



CRITERIOS PARA LOS ASISTENTES CONVERSACIONALES EN EL ÁMBITO DE LA SALUD






Prólogo

En los últimos años, la irrupción de la inteligencia artificial generativa ha transformado profundamente la manera en que concebimos la atención sanitaria, la gestión del conocimiento y la relación entre profesionales y pacientes. En un contexto marcado por el incremento de la cronicidad, la sobrecarga asistencial y la necesidad de modelos más equitativos y sostenibles, los asistentes conversacionales emergen como una herramienta con un enorme potencial, pero también esto supone un horizonte con nuevos desafíos éticos, comunicativos y técnicos que requieren una reflexión colectiva para afrontar el reto.

El I Consenso IA Generativa en Dolor, celebrado el 19 de octubre, surge precisamente de esta necesidad de establecer criterios claros que garanticen un uso responsable, efectivo y seguro de estos sistemas. A lo largo del encuentro, personas expertas de diferentes disciplinas analizaron las oportunidades y riesgos de estas tecnologías y elaboraron un marco estructurado en torno a tres pilares fundamentales: la información veraz, la comunicación y la ética. Este documento recoge las conclusiones de las personas expertas y ofrece una guía práctica para orientar el desarrollo, implementación y evaluación de asistentes conversacionales en el ámbito de la salud, con especial atención al manejo del dolor crónico.

Más allá de la dimensión técnica, este consenso pone el foco en lo verdaderamente esencial: asegurar que la IA complementa, pero nunca sustituye, la labor de las y los profesionales sanitarios; que amplía la capacidad de respuesta del sistema sin comprometer la dignidad, la autonomía y la seguridad de la ciudadanía; y que se construye sobre principios sólidos de equidad, transparencia y humanidad.



El presente documento pretende servir como referencia para equipos clínicos, desarrolladores, responsables institucionales y cualquier agente con implicación en la implantación de herramientas conversacionales en salud. Su propósito es doble: por un lado, impulsar una IA útil, rigurosa y alineada con las necesidades reales de la población; por otro, fomentar una reflexión ética continua que permita anticipar riesgos y asegurar que la tecnología avanza siempre al servicio de las personas.

Aurora María Araújo Narváez
Decana ICPFCM


Introducción

El Instituto #SaludsinBulos y el Colegio Profesional de Fisioterapeutas de la Comunidad de Madrid (CPFCEM) reunimos el pasado año a sociedades científicas, colegios profesionales y asociaciones de pacientes para crear el primer consenso sobre asistentes conversacionales en salud. Nació como un intento de definir unas líneas estratégicas para introducir de forma explícita la ética en la IA aplicada a la asistencia sanitaria. El consenso fija criterios para el desarrollo y validación de chatbots basados en grandes modelos de lenguaje (LLM): transparencia de fuentes e intereses, veracidad, supervisión profesional, respeto a la privacidad y enfoque empático hacia los pacientes, entre otros requisitos.

Nuestra intención es que sea una herramienta práctica para aprovechar el potencial de los asistentes conversacionales (mejorar el diagnóstico y tratamiento, orientar frente a los bulos, reforzar la educación sanitaria) sin perder de vista sus riesgos: sesgos, alucinaciones, autodiagnóstico irresponsable o vulneración de la privacidad.

Este marco ético ha servido de guía directa para la creación de la iniciativa Mírame a los Ojos, que busca fomentar una comunicación clínica eficaz con la ayuda de la IA. En colaboración con sociedades científicas, colegios profesionales y asociaciones de pacientes, está orientada a devolver la mirada al paciente con el apoyo de la IA. Se apoya en escribas digitales (sistemas de ayuda al diagnóstico que transcriben y resumen la conversación clínica), asistentes conversacionales y agentes de voz que facilitan la escucha activa, la explicación comprensible y la toma de decisiones compartida.





La iniciativa incorpora líneas de trabajo que desarrollan y aterrizan el consenso: elaboración de guías para un uso responsable de la IA en la consulta, programas de formación para profesionales sanitarios y pacientes, y proyectos piloto para mejorar la relación asistencial, reducir la brecha digital y adaptar la comunicación a las necesidades de personas mayores y colectivos vulnerables.

La IA está transformando la asistencia sanitaria a pasos agigantados y, aunque parezca paradójico, ayudando a humanizar la comunicación en salud, pero necesita marcos éticos como los que propone este consenso. En el último año se han producido avances en la legislación, la ética y la tecnología que hacen más necesario que nunca la colaboración de profesionales sanitarios, pacientes, gestores, empresas y Administración, como propone el consenso.

ÉTICA EN LA IA

En el campo de la ética, la Organización Mundial de la Salud ha publicado la guía *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*¹, que analiza usos potenciales de los LMM en diagnóstico, asistencia clínica, administración sanitaria, investigación y desarrollo de fármacos, y detalla riesgos como las alucinaciones, los sesgos, la degradación de habilidades clínicas o la deshumanización de la relación asistencial.

Al mismo tiempo, revisiones² recientes sobre LLM en medicina insisten en la necesidad de marcos robustos de evaluación, métricas específicas para tareas médicas y supervisión humana estrecha antes de su despliegue rutinario, tal como recogíamos en el consenso.





AVANCES LEGALES

En el terreno legal, el AI Act³ europeo (Reglamento (UE) 2024/1689) que ha entrado en vigor, establece el primer marco legal integral sobre IA en el mundo. Su enfoque está basado en riesgos, y clasifica como alto riesgo múltiples usos sanitarios y fija obligaciones específicas de transparencia, entre ellas informar cuando una persona interactúa con un chatbot y garantizar que el contenido generado por IA (incluidos los deepfakes o vídeos ultrarrealistas generados con IA) sea identificable.

Por su parte, en España ha avanzado en la regulación del contenido generado por IA con una propuesta normativa que prevé sanciones muy elevadas si no se etiquetan adecuadamente los contenidos sintéticos, y atribuye a la recientemente creada Agencia Española de Supervisión de la Inteligencia Artificial (AESIA) un papel relevante en la supervisión. Estos cambios normativos refuerzan la importancia de los principios de transparencia, trazabilidad y rendición de cuentas recogidos en el consenso.

MERCADO

En el último año, la IA en salud ha pasado de los pilotos a convertirse en una pieza estructural de la transformación digital, y los agentes conversacionales son uno de los motores principales de ese cambio. El informe más reciente de Future Market Insights⁴ estima que el mercado global de IA conversacional en salud pasará de 17,2 mil millones de dólares en 2025 a 169,5 mil millones en 2035, con los chatbots concentrando alrededor del 35,7 % del mercado y aplicaciones clave en patient engagement, apoyo a la salud mental, monitorización remota, telemedicina y automatización administrativa. En paralelo, los chatbots sanitarios han dejado de ser simples FAQ: hoy se integran con la historia clínica electrónica, ayudan a cribar síntomas, monitorizan enfermedades crónicas y detectan incluso malestar emocional, apoyados en modelos de lenguaje generativos y agentes capaces de planificar tareas y escalar al profesional cuando es necesario.

AGENTES CONVERSACIONALES


Es esa una de las grandes revoluciones en marcha de la IA, los agentes conversacionales. Un asistente conversacional en salud es un sistema que responde a las preguntas del paciente o profesional mediante diálogo (normalmente texto o voz), ofreciendo información, orientación básica o apoyo en tareas muy concretas y predefinidas. Un agente conversacional, en cambio, no solo “habla”, sino que toma decisiones y ejecuta acciones: puede planificar varios pasos, llamar a herramientas (por ejemplo, consultar la historia clínica, generar un informe, programar una cita, lanzar un recordatorio), comprobar resultados y volver a interactuar con la persona en función de lo que ha hecho. Hace un año aún no estaban maduros los grandes modelos multimodales con contextos largos, la infraestructura para integrar de forma segura LLM, datos clínicos y herramientas. Ahora sí contamos con modelos más potentes y eficientes, APIs pensadas para agentes, y experiencias acumuladas que hacen viable pasar del simple chatbot al agente conversacional clínico.

BENEFICIOS

La evidencia reciente señala tres líneas especialmente relevantes de la IA en comunicación clínica: salud mental, formación y triaje/derivación avanzada. Un metaanálisis⁵ de 15 ensayos controlados con 1.974 jóvenes encontró que los agentes conversacionales de IA logran un efecto moderado-alto en la reducción de síntomas depresivos.

En educación médica, un estudio piloto realizado en la Universidad de Málaga con “pacientes virtuales” basados en LLM ha mostrado que se perciben como herramientas de alta usabilidad y útiles para practicar entrevistas clínicas complejas y recibir feedback estructurado sobre habilidades de comunicación⁶.





Al mismo tiempo, empiezan a desplegarse plataformas agénticas de triaje como CLARITY⁷, que combinan flujos estructurados con agentes LLM colaborativos: en su implantación en una red interhospitalaria nacional ha gestionado más de 55.000 diálogos en dos meses y ha superado al triaje humano en precisión de derivación inicial, con consultas hasta tres veces más breves.

En conjunto, estas tendencias consolidan a los agentes conversacionales como una pieza central tanto en la atención directa como en la formación de competencias comunicativas, y refuerzan la necesidad de marcos éticos y guías

RIESGOS

Uno de los riesgos que ya se apuntaban en el consenso y que se ha consolidado en 2025 es el de la propagación de los vídeos falsos o deep fakes en salud. Un estudio publicado en Internal Medicine Journal⁸ comprobó cómo los deepfakes generados por IA están empezando a afectar de forma directa a la salud. Los autores describen casos reales en los que se han creado vídeos manipulados de reconocidos especialistas en diabetes para promocionar suplementos o dispositivos no validados, sin autorización de los propios expertos, y advierten de que este tipo de contenidos puede llevar a algunos pacientes a abandonar tratamientos basados en la evidencia, gastar dinero en productos inútiles y perder confianza en sus médicos y en las instituciones sanitarias. El estudio subraya tres grandes riesgos: daño clínico potencial (por cambios en la medicación guiados por bulos), erosión de la confianza pública en la información sanitaria legítima y daño reputacional y emocional para los profesionales cuya imagen se suplanta. También propone respuestas coordinadas que van desde la formación de los sanitarios para reconocer y denunciar deepfakes, en línea con lo que hacemos en #SaludsinBulos, hasta el refuerzo de la regulación y el desarrollo de mejores herramientas de detección por parte de plataformas y desarrolladores de IA.

Desde el Instituto #SaludsinBulos seguiremos trabajando con sociedades científicas, colegios profesionales y asociaciones de pacientes para la comunicación clínica sea una de las herramientas más importantes contra esta y otras amenazas a la información veraz en salud.

Carlos Mateos, coordinador del Instituto #SaludsinBulos



BIBLIOGRAFÍA

¹World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. Geneva: World Health Organization; 2025.

²Chen X, Xiang J, Lu S, et al. Evaluating large language models and agents in healthcare: key challenges in clinical applications. *Intelligent Medicine*. 2025;5(3):151-163. doi:10.1016/j.imed.2025.03.002

³Unión Europea. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. 12 Jul 2024.

⁴Future Market Insights. Conversational AI in Healthcare Market Size and Share Forecast Outlook 2025 to 2035. 9 Oct 2025.

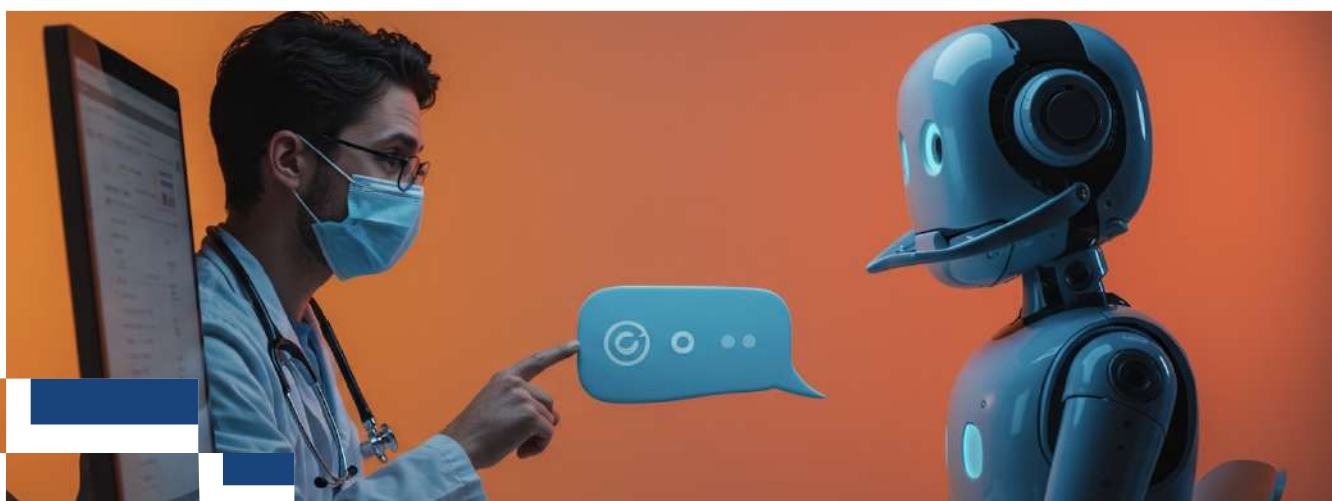
⁵Feng Y, Wu W, Hang Y, et al. Effectiveness of AI-Driven Conversational Agents in Improving Mental Health Among Young People: Systematic Review and Meta-Analysis. *J Med Internet Res*. 2025;27:e69639. doi:10.2196/69639

⁶Zambrano-Serrano B, Cueto-Galán R, Escamilla-Sánchez A, et al. AI-Driven Virtual Patient Simulations for Communication Training in Medical Students: A Pilot Study. *Preprint*, 2025; doi:10.21203/rs.3.rs-7270436/v1.

⁷Shaposhnikov V, Nesterov A, Kopanichuk I, et al. CLARITY: Clinical Assistant for Routing, Inference, and Triage. *arXiv preprint arXiv:2510.02463*. 2025.

⁸Jenkins AJ, Shaw JE, Cohen ND, et al. Artificial intelligence-generated deepfakes in healthcare and biomedical research: major adverse implications for patients, public and professionals. *Internal Medicine Journal*. 2025;55(9):1583-1586. doi:10.1111/imj.70152

En octubre de 2024 se celebró el I Consenso IA generativa en Salud. Los participantes debatieron los criterios que debe cumplir un asistente conversacional (chatbot) y llegaron a tres puntos sustantivos: la información veraz, la comunicación y la ética. Para llegar a esta conclusión primero se trataron las oportunidades y riesgos de estas tecnologías.



Oportunidades

- **Equidad en la atención.** Los chatbots pueden estar disponibles para cualquier persona con acceso a internet, garantizando que la herramienta sea accesible sin importar la ubicación del paciente, su lengua, su grado de alfabetización digital, cultura y diversidad funcional.
 - **Atención inmediata.** La herramienta está disponible siempre que el paciente lo necesita.
 - **Control continuo del dolor crónico,** permitiendo un monitoreo constante en este tipo de pacientes, cuya atención representa el 80% del gasto sanitario y facilitando un seguimiento que el personal sanitario a menudo no puede ofrecer por falta de tiempo.
 - **Apoyo al diagnóstico,** optimizando el tiempo en el análisis e interpretación de pruebas y apoyando la toma de decisiones clínicas.
-

- **Sostenibilidad del sistema sanitario.** Con la jubilación masiva de profesionales en los últimos años, (y que aumentará en los siguientes), la IA puede ayudar a reducir la carga de trabajo y ayudar a paliar el déficit de trabajadores.

- **Reinvención profesional.** La IA fomenta el aprendizaje continuo y el desarrollo profesional de los sanitarios, creando nuevas dinámicas que facilitan y enriquecen su labor.

- **Ampliación de la información,** facilitando la recopilación de datos de diversas fuentes y ofreciendo una base de información más completa para la atención del paciente, con un formato que pueda entender.



Riesgos

- **Sesgos económicos.** La persistencia del dolor crónico genera una demanda continua de soluciones, lo que algunas empresas podrían ver como una oportunidad para ofrecer sus servicios con fines comerciales, influyendo potencialmente en la imparcialidad de las recomendaciones.
 - **Ausencia de normativa internacional** detallada y unificada en todos los países que asegure el uso seguro y ético de estos asistentes conversacionales.
 - **Infravalorar la competencia de los profesionales.** Las respuestas y diagnósticos de los chatbots no deben reemplazar la experiencia y el juicio clínico de los profesionales de la salud.
 - **Falta de formación y preparación del paciente** en el uso de este tipo de herramientas puede generar frustración y malentendidos.
-

- **Falsa sensación de seguridad y reducción de la responsabilidad individual** debido a la excesiva confianza en la herramienta y a su capacidad de responder de forma correcta, lo que podrían llevar a la toma de malas decisiones.
- **Falta de personalización.** Si la herramienta tiende a basarse solo en los síntomas, sin preguntar sobre características individuales (como edad, género, raza), sus respuestas pueden no estar adecuadas al contexto personal de cada paciente.
- **Falta de consideración al impacto de la información** al comunicar información sobre el estado de salud sin tener en cuenta cómo esta podría afectar emocionalmente al paciente.
- **Falta de contexto.** Si el chatbot solo reconoce los síntomas físicos y no capta el contexto completo de la dolencia, incluidas las circunstancias sociales y emocionales del paciente, se verá limitada la comprensión de las vulnerabilidades individuales.

Requisitos de los asistentes conversacionales en dolor:

INFORMACIÓN VERAZ:

- **Adaptación a las necesidades del paciente** proporcionando información precisa que responda a las necesidades individuales del paciente, sin comprometer la seguridad de sus datos.
- **Educación en salud.** Contribución a la educación en salud, promoviendo la identificación y difusión de información confiable para contrarrestar la desinformación.



- **Amplitud de documentación.** Aprovechando la gran capacidad de almacenamiento de estos sistemas se contará con una amplia base de información que mejorará el apoyo al diagnóstico.
- **Confiabilidad de los datos e identificación de las fuentes.** Los conjuntos de datos se crearán con la colaboración de profesionales de la salud y se basarán en fuentes confiables, señalando claramente el origen de la información y referencias utilizadas.
- **Diversidad de fuentes.** Es esencial contar con múltiples fuentes para contrastar la información y facilitar un diagnóstico más preciso.
- **Actualización constante de la información** en la que se basan los asistentes conversacionales para mantenerlos actualizados y evitar obsolescencia y sesgos.
- **Dinamismo en la información del paciente.** Incorporación constante de nueva información sobre el paciente y concepción de la ayuda diagnóstica como un proceso de mejora continua, en función de los cambios en la salud del paciente y las actualizaciones clínicas.
- **Veracidad.** Integración de mecanismos que aseguren la exactitud y veracidad de la información.

COMUNICACIÓN:

- **Escucha activa.** Captación tanto de las necesidades como las circunstancias específicas del paciente para poder orientarlo de manera adecuada.
- **Empatía.** Comunicación con empatía, adaptándose al estado emocional y percepción del paciente para fomentar una buena adherencia al tratamiento y al seguimiento del plan terapéutico.



- **Intercambio de información recíproca.** El paciente podrá proporcionar detalles sobre su estado y condicionantes de salud y confirmar la interpretación que realiza el chatbot, lo que permitirá un diagnóstico y tratamiento personalizados con menor riesgo de error.
- **Individualización.** Información adaptada a los datos específicos que el paciente ofrece voluntariamente.
- **Adaptabilidad.** Información ajustada según la rutina y el ritmo de vida del paciente.
- **Enfoque integral.** Las recomendaciones promoverán mejoras en las actividades diarias del paciente y no limitarse a recordatorios sobre citas o medicación.

ÉTICA:

- **Cumplimiento legal.** Operativa de acuerdo con las normativas vigentes para garantizar el uso ético y seguro.
- **Responsabilidad.** Mecanismos que aseguren la responsabilidad por la exactitud o veracidad de la información proporcionada, con actualizaciones constantes y la verificación de un equipo médico. Reconocimiento explícito sobre quién recae la responsabilidad y en qué medida (incluida la jurídica) y el establecimiento de mecanismos regulatorios que prevengan potenciales efectos adversos en los pacientes.
- **Equidad en el acceso.** Igualdad de oportunidades de acceso, sin discriminación geográfica, social, funcional o de cualquier otra índole.
- **Prevención.** Prevención o remedio de efectos sistémicos desfavorables en poblaciones específicas de pacientes.



- **Trazabilidad.** Documentación de forma exhaustiva de la información técnica relativa al desarrollo del sistema, incluido su conjunto de datos y la evaluación de su rendimiento.
- **Transparencia.** Explicación, en la medida de lo posible, de la lógica de los procesos subyacentes sobre los que se sustentan los resultados de los modelos.

- **Ausencia de sesgos.** Garantía de ausencia de sesgos que perjudiquen la calidad de la información disponible, sean individuales, sociales o comerciales.

- **Autonomía.** Preservación de la autonomía de los pacientes, respetando sus decisiones.



- **Protección de la dignidad de los pacientes,** sus derechos de autodeterminación y su capacidad para tomar decisiones.
 - **Disponibilidad de información comprensible** para que los pacientes puedan utilizarla de acuerdo con sus valores, evitando el riesgo de la sobreutilización e incluso de conductas adictivas.
 - **Supervisión e intervención del profesional sanitario** en tiempo y forma. Fomento de la comprensión de que la IA es una herramienta complementaria, que no reemplaza a los profesionales.
 - **Privacidad/Confidencialidad.** Preservación de la protección de la información sensible de los pacientes frente al acceso ilegítimo, así como de la confidencialidad de los datos con el fin de preservar su intimidad e impedir su identificación si lo desea.
-

- **Seguridad.** Mecanismos robustos que garanticen la seguridad de los datos privados mediante evaluación continua de las vulnerabilidades. Medidas preventivas ante filtraciones, ciberataques u otras amenazas.
- **No-maleficencia.** Prevención de daños y gestión de riesgos potenciales a los pacientes debido a información incorrecta o engañosa (alucinaciones).
- **Promoción de la salud pública.** Desarrollo de sistemas seguros, precisos, eficaces y de alta calidad para priorizar la salud pública y el bienestar de las personas.
- **Propiedad intelectual.** Reconocimiento legítimo de la propiedad del trabajo intelectual.
- **Colaboración Humano-Máquina.** Promoción del uso de los sistemas de IA como herramientas al servicio del conocimiento facultativo.



Panelistas



Patricia Llaque

Directora de Bioética y
Neuroética de Odisela



Juan Iniesta

Farmacéutico técnico



Esther Nieto

Presidenta de FAECAP



Manuela Monleón

Vpta. y coord. del grupo de
atención a personas con dolor
desde AP de FAECAP



Eduardo Fondevila

Presidente de la Sociedade
Galega de Fisioterapeutas
(SOGAFI)



Héctor Beltrán

Profesor titular en
Universidad de
Castilla - La Mancha



José Miguel Cacho

Digital health advisor



José Luis Baquero

Responsabilidad Social
Corporativa en OAFI y
AECOSAR



Javier Resa

Coord. Nacional del Grupo
de Trabajo en Dolor y
Cuidados Paliativos de
SEMergen

Panelistas



Juana Sánchez
Responsable Grupo de
Trabajo de Dolor



Elena Plaza
Enfermera experta en
IA



Dr. Diego Benavent
Servicio de
Reumatología del
Hospital Universitario
de Bellvitge



Raúl Ferrer Peña
Vocal II del Colegio Profesional
de Fisioterapeutas de la
Comunidad de Madrid



Carlos Royo
Director de Estrategia de
Salud



Ángeles Ruiz
Farmacéutica /
divulgadora de salud



Esther Mahillo
Directora científica y manager
general del Grupo de Tratamiento
de los Tumores Digestivos (TTD)



Carlos Mateos
coordinador de #SaludsinBulos
y director de la agencia COM
Salud