5.1. Limites de los chatbots

Antes de ver el funcionamiento de los chatbots debemos hacer una pequeña reflexión sobre su uso. Para ello nos basaremos en

> Martínez Comeche, J. A. (2025). *Consideraciones éticas en el empleo de chatbots: Redacción de referencias bibliográficas de textos creados por aplicaciones de Inteligencia Artificial generativa*. Universidad Complutense de Madrid. https://docta.ucm.es/rest/api/core/bitstreams/02323257-5546-4bd4-89d2-ea1abb971dc2/content [Visitado: 15/01/2026]

Este artículo analiza de forma exhaustiva las implicaciones éticas y los desafíos técnicos derivados del uso de chatbots basados en Modelos de Lenguaje Grandes (LLM), como ChatGPT o Gemini,. El autor define estas herramientas como agentes conversacionales que, mediante una arquitectura de red neuronal denominada Transformer, emulan conversaciones humanas para proporcionar información o servicios. A pesar de su éxito masivo, el texto advierte sobre la necesidad de abordar sus vulnerabilidades para garantizar un uso correcto y leal por parte de los usuarios.

Las **principales vulnerabilidades** identificadas en las fuentes se dividen en varias categorías críticas:

- Alucinaciones: Los chatbots generan a menudo información que parece verídica pero es falsa o carece de sentido, incluyendo la invención de referencias bibliográficas inexistentes. Se ha documentado que, en algunos casos, solo el 7% de las referencias proporcionadas por un chatbot eran auténticas. Esto resulta especialmente peligroso en ámbitos como el sanitario o el académico, donde el rigor es esencial.
- Toxicidad y sesgos: Estas herramientas pueden perpetuar y amplificar estereotipos sociales, raciales o de género presentes en sus datos de entrenamiento. Por ejemplo, se han detectado sesgos culturales donde se asocia más frecuentemente a ciertos grupos religiosos con el terrorismo. Además, es posible inducir comportamientos tóxicos si se pide al sistema que simule la personalidad de figuras históricas polémicas.
- Privacidad: Existe un riesgo real de filtración de información sensible (nombres, cuentas bancarias) cargada durante el entrenamiento o proporcionada por los usuarios en sus interacciones,. Esto ha llevado a diversas empresas y países a desarrollar normativas estrictas de protección de datos, como el RGPD en la Unión Europea.

Copyright 2026

- Sostenibilidad: El entrenamiento y mantenimiento de estos modelos supone un alto coste ambiental debido al consumo masivo de electricidad y agua para enfriar los centros de datos. Se estima que para 2027 el consumo energético de la IA podría ser comparable al de países enteros como Argentina.

En cuanto a la originalidad y autoría, el artículo señala que los contenidos generados suelen ser combinaciones de ideas existentes y no investigaciones inéditas, lo que plantea riesgos de plagio y una posible disminución del pensamiento crítico en los estudiantes. Revistas de prestigio como *Nature* han establecido que los LLM no pueden ser considerados autores, ya que no pueden asumir responsabilidades legales.

Como aportación principal, el autor propone una metodología ética para la redacción de referencias bibliográficas de textos creados por IA. Debido a que las respuestas de los *chatbots* no son reproducibles y carecen de un autor físico, sugiere adaptar la normativa APA (7ª edición). La propuesta consiste en utilizar el *prompt* (la petición del usuario) como título de la referencia, indicando entre corchetes que es un texto generado por IA e incluyendo la fecha y la aplicación utilizada. Finalmente, recomienda incluir el texto íntegro generado por el *chatbot* en un apéndice del trabajo para garantizar la transparencia y permitir la verificación de la información.

Este mismo artículo contiene el siguiente apéndice:

> **Apéndice**
> [Índice de aspectos éticos en el empleo de chatbots] (2025, 5 de abril). [Texto generado por Inteligencia Artificial]. Gemini. https://gemini.google.com/app

El texto generado fue el siguiente:
Este índice detalla los aspectos éticos fundamentales que deben considerarse al diseñar, desarrollar, implementar y utilizar aplicaciones chatbot en diversos contextos. Su objetivo es promover un uso responsable y beneficioso de esta tecnología, minimizando los riesgos y maximizando los beneficios para la sociedad.

I. Impacto en el Empleo y la Economía Laboral

A. Desplazamiento Laboral:

- Evaluación del potencial de los chatbots para automatizar tareas y roles laborales existentes.
- Análisis del impacto en diferentes sectores y niveles de cualificación.
- Consideración de estrategias para la transición y la requalificación de la fuerza laboral afectada.

B. Creación de Nuevos Roles:

- Identificación de nuevas oportunidades laborales relacionadas con el diseño, desarrollo, mantenimiento y supervisión de chatbots.
- Análisis de las habilidades y la formación necesarias para estos nuevos roles.

C. Cambios en la Naturaleza del Trabajo:

- Evaluación de cómo los chatbots pueden modificar las tareas y responsabilidades de los trabajadores humanos.
- Consideración del potencial para la colaboración humano-máquina y la mejora de la productividad.

D. Implicaciones para los Salarios y las Condiciones Laborales:

- Análisis del posible impacto de la automatización con chatbots en los niveles salariales. Consideración de las implicaciones para la seguridad laboral y las condiciones de trabajo.

II. Sesgos y Discriminación

A. Incorporación de Sesgos en los Datos de Entrenamiento:

- Evaluación de los datos utilizados para entrenar los chatbots en busca de sesgos históricos, sociales o culturales.
- Identificación de cómo estos sesgos pueden manifestarse en las respuestas y el comportamiento del chatbot.

B. Perpetuación y Amplificación de Sesgos:

- Análisis de cómo los chatbots pueden perpetuar y amplificar los sesgos presentes en sus datos de entrenamiento o en las interacciones con los usuarios.
- Evaluación del impacto de estas respuestas sesgadas en diferentes grupos de usuarios.

C. Diseño para la Equidad y la Inclusión:

- Implementación de estrategias para mitigar y corregir los sesgos en los datos y los algoritmos de los chatbots.
- Diseño de chatbots que sean inclusivos y que traten a todos los usuarios de manera justa y equitativa.

D. Monitoreo y Auditoría de Sesgos:

Copyright 2026

- Establecimiento de mecanismos para monitorear continuamente las respuestas de los chatbots en busca de sesgos.
- Realización de auditorías periódicas para evaluar y corregir los sesgos identificados.

III. Privacidad y Protección de Datos del Usuario

A. Recopilación y Almacenamiento de Datos Personales:

- Evaluación de la cantidad y el tipo de datos personales que los chatbots recopilan y almacenan durante las interacciones.
- Consideración de la necesidad y la proporcionalidad de la recopilación de estos datos.

B. Uso y Compartición de Datos Personales:

- Análisis de cómo se utilizan los datos personales recopilados por los chatbots.
- Evaluación de las políticas y prácticas de compartición de datos con terceros.

C. Seguridad de los Datos:

- Evaluación de las medidas de seguridad implementadas para proteger los datos personales de los usuarios contra accesos no autorizados, pérdidas o filtraciones.
- Cumplimiento de las regulaciones de protección de datos (GDPR, CCPA, etc.).

D. Consentimiento Informado y Transparencia:

- Asegurar que los usuarios estén informados de manera clara y comprensible sobre la recopilación, el uso y el almacenamiento de sus datos personales.
- Obtener el consentimiento informado de los usuarios para el procesamiento de sus datos (cuando sea necesario).

IV. Transparencia y Explicabilidad

A. Claridad sobre la Naturaleza del Chatbot:

- Asegurar que los usuarios sean conscientes de que están interactuando con una inteligencia artificial y no con un ser humano.
- Evitar la simulación engañosa de emociones o conciencia humana.

B. Explicabilidad de las Decisiones y Respuestas:

- En contextos donde las decisiones del chatbot tienen un impacto significativo en los usuarios, considerar la necesidad de proporcionar explicaciones sobre cómo se llegaron a esas decisiones.
- Evaluación de la viabilidad y la utilidad de las explicaciones para diferentes tipos de usuarios.

Copyright 2026

C. Transparencia sobre las Limitaciones:

- Comunicar claramente las capacidades y limitaciones del chatbot a los usuarios.
- Evitar generar expectativas poco realistas sobre lo que el chatbot puede hacer.

V. Responsabilidad y Rendición de Cuentas

A. Definición de Responsabilidades:

- Establecer claramente quién es responsable del diseño, desarrollo, implementación y mantenimiento del chatbot.
- Determinar la responsabilidad en caso de errores, daños o consecuencias negativas derivadas del uso del chatbot.

B. Mecanismos de Retroalimentación y Reclamación:

- Proporcionar a los usuarios canales claros para ofrecer retroalimentación sobre el rendimiento y el comportamiento ético del chatbot.
- Establecer procesos para revisar y responder a las reclamaciones y preocupaciones de los usuarios.

C. Auditoría y Evaluación Ética:

- Realizar auditorías periódicas para evaluar el cumplimiento de los principios éticos en el diseño y la operación del chatbot.
- Involucrar a expertos en ética y a las partes interesadas en el proceso de evaluación.

VI. Uso Malicioso y Daño Potencial

A. Prevención de la Generación de Contenido Dañino:

- Implementar medidas para evitar que los chatbots generen o difundan contenido ofensivo, discriminatorio, falso, engañoso o perjudicial.
- Monitoreo continuo y actualización de los filtros de contenido.

B. Detección y Manejo de Solicitudes Maliciosas:

- Diseño de chatbots para identificar y responder de manera segura a intentos de manipulación o uso malintencionado.
- Evitar la divulgación de información sensible o la realización de acciones perjudiciales a solicitud de usuarios con intenciones maliciosas.

C. Uso en Contextos Sensibles:

Copyright 2026

- Evaluar cuidadosamente los riesgos éticos asociados con el empleo de chatbots en áreas sensibles como la salud, la justicia, la educación o la seguridad.
- Considerar la necesidad de salvaguardias adicionales y supervisión humana en estos contextos.

D. Impacto en la Confianza y la Cohesión Social:

- Considerar cómo el uso generalizado de chatbots podría afectar la confianza en la información y en las interacciones sociales.
- Evaluar el potencial para la difusión de desinformación o la manipulación de la opinión pública.

> Este índice proporciona un marco exhaustivo para abordar los aspectos éticos relacionados con el empleo de chatbots. Su aplicación requiere un compromiso continuo con la reflexión ética, la colaboración multidisciplinaria y la adaptación a medida que la tecnología y su impacto social evolucionan.

Revision #3
Created 2026-01-15 17:25:55 CET by Javier Anzano
Updated 2026-03-26 23:10:49 CET by Miguel Serrano Larraz

Copyright 2026