



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS
Acreditación Institucional de Alta Calidad

ARTÍCULO DE REFLEXIÓN

TECNOLOGÍAS EMERGENTES Y LENGUAJE

Chatbots educativos y ética conversacional

Educational Chatbots and Conversational Ethics

Chatbots educacionais e ética conversacional

Hector Ariel Feruglio Ortiz¹  

Highlights

- La IA conversacional redefine la relación pedagógica humano-máquina.
- Se introduce la ética conversacional como marco para educación con inteligencia artificial.
- Los *chatbots* operan como mediadores activos en procesos formativos.
- Se propone la figura del *cyborg educador* en contextos educativos.
- Este artículo aporta un enfoque crítico frente al uso educativo de inteligencia artificial.

Resumen

Este artículo analiza la incidencia de la inteligencia artificial conversacional en la educación superior, en un contexto donde los sistemas generativos modifican las prácticas sociales y pedagógicas. Se plantea como problema central la redefinición de los vínculos entre humanos y máquinas y sus implicaciones éticas y formativas. Esta investigación adopta un enfoque teórico-conceptual e interdisciplinario, integrando aportes de la filosofía de la técnica, los estudios de la comunicación y la pedagogía crítica. Se realizó un análisis de marcos teóricos y experiencias documentadas para problematizar la interacción pedagógica mediada por sistemas conversacionales. Se observa una transformación en los modelos de interacción educativa, marcada por relaciones parasociales unilaterales, atribución de agencia a sistemas no humanos y nuevas formas de mediación tecnológica. Se propone la figura del *cyborg educador* y se introduce la noción de *ética conversacional* para orientar prácticas más responsables. Se concluye que una pedagogía política, inclusiva y contextualizada resulta clave para promover acoplamientos virtuosos entre humanos y máquinas.

Palabras clave: inteligencia artificial; comunicación; educación; ética; tecnología; aprendizaje.

Abstract

This article examines the impact of conversational artificial intelligence in higher education, within a context where generative systems are reshaping social and pedagogical practices. The central problem addressed is the redefinition of human-machine relations and their ethical and educational implications. This research adopts a theoretical-conceptual and interdisciplinary approach, integrating contributions from the philosophy of technology, communication studies, and critical pedagogy. A review of theoretical frameworks and documented experiences was conducted in order to problematize pedagogical interactions mediated by conversational systems. The findings reveal significant transformations in educational

¹ Doctor en Ciencias Humanas de la Universidad Nacional de Catamarca (UNCA), Argentina. Docente del Departamento de Filosofía – FH-UNCA. Correo electrónico: haferuglio@huma.unca.edu.ar

Cómo citar: Feruglio, H. (2026). Chatbots educativos y ética conversacional. *Enunciación*, 31(1), e24041. <https://doi.org/10.14483/22486798.24041>

Artículo recibido: 22 de agosto de 2025; aprobado: 13 de noviembre de 2025

interaction models, including the emergence of unilateral parasocial relations, the attribution of agency to non-human systems, and new forms of technological mediation. The figure of the *cyborg educator* is proposed, along with the notion of *conversational ethics* to guide more responsible practices. It is concluded that an inclusive and context-sensitive political pedagogy is essential to foster virtuous couplings between humans and machines.

Keywords: artificial intelligence; education; communication; technology; ethics; learning

Resumo

Este artigo analisa o impacto da inteligência artificial conversacional no ensino superior, em um contexto em que os sistemas generativos reconfiguram as práticas sociais e pedagógicas. O problema central abordado é a redefinição das relações entre humanos e máquinas e suas implicações éticas e educativas. A pesquisa adota uma abordagem teórico-conceitual e interdisciplinar, integrando contribuições da filosofia da técnica, dos estudos da comunicação e da pedagogia crítica. Realizou-se uma revisão de marcos teóricos e experiências documentadas para problematizar a interação pedagógica mediada por sistemas conversacionais. Os resultados revelam transformações significativas nos modelos de interação educativa, incluindo o surgimento de relações parasociais unilaterais, a atribuição de agência a sistemas não humanos e novas formas de mediação tecnológica. Propõe-se a figura do *educador ciborgue*, juntamente com a noção de *ética conversacional*, para orientar práticas mais responsáveis. Conclui-se que uma pedagogia política inclusiva e sensível ao contexto é essencial para promover acoplamentos virtuosos entre humanos e máquinas.

Palavras-chave: inteligência artificial; comunicação; educação; ética; tecnologia; aprendizagem

Introducción

Este artículo ofrece una revisión crítica y multidisciplinaria sobre la evolución histórica, filosófica, técnica y educativa de la inteligencia artificial (IA) conversacional, centrándose en los agentes conversacionales como una interfaz clave entre humanos y máquinas. Se parte de la premisa de que los asistentes conversacionales, lejos de ser simples herramientas automatizadas, representan operadores técnicos con capacidad de modificar las condiciones de subjetivación, las prácticas pedagógicas y las mediaciones comunicacionales en el ámbito educativo. El propósito central consiste en defender la necesidad de desarrollar una pedagogía política y una ética conversacional que permita orientar el diseño, la implementación y el uso de estos sistemas de manera situada, inclusiva y reflexiva. La importancia del tema radica en el creciente protagonismo de la IA en los sistemas de educación superior, así como en los desafíos que plantea su expansión acelerada para la subjetividad, la agencia humana, la equidad, la autonomía docente y la calidad de los procesos formativos.

El texto parte de un recorrido que va desde las representaciones míticas y literarias de seres artificiales hasta el desarrollo contemporáneo de sistemas de lenguaje generativo como ChatGPT, analizando las transformaciones en la relación humano-máquina, especialmente en el ámbito de la comunicación, la educación y la subjetividad. Se emplea una metodología de carácter teórico-conceptual, con enfoque interdisciplinario, que articula aportes de la filosofía de la técnica, los estudios de comunicación y la pedagogía crítica. La estrategia analítica se basa en una reconstrucción histórica y conceptual de la IA conversacional, considerando su evolución técnica y sus implicaciones éticas y pedagógicas. [Balmaceda \(2024\)](#) recuerda que Descartes atribuyó al diálogo una capacidad exclusivamente humana que las máquinas no podían superar. Sin embargo, con el *test* de Turing se abrió la posibilidad de trascender esta limitación mediante habilidades como el engaño, entendidas por algunas teorías de la mente como indicio de inteligencia. Este paso de una visión esencialista a una evaluación funcional del comportamiento

comunicativo marcó el inicio de una nueva etapa en la relación entre humanos y máquinas, profundizada en el siglo XXI con los sistemas de IA conversacional.

En este trabajo se propone una reconstrucción histórica de la evolución de los agentes conversacionales, desde ELIZA ([Weizenbaum, 1966](#)) hasta sistemas contemporáneos como ChatGPT (OpenAI) o Bard (Google). Se identifican tres momentos clave en la evolución del procesamiento del lenguaje natural (PLN): el enfoque simbólico de reglas programadas, el enfoque estadístico basado en probabilidades y el enfoque neuronal, fundado en redes de aprendizaje profundo. Los modelos de lenguaje de gran tamaño (LLM), introducidos en 2018, marcan un punto de inflexión en la capacidad de los sistemas para generar respuestas contextualmente relevantes con un alto grado de autonomía. En este proceso histórico se recuperan hitos como ELIZA, Parry, Jabberwacky y CleverBot, que anticiparon fenómenos actuales como la tendencia humana a antropomorfizar sistemas tecnológicos que simulan empatía o comprensión, un fenómeno que refuerza la idea de que la percepción humana en torno a la IA está fuertemente mediada por la apariencia lingüística de comprensión, aunque el sistema no tenga ‘conciencia’ ni intencionalidad.

Este artículo se organiza en cinco apartados: primero, se reconstruye el marco teórico y la genealogía de los agentes conversacionales; segundo, se problematizan los desafíos éticos y comunicacionales de la IA en educación; tercero, se introduce el concepto del educador *cyborg* como figura crítica; cuarto, se propone el marco de la ética conversacional; y, finalmente, se presentan las conclusiones generales del trabajo.

Tesis y argumentos

La tesis central de este trabajo es que la IA conversacional representa un cambio de paradigma en la comunicación humano-máquina y en la educación, el cual abre posibilidades pedagógicas que requieren de un marco ético y crítico. Su evolución, desde los primeros programas simbólicos como ELIZA hasta los modelos neuronales actuales, evidencia un salto técnico hacia sistemas más adaptativos y coherentes.

Sin embargo, estos avances conviven con límites estructurales—sesgos, problemas de privacidad y restricciones en la comprensión contextual—que exigen regulación y análisis crítico. A ello se suma el efecto ELIZA, que muestra cómo los usuarios proyectan intencionalidad o empatía en las máquinas, generando el riesgo de sobrestimar sus capacidades. En el ámbito educativo, los *chatbots* ofrecen oportunidades para personalizar y ampliar el aprendizaje, pero solo si se diseñan bajo una pedagogía política que garantice inclusión, autonomía y justicia social.

Chatbots como interfaz del lenguaje natural

[Ciesla \(2024\)](#) sostiene que comprender la terminología y la historia de los *chatbots* es un requisito para adentrarse en este campo. Su trabajo expone los conceptos fundamentales y los primeros desarrollos de la IA, señalando que proyectos iniciales como ELIZA ofrecieron solo una simulación restringida del lenguaje humano. En cambio, sistemas conversacionales más recientes, como LaMDA de Google y ChatGPT de OpenAI (2022-2023), lograron superar estas limitaciones.

Sin embargo, el *test* de Turing no evalúa la inteligencia de las máquinas, sino su capacidad para imitar la conversación humana. Con el paso del tiempo, las limitaciones del enfoque simbólico se hicieron evidentes: su rigidez, dependencia de reglas predefinidas y escasa capacidad de adaptación al contexto impedían una interacción fluida y realista. Estas carencias impulsaron el desarrollo de nuevos métodos más flexibles y autónomos. Así surgieron, a partir de la década de 1980, los enfoques estadísticos, que introdujeron técnicas de aprendizaje automático basadas en el análisis probabilístico de grandes volúmenes de datos lingüísticos. Esta transición permitió que los sistemas comenzaran a inferir patrones del lenguaje en lugar de depender exclusivamente de reglas fijas.

Posteriormente, con el avance del poder computacional y el desarrollo de redes neuronales profundas, se consolidó el enfoque neuronal, que hoy domina el campo del PLN. Este modelo, inspirado en el funcionamiento del cerebro humano, permite una mayor comprensión del contexto, adaptabilidad semántica y la generación de respuestas con mayor coherencia, y constituye la base de los LLMs que operan en los *chatbots* más avanzados de la actualidad, como ChatGPT. En comparación con enfoques más recientes, el PLN simbólico es mucho más lento y depende en gran medida de la intervención humana. Un ejemplo destacado de este enfoque fue el experimento de Georgetown en 1954, que logró traducir automáticamente más de 60 frases del ruso al inglés. Aunque hoy en día este logro parece modesto, en su momento marcó un hito.

En la década de 1980, el enfoque simbólico comenzó a ser desplazado debido a su dependencia excesiva del trabajo humano. Surgió entonces el PLN estadístico, que utiliza modelos probabilísticos y álgebra lineal para procesar el lenguaje. Este método incorpora aprendizaje automático (*machine learning*) para que los ordenadores adquieran un mayor grado de autonomía en el aprendizaje de idiomas.

Gracias al aumento de la capacidad computacional, las máquinas comenzaron a analizar datos de manera más independiente, asignando probabilidades a las palabras y enfocándose en patrones recurrentes que emplearan la inducción para procesar grandes volúmenes de datos de manera eficiente. En la actualidad, el enfoque más avanzado en el campo es el PLN neuronal, que utiliza aprendizaje profundo (*deep learning*) para procesar información en múltiples capas. Este método, inspirado en el funcionamiento del cerebro humano, permite comprender el contexto con mayor precisión que los enfoques simbólicos o estadísticos.

Desde 2023, el PLN neuronal lidera el desarrollo en IA y es la base de los *chatbots* modernos, como ChatGPT. Este enfoque permite un alto grado de autonomía en los sistemas, que requieren poca intervención humana para operar. Gran parte del PLN neuronal se ejecuta en centros de datos con unidades de procesamiento gráfico (GPUs), lo que permite analizar y procesar enormes conjuntos de datos de manera extremadamente eficiente. Un modelo lingüístico es, esencialmente, una representación probabilística de las palabras en un lenguaje. Los algoritmos diseñados para analizar enormes conjuntos de datos lingüísticos son lo que se conoce como LLMs.

Ciesla (2024) señala que los modelos de lenguaje de gran escala constituyen el fundamento de los *chatbots* actuales, y que su desempeño se evalúa según variables como el tamaño del modelo, la cantidad de datos de entrenamiento, los costos de procesamiento y los resultados obtenidos. Estos parámetros,

representables de manera cuantitativa, se articulan con la llamada *ley de escala neuronal*, que establece relaciones proporcionales entre las variables técnicas y el desempeño alcanzado. Tal formalización hace posible comparar, optimizar y escalar los modelos de manera sistemática conforme a sus capacidades.

A partir de este marco técnico, se puede distinguir entre dos grandes categorías de IA: la débil y la fuerte. La primera se orienta a la realización de tareas específicas dentro de un entorno limitado, con escasa adaptabilidad. Por el contrario, la IA fuerte busca emular el razonamiento humano, lo que implica operar con mayor autonomía, aprender continuamente y tomar decisiones en contextos nuevos. Estos sistemas, en su versión más avanzada, serían capaces no solo de analizar datos o generar texto, sino también de interpretar emociones y actuar en consecuencia, superando con creces las capacidades acotadas de la IA débil. Con el tiempo, esta última categoría, tal como la entendemos hoy, podría desaparecer al ser completamente integrada por versiones más avanzadas. Esto haría que la distinción entre ambas deje de existir. Por ende, una adopción generalizada de la IA debe tener en cuenta diversos aspectos éticos y sociales.

El efecto ELIZA y la ilusión del diálogo

ELIZA fue uno de los primeros *chatbots* en utilizar un enfoque básico de procesamiento del PLN. Desarrollado por Joseph Weizenbaum a mediados de los años 60, este programa empleaba un conjunto de reglas lingüísticas simples para simular conversaciones, especialmente en el contexto de una supuesta psicoterapia rogeriana. Su funcionamiento se basaba en la repetición y reformulación de frases del usuario, sin comprender verdaderamente su contenido.

A pesar de sus limitaciones técnicas, ELIZA marcó un hito al demostrar que era posible establecer una interacción aparentemente coherente entre un humano y una máquina. Este avance no solo abrió nuevas posibilidades en el campo de la IA; también generó interrogantes sobre la tendencia humana a atribuir intencionalidad o empatía a sistemas sin conciencia.

Ciesla (2024) sostiene que la relevancia de ELIZA superó lo estrictamente técnico, ya que inauguró una serie de desarrollos posteriores caracterizados por mayor complejidad algorítmica y capacidad adaptativa. Su impacto inicial abrió el camino para la evolución de *chatbots* más sofisticados, como Parry, Jabberwacky o CleverBot, hasta llegar a los actuales modelos neuronales como ChatGPT. Aunque concebido con fines recreativos y como un terapeuta virtual simplificado, ELIZA apenas podía contextualizar información, lo que restringía su utilidad a un nivel básico.

A diferencia de los *chatbots* actuales, ELIZA no tiene capacidad para evolucionar por sí misma. Utiliza el mismo enfoque conversacional básico en todas sus interacciones, representando una forma temprana y limitada de IA sin aprendizaje automático. Su base de datos no reconoce conceptos culturales como The Beatles o The Rolling Stones, tratándolos simplemente como sustantivos. En su tiempo, los *chatbots* como ELIZA no contaban con acceso a Internet, que apenas comenzaba a desarrollarse, lo que limitaba drásticamente su capacidad de actualización, contextualización o aprendizaje. Toda la información debía ser incorporada manualmente por programadores, lo que restringía sus respuestas a un repertorio fijo y estático. Como resultado, la interacción con estos sistemas era predecible y carente de adaptación a los intereses o referencias culturales del usuario, evidenciando las restricciones técnicas propias de su época.

Por el contrario, los sistemas modernos aprovechan ampliamente los datos en línea. Las mejoras en *chatbots* históricos como ELIZA requerían la intervención directa de programadores para actualizar manualmente sus *scripts*. Pese a sus limitaciones, ELIZA marcó un hito en el desarrollo de *chatbots* y estableció una base para los avances posteriores en IA. Desde una perspectiva comunicacional, no es necesario que una IA posea emociones genuinas para generar una respuesta empática; basta con que simule comprensión para que los usuarios tiendan a atribuirle intencionalidad o afectividad. Este fenómeno, ampliamente explorado en el cine, la literatura y el *marketing*, se conoce como el *efecto ELIZA*. Este efecto ilustra la facilidad con la que los humanos proyectan rasgos antropomorfos sobre las máquinas, especialmente cuando estas generan respuestas lingüísticamente coherentes que aparentan relevancia. Una tendencia puede inducir percepciones erróneas y llevar a una sobreestimación de la inteligencia o comprensión real del sistema.

En el ámbito de la IA y la interacción humano-computadora, Ciesla (2024) destaca que el denominado *efecto ELIZA* resalta la necesidad de una mirada crítica sobre el diseño y la evaluación de los sistemas conversacionales. Advierte que la simulación de empatía o conocimiento por parte de un *chatbot* no implica su posesión real, y que esta ilusión puede resultar tan convincente como problemática.

En este sentido, se vuelve imprescindible asumir una actitud crítica y reflexiva que permita evitar la sobrevaloración de sistemas que, si bien operan mediante lenguaje, carecen de comprensión, intención o conciencia. En 1972 se presentó Parry, un *chatbot* diseñado para simular a un paciente paranoico. A diferencia de ELIZA, este sistema gestionaba variables emocionales como ira, miedo y desconfianza. Por ejemplo, personificaba a un trabajador postal con temores relacionados con el crimen organizado, ofreciendo respuestas cargadas de emociones según las interacciones del usuario.

Más adelante, en 1988 se creó Jabberwacky, un *chatbot* orientado a lograr conversaciones más ‘humanas’ al aprender de interacciones previas con los usuarios. Este sistema utilizaba coincidencias de patrones contextuales, lo que le permitía formular respuestas más naturales y adaptativas. En 2008, Jabberwacky evolucionó a CleverBot, mejorando sus capacidades conversacionales, aunque aún dependía de patrones lingüísticos predefinidos y datos de entrenamiento recopilados durante las conversaciones con los usuarios.

Este proceso de evolución técnica se articuló con una dimensión simbólica: la *antropomorfización*, entendida como la atribución de características humanas a entidades no humanas, desempeñó un papel central en el diseño y la recepción de estos sistemas. Este recurso no solo mejora la experiencia del usuario, sino que también favorece la creación de vínculos afectivos con la máquina, transformando la interacción en algo más que un simple intercambio funcional. En este contexto, Parry, desarrollado en la década de 1970, ha sido considerado el primer *chatbot* en superar el *test* de Turing, ya que su perfil simulado de paciente paranoico permitía justificar respuestas incoherentes como parte del trastorno que representaba.

Asimismo, Jabberwacky introdujo elementos como el humor y la improvisación, anticipando el aprendizaje en tiempo real. Ciesla (2024) sostiene que, aunque sistemas como ELIZA, Parry y Jabberwacky son rudimentarios frente a los modelos neuronales actuales, sus contribuciones fueron decisivas. ELIZA abrió los primeros ensayos de interacción humano-máquina, Parry incorporó la simulación de estados

emocionales y Jabberwacky introdujo elementos de espontaneidad y adaptabilidad, estableciendo así las bases de la sofisticación presente en asistentes como ChatGPT.

La conversación como frontera de la IA

En la actualidad, los avances en los *chatbots* basados en IA están reorientando las líneas de investigación y redefiniendo los límites de la comunicación humano-máquina (HMC). Tal como lo señala [Balmaceda \(2024\)](#), esta transformación se manifiesta en el hecho de que

para sorpresa de Descartes, ChatGPT, y hoy sus numerosas inspiraciones y copias (como LLaMA de Meta, Bard de Google, Claude 2 de Anthropic y GPT-J de EleutherAi) se basan en la conversación: plataformas cuya interfaz es la interacción en tiempo real por texto mediante lenguaje natural. (p. 51)

Esta afirmación subraya el carácter conversacional como núcleo técnico y simbólico de los nuevos agentes artificiales.

En consecuencia, este tipo de desarrollos tecnológicos plantea dilemas éticos relevantes, especialmente en el marco de la IA generativa, que permite la creación de entidades capaces de sostener interacciones aparentemente empáticas. Los resultados que arrojan estas entidades son muchas veces impredecibles, y su percepción oscila entre el entusiasmo por sus posibilidades pedagógicas y el escepticismo frente a sus riesgos sociales y subjetivos. [Guzmán y Lewis \(2020\)](#) sugieren que las interacciones con *chatbots* sociales no encajan en los paradigmas actuales de la teoría de la comunicación, que se han centrado en la comunicación humano-humano.

Sin embargo, los humanos también pueden formar relaciones parasociales unilaterales con entidades no humanas o con la tecnología en general. Desde la perspectiva de estos autores, la motivación para estudiar la comunicación entre personas y máquinas surge de un cambio fundamental en el diseño y la función de la tecnología. Las explicaciones del alcance de la comunicación humano-máquina incluyen la comunicación con interlocutores digitales, agentes virtuales e IAs, y personas tecnológicamente aumentadas en entornos reales o virtuales y aumentados. Definir los límites de la HMC implica, en primer lugar, decidir qué tecnologías y aplicaciones deben ser incluidas dentro de su campo de estudio.

Tal como lo advierten [Guzmán y Lewis \(2020\)](#), restringir estos límites únicamente a formas de interacción que imiten el comportamiento humano permitiría una delimitación clara, pero implicaría descartar otros fenómenos emergentes antes de haber comprendido su potencial valor comunicativo. Según los autores, es necesario ampliar el marco analítico para incluir también aquellas formas de vinculación donde la reciprocidad, la intencionalidad o la agencia no están claramente definidas, pero donde igualmente se genera sentido.

En consecuencia, la definición de *comunicación* se vuelve una categoría estratégica: según cómo se conceptualice, se trazan las fronteras de lo que se considera interacción significativa con sistemas no humanos. A medida que las tecnologías evolucionan, comprender lo que ocurre ‘detrás de escena’—en los

procesos técnicos, sociomateriales y semióticos que configuran dichas interacciones—se torna fundamental para teorizar sobre la HMC (Guzmán y Lewis, 2020).

Ahora bien, aunque los sistemas de IA han transformado profundamente las modalidades de interacción, los *chatbots* conversacionales aún enfrentan importantes límites estructurales. Uno de los más relevantes es su limitada capacidad para comprender el contexto. Ciesla (2024) señala que modelos como GPT-4 operan con ventanas de contexto limitadas—e.g., de 32 000 tokens—, lo que restringe su capacidad para sostener información de largo alcance o interpretar referencias implícitas en marcos más amplios. A estas limitaciones se suma la dificultad para manejar ambigüedad semántica, ironía o emociones implícitas, lo que puede derivar en respuestas gramaticalmente correctas pero carentes de sentido común o culturalmente inadecuadas.

Otro aspecto crítico son los sesgos en los datos de entrenamiento. Debido a que los LLM aprenden a partir de grandes *corpus* extraídos de Internet, reproducen estereotipos, errores y patrones lingüísticos discriminatorios. A estos desafíos se suman problemas relativos a la privacidad y la seguridad de los datos. Jain y Ghanavati (2020) destacan que la arquitectura actual de recolección y procesamiento masivo de información personal supera las capacidades de control que ofrecen enfoques tradicionales como la privacidad diferencial. En este sentido, subrayan la necesidad urgente de establecer marcos normativos actualizados, capaces de proteger a los usuarios frente al uso indebido o la explotación comercial de sus datos sensibles. Frente a este panorama, una revisión crítica del diseño de *chatbots* educativos se vuelve indispensable. Estos sistemas no deberían ser concebidos como simples dispositivos automatizados, sino como entornos potenciales de mediación pedagógica.

Potencialidades y desafíos de la IA conversacional

Bradley (2025) sostiene que la incorporación de la IA en la educación no puede abordarse únicamente desde lo técnico, sino que exige una transformación del pensamiento. Esta advertencia, vinculada con la idea de entropía del pensamiento, resalta la necesidad de contrarrestar los automatismos mediante prácticas formativas orientadas al cuidado, el deseo y la inteligencia colectiva. En una línea complementaria, Crawford (2023) argumenta que tampoco es suficiente reducir la IA a una definición técnica basada en redes neuronales y el reconocimiento de patrones estadísticos, ya que su desarrollo involucra dimensiones materiales, políticas y sociales más amplias. Por ello, enfatiza la importancia de interrogar qué se optimiza, para quién y bajo qué criterios de decisión.

El uso de IA en la educación superior abarca procesos de asesoramiento sobre la base de datos de aprendizaje y asistentes de IA de acompañamiento a los estudiantes. En el ámbito educativo, los *chatbots* se utilizan para proporcionar tutorías, responder preguntas frecuentes y facilitar el aprendizaje interactivo. Ithurburu (2023) indica que, según el consenso y los documentos complementarios de la UNESCO (2021), los principales desafíos de la política pública en IA son garantizar la inclusión y la equidad, promover la formación docente en competencias digitales para un uso pedagógico de la IA y desarrollar sistemas de datos que fortalezcan la gestión educativa.

En esta línea, [Martins \(2023\)](#) resalta que los *chatbots* representan un recurso con gran potencial para mejorar los procesos de enseñanza y aprendizaje, ya que permiten brindar retroalimentación y atender a grandes grupos de estudiantes, como en los cursos masivos en línea, lo que favorece tanto la satisfacción del alumnado como su rendimiento académico. Sin embargo, advierte que todavía no existen estudios exhaustivos que evalúen de manera integral los riesgos y limitaciones de estas herramientas, aunque se reconocen dificultades en su implementación.

La eficacia de estos sistemas no solo depende de factores individuales del estudiantado, como su motivación, nivel de alfabetización digital o estrategias de aprendizaje, sino también de las características técnicas del programa, especialmente su diseño, su arquitectura y el grado de transparencia de sus algoritmos. En este marco, la personalización se plantea como un requisito clave para mejorar la efectividad pedagógica. Sin embargo, este énfasis en la adaptación también conlleva riesgos significativos: puede generar desigualdades entre quienes acceden a sistemas más sofisticados y quienes dependen de versiones limitadas, reforzar la dependencia tecnológica en detrimento de la autonomía crítica y abrir interrogantes éticos en torno a la privacidad de los datos que sustentan dicha personalización. Por ello, más que una solución automática, la incorporación de *chatbots* exige un abordaje crítico que pondere sus beneficios formativos frente a los riesgos de inequidad y control tecnológico.

No obstante, el estado actual de la tecnología no permite implementar plenamente este tipo de herramientas. Otro desafío central en la implementación de *chatbots* educativos es garantizar la privacidad de los datos personales y abordar las implicaciones éticas de su uso en contextos formativos. Cuando la dimensión técnica se entrelaza con aspectos psicológicos y simbólicos, resulta especialmente relevante observar cómo los usuarios tienden a antropomorfizar a los agentes conversacionales, atribuyéndoles cualidades humanas como personalidad, género o incluso intencionalidad comunicativa.

Si bien puede favorecer la generación de vínculos de confianza y cercanía que potencien la motivación y el compromiso en contextos educativos, esta tendencia también entraña riesgos importantes: puede inducir percepciones erróneas sobre la capacidad real de la máquina, fomentar una sobreestimación de su agencia y diluir la responsabilidad humana en la toma de decisiones. La antropomorfización, entonces, constituye un fenómeno ambivalente: por un lado, abre oportunidades pedagógicas y comunicacionales al facilitar la interacción y la aceptación de la tecnología; pero, por otro, exige un abordaje crítico para evitar que la ilusión de comprensión empañe la conciencia de sus límites técnicos, éticos y sociales. [Ciesla \(2024\)](#) advierte que esta ilusión puede obstaculizar una evaluación crítica de la tecnología, en particular en contextos educativos donde la interacción con asistentes virtuales influye en la autopercepción, la motivación y la autorregulación del aprendizaje.

Por ello, no basta con desarrollar *chatbots* funcionales: es esencial que su diseño esté fundamentado en teorías pedagógicas sólidas que incluyan no solo criterios técnicos de eficiencia, sino también enfoques políticos y éticos que reconozcan el poder formativo de las tecnologías y su impacto sobre las subjetividades. En este sentido, se vuelve fundamental que los agentes conversacionales destinados a contextos educativos no solo asistan o automaticen tareas, sino que se diseñen como herramientas críticas, orientadas a fomentar la autonomía del estudiantado, el pensamiento reflexivo y la justicia educativa. Un enfoque de este tipo requiere articular el diseño tecnológico con una pedagogía política que comprenda la educación como un proceso de individuación colectiva, y no como una simple transferencia de contenidos.

Perspectiva metodológica que sustenta la reflexión expuesta

La propuesta metodológica que articula los apartados *Del profesor automatizado al cyborg educador como pedagogía política* y *Ética conversacional como paradigma de inteligibilidad para la formación educativa* se inscribe en un enfoque filosófico-crítico inspirado en [Gilbert Simondon \(2007, 2017\)](#), complementado con aportes contemporáneos de [Sandrone \(2020\)](#), [Shah \(2023\)](#) y [Ciesla \(2024\)](#). Este marco entiende la educación como un espacio de acoplamiento técnico-subjetivo, donde las tecnologías no son simples instrumentos, sino actores activos en la constitución de entornos formativos.

Del profesor automatizado al cyborg educador

Desde la perspectiva de [Gilbert Simondon \(2017\)](#), la máquina no es un mero instrumento, sino un ser técnico relacional que participa en la transformación de información y en la configuración de vínculos con los usuarios. En este sentido, la relación humano-máquina debe entenderse como un proceso de individuación, donde ambos se co-constituyen en dinámicas de acoplamiento técnico y subjetivo. Desde esta perspectiva, el vínculo entre humanos y máquinas no debe pensarse en términos de exterioridad entre humano y artefacto, sino como un proceso de individuación en el que ambos co-constituyen nuevas formas de relación y sentido. [Sandrone \(2020\)](#) desarrolla la figura del *cyborg* educador para comprender la enseñanza mediada por tecnologías como una práctica situada que incorpora los dispositivos en dinámicas pedagógicas emancipadoras. Este planteo trasciende la mera instrumentalización técnica y se orienta hacia una mayéutica tecnológica que posibilita la creación de sintagmas sociotécnicos, configurando la experiencia educativa como un espacio de coindividuación entre sujetos y máquinas. En este marco, introduce la distinción entre autómatas y *cyborg*: mientras Babbage, en el siglo XIX, concebía el trabajo humano como cooperación con la máquina, Ure lo reducía a la simple supervisión de un sistema técnico autónomo. Esta tensión permite repensar críticamente las relaciones contemporáneas entre tecnología, educación y trabajo.

De acuerdo con este enfoque, la tarea docente se concibe como un ejercicio flexible y contextual, estructurado por ensamblajes técnico-humanos en los que los dispositivos actúan como agentes pedagógicos activos y no como elementos neutrales. En esta línea, [Sandrone \(2020\)](#) subraya que cada docente o comunidad de docentes genera su propio “*cyborg* local” (p. 29) al organizar las interacciones entre individuos humanos y técnicos para producir aprendizajes que difícilmente podrían lograrse de manera separada. Esta propuesta dialoga con la visión de [Simondon \(2007\)](#), pues el educador no solo media el conocimiento, sino que coordina entornos de acoplamiento virtuoso entre humanos y tecnologías. Desde esta perspectiva, el principal desafío de las instituciones educativas consiste en promover una mentalidad técnica que permita acoplar las tecnologías a intereses y necesidades locales, y desacoplarlas de los esquemas dominados por las grandes corporaciones. Solo así será posible diseñar entornos educativos híbridos que favorezcan la participación, la crítica y la creatividad de los estudiantes mediante formas de interacción con agentes conversacionales, asistentes inteligentes y plataformas digitales que no se limiten a una función instrumental, sino que expandan las posibilidades de la formación. En este sentido, el desarrollo de una ética conversacional con las máquinas supone construir marcos normativos y simbióticos que reconozcan tanto los derechos y responsabilidades de los usuarios como el rol activo de los sistemas automatizados en los procesos de aprendizaje. La noción de *cyborg educador* puede enriquecerse con los

desarrollos recientes sobre gemelos digitales, entendidos como réplicas virtuales conectadas a entornos físicos mediante sensores e IA. [Sermakani et al. \(2024\)](#) definen el *gemelo digital* como una representación lógica precisa y en constante actualización de un producto o proceso real, utilizada para probar distintos escenarios, anticipar problemas e identificar oportunidades de optimización. Trasladada al ámbito educativo, esta noción permitiría concebir un gemelo digital docente capaz de modelar y simular escenarios pedagógicos complejos a fin de prever desafíos y optimizar la toma de decisiones.

La ética conversacional como marco para la inteligibilidad de los *chatbots*

[Shah \(2023\)](#) sostiene que la IA generativa, cuando se utiliza de manera creativa y pedagógica, tiene el potencial de transformar el rol docente y promover una educación centrada en el estudiante. En lugar de rechazar estas herramientas, plantea que deben integrarse críticamente en los procesos de enseñanza, ya que, si se aplican de forma reflexiva, pueden ayudar a docentes y administradores a preparar mejor a sus estudiantes para afrontar los desafíos futuros.

Esta perspectiva refuerza la idea planteada en el artículo sobre la necesidad de una pedagogía política que no solo tolere la IA, sino que la reinscriba en marcos educativos reflexivos y colaborativos. Por ejemplo, interactuar con un *chatbot* como Jabberwacky permite abordar diversas dimensiones del vínculo entre humanos y tecnologías. En primer lugar, resalta la importancia del lenguaje y la comunicación, al ofrecer un espacio para explorar los matices de las conversaciones, el papel del contexto y los desafíos inherentes a la transmisión efectiva de significados. Además, los sistemas de IA conversacional pueden reflejar e incluso amplificar los sesgos, estereotipos y patrones de comportamiento humanos presentes en los datos utilizados durante su entrenamiento.

[Ciesla \(2024\)](#) subraya que los LLMs, al procesar grandes volúmenes de lenguaje humano, manifiestan conductas emergentes que permiten estudiar cómo responden a patrones lingüísticos con connotaciones ideológicas, emocionales o afectivas. Este fenómeno no solo revela la influencia cultural en sus respuestas, sino que también ofrece a investigadores y desarrolladores una oportunidad para analizar la manera en que diferentes entradas modifican el comportamiento del modelo. Del mismo modo, las respuestas de estos sistemas pueden impactar emocional y psicológicamente en los usuarios, sobre todo cuando se interpretan como empáticas, personalizadas o significativas en el plano afectivo. Este tipo de interacción plantea preguntas importantes sobre la autenticidad de los vínculos, la percepción de agencia por parte del usuario y el impacto subjetivo de las interacciones simuladas. Por lo tanto, el estudio de estos sistemas se presenta como un campo fértil para investigar fenómenos emocionales, mecanismos de proyección antropomórfica y condiciones de posibilidad de comunicación simbólica entre humanos y máquinas. Esto exige un abordaje interdisciplinario que combine saberes de la psicología, la lingüística, la teoría de la comunicación y la filosofía de la técnica.

[Ciesla \(2024\)](#) advierte que el uso de tecnologías como Jabberwacky plantea desafíos éticos vinculados con la privacidad, la seguridad de los datos, el consentimiento informado y la responsabilidad en la aplicación de la IA. Subraya, además, la necesidad de que el desarrollo tecnológico se oriente por principios éticos y al bienestar humano. En esta línea, los agentes conversacionales muestran que la conversación trasciende el simple intercambio de información al constituirse también en un medio para construir

relaciones, expresar emociones y negociar significados. El análisis de las limitaciones y fortalezas de estos sistemas permite comprender mejor la propia capacidad comunicativa humana y el modo en que la tecnología influye en las relaciones interpersonales. En una consulta realizada a ChatGPT (OpenAI, 2024) identificó, como principios fundamentales de una ética conversacional con las máquinas, una serie de lineamientos orientados a regular dicha interacción:

Tabla 1.

Principios fundamentales de la ética conversacional

Principio	Descripción
<i>Transparencia</i>	Las máquinas deben identificar claramente su naturaleza no humana, evitando confundir a los usuarios sobre su identidad o capacidades. Es importante informar sobre el propósito del chatbot o sistema de IA y sobre la manera en que se procesarán y utilizarán los datos generados durante la interacción.
<i>Respeto y dignidad</i>	Los sistemas de IA deben tratar a los usuarios con respeto, evitando respuestas ofensivas, discriminatorias o sesgadas. Los diseñadores y desarrolladores deben garantizar que las interacciones no promuevan la deshumanización ni conductas perjudiciales.
<i>Privacidad y seguridad</i>	Las conversaciones deben ser protegidas, garantizando la confidencialidad y seguridad de los datos. Es fundamental obtener consentimiento informado para recopilar, almacenar o utilizar información personal.
<i>Inclusión y equidad</i>	Los sistemas conversacionales deben diseñarse para ser inclusivos y accesibles, considerando la diversidad cultural, lingüística y de capacidades de los usuarios. Deben minimizarse los sesgos en los datos de entrenamiento para evitar perpetuar desigualdades o prejuicios.
<i>Intencionalidad y claridad en los límites</i>	Los sistemas deben ser diseñados para cumplir funciones claras, evitando generar expectativas irreales en los usuarios. Debe establecerse un marco ético que delimite el alcance de las interacciones, evitando intrusiones innecesarias o manipulaciones.
<i>Beneficencia y no maleficencia</i>	Las interacciones deben priorizar el bienestar del usuario, asegurándose de que el sistema no cause daño emocional, psicológico o social. Es necesario prevenir el uso malintencionado de los sistemas conversacionales.
<i>Capacidad de respuesta y adaptabilidad</i>	Las máquinas deben ser programadas para adaptarse a las necesidades y contextos de los usuarios, respondiendo de manera apropiada y empática. Los sistemas deben incluir mecanismos para corregir errores, mitigar malentendidos y recibir retroalimentación.
<i>Fomento de la autonomía</i>	Las interacciones deben empoderar a los usuarios, ofreciéndoles información y opciones claras para tomar decisiones informadas. Los sistemas no deben manipular ni coaccionar a los usuarios hacia determinadas acciones o comportamientos.
<i>Fomento del bien común</i>	Se deben usar los sistemas para promover bienestar social en áreas como la educación, la salud y la información.
<i>Mejora continua</i>	Se deben adaptar los sistemas éticos a nuevos desafíos tecnológicos y sociales.
<i>Colaboración simbiótica</i>	Se deben diseñar máquinas que complementen y potencien las habilidades humanas, promoviendo una relación colaborativa y transformadora.

Estos principios buscan garantizar que la relación entre los humanos y las máquinas sea respetuosa, justa, equitativa y beneficiosa, preservando siempre la dignidad y autonomía humanas mientras se capitalizan los avances tecnológicos. Un enfoque de esta naturaleza requiere no solo un diseño ético y responsable de la tecnología, sino también la configuración de un marco relacional en el que la interacción con las máquinas favorezca el desarrollo humano y social.

En este sentido, Serres (2015) propone comprender dicha relación a través de la noción de simbiosis, entendida como un vínculo dinámico en el que ambas partes se transforman y enriquecen mutuamente, más allá de la mera coexistencia. En su obra *El parásito*, distingue tres modalidades: el mutualismo, en el

que ambas especies se benefician; el comensalismo, donde una obtiene ventajas sin afectar a la otra; y el parasitismo, cuando una se beneficia en detrimento de la otra. Aplicadas a la convivencia entre humanos y máquinas, estas formas permiten reconocer que no solo las tecnologías se adaptan a las necesidades humanas, sino que también las personas modifican sus prácticas, saberes y modos de vida en función de las relaciones que establecen con ellas en la cotidianidad.

Conclusión

La revisión crítica sobre la evolución histórica de las IA conversacionales nos permitió evaluar sus riesgos, sus oportunidades y sus límites en el campo educativo. Describir sus transformaciones nos permitió vislumbrar la necesidad de redefinir el vínculo humano-máquina en los procesos de formación. Aunque los *chatbots* y las IA conversacionales han recorrido un largo camino y ofrecen muchas ventajas, todavía enfrentan desafíos significativos. Comprender estos desafíos es crucial para el futuro de la interacción humano-IA.

A medida que la tecnología avanza, es importante que los desarrolladores de *chatbots* no solo sean técnicamente competentes, sino también éticamente responsables y centrados en una ética conversacional. Solo entonces podremos aprovechar plenamente su potencial y mejorar la calidad de nuestras interacciones diarias con estas tecnologías. La implementación de una ética conversacional implica redefinir las formas de comunicación para reubicar a las máquinas en un rol diferente, antes reservado solo para los humanos. Podemos observar que algunos diseñadores operan desde un modelo en el que la tecnología, incluida la IA, continúa siendo concebida como mediadora de la interacción humana, aunque con un grado creciente de agencia.

Desde esta perspectiva, la IA puede intercambiar mensajes con personas, pero sus acciones están situadas dentro del contexto más amplio de mediación de interacciones humanas. Otras aproximaciones proponen una definición general de la comunicación que no depende de quién o qué esté comunicando. En este caso, la comunicación se define como un proceso de "creación de significado" logrado a través del intercambio de mensajes con implicaciones para el individuo y la sociedad, limitando el proceso a partir de un cierto antropocentrismo.

Desde nuestra perspectiva, la comunicación puede ocurrir entre humanos y entre humanos y máquinas o entre máquinas, reconociendo que las máquinas pueden ocupar roles de comunicadores y mediadores. De este modo, se evita calificar la comunicación humano-máquina como inferior. Comunicarse desde esta perspectiva implica establecer un vínculo no instrumental con las tecnologías digitales, que no se limitan a ser simples herramientas, sino que configuran un entorno compuesto por agentes técnicos que poseen distintos niveles de normatividad y capacidad de acción. Por ello, proponemos un modo de relación donde las máquinas no sean vistas como meras herramientas, sino como socios cognitivos capaces de ampliar las capacidades humanas, sin comprometer la ética ni la autonomía.

Desde esta perspectiva, consideramos importante reflexionar sobre cómo diseñar tecnologías que no solo complementen al ser humano, sino que también fomenten un diálogo creativo y transformador, respetando las singularidades de ambas partes. Finalmente nos parece interesante considerar las definiciones

elaboradas por ChatGPT en torno a la ética de la conversación para pensar la convivencia simbiótica entre humanos y máquinas.

Reconocimientos

El proyecto *Chatbot de gestión educativa para el acompañamiento pedagógico de estudiantes de nivel superior FH-UNCA*, financiado por el Ministerio de Ciencia y Tecnología de la Nación en el marco del Programa Federal de Innovación, desarrolla un asistente digital que optimiza la comunicación institucional, agiliza trámites y fortalece la continuidad pedagógica, favoreciendo retención y egreso estudiantil.

Referencias

- Andreoli, S., Batista, A., Fucksman, B., Gladkoff, L., Martínez, K., y Perillo, L. (2022). *Inteligencia artificial y educación: Un marco para el análisis y la creación de experiencias en el nivel superior*. Centro de Innovación en Tecnología y Pedagogía (Citep). <https://citep.rec.uba.ar/publicaciones>
- Balmaceda, T. (2024). IA generativa y disrupciones. En J. Cruz Balian, J. I. Cuiule, L. Sendra y P. González (Eds.), *OK, Pandora. Seis ensayos sobre inteligencia artificial* (pp. 45-68). El Gato y la Caja.
- Bradley, J. P. N. (2025). Bernard Stiegler and the philosophy of education III: AI and the entropy of thought. *Educational Philosophy and Theory*, 57(5), 389-394. <https://doi.org/10.1080/00131857.2025.2475443>
- Ciesla, R. (2024). *The book of chatbots: From Eliza to ChatGPT*. Springer. <https://doi.org/10.1007/978-3-031-51004-5>
- Crawford, K. (2023). *Atlas de la IA: poder, política y costes planetarios de la inteligencia artificial*. NED Ediciones.
- Guzmán, A. L., y Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media y Society*, 22(1), 70-86. <https://doi.org/10.1177/1461444819858691>
- Ithurburu, V. (2023). Inteligencia artificial en el decir de política educativa global: Derecho a la educación, escuela y parresia. En S. Grimberg & A. Grinberg (Eds.), *Educación de plataforma. Sociedad postmedia y pedagogía porvenir* (pp. 91-105). Miño Dávila.
- Martins, L. (2023). Una introducción a los chatbots y sus aplicaciones en educación. En M. Fernández-Ferrer (Ed.), *Chatbots en educación. Tendencias actuales y desafíos futuros* (pp. 15-38). LMI.
- OpenAI. (2024). *ChatGPT* (Versión GPT-4) [Modelo de lenguaje grande]. <https://chat.openai.com>
- Sandrone, D. (2020). Cyborg educador. *Propuesta Educativa*, 29(54), 18-30. <https://www.propuestaeducativa.flacso.org.ar/index.php/propuestaeducativa/article/view/54>

Sermakani, A. M., Radhika, R., y Sujatha, V. (2024). Digital Twin: Past, Present, and Future. En A. Daniel, S. Sriramulu, N. Partheeban y S. Jayagopalan (Eds.), *Digital Twin Technology and Applications* (1a ed.). Auerbach Publications. <https://doi.org/10.1201/9781003469612>

Serres, M. (2015). *El parásito*. Fondo de Cultura Económica.

Shah, P. (2023). *AI and the future of education: Teaching in the age of artificial intelligence*. Jossey-Bass. [https://newteacherlibraryandtools.square.site/uploads/b/e4443390-f724-11ec-9291-b5b724deb1dc/AI%20and%20the%20Future%20of%20Education%20Teaching%20in%20the%20Age%20of%20Artificial%20Intelligence%20\(Priten%20Shah\)%20\(Z-Library\).pdf](https://newteacherlibraryandtools.square.site/uploads/b/e4443390-f724-11ec-9291-b5b724deb1dc/AI%20and%20the%20Future%20of%20Education%20Teaching%20in%20the%20Age%20of%20Artificial%20Intelligence%20(Priten%20Shah)%20(Z-Library).pdf)

Simondon, G. (2007). *El modo de existencia de los objetos técnicos*. Prometeo.

Simondon, G. (2017). *Sobre la técnica*. Cactus.

