Humanidades, IA generativa y producción académica: una perspectiva sociotécnica

Revista Latinoamericana de Economía y Sociedad Digital

Issue 6

DOI: https://doi.org/10.53857/RLESD.06.2026.01

Publicado: 14 octubre, 2026

Recibido: 30 noviembre, 2025

Cita sugerida: Lagos, J. (2026). Humanidades, IA generativa y producción académica: una perspectiva sociotécnica. Revista Latinoamericana de Economía y Sociedad, 6, 1-13.

Licencia: Creative Commons Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0)

Tipo: Ensayo

Palabras clave: Academia, Artificial intelligence

Resumen

Este ensayo examina, desde un enfoque sociotécnico, dos efectos articulados de la irrupción de la Inteligencia Artificial (IA) generativa en el ámbito académico: por un lado, su capacidad para exponer las vulnerabilidades del modelo academicista contemporáneo —tensionado por la productividad cuantificable, la hiperespecialización y el principio de *publish or perish*—; por otro, su capacidad para reafirmar, por contraste, el valor formativo de las humanidades en la educación ética del pensamiento. El texto describe el funcionamiento técnico de los modelos extensos de lenguaje, analiza las dinámicas institucionales que median su adopción y discute por qué la detección automática no constituye una respuesta suficiente al problema de la integridad académica. Cierra con cuatro recomendaciones orientadas a revitalizar las prácticas formativas humanísticas y a orientar un uso ético y genuinamente humano de la IA generativa.

Abstract

This essay examines, from a sociotechnical perspective, two articulated effects of the irruption of generative artificial intelligence in academic settings: on the one hand, its capacity to expose the vulnerabilities of the contemporary academicist model —strained by

quantifiable productivity, hyperspecialization, and the *publish or perish* principle—; on the other, its capacity to reaffirm, by contrast, the formative value of the humanities in the ethical education of thought. The text describes the technical functioning of large language models, analyzes the institutional dynamics that mediate their adoption, and discusses why automatic detection does not constitute a sufficient response to the problem of academic integrity. It closes with four recommendations aimed at revitalizing humanistic formative practices and guiding an ethical and genuinely human use of generative AI.

Resumo

Este ensaio examina, a partir de uma abordagem sociotécnica, dois efeitos articulados da irrupção da inteligência artificial generativa no âmbito acadêmico: por um lado, sua capacidade de expor as vulnerabilidades do modelo academicista contemporâneo —tensionado pela produtividade quantificável, pela hiperespecialização e pelo princípio do *publish or perish*—; por outro, sua capacidade de reafirmar, por contraste, o valor formativo das humanidades na educação ética do pensamento. O texto descreve o funcionamento técnico dos modelos de linguagem de grande escala, analisa as dinâmicas institucionais que mediam sua adoção e discute por que a detecção automática não constitui uma resposta suficiente ao problema da integridade acadêmica. Encerra com quatro recomendações voltadas a revitalizar as práticas formativas humanísticas e a orientar um uso ético e genuinamente humano da IA generativa.

Introducción[1]

Según muestra la creciente evidencia (Arar et al., 2025), lla irrupción de la Inteligencia Artificial generativa en el ámbito académico está transformando aceleradamente los modos que producimos, evaluamos y comprendemos el conocimiento. Sus efectos no parecen homogéneos ni restringidos a una disciplina: en las ciencias empíricas, la automatización de resúmenes y el uso de herramientas para el análisis de datos han comenzado a producir distorsiones significativas, reflejadas en el aumento de retractaciones motivadas por resultados inventados, correlaciones inexistentes o errores metodológicos que pasan inadvertidos para los revisores (Kocyigit et al., 2025, p. 8). Estas fallas no constituyen simples incidentes aislados, sino síntomas estructurales que afectan a disciplinas que dependen de la confiabilidad acumulada de bases de datos, metaanálisis y repositorios colectivos. Allí donde un error generado por IA se incorpora sin ser detectado, se afectan no solo líneas completas de investigación, sino también decisiones de política pública y evaluaciones científicas de alto impacto.

En las Humanidades, esta misma vulnerabilidad toma una forma distinta. Estas disciplinas no se organizan en torno a la validación empírica de datos, sino en torno a la interpretación, la elaboración teórica y la construcción de argumentos. Aquí, los modelos extensos de

lenguaje (LLM) que hoy exhiben capacidades sin precedentes para generar texto de apariencia humana y para procesar, reorganizar o sintetizar contenidos complejos con notable fluidez (Arriagada-Bruneau et al., 2026, p. 161) plantean interrogantes de otra índole. Su aptitud para producir pasajes que simulan comprensión, reconstruir argumentos filosóficos o formular análisis críticos en cuestión de segundos introduce una incertidumbre particular: no solo sobre la autenticidad de los textos generados, sino también sobre la forma en que evaluamos, enseñamos y producimos conocimiento en disciplinas cuyo núcleo es la lectura atenta, la reflexión situada y la articulación cuidadosa de las ideas.

Estudiar este fenómeno resulta relevante porque la adopción acelerada de LLM en contextos académicos puede modificar tanto las condiciones materiales de la investigación como las expectativas sociales respecto del trabajo intelectual. Desde un punto de vista técnico, estas herramientas amplifican la capacidad de generar textos, revisar literatura o sintetizar argumentos con una velocidad inédita, lo que altera los ritmos y métodos de producción académica. Desde un punto de vista social, su uso plantea tensiones sobre la autoría, la originalidad, la formación de estudiantes y la confianza en los mecanismos de evaluación y publicación, todos elementos fundamentales para la credibilidad de nuestras instituciones de conocimiento. Esta preocupación adquiere una dimensión adicional si se considera la crisis sostenida de las humanidades, frecuentemente cuestionadas por su utilidad práctica, su capacidad de atraer estudiantes y su lugar en universidades orientadas a la productividad inmediata (Nussbaum, 2010, pp. 19-23) En ese contexto, la irrupción de la inteligencia artificial generativa aparece para muchos como un posible clavo final en dicha crisis, pues amenaza con replicar —o simular— tareas que antes constituían evidencia del trabajo humanístico: leer, interpretar, argumentar y comprender textos complejos.

El presente ensayo se propone analizar la relación entre la IA generativa y el trabajo académico en Humanidades desde una perspectiva sociotécnica. La pregunta que orienta el análisis es la siguiente: ¿de qué manera la irrupción de la IA generativa pone en evidencia las vulnerabilidades del modelo academicista y, al mismo tiempo, reafirma el valor de las Humanidades en la formación ética del pensamiento? El ensayo se inscribe específicamente en la intersección entre las humanidades como disciplina académica y el sistema institucional de producción y circulación del conocimiento.

Para abordar esta pregunta no basta con describir cómo funcionan los modelos ni con discutir sus efectos sociales por separado. La IA generativa transforma simultáneamente prácticas de investigación, expectativas pedagógicas, normativas institucionales y modos de producción del conocimiento. Por eso, el análisis se ordena a partir de un enfoque sociotécnico que permite comprender estos sistemas no como herramientas aisladas, sino como parte de un entramado en el que técnicas, instituciones, usuarios y valores se moldean mutuamente. La sección siguiente explicará por qué esta perspectiva es imprescindible para entender el fenómeno en su complejidad.

I. ¿Por qué adoptar un enfoque sociotécnico?

Analizar la Inteligencia Artificial Generativa exige abandonar la idea de que estas herramientas operan de manera aislada. Un enfoque sociotécnico parte de la premisa de que la IA es inseparable del entramado humano e institucional donde se diseña, implementa y utiliza (Kudina & van de Poel, 2024, p. 1). Esta premisa es heredera de una tradición consolidada en los estudios sociales de la ciencia y la tecnología (STS), que ha mostrado de manera sostenida cómo los artefactos técnicos se construyen socialmente y, a su vez, configuran formas de vida social (Arriagada-Bruneau et al., 2026, p. 161).En este marco, la IA no es, por tanto, un objeto neutral, sino un componente más de un sistema complejo en el que decisiones humanas, infraestructuras digitales y marcos normativos se condicionan mutuamente (Arriagada-Bruneau et al., 2026, p. 26).

Este enfoque entiende que un sistema de IA está compuesto tanto por elementos tradicionales (artefactos técnicos, agentes humanos e instituciones) como por componentes propios de la IA contemporánea: agentes artificiales con capacidad de adaptación y normas técnicas que regulan su comportamiento (Kudina & van de Poel, 2024, p. 3; Arriagada-Bruneau et al., 2026, p. 26). A falta de intencionalidad humana, estos agentes operan según reglas inscritas en su diseño, lo que significa que cada decisión técnica incorpora criterios sociales, valores implícitos y supuestos culturales (Arriagada-Bruneau, 2025, p. 40). De ahí que la IA deba ser estudiada no solo por lo que hace, sino por el modo en que sus acciones se insertan en relaciones sociales, jerarquías institucionales y dinámicas laborales preexistentes.

Los autores convocados en este marco no son intercambiables, sino complementarios. Kudina y van de Poel (2024) aportan la categoría de sistema sociotécnico aplicada específicamente a la IA, incorporando los agentes artificiales como nuevo elemento constitutivo. Arriagada-Bruneau (2025), junto también con López y Mendoza (2026), ofrecen el aterrizaje latinoamericano y crítico de esa categoría, mostrando cómo los modelos son situados y cómo los valores implícitos pueden reproducir desigualdades coloniales y culturales. Miller (2019), por su parte, desplaza la atención desde el sistema hacia la práctica comunicativa: la explicabilidad, sostiene, no es una propiedad técnica del modelo, sino una relación epistémica entre quienes lo desarrollan, quienes lo usan y quienes son afectados por sus resultados. Estas tres voces, en conjunto, permiten leer el fenómeno de la IA generativa en humanidades en dos dimensiones simultáneas: como entramado institucional susceptible de exponer sus propias vulnerabilidades, y como práctica comunicativa que reclama formas humanas de juicio y responsabilidad.

Justificar esta aproximación es indispensable para comprender adecuadamente los riesgos y oportunidades de la IA. Problemas como la opacidad, la explicabilidad, la autoría o la integridad académica no pueden resolverse únicamente desde el código, porque dependen de contextos sociopolíticos y de los grupos afectados por sus resultados (Arriagada-Bruneau et al., 2026, pp. 31-32). Los modelos son “situados”: reproducen, y a veces amplifican, los marcos culturales y las desigualdades presentes en los datos y en las instituciones que los utilizan (Arriagada-Bruneau, 2025, p. 40).

Por esta razón, el ensayo continúa con tres niveles complementarios:

1. Una descripción técnica, necesaria para entender qué pueden y qué no pueden hacer los modelos;

2.Un análisis social, que examina cómo estas herramientas exponen las tensiones del entramado academicista;

3. Finalmente, conclusiones y recomendaciones, orientadas a iluminar cómo las Humanidades pueden guiar un uso verdaderamente humano y ético de la Inteligencia Artificial. Las dos dimensiones de la pregunta,vulnerabilidad institucional y reafirmación formativa, se leen, desde aquí en adelante, como dos efectos del mismo entramado sociotécnico, no como cuestiones independientes.

II. Descripción técnica

Para comprender el impacto que las inteligencias artificiales generativas pueden tener en la investigación académica, especialmente en las Humanidades, es necesario entender en términos generales qué tipo de sistemas son los modelos extensos de lenguaje (LLM) y cómo funcionan. Aunque su funcionamiento interno es altamente complejo, algunas características fundamentales permiten tener una visión clara de sus capacidades y limitaciones.

En primer lugar, los LLM son modelos de aprendizaje automático entrenados con enormes cantidades de texto. A partir de estos datos, aprenden a reconocer patrones lingüísticos que les permiten resumir, traducir, clasificar, predecir y generar lenguaje humano con notable fluidez (Okerlund et al., 2022, p. 8). Su rasgo más distintivo no es solo su tamaño, sino la sorprendente capacidad de generalización que han adquirido: pueden enfrentar tareas nuevas con apenas unos ejemplos, sin necesidad de un entrenamiento específico para cada una. Esta habilidad, conocida como *few-shot learning*, marca una diferencia importante respecto de tecnologías anteriores (Arriagada-Bruneau et al., 2026, p. 161) y ha llevado a algunos expertos a considerarlos como un avance significativo hacia formas más amplias de inteligencia artificial.

Su rendimiento se explica en gran medida por la arquitectura *transformer*, introducida originalmente por Vaswani et al. (2017), que les permite procesar el lenguaje de manera no secuencial, identificando la relación entre palabras que no necesariamente están una al lado de la otra. En lugar de comprender una frase paso a paso, estos modelos trabajan con representaciones matemáticas de las palabras llamadas vectores, que sitúan cada término en un espacio de muchas dimensiones para captar matices de significado (Okerlund et al., 2022, p. 35). La complejidad del sistema se mide en parámetros, que pueden entenderse como elementos análogos a neuronas: unidades que almacenan patrones, activan relaciones y contribuyen a la producción final del texto (Okerlund et al., 2022, p. 36).

El proceso mediante el cual generan respuestas también tiene componentes aleatorios.

Cuando el modelo escribe, no simplemente “elige la palabra correcta”, sino que utiliza estrategias de decodificación diseñadas para evitar respuestas demasiado obvias o repetitivas (Arriagada-Bruneau et al., 2026, p. 164). Esto explica por qué dos respuestas del mismo modelo pueden ser distintas incluso cuando se formula la misma pregunta. Este procedimiento, sin embargo, también introduce limitaciones: algunos modelos procesan el texto de manera unidireccional y ciertas capas internas enfrentan bloqueos matemáticos,como el llamado *softmax bottleneck*, que restringen la expresividad de las representaciones más abstractas del modelo (Arriagada-Bruneau et al., 2026, p. 164).

La calidad del sistema depende de los datos con los que se entrena. Estos provienen de corpus masivos que incluyen libros digitalizados, repositorios públicos y grandes volúmenes de texto extraídos de internet (Okerlund et al., 2022, pp. 32-33). En muchos casos, estos conjuntos de datos son tan extensos y heterogéneos que ni siquiera los propios desarrolladores pueden identificar con precisión qué contiene el corpus completo. Además, gran parte del entrenamiento se apoya en recursos como *Common Crawl*, donde predominan voces y estilos asociados a ciertos grupos geográficos y demográficos, lo que introduce sesgos culturales y lingüísticos sistemáticos (Bender et al., 2021, pp. 613-615). A ello se suma lo que algunos investigadores llaman “fronteras temporales”: como los modelos trabajan con datos históricos, pueden quedar desactualizados y fabricar información reciente cuando no la encuentran en su entrenamiento (Arriagada-Bruneau et al., 2026, p. 165).

Un rasgo fundamental de estos sistemas es su marcada opacidad. Hoy en día, no existe un procedimiento consensuado que permita determinar con claridad la calidad o confiabilidad de las respuestas generadas por un LLM, y los desarrolladores no están obligados a revelar cómo los prueban o validan (Okerlund et al., 2022, p. 9). Esta falta de transparencia dificulta enormemente cualquier evaluación externa, y resulta aún más problemática cuando se considera la amplia variedad de alucinaciones que pueden producir (Arriagada-Bruneau et al., 2026, pp. 162-164). Algunas consisten en presentar como hechos contenidos que no pueden ser corroborados o que directamente son inventados; otras aparecen cuando el modelo no logra seguir la lógica que el usuario le plantea; y existe también la tendencia de ciertos sistemas a ajustar sus respuestas para coincidir con expectativas humanas detectadas durante el entrenamiento, incluso cuando ello contradice la realidad.

En resumen, aunque los LLM han ampliado de manera notable lo que la Inteligencia Artificial puede hacer con el lenguaje, lo han hecho sobre una base técnica que combina capacidades excepcionales con limitaciones estructurales y riesgos importantes. Comprender estas características es indispensable para examinar de qué modo se acoplan al modelo academicista contemporáneo y qué implicancias tienen, en particular, para disciplinas que dependen profundamente de la interpretación, el análisis crítico y la autenticidad intelectual.

III. Análisis social

El uso de modelos generativos de inteligencia artificial en el mundo académico se desarrolla dentro de un entramado institucional que condiciona fuertemente su impacto. Universidades, revistas científicas, agencias de financiamiento y organismos de evaluación han configurado, a lo largo de décadas, un ecosistema en el que la productividad cuantificable se ha convertido en el principal indicador de mérito. En este contexto, las grandes empresas tecnológicas que diseñan los modelos terminan influyendo indirectamente en las prácticas académicas, pues sus herramientas se integran a procesos de escritura, revisión y difusión del conocimiento sin una reflexión suficiente sobre sus implicancias.

Las instituciones académicas enfrentan el desafío de incorporar estas tecnologías mientras operan bajo presiones intensas: financiamiento competitivo, acreditaciones periódicas, rankings internacionales y exigencias administrativas que privilegian la rapidez y la eficiencia (García-Huidobro, 2018, pp. 308-311). El resultado es una adopción irregular, en ocasiones improvisada, que no siempre considera el modo en que estas herramientas transforman la naturaleza del trabajo intelectual.

En este escenario, estudiantes e investigadores se encuentran en una posición especialmente vulnerable. Para los estudiantes, la facilidad con que la IA puede producir explicaciones, resúmenes o comentarios críticos plantea interrogantes sobre cómo se evalúa el aprendizaje y cómo se desarrolla la capacidad de análisis personal. Para los investigadores la irrupción de estos modelos tensiona ideas fundamentales sobre la autoría, la originalidad y el proceso de construcción intelectual. La escritura, que tradicionalmente ha sido el medio a través del cual se forma el pensamiento, puede verse reducida a un producto más, moldeado por herramientas que simulan comprensión sin realmente poseerla.

Sin embargo, ninguna de estas tensiones puede entenderse sin considerar el peso del *publish or perish*, el principio que estructura la carrera académica moderna (Miranda Montecinos, 2013, p. 711). La estabilidad laboral, los ascensos y la legitimidad profesional dependen, en gran medida, de la cantidad de artículos publicados en revistas indexadas. Este sistema, concebido para incentivar la investigación, ha generado consecuencias no deseadas: la fragmentación de resultados, la proliferación de publicaciones de escaso impacto real y una hiperespecialización que encierra a los investigadores en nichos cada vez más estrechos. En lugar de nutrir un espacio de conversación amplia, las humanidades terminan multiplicando textos que casi nadie lee, desconectándose de la formación generalista que tradicionalmente justificaba su existencia en la universidad.

La IA generativa entra en este ecosistema como un elemento que puede acelerar esas tendencias. Al permitir producir borradores con rapidez, sintetizar literatura o generar estructuras argumentativas (Talukdar et al., 2025, p. 50), la tecnología facilita responder a las exigencias de productividad sin necesariamente fortalecer el razonamiento crítico o la reflexión propia. La herramienta, en sí misma, no degrada la investigación, pero interactúa

con un sistema ya marcado por incentivos que valoran más la frecuencia de publicación que la profundidad intelectual. En consecuencia, en lugar de aliviar la presión sobre académicos y estudiantes, puede intensificarla al elevar los estándares de velocidad y volumen, sin mejorar los fundamentos de la formación humanística.

En este sentido, la IA generativa actúa como un espejo incómodo del modelo academicista. Obliga a preguntar si las Humanidades pueden sostener su promesa formativa (desarrollar pensamiento crítico, sensibilidad histórica, interpretación cuidadosa y juicio reflexivo) en un entorno donde los incentivos institucionales empujan hacia la velocidad, la especialización extrema y la desconexión social (Klein et al., 2025, pp. 12-13). Lo que ese espejo revela es que el modelo academicista contemporáneo es vulnerable porque ha tendido a confundir la cantidad con la profundidad, la métrica con el sentido y la repetición formal con el pensamiento. La IA generativa no produce esa vulnerabilidad: la expone.

IV. Conclusiones y recomendaciones

El recorrido realizado permite volver a la pregunta inicial: ¿de qué manera la irrupción de la IA generativa pone en evidencia las vulnerabilidades del modelo academicista y, al mismo tiempo, reafirma el valor de las humanidades en la formación ética del pensamiento? Las secciones anteriores han ofrecido elementos técnicos, sociales e interpretativos para responder a la pregunta de manera articulada.

El primer hallazgo es que la IA generativa expone, antes que las debilita, las fragilidades preexistentes del modelo academicista. Un estilo de producción intelectual saturado de tecnicismos, citas automáticas y textos escritos más para cumplir métricas que para iluminar un problema resulta efectivamente vulnerable a la automatización, porque la Inteligencia Artificial puede imitarlo con notable facilidad. Lo que se observa, entonces, no es una sustitución de las humanidades por la IA, sino el revelado, por contraste técnico, de prácticas institucionales que ya antes habían perdido fecundidad.

El segundo hallazgo es que, al exponer estas vulnerabilidades, la IA generativa reafirma aquello que las humanidades sí pueden ofrecer y que ningún sistema técnico puede sustituir. Conviene aquí considerar, con ánimo de justicia interpretativa, un posible contraargumento: que la forma más directa de resguardar la integridad académica sería desarrollar mecanismos capaces de distinguir de manera fiable entre escritura humana y escritura generada por IA, otorgando un “sello de autenticidad”. Sus limitaciones, sin embargo, son tres. Primero, la detección perfecta supone un nivel de transparencia que no existe: la producción textual mediada por IA está atravesada por contextos, roles y decisiones opacas que no pueden reducirse a un certificado técnico (Alvarado, 2023, p. 23). Segundo, la explicabilidad no es una propiedad del sistema que pueda activarse desde fuera, sino una práctica comunicativa orientada a permitir que los usuarios actúen con criterio (Páez, 2019, p. 446). Tercero, la confianza no se construye revelando código o entregando probabilidades estadísticas, sino estableciendo relaciones epistémicas comprensibles sobre cuándo, cómo y para qué se emplea la IA (Miller, 2019, p. 3). La

respuesta al problema, por tanto, no es técnica sino formativa: lo que reafirma a las humanidades es justamente su capacidad para sostener el juicio, la interpretación situada y la responsabilidad intelectual que ninguna métrica ni detector puede asegurar.

Los dos hallazgos responden a la pregunta de manera articulada: la IA generativa expone la vulnerabilidad del modelo academicista en la misma operación en que reafirma el valor de las humanidades, porque la fragilidad de uno y la pertinencia de las otras son dos efectos del mismo fenómeno. Lo que la IA puede replicar es lo que el academicismo había dejado vacío de pensamiento; lo que no puede replicar es lo que las humanidades, cuando se ejercen con seriedad, todavía custodian.

El aporte de este ensayo al debate actual reside en mostrar que la discusión sobre IA y humanidades no se zanja eligiendo entre el alarmismo y el entusiasmo tecnológico, sino re-orientándola hacia el modelo institucional que las contiene. De este desplazamiento se desprenden cuatro recomendaciones prácticas, todas orientadas a fortalecer las prácticas formativas humanísticas en un entorno académico transformado por la IA generativa.

Finalmente, se presentan algunas recomendaciones orientadas hacia el uso de la IA generativa en la producción académica. La primera de ellas es adoptar políticas editoriales explícitas. Que las revistas científicas establezcan claramente qué formas de asistencia de IA son aceptables, cuáles vulneran la integridad académica y cómo deben declararse permite construir un marco realista de transparencia sin recaer en promesas tecnológicas imposibles ni en mecanismos punitivos desproporcionados. Algunas publicaciones de referencia, como *Ethics* de la University of Chicago Press, ya exigen declarar el uso de herramientas de IA y reafirman que la responsabilidad intelectual recae exclusivamente en los autores humanos. La política editorial no es un trámite administrativo sino una manera de custodiar la convicción de que la creatividad, entendida como capacidad de proponer interpretaciones nuevas, formular problemas inéditos y asumir responsabilidad por las ideas propias, no puede delegarse en un modelo estadístico.

La segunda recomendación tiene que ver con la formación. La labor humanística se construye en el recorrido intelectual que la hace posible: la lectura cuidadosa que obliga a detenerse en un argumento difícil, la recolección rigurosa de citas que enseña responsabilidad con las fuentes, la escritura pausada que permite descubrir y depurar las propias ideas, y el diálogo crítico con los pares que somete las intuiciones a contraste. Incluso prácticas que pueden parecer anacrónicas, como la escritura a mano o la elaboración de fichas, cumplen una función pedagógica irremplazable, pues ralentizan el pensamiento y permiten que la comprensión madure. Ninguna IA participa de este proceso: puede producir textos convincentes, pero no del conflicto interpretativo ni de la transformación personal que ocurre al estudiar algo con profundidad.

La tercera recomendación tiene que ver con la recuperación de un canon humanístico de calidad. Leer no equivale a procesar textos, mientras que comprender no equivale a resumir. Una formación humanística sólida requiere un canon flexible, revisable pero

existente, de grandes libros y problemas fundamentales que permita a los estudiantes adquirir un trasfondo común de lecturas (Torralba López, 2022, pp. 32-34). Este trasfondo es indispensable para evaluar críticamente las respuestas generadas por una IA, pues permite reconocer cuándo una formulación es superficial y cuándo, en cambio, surge de tradiciones filosóficas, literarias o históricas profundas.

Finalmente, se recomienda capacitar en el uso honesto de la IA. Demonizar estas herramientas solo genera hipocresía y las empuja al uso oculto. La alternativa ética consiste en enseñar a emplearlas como instrumentos de apoyo para orientar lecturas, organizar ideas preliminares o esbozar estructuras argumentativas, sin convertirlas en sustitutos del razonamiento propio. Es necesario, además, mostrar que la apariencia de buenos resultados puede atrofiar habilidades fundamentales: enfrentarse a la página en blanco, decidir cómo comenzar un argumento, revisar críticamente un texto propio son procesos que moldean la mente y fortalecen la creatividad intelectual. La IA puede ayudar a avanzar, pero no puede reemplazar el trabajo personal que convierte la información en comprensión ni la escritura en un acto de pensamiento.

Cabe reconocer, finalmente, los límites de este ensayo. Por su carácter argumentativo, no aporta evidencia empírica sistemática sobre la magnitud o la distribución concreta de los efectos descritos.Trabajos futuros podrían complementarlo con revisiones sistemáticas de la literatura o estudios de caso que documenten, a nivel disciplinar y regional, los usos efectivos de la IA generativa en la enseñanza e investigación humanísticas. Otras líneas pendientes incluyen el análisis comparativo entre tradiciones disciplinares (filosofía, historia, literatura, estudios culturales), el seguimiento longitudinal de las políticas editoriales que están emergiendo y la elaboración de criterios pedagógicos concretos para una alfabetización crítica en IA en el aula universitaria.

En definitiva, las Humanidades no son las grandes amenazadas por la Inteligencia Artificial, sino la condición de posibilidad para su uso plenamente humano. Allí donde la IA solo combina patrones, las Humanidades enseñan a comprender, a interpretar y a ejercer un juicio que ninguna máquina puede replicar. El desafío no es proteger a las humanidades de la tecnología, sino permitir que ellas iluminen su empleo ético y responsable. Toda interacción valiosa con la IA depende, en último término, de la calidad de la inteligencia humana que la guía.

Bibliografía

Alvarado, R. (2023). *AI as an epistemic technology*. Science and Engineering Ethics, 29, Article 32.

Arar, K. H., Özen, H., Polat, G., & Turan, S. (2025). *Artificial intelligence, generative artificial intelligence and research integrity: A hybrid systemic review*. Smart Learning Environments, 12, artículo 44.

Arriagada-Bruneau, G. (2025). *Descolonizar la ética de inteligencia artificial para políticas públicas inclusivas: Reflexiones desde Latinoamérica*. Revista Estudios y Políticas Públicas, 24, 35–50.

Arriagada-Bruneau, G., López, C., & Mendoza, M. (2026). *Ethics in artificial intelligence and information technologies*. CRC Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). Association for Computing Machinery.

Bijker, W. E., Hughes, T. P., & Pinch, T. (Eds.). (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT Press.

García-Huidobro, J. (2018). *Algunas observaciones críticas sobre los procesos chilenos de acreditación*. En J. A. Guzmán, J. I. Brito, & I. Illanes (Eds.), *La universidad en debate: 18 miradas sobre una controversia* (pp. 295–311). Universidad de los Andes.

Klein, L., Martin, M., Brock, A., Antoniak, M., Walsh, M., Johnson, J. M., Tilton, L., & Mimno, D. (2025). *Provocations from the humanities for generative AI research*. arXiv. https://arxiv.org/abs/2502.19190

Kocyigit, B. F., Okyay, R. A., Seiil, B., Qumar, A. B., & Sumbul, H. E. (2025). *Analysis of retracted publications on artificial intelligence: Trends, ethical concerns, and scientific integrity*. Journal of Korean Medical Science, 40(44): 1-12.

Kudina, O., & van de Poel, I. (2024). *A sociotechnical system perspective on AI*. Minds and Machines, 34, Article 21.

MacKenzie, D., & Wajcman, J. (1999). *The Social Shaping of Technology*. Oxford University Press.

Miller, T. (2019). *Explanation in artificial intelligence: Insights from the social sciences*. Artificial Intelligence, 267, 1–38.

Miranda Montecinos, A. (2013). Plagio y ética de la investigación científica. Revista Chilena de Derecho, 40(2), 711–726.

Nussbaum, M. C. (2010). *Sin fines de lucro: Por qué la democracia necesita de las humanidades* (M. V. Rodil, Trad.). Katz Editores.

Okerlund, J., Klasky, E., Middha, A., Kim, S., Rosenfeld, H., Kleinman, M., & Parthasarathy, S. (2022). *What's in the chatterbox? Large language models, why they matter, and what we should do about them*. University of Michigan, Gerald R. Ford School of Public Policy.

Páez, A. (2019). *The pragmatic turn in explainable artificial intelligence (XAI)*. Minds and

Machines, 29, 441–459.

Talukdar, S. R., Lee, L. M., & Jandreau, C. B. (2025). Does generative artificial intelligence pose a threat to the humanities? Journal of AI, Humanities, and New Ethics, 1(2), 43–59.

Torralba López, J. M. (2022). *Una educación liberal: Elogio de los grandes libros*. Encuentro.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *31st Conference on Neural Information Processing Systems.*

Notas

↑1 El autor agradece los comentarios de la profesora Gabriela Arriagada-Bruneau.