comunicaciones, y una aplicación desktop Python para el operador WoZ (Cubero et al., 2024). Esta arquitectura permitió evaluaciones remotas de interacciones usuario-robot eliminando barreras logísticas del despliegue físico.

La evaluación mediante panel de expertos identificó tres limitaciones críticas que motivaron la re-arquitectura completa. Entre las limitaciones clave, destacan (1) latencia del sistema que comprometía la naturalidad de las interacciones, factor crítico en asistencia geriátrica; (2) ausencia de respuestas pre-generadas que obligaba a redacción manual de cada respuesta del robot; y (3) limitaciones de expresividad emocional. Además, se detectaron problemas técnicos en detección de silencios que provocaban pérdida de mensajes del usuario.

Estas limitaciones evidenciaron la necesidad de priorizar optimización de rendimiento, automatización inteligente de respuestas con preservación de control humano, y modelado emocional sofisticado.

3.2 Sistema mejorado (v2.0.0): Arquitectura y componentes

El sistema SHARA-WoZ v2.0.0 implementa una arquitectura modular rediseñada que aborda sistemáticamente las limitaciones identificadas, manteniendo el modelo tripartito de la figura 2, pero con mejoras sustanciales en cada componente.

3.2.1 Aplicación de escritorio del operador

La interfaz del operador fue completamente reimplementada utilizando PyQt6, *framework* nativo que proporciona mayor rendimiento y capacidades de integración que las tecnologías web empleadas en la primera versión. El sistema se modela organizando sus funcionalidades en servicios especializados que gestionan aspectos específicos del sistema. Estos pueden ser identificados como funciones clave de todo sistema de Mago de Oz para SARs:

- **Gestión de eventos:** Coordina la comunicación entre componentes mediante un sistema de publicación-suscripción desacoplado.
- **Gestión de estado:** Mantiene el estado global del sistema incluyendo conexión, modo de operación y usuario activo.
- **Comunicación en tiempo real:** Maneja la conexión bidireccional con el servidor mediante sockets con reconexión automática.
- **Procesamiento de mensajes:** Coordina el flujo conversacional entre modos automático y manual.
- **Streaming de vídeo:** Procesa y renderiza el *feed* de vídeo del usuario en tiempo real.

La implementación particular de la interfaz presenta tres paneles principales (Figura 3): un panel de chat con historial conversacional y controles de estado emocional del robot, un panel de vídeo que muestra al usuario en tiempo real, y un navegador web integrado que replica la interfaz que ve el usuario final. Se incluye un panel de control inferior que permite la manipulación de la simulación fuera del flujo conversacional habitual, de forma que el operador puede manipular la simulación a todos los niveles.

3.2.2 Integración con inteligencia artificial

La integración con servicios de OpenAI representa la innovación más significativa del sistema en esta versión. La propuesta arquitectónica aprovecha los avances de la IA generativa y los sistemas basados en LLMs para una generación dinámica de posibles respuestas del robot. El servidor implementa dos estrategias complementarias de generación de respuestas:

- **Generación de respuesta principal:** El sistema genera una respuesta conversacional utilizando un contexto temporal e identificación de usuario para mayor personalización en las conversaciones, de la misma forma que el robot físico.
- **Generación de variantes emocionales:** La innovación consiste en generar simultáneamente ocho variantes de la respuesta principal, cada una adaptada a un estado

emocional específico del robot (sorpresa, neutral, negación, afirmación, enfado, tristeza, alegría, sonrojada). Las variantes se generan en paralelo para minimizar latencia, proporcionando al operador opciones contextualizadas listas para envío inmediato.

Esta arquitectura de respuestas múltiples reduce en gran medida el tiempo de respuesta en modo semi-automático mientras que preserva la capacidad de supervisión y modificación humana (Figura 4).

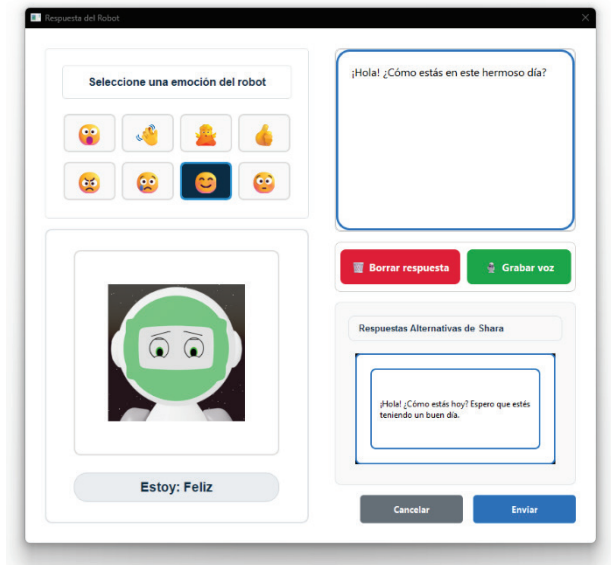


Figura 4: Interfaz de respuesta del operador. Muestra la respuesta generada por la simulación y para cada uno de sus estados aporta una respuesta alternativa. El operador puede modificar las respuestas o aceptarlas.

3.2.3 Servidor central y servicios cloud

Se dispone de un servidor Node.js como centro de comunicación, que incorpora optimizaciones significativas. Gestiona conversaciones por usuario con persistencia de historial completo, coordina el envío diferencial de mensajes según el modo operacional activo (automático, semi-automático o manual), e integra servicios cloud de Google para transcripción de audio2 y síntesis de voz3 con configuración optimizada como el robot físico.

El servidor implementa lógica de coordinación que en modo automático envía respuestas directamente sin intervención del operador, en modo semi-automático presenta opciones al operador para aprobación/modificación, y en modo manual permite composición completa por el operador.

²<https://cloud.google.com/speech-to-text> Último acceso: 04/11/2025

³<https://cloud.google.com/text-to-speech> Último acceso: 04/11/2025

3.2.4 Interfaz web del usuario final

La interfaz React/Three.js del usuario final se encarga de la simulación 3D validada del robot con animaciones faciales y corporales sincronizadas con estados emocionales. Esta continuidad garantiza comparabilidad con evaluaciones previas (Figura 5).

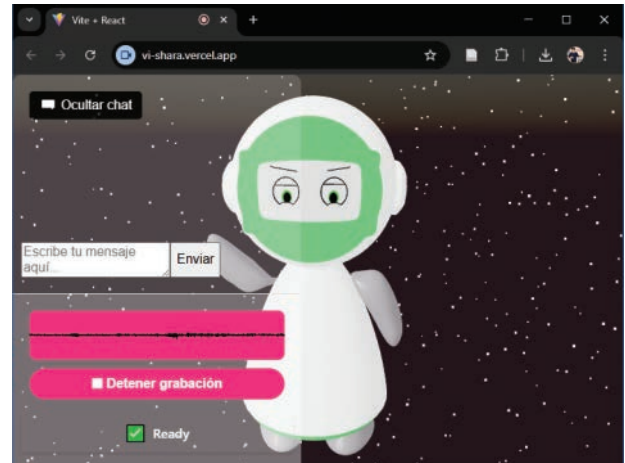


Figura 5: Interfaz web de la simulación del robot SHARA. Muestra una representación 3D y da la opción de ver un historial conversacional en forma de chat al igual que un indicador del estado de la simulación.

3.3 Justificación de decisiones de diseño

Las decisiones arquitectónicas responden sistemáticamente a las limitaciones identificadas y constituyen las contribuciones principales del artículo al aportar un modelo arquitectónico y basado en evidencias de los sistemas de Mago de Oz para SARs:

Migración a PyQt6 nativo: Reduce latencia mediante eliminación de sobrecarga de comunicación navegador-aplicación y proporciona procesamiento multi-hilo eficiente para vídeo y mensajes sin bloqueo de interfaz. Las evaluaciones internas mostraron una reducción del 40% en tiempo de renderizado comparando con la implementación web híbrida de la v1.0.0.

Sustitución de servicio de generación de respuestas: El sistema previo usaba la combinación de los servicios de Watson Assistant, Natural Language Understanding (NLU) y los servicios de traducción de Google para generar las respuestas del robot, mientras que la nueva implementación usa los servicios de GPT como sustitución, ya que elimina la necesidad de traducción y engloba las dos funcionalidades que se empleaban con los servicios de Watson.

Sistema de respuestas múltiples: Aborda simultáneamente las limitaciones de "respuestas pre-generadas" y "protocolo de comunicación social" proporcionando opciones

contextualizadas que mantienen al operador en el bucle de decisión para juicios clínicos sobre apropiación emocional.

Arquitectura modular: Facilita extensibilidad para incorporación futura de funcionalidades identificadas por expertos (detección de emergencias, análisis de comportamiento no verbal).

Esta arquitectura establece una base extensible para mejoras futuras mientras resuelve las limitaciones identificadas por expertos en el sistema original.

4. Resultados

Esta sección presenta una evaluación comparativa entre las versiones v1.0.0 y v2.0.0 del sistema SHARA-WoZ, analizando métricas de rendimiento para aplicaciones de robótica social asistencial. Las mediciones se realizaron en condiciones controladas para garantizar comparabilidad entre ambas arquitecturas.

4.1 Métricas de latencia conversacional

La latencia de respuesta conversacional constituye un factor crítico en interacciones usuario-robot, especialmente en contextos geriátricos donde la naturalidad de la comunicación impacta directamente en la aceptación del sistema (Wang, Y. L., & Lo, C. W. 2025). La tabla 1 presenta una comparativa detallada de los tiempos de respuesta entre los servicios de procesamiento de lenguaje natural empleados en cada versión.

Tabla 1: Comparativa de latencia en servicios de procesamiento conversacional

Componente	V1.0.0	V2.0.0	Reducción
Transcripción de voz	Google STT: ~1000 ms	Google STT: ~1000 ms	0%
Traducción	Google Translate: 800-1300 ms	No requerido	100%
Síntesis de voz	Google TTS: 500-800 ms	Google TTS: 500-800 ms	0%
Procesamiento IA	IBM Watson Assistant + IBM NLU: 1500-3000 ms	GPT-4o-mini: 200-500 ms	80-85%
Latencia total (promedio)	3800-5100 ms	1700-2300 ms	55%
Latencia p95	6500 ms	3000 ms	54%

La implementación de GPT-4o-mini representa la mejora más significativa, eliminando la necesidad de servicios de traducción intermedios y proporcionando respuestas con latencias optimizadas para aplicaciones interactivas en tiempo real. La generación paralela de respuestas

emocionales mantiene tiempos competitivos mediante procesamiento concurrente.

4.2 Rendimiento de interfaz del operador

La migración de una arquitectura híbrida web-nativa a PyQt6 puro impacta significativamente en la velocidad de respuesta de la interfaz del operador, como se detalla en la tabla 2.

Tabla 2: Métricas de rendimiento de interfaz del operador

Métrica	V1.0.0	V2.0.0	Mejora
Tiempo de inicialización	3000-5000 ms	800-1200 ms	73%
Renderización inicial UI	450-600 ms	250-350 ms	42%
Actualización de chat	120-180 ms	50-80 ms	56%
Procesamiento de vídeo (frame)	45-60 ms	25-35 ms	42%
Uso de memoria (base)	280-350 MB	120-180 MB	49%
Uso de CPU (idle)	8-12%	3-5%	58%

La eliminación de la sobrecarga del renderizado web y la utilización de bibliotecas nativas de Qt resulta en mejoras sustanciales de rendimiento.

4.3 Comparativa de arquitectura de despliegue

Por último, la tabla 3 analiza las diferencias en latencia y disponibilidad entre el despliegue local de v1.0.0, para el cual se plantea un despliegue en un servidor local, y la infraestructura cloud de v2.0.0.

Tabla 3: Métricas de infraestructura y despliegue

Aspecto	V1.0.0 (Local)	V2.0.0 (Cloud)	Diferencia
Latencia red	<5 ms (LAN)	15-30 ms	+10-25 ms
Disponibilidad	95-98%	99.9%	+2-4%
Tiempo de despliegue	15-30 min	2-5 min	85% reducción
Escalabilidad	Manual	Automática	-
Cold start (primera solicitud)	0 ms	~500 ms	+~500 ms
Coste/mes	>0€ (servidor local)	0€ (plan gratuito)	Variable

Aunque la arquitectura *cloud* introduce latencias de red adicionales, estas son compensadas por las mejoras en el procesamiento de IA y la disponibilidad global. Los servicios *edge* de Vercel minimizan la latencia para usuarios

geográficamente distribuidos mediante caching y distribución global.

5. Conclusiones

El presente trabajo ha descrito el desarrollo e implementación del sistema SHARA-WoZ v2.0.0, una evolución de la plataforma de evaluación para robots sociales asistenciales que aborda sistemáticamente las limitaciones identificadas en su versión anterior. Los resultados obtenidos validan las decisiones arquitectónicas adoptadas y demuestran mejoras cuantificables en múltiples dimensiones críticas para la evaluación efectiva de interacciones humano-robot para adultos mayores.

5.1 Lecciones aprendidas arquitecturales

Es posible identificar cinco aspectos fundamentales en sistemas de Mago de Oz para SARs que se han validado exitosamente a través de métricas objetivas. La migración a PyQt6 nativo ha resultado en una reducción del 42% en tiempos de renderizado y del 49% en uso de memoria, confirmando las ventajas de arquitecturas nativas sobre soluciones híbridas web. La integración de GPT-4o-mini ha transformado el sistema conversacional, reduciendo la latencia total en un 55% y eliminando la necesidad de servicios de traducción intermedios. El sistema de generación de respuestas múltiples por estado emocional ha reducido el tiempo de respuesta en modo semi-automático de 15-30 segundos a 2-3 segundos, representando una mejora del 87-93% mientras preserva el control humano esencial en aplicaciones clínicas.

La arquitectura modular implementada facilita la extensibilidad del sistema sin refactorización mayor, como evidencian los servicios especializados desacoplados que gestionan eventos, estado, comunicación y procesamiento de manera independiente. El rediseño de la interfaz del operador, basado en patrones de diseño modernos, ha mejorado significativamente la supervisión simultánea de múltiples canales de interacción.

5.2 Implicaciones para la robótica social asistencial

Este trabajo contribuye al campo emergente de la robótica social asistencial proporcionando una plataforma de evaluación que equilibra sofisticación técnica con aplicabilidad práctica. La experiencia acumulada durante el desarrollo del sistema permite identificar aspectos fundamentales que todo sistema de Mago de Oz para robots sociales asistenciales debería contemplar.

En primer lugar, la minimización de la latencia conversacional emerge como factor clave. Los resultados demuestran que latencias superiores a 3 segundos

comprometen significativamente la naturalidad de las interacciones, especialmente en contextos geriátricos donde la fluidez comunicativa impacta directamente en la aceptación del sistema. La arquitectura debe priorizar la optimización de cada componente del flujo conversacional, desde la transcripción de voz hasta la síntesis de respuestas.

En segundo lugar, la gestión modular de funcionalidades mediante servicios especializados facilita tanto el desarrollo como el mantenimiento del sistema. La separación de responsabilidades en servicios independientes para gestión de eventos, estado, comunicación en tiempo real, procesamiento de mensajes y streaming de vídeo permite escalabilidad, extensibilidad y facilita la incorporación de nuevas capacidades sin comprometer la estabilidad del sistema existente.

En tercer lugar, la integración de inteligencia artificial conversacional contextualizada representa un elemento diferenciador. La capacidad de generar respuestas personalizadas considerando el historial conversacional y el contexto temporal del usuario, combinada con la generación simultánea de variantes emocionales, proporciona al operador herramientas para mantener interacciones naturales y coherentes sin sacrificar la supervisión humana en aplicaciones clínicas.

Finalmente, la capacidad del sistema para operar en modos automático, semi-automático y manual responde a la necesidad identificada en la literatura de transiciones graduales desde control humano completo hacia autonomía robótica. Esta flexibilidad operacional resulta esencial en contextos geriátricos donde la adaptabilidad y sensibilidad contextual son cruciales para la aceptación y efectividad terapéutica del sistema.

5.3 Trabajo futuro

Las direcciones futuras de desarrollo se centran en tres áreas prioritarias identificadas durante la evaluación del sistema.

La implementación de un repertorio gestual ampliado y expresiones faciales mejoradas aumentará la naturalidad de las interacciones. Se desarrollarán comportamientos proactivos que permitan al robot iniciar conversaciones contextualmente apropiadas, junto con un sistema de memoria conversacional a largo plazo para mantener continuidad entre sesiones.

El rediseño de la interfaz seguirá principios de diseño centrado en la tarea, simplificando componentes y facilitando el control de la simulación y de la propia interfaz del operador.

Estas mejoras se implementarán iterativamente con validación continua por usuarios finales y expertos del

dominio, manteniendo el equilibrio entre funcionalidad avanzada y simplicidad operacional que caracteriza al sistema actual.

Además, se plantea la evaluación y validación de este sistema con usuarios expertos en el ámbito de la interacción persona ordenador, persona robot y expertos en el ámbito sanitario para evaluar la calidad conversacional de la simulación. La evaluación se centraría en evaluar los componentes de la interfaz en su usabilidad, y cómo de sencillo resultaría emplear esta forma de evaluar robots sociales durante su desarrollo.

5.4 Contribución

El sistema SHARA-WoZ v2.0.0 representa un avance significativo en la evolución de plataformas de evaluación para robótica social asistencial. Al abordar sistemáticamente las limitaciones técnicas mientras mantiene la flexibilidad operacional, este trabajo establece un precedente para el desarrollo de herramientas de evaluación que equilibran rigurosidad técnica con aplicabilidad práctica. La naturaleza modular y extensible de la arquitectura propuesta facilita su

adaptación a diversos contextos de investigación y dominios de aplicación más allá del cuidado geriátrico.

Los resultados demuestran que es posible desarrollar sistemas WoZ que proporcionen tanto la sofisticación técnica requerida por investigadores como la accesibilidad necesaria para implicados técnicos. Esta democratización del acceso a herramientas de evaluación robótica constituye un paso importante hacia la adopción de robots sociales en contextos asistenciales, contribuyendo en última instancia a mejorar la calidad de vida de poblaciones vulnerables mediante tecnología centrada en el humano.

Agradecimientos

Esta investigación fue financiada por la JUNTA DE COMUNIDADES DE CASTILLA-LA MANCHA mediante las ayudas números SBPLY/21/180501/000160 (SHARA3) y SBPLY/24/180225/000176 (AKAI-SHARA); y por el MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES mediante la ayuda número FPU22/00839 (contrato predoctoral).

Referencias

- Abdi, J., Al-Hindawi, A., Ng, T., & Vizcaychipi, M. P. (2018). Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open*, 8(2), e018815. <https://doi.org/10.1136/bmjopen-2017-018815>
- Alenjung, B., Lindblom, J., Andreasson, R., & Ziemke, T. (2019). User experience in social human-robot interaction. In *Rapid automation: Concepts, methodologies, tools, and applications* (pp. 1468-1490). IGI Global.
- Asi, A. M., Ulate, M. M., Martin, M. F., & van der Roest, H. (2022). Methodologies used to study the feasibility, usability, efficacy, and effectiveness of social robots for elderly adults: Scoping review. *Journal of Medical Internet Research*, 24(8), e37434. <https://doi.org/10.2196/37434>
- Bevilacqua, R., Di Rosa, M., Riccardi, G. R., Pelliccioni, G., Lattanzio, F., Felici, E., Margaritini, A., Amabili, G., & Maranesi, E. (2022). Design and development of a scale for evaluating the acceptance of social robotics for older people: The robot era inventory. *Frontiers in Neurorobotics*, 16, 883106. <https://doi.org/10.3389/fnbot.2022.883106>
- Coronado, E., Kiyokawa, T., Garcia Ricardez, G. A., Ramirez-Alpizar, I. G., Venture, G., & Yamanobe, N. (2022). Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0. *Journal of Manufacturing Systems*, 63, 392-410. <https://doi.org/10.1016/j.jmsy.2022.04.007>
- Cubero, G., Villa, L., Favela, J., Ochoa, S., Díaz-Fernández, C., & Hervás, R. (2024, November). 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 *International Conference on Ubiquitous Computing and Ambient Intelligence* (pp. 87-98). Cham: Springer Nature Switzerland. Feil-Seifer, D., & Matarić, M. J. (2005). Defining socially assistive robotics. In *Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics* (pp. 465-468). IEEE. <https://doi.org/10.1109/ICORR.2005.1501143>
- Hoffman, G. (2016). OpenWoz: A runtime-configurable wizard-of-oz framework for human-robot interaction. In *AAAI Spring Symposium - Technical Report* (Vol. SS-16-01-07, pp. 121-126).
- Kachouie, R., Sedighadeli, S., Khosla, R., & Chu, M. T. (2014). Socially assistive robots in elderly care: A mixed-method systematic literature review. *International Journal of Human-Computer Interaction*, 30(5), 369-393. <https://doi.org/10.1080/10447318.2013.873278>
- Koh, W. Q., Felding, S. A., Budak, K. B., Toomey, E., & Casey, D. (2021). Barriers and facilitators to the implementation of social robots for older adults and people with dementia: A scoping review. *BMC Geriatrics*, 21(1), 351. <https://doi.org/10.1186/s12877-021-02277-9>
- Porfirio, D., Saupé, A., Albarghouthi, A., & Mutlu, B. (2018). Authoring and verifying human-robot interactions. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (pp. 75-86). <https://doi.org/10.1145/3242587.3242634>
- Riek, L. D. (2012). Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1), 119-136. <https://doi.org/10.5898/JHRI.1.1.Riek>

- Rigaud, A. S., Dacunha, S., Harzo, C., Lenoir, H., Sfeir, I., Piccoli, M., & Pino, M. (2024). Implementation of socially assistive robots in geriatric care institutions: Healthcare professionals' perspectives and identification of facilitating factors and barriers. *Journal of Rehabilitation and Assistive Technologies Engineering*, 11, 20556683241284765.
- Sequeira, P., Alves-Oliveira, P., Ribeiro, T., Di Tullio, E., Petisca, S., Melo, F. S., Castellano, G., & Paiva, A. (2016). Discovering social interaction strategies for robots from restricted-perception wizard-of-oz studies. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 197-204). IEEE. <https://doi.org/10.1109/HRI.2016.7451752>
- Tobis, S., Piasek-Skupna, J., Neumann-Podczaska, A., Suwalska, A., & Wieczorowska-Tobis, K. (2023). The effects of stakeholder perceptions on the use of humanoid robots in care for older adults: Post-interaction cross-sectional study. *Journal of Medical Internet Research*, 25, e46617. <https://doi.org/10.2196/46617>
- Villa, L., Hervás, R., Cabañero, L., Fontecha, J., López, G., & Favela, J. (2025). Evaluating proactivity levels in socially assistive robots for elderly care: a user adoption assessment. *Behaviour & Information Technology*, 1-26.
- Villa, L., Hervás, R., Dobrescu, C. C., Cruz-Sandoval, D., & Favela, J. (2022, June). Incorporating affective proactive behavior to a social companion robot for community dwelling older adults. In *International Conference on Human-Computer Interaction* (pp. 568-575). Cham: Springer Nature Switzerland.
- Villa, L., Hervás, R., Cruz-Sandoval, D., & Favela, J. (2022, November). Design and evaluation of proactive behavior in conversational assistants: approach with the eva companion robot. In *International Conference on Ubiquitous Computing and Ambient Intelligence* (pp. 234-245). Cham: Springer International Publishing.
- Zhao, D., Sun, X., Shan, B., Yang, Z., Yang, J., Liu, H., Jiang, Y., & Hiroshi, Y. (2023). Research status of elderly-care robots and safe human-robot interaction methods. *Frontiers in Neuroscience*, 17, 1291682. <https://doi.org/10.3389/fnins.2023.1291682>
- Papadopoulos, C., Castro, N., Nigath, A., Davidson, R., Faulkes, N., Menicatti, R., ... & Sgorbissa, A. (2022). The CARESSES randomised controlled trial: exploring the health-related impact of culturally competent artificial intelligence embedded into socially assistive robots and tested in older adult care homes. *International Journal of Social Robotics*, 14(1), 245-256.
- Wang, Y. L., & Lo, C. W. (2025). The effects of response time on older and young adults' interaction experience with Chatbot. *BMC psychology*, 13(1), 150.