

FANTASMAS DIGITALES, ALGORITMOS MORALES Y EL DESAFÍO DE ENSEÑAR ÉTICA EN LA ERA POSHUMANA

Digital Ghosts, Moral Algorithms, and the Challenge of Teaching Ethics in the Posthuman Age

AMITABH VIKRAM DWIVEDI*

amitabh.vikram@smvdu.ac.in

<https://orcid.org/0000-0002-7779-2654>

Shri Mata Vaishno Devi University, UT of Jammu & Kashmir, India

<https://ror.org/036x6w630>

Forma sugerida de citar: Dwivedi, Amitabh Vikram. (2026). Fantasmas digitales, algoritmos morales y el desafío de enseñar ética en la era poshumana. *Sophia, Colección de Filosofía de la Educación*, (40), pp. 83-123.

Resumen

Este artículo analiza las implicaciones éticas derivadas de las tecnologías digitales (TIC) y las infraestructuras algorítmicas en la educación contemporánea, especialmente la inteligencia artificial (IA), los sistemas de clasificación basados en datos y los incentivos orientados al beneficio económico. Examina cómo estos factores transforman la conducta moral de estudiantes y docentes. A partir del concepto de “fantasmas digitales”, entendido como las huellas de autoría, intención y autenticidad que persisten en obras mediadas o generadas por IA, se sostiene que los modelos educativos actuales propician una forma de responsabilidad poshumana marcada por la erosión de la rendición de cuentas tradicional. Mediante viñetas empíricas, se exploran ejemplos de decadencia ética, como la inflación algorítmica de citas, las tesis redactadas con modelos lingüísticos de gran escala y un escándalo ocurrido en julio de 2025: la compraventa de filiaciones institucionales a través de LinkedIn. Estos casos evidencian el rumbo preocupante que sigue la educación, donde la ética se diluye en los intersticios de un sistema tecnológicamente integrado. Cuanto más dependen las instituciones de plataformas y métricas digitales, mayor es el riesgo para la integridad y la responsabilidad académica. En este contexto, urge reconsiderar el significado práctico de la integridad y redefinir los marcos de responsabilidad ante las transformaciones tecnológicas que reconfiguran la idea misma de lo humano.

Palabras clave

Ética digital, integridad académica, educación poshumana, algoritmos morales, capitalismo de vigilancia, ética de la virtud.

* Es decano de Humanidades y Ciencias Sociales y profesor de la Facultad de Lenguas y Literatura de la Universidad Shri Mata Vaishno Devi (India), además de experto certificado por Harvard en Escritura y Oratoria. Su investigación se centra en la lingüística, la documentación lingüística y la preservación de las lenguas, la literatura y la cultura del sur de Asia en peligro de extinción. Google Académico: <https://scholar.google.com/citations?user=R1bv5L4AAAAAJ&hl=en>
Índice h: 9

Abstract

The article tracks how AI sorting tools and profit-seeking metrics quietly rewrite the ethical rulebook of the university. Introducing the concept of “digital ghosts” to describe the artefactual traces of authorship, intent, and authenticity produced by technology-mediated work, it contends that current educational models encourage a posthuman form of responsibility characterized by weakened accountability. Through empirical vignettes—including algorithm-driven citation inflation, dissertations constructed with large language models, and a July 2025 incident in which institutional affiliations were openly sold on LinkedIn—the article documents an academy where ethical judgment is displaced by platform logic and metrics, and where growing dependence on dashboards increases integrity risks for students and faculty alike. It explains how algorithmic infrastructures reshape authorship and evaluation, examines the consequences across academic roles, and proposes a program to restore accountability that pairs provenance and audit-ready practices with pedagogy integrating technical literacy and ethical reasoning. The article concludes that academic integrity must be deliberately redesigned for digitally mediated scholarship rather than assumed to remain intact.

Keywords

Digital Ethics, Academic Integrity, Posthuman Education, Moral Algorithms, Surveillance capitalism, Virtue Ethics

84



Introducción

Las últimas investigaciones han sugerido que se preste más atención a las infraestructuras algorítmicas que informan y producen a la investigación, en una academia contemporánea mediada digitalmente (Recio Sastre, 2025). La inteligencia artificial (IA), los sistemas de clasificación basados en datos y la motivación por el lucro tras la llegada de la IA, han transformado las universidades en simples lugares de interacción entre profesores y estudiantes. Este cambio plantea temas cruciales sobre la responsabilidad de proteger a todos los afectados: estudiantes, profesores e instituciones. Este auge de los “fantasmas digitales” socava la concepción convencional de la autoría y nos obliga a replantearnos lo que gestionamos y lo que no, de lo que somos responsables y en lo que creemos.

Este artículo explora los aspectos éticos que rodean estas tecnologías digitales en la educación moderna. Mediante el estudio de cómo los cambios en estos factores afectan a la conducta ética, tanto de los estudiantes como del profesorado, el estudio pretende mostrar cómo los paradigmas educativos actuales han desempeñado un papel importante en la generación de un modo atenuado de responsabilidad normativa caracterizado por ese viejo tópico poshumano, la responsabilidad moral.

Aunque las instituciones educativas se están volviendo cada vez más digitales, el núcleo de la integridad académica parece estar perdiendo su esencia. Todo ello plantea importantes temas sobre la veracidad de la

generación de conocimiento en un entorno cada vez más caracterizado por las publicaciones algorítmicas y el aislamiento de los espíritus humanos que participan en la producción del conocimiento. En la raíz de este dilema se encuentra la necesidad de comprender cómo un sistema de información poshumano desafía las prácticas éticas y desplaza la carga de la responsabilidad de los actores humanos.

El estudio propone que la proliferación de fantasmas digitales — productos de la IA y la algometría— provoca una irresponsabilidad epistémica en el mundo académico. Un cambio de tal magnitud nos obliga a replantearnos la integridad y la responsabilidad, y a rematerializar la obligación ética de acuerdo con los mundos creados por nuestras tecnologías digitales.

A medida que nuestras prácticas educativas se entrelazan cada vez más con las innovaciones digitales, es fundamental que los académicos, los responsables políticos y la sociedad en general participen en debates sobre las implicaciones éticas de esta transformación. Esto lleva a reflexionar sobre la integridad académica y las rutinas éticas, revelando en última instancia lo que podríamos esperar cuando los algoritmos digitales configuran cada vez más las actividades educativas como parte de la *gubernamentalidad algorítmica*.

Este tema se vuelve aún más relevante teniendo en cuenta los recientes escándalos relacionados con filiaciones institucionales, tecnologías digitales y fraude académico. En el contexto de un entorno pedagógico en rápida evolución, el conocimiento de este asunto se ha vuelto esencial, en gran parte porque los marcos éticos existentes deben cambiar para adaptarse a las nuevas realidades facilitadas por las tecnologías. A medida que más partes interesadas se comprometen con la visibilidad y las métricas cuantitativas en lugar del compromiso real, resulta esencial abordar esta problemática para la educación y el futuro de la investigación académica.

En este artículo se aplica un enfoque de investigación cualitativa, importando viñetas empíricas para representar estas formas de declive que surgen debido a las influencias algorítmicas y los resultados generados por la IA. El análisis se inscribe en una revisión bibliográfica exhaustiva y utiliza la técnica de los estudios de caso junto con marcos teóricos aplicables. Los enfoques van desde una revisión detallada de los principales escándalos de corrupción hasta los ensamblajes algorítmicos y computacionales; así como desde consideraciones sobre las transformaciones neoliberales en las prácticas de autoría y citación hasta un análisis de la economía moral en la que surgen tales revelaciones y una reflexión sobre sus consecuencias. La recopilación de datos consiste en una revisión de



la bibliografía y los informes relacionados con las tecnologías digitales (TIC) en la educación, combinada con la integración de las opiniones expresadas en los debates actuales sobre la ética digital y otros asuntos.

El artículo introduce los conceptos de *fantasma digital* y *limbo epistémico*, para proporcionar un marco conceptual que presenta definiciones junto con una exploración de estos fenómenos. Se presentan estudios de caso y ejemplos de la vida real que ilustran la degradación ética en el mundo académico contemporáneo. Las secciones siguientes amplían el análisis de la conversación y el discurso, examinando sus efectos y su relación con la honestidad académica y la responsabilidad moral. Por último, la conclusión ofrece ideas y recomendaciones para recalibrar la responsabilidad en los contextos educativos poshumanos, junto con sugerencias para futuras investigaciones.

86



El mundo de los fantasmas digitales

Esta sección define los fantasmas digitales y localiza su importancia ontológica en la erudición contemporánea, para luego mostrar cómo los sistemas de IA, las plataformas métricas y los motores de citas erosionan la responsabilidad epistémica. El texto aísla cuatro modos arraigados de fantasmas académicos —manuscritos generados por modelos grandes de lenguaje (LLM), círculos de citas recursivas, puestos de autoría intercambiables y afiliaciones institucionales alquilables— y muestra cómo cada uno de ellos cambia la economía del prestigio de la veracidad a la visibilidad. A continuación, rastrea las consecuencias morales una vez que la deliberación humana es cedida a algoritmos de evaluación opacos: la responsabilidad se diluye en los circuitos y lo “bueno” se convierte en lo que el panel de control puede contar. Un análisis conciso del escándalo del “alquiler de afiliaciones” de Pune, llevado a cabo a la vista de todo el público, ilustra la mecánica de esta erosión. La sección termina señalando dos gráficos que se publicarán próximamente: una matriz de evaluación de riesgos de los daños algorítmicos y una cronología que reconstruye paso a paso el incidente de Pune.

Las universidades modernas actuales están llenas de fantasmas digitales, agentes fantasmales cuya presencia se manifiesta en contenido generado automáticamente, listas de lectura generadas algorítmicamente y gráficos de citas, que se difunden silenciosamente en las redes académicas de actores no humanos que escapan a cualquier autoría humana responsable. Pero estos ecos de innovación no son inocentes, si no que marcan

una preocupante “erosión de la responsabilidad epistémica”, el deber fundamental de rendir cuentas por lo que decimos, observamos, medimos e interpretamos. Cada vez que un modelo lingüístico presenta una revisión bibliográfica en cuestión de segundos o que los servicios de métricas fabrican un Índice h en cuestión de horas, el juicio humano se cede a los *doppelgängers* computacionales de la caja negra (Pasquale, 2015; Zuboff, 2019). En la denominada “sociedad de la transparencia” (Han, 2015), la ironía es que, a medida que los resultados se han vuelto más visibles, la agencia moral que hay detrás de ellos ha desaparecido de la vista.

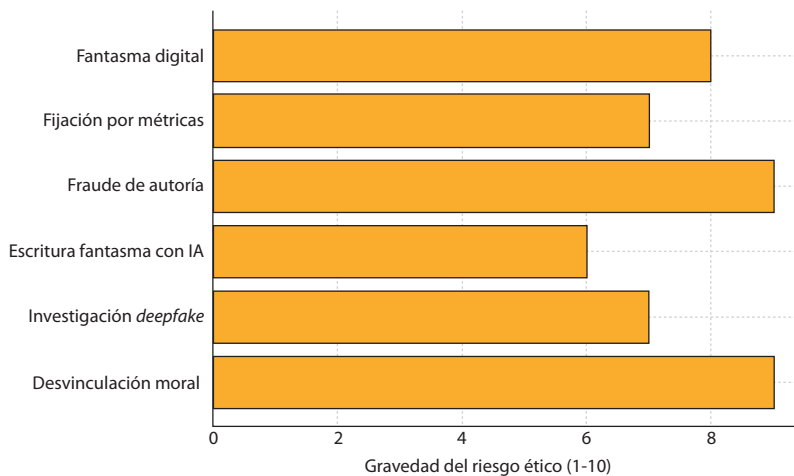
El problema se ve agravado por el hecho de que los fantasmas digitales no se limitan al simple plagio, que es el objetivo más fácil para el *software* de detección. Señalan un cambio ontológico más profundo en qué es la erudición y cómo entra en el mundo académico, una transición que Hayles (1999) previó al hablar del “poshumano”. Los objetos de conocimiento circulan ahora separados del trabajo encarnado y de la responsabilidad, asumiendo al menos cuatro formas recurrentes. Por ejemplo:

- Los manuscritos completos redactados por LLM se adaptan a los autores humanos nominales, lo que se denomina *retórica de la automatización*.
- Los algoritmos de citación desencadenan bucles de retroalimentación recursivos que magnifican la autoridad aparente de obras que pocos han leído realmente (Fong & Wilhite, 2017).
- La autoría se vuelve fluida: los nombres pueden comprarse, venderse o intercambiarse en mercados negros que se parecen menos a una colaboración colegiada y más a un comercio (Hosseini *et al.*, 2018).
- Mientras tanto, los vínculos institucionales se convirtieron en activos rentables que se vendían para mejorar las métricas de clasificación global, lo que supone una inversión de las normas de lo común y escrutinio desinteresado de Merton (1973).

En conjunto, estas expresiones ayudan a indicar la forma de un ecosistema académico en el que la visibilidad y el cálculo, e incluso la fiscalidad, superan el tipo de responsabilidad que solía encarnar la persona que ocupaba el escritorio al final del pasillo y el peso de los fantasmas digitales que escriben cada vez más la historia de la vida académica. En conjunto, estas manifestaciones apuntan a una economía académica poshumana cuya moneda principal no es la verdad, sino la apariencia transmisible. La erudición se recompensa no por la profundidad de la investigación que encarna, sino por su resonancia algorítmica, es decir,

su capacidad para registrarse como un evento cuantificable en los paneles de clasificación, los índices de citas y los análisis de las redes sociales. En un entorno así, el *fantasma digital* no es una anomalía que haya que exorcizar, sino la norma estructural que revela lo que se ha recodificado en la era de los algoritmos morales.

Figura 1
Clasificación de la gravedad de los riesgos éticos en la era digital en una escala de 1 a 10



88



Toma de decisiones poshumana y erosión moral

La noción de Braidotti (2013) sobre la toma de decisiones poshumana captura una coyuntura histórica en la que el locus de la agencia migra de los sujetos encarnados a los ensamblajes maquínicos: agentes de *software*, modelos predictivos y arquitecturas de datos que siguen su propia lógica, internamente coherente, pero socialmente sin interpretación. En la educación superior, esta migración se concreta en una infraestructura interconectada de paneles de clasificación algorítmica, índices de citas que se actualizan automáticamente, *suites* de detección de plagio, algoritmos de puntuación de subvenciones y motores de recomendación que aconsejan a los comités de selección a quién contratar y a las bibliotecas qué comprar. El efecto acumulativo es precisamente lo que Pasquale (2015) describe como la “sociedad de la caja negra”: los juicios trascendentales sobre el mérito académico, el avance profesional y el prestigio institucional se realizan mediante procedimientos computacionales oscuros cuyas

suposiciones, ponderaciones y datos de entrenamiento rara vez se revelan o se debaten.

Esta reconfiguración de la agencia cambia el terreno moral de la academia. El retrato que hace Arendt (1963) de la “banalidad del mal” advierte que las catástrofes morales suelen ser el resultado, no de una intención diabólica, sino del fracaso habitual a la hora de ejercer el juicio reflexivo, una falta de reflexión basada en la rutina procedimental. La gobernanza algorítmica intensifica este peligro al proporcionar una coartada ya preparada: la responsabilidad puede difuminarse en el código, externalizarse a conjuntos de datos o atribuirse al “sistema”. Mittelstadt *et al.* (2016) señalan que la mayoría de los procesos de aprendizaje automático están diseñados para la optimización y la predicción, no para la deliberación normativa, pues carecen de cualquier capacidad intrínseca para sopesar la justicia, la equidad o la integridad académica. Cuando los comités de promoción se remiten a los límites del Índice h generados automáticamente que se actualizan cada noche, o los editores de revistas confían en las previsiones del impacto de las citas, toman decisiones con implicaciones morales aún no determinadas, cediendo el manejo ético a la correlación estadística. La consecuencia de esta evolución ha sido una iteración pseudovoluntarista de la reflexión de Arendt: las personas pulsán los botones, pero toda la indagación moral que debería dirigir su comportamiento se extradita tácticamente a algoritmos prácticamente indescifrables.



El incidente de Pune: un estudio de caso de ghosting académico

El 7 de julio de 2025, una búsqueda rutinaria por LinkedIn reveló una publicación de un profesor de economía en mitad de su carrera en un instituto privado de gestión en Pune. El mensaje era transaccional: “40 000 rupias por artículos únicos o de primer autor; 25 000 rupias por artículos de autor intermedio, solo hay que añadir el nombre de nuestro instituto. Envíe un mensaje directo para obtener más detalles”. En cuestión de horas, la oferta había acumulado miles de visitas, docenas de “me gusta” y un hilo de comentarios que parecía una etnografía de la cultura académica contemporánea, antes de que el autor borrara apresuradamente las pruebas. Sin embargo, las capturas de pantalla ya habían comenzado a circular en Twitter, ResearchGate y grupos departamentales de WhatsApp, lo que garantizó la posteridad de la publicación como un caso de estudio de lo que podría llamarse la *financiación de la afiliación*: tratar la marca institucional como un complemento comercial para cualquier manuscrito a punto de ser presentado.

Figura 2
Publicación censurada en LinkedIn que hace referencia
a una ética de investigación cuestionable



Lo que hizo que el incidente fuera especialmente revelador no fue la solicitud fraudulenta en sí misma —pues anuncios similares han circulado durante mucho tiempo en los pasillos de conferencias— sino la amplitud y el tono de las reacciones que provocó. Un doctor experto en *marketing* observó con pragmatismo: “El sector educativo tiene nuevos verticales de negocio. Muchos profesores adjuntos han dejado la docencia para dedicarse a esto a tiempo completo” (comentario en LinkedIn, 9 de julio de 2025). Otros expresaron abiertamente su disgusto o su humor negro, pero en el hilo se percibía un trasfondo de resignación: así es como se juega ahora el juego de las métricas. El espectro de respuestas se ajusta perfectamente a la taxonomía de la desconexión moral de Bandura (2016). Algunos comentaristas emplearon comparaciones ventajosas (“al menos no está falsificando datos”), otros desplazaron la responsabilidad (“las clasificaciones del NIRF¹ no nos dejan otra opción”) y otros minimizaron las consecuencias (“un ejercicio de marca inofensivo”). La renuncia más radical vino del profesor M. R. Saeed: “¿A quién le importa la ética? ¿Existe en la política, en la vida empresarial o, lo que es peor, en el mundo académico? ¡Las facultades se ven obligadas a hacer esto o perecer!” (comentario en LinkedIn, 9 de julio de 2025).

En una sola tarde, la publicación de Pune sacó a la luz toda una economía sumergida de intercambio de autorías y alquiler de afiliaciones que suele operar fuera de la plataforma, cristalizando la crisis más amplia de integridad a la que se enfrenta ahora el mundo académico digital. Demostró lo rápido que las normas morales pueden reformularse como oportunidades de mercado una vez que la reputación académica está mediada por clasificaciones algorítmicas y paneles de citas, y lo fácil que resulta para los actores individuales racionalizar su participación en esos mercados cuando las presiones sistémicas hacen que la desviación ética parezca no sólo normal, sino necesaria.

La crisis de la verdad en el mundo académico

En esta sección se describe cómo la preponderancia de las métricas redefine el valor académico a través de los *rankings* (NIRF, Shanghái, REF, SJR, etc.), interpretándolos con el poder disciplinario de Foucault para mostrar cómo la visibilidad cuantitativa desplaza el propósito ético. A continuación, se analiza la monetización de la autoría —fábricas de artículos, cárteles de citas y comercio de firmas— a través de la solicitud de LinkedIn como un eje concreto entre las estructuras de incentivos y la práctica cotidiana. Luego, se realiza un breve análisis global (exclusiones de Clarivate, retractaciones de Hindawi, fábricas de artículos chinos, círculos de citas brasileños) para ilustrar que el problema de la integridad es sistémico y no local. La sección culmina con una descripción del capitalismo de vigilancia en el mundo académico, donde las métricas de impacto, las descargas y las métricas alternativas se convierten en recursos extractivos, alejando el trabajo intelectual de la investigación y normalizando la manipulación de datos por encima del descubrimiento.



Presiones institucionales en las clasificaciones y métricas

Durante la última década, la educación superior se ha reconvertido en lo que Muller (2018) denomina un régimen de “fijación métrica”, un entorno en el que el valor de una persona, un programa o una institución se deduce casi exclusivamente a partir de sustitutos numéricamente manejables —factores de impacto, índices h, porcentajes de colocación y recuentos de citas— mientras que el juicio contextual y el propósito ético pasan a un segundo plano. En ningún lugar es más visible esta lógica que en el NIRF de la India, cuyas puntuaciones anuales determinan la financiación gubernamental, la demanda de estudiantes y el prestigio mediá-

tico. Al recompensar el volumen de publicaciones y el número de citas, el marco fabrica lo que un miembro del cuerpo docente de una escuela de negocios, A. S. Chandel, denomina “un juego que se juega solo para subir en la escala del NIRF y atraer admisiones” (comentario en LinkedIn, 9 de julio de 2025).

Y los *rankings* análogos al NIRF (Shanghái desde China, REF desde Reino Unido o SJR desde Norteamérica) propagan los mismos incentivos a nivel mundial, animando a los académicos a dividir sus estudios en múltiples “unidades mínimas publicables”, cortejar a los cárteles de citas o comprar directamente la autoría.

Una perspectiva útil aquí es la genealogía del poder disciplinario de Foucault (1977), que considera las clasificaciones como tecnologías panópticas que observan, comparan y normalizan los organismos académicos a través de métricas. Bajo su mirada, los investigadores se convierten en lo que Foucault denominaba “cuerpos dóciles”: agentes autorregulados que adaptan su comportamiento a los mínimos y máximos de los indicadores, transformando su labor intelectual en resultados susceptibles de ser plasmados en hojas de cálculo en lugar de investigaciones basadas en principios. “Las instituciones y las disciplinas lamentan la obsesión de las clasificaciones por el número de publicaciones, en detrimento de la propia misión de enseñanza-aprendizaje a la que se dedican” (comentario publicado en LinkedIn, 9 de julio de 2025). El efecto es un alejamiento de la curiosidad epistémica hacia la necesidad estructural de la congruencia métrica, que prioriza la visibilidad cuantitativa sobre la honestidad intelectual y la profundidad cultural.

Pluriempleo académico y monetización de la autoría

La ahora infame solicitud de LinkedIn expone una lógica de mercado más amplia en la que la autoría académica en sí misma funciona como un “Uber intelectual”, una “economía *gig* de la erudición” en la que la visibilidad se puede comprar, vender o alquilar con un clic. Lejos de ser un desliz aislado, el caso analizado pertenece a un ecosistema de atajos monetizados que opera a través de varios canales bien definidos.

- *Fábricas de artículos.* Estas entidades sin escrúpulos (es decir, servicios de tutoría, agencias de edición y las infames fábricas de artículos) producen manuscritos listos para usar con conjuntos de datos falsificados y figuras personalizadas, e incluso títulos de revistas preimpresos a cambio de un pago. Por ejemplo, un análisis reciente de las bases de datos de las editoriales

descubrió miles de estos artículos, muchos de ellos con revisiones por pares casi idénticas y gráficos copiados y pegados, lo que suscitó dudas sobre la autenticidad de las publicaciones académicas (Else & Van Noorden, 2021).

- *Cárteles de citas.* En los entresijos del mundo académico, los investigadores se reúnen en grupos de Slack, canales privados de Telegram y editoriales invitadas en números especiales de gran impacto para crear “círculos de citas”. De igual forma, utilizan redes clandestinas para ayudarse mutuamente a reciclar artículos, reciclaje que se realiza en forma de una lista de referencias que aumenta las métricas de impacto sin aportar una contribución académica significativa (Fong & Wilhite, 2017). Estas actividades pervierten el valor real del trabajo académico y erosionan los cimientos de la rigurosidad académica.
- *Comercio de autorías.* Se trata de una tendencia increíblemente peligrosa, pero parece que la autoría se puede intercambiar de forma similar a los favores políticos o a las titulaciones de “intercambios de visibilidad”. Hosseini *et al.* (2018) revelan las tarifas en las que se establece el precio de las firmas: los créditos de primer autor son altos, los agradecimientos son más bajos y se genera una ficha de servicio de escritura fantasma por cada artículo escrito por un escritor fantasma. La autoría se ha convertido en una mercancía que devalúa la investigación al introducir amplios intereses lucrativos en el corazón del consorcio intelectual académico.

La ingeniosidad práctica de este mercado ha sido captada por el profesor de métodos cuantitativos Pratik B., quien señala en un hilo de LinkedIn que los académicos “secuestran las sugerencias de Google Scholar de personas con el mismo nombre para inflar su número de citas, mientras que otros compran citas a 1000 rupias cada una, como si se tratara de una venta en línea” (comunicación personal, 9 de julio de 2025). Desde un punto de vista marxista, estas prácticas ejemplifican la alienación del trabajo intelectual: el objeto académico permanece separado de su productor y vuelve a circular como un valor de cambio fungible, evaluable mediante clasificaciones, factores de impacto y dinero de matrícula. El oficio del escritor, asociado desde hace tiempo a la investigación y la evidencia, es apto para ser comercializado en los mercados de datos.



Ejemplos globales de cárteles de citas y fábricas de artículos

Como ya adelantamos, el problema de la integridad no está limitado geográficamente, sino que es estructural, mundial. En 2023, Clarivate Analytics² eliminó 82 títulos de Web of Science después de que los analistas forenses descubrieran “patrones de citas anómalos”, un eufemismo para referirse a redes de artículos cuyas referencias fueron manipuladas para inflar las puntuaciones del factor de impacto (Oransky, 2023). De igual forma, siguen apareciendo irregularidades similares en los expedientes del Comité de Ética Editorial (COPE), cuyos informes de incidentes parecen un catálogo de engaños a escala industrial:

- *Cadenas de producción de artículos falsos en China.* La minería de textos realizada por investigadores independientes rastreó más de 400 artículos falsificados, hasta dar con tiendas comerciales de redacción fantasma en China. Todos fueron publicados entre 2019 y 2021 en revistas de Elsevier, con revisores falsos por correo electrónico e imágenes microscópicas copiadas y pegadas (Schneider, 2021).
- *Retiradas masivas de Hindawi.* En 2023, la editorial de acceso abierto Hindawi retiró más de 300 artículos después de que los auditores encontraran plantillas de informes de revisión por pares, cifras recicladas y listas de autores que cambiaron entre la aceptación y la impresión, un patrón que la propia empresa calificó de “manipulación sistemática del proceso de publicación” (Retraction Watch, 2023).
- *Círculos de citas brasileños.* Ya en 2013, el mapeo bibliométrico reveló la existencia de grupos de investigadores brasileños que canalizaban cientos de referencias recíprocas a través de números especiales y actas de congresos para impulsar sus revistas en las tablas de clasificación nacionales (Van Noorden, 2013).

En conjunto, estos acontecimientos revelan una variante académica de lo que Zuboff (2019) denomina “capitalismo de vigilancia”. Las métricas generadas a partir de citas, descargas y señales altimétricas sirven como recursos principales para la extracción de valor, mientras que el conocimiento que estas métricas pretenden reflejar pasa a un segundo plano. Las revistas persiguen factores de impacto para asegurarse fuentes de ingresos por APC;³ las universidades recopilan cuadros de mando de citas para ascender en las listas de clasificación mundiales; los académicos tratan los puestos de autoría como moneda de cambio. En este circui-



to económico, la manipulación de datos —y no el descubrimiento— se convierte en el principal modo de producción, lo que aleja aún más el trabajo intelectual de su propósito epistémico y afianza un mercado en el que la atención y la credibilidad pueden fabricarse algorítmicamente.

Delegación moral a las máquinas

Esta sección traza la reconfiguración de la autoría y la responsabilidad en la práctica académica bajo la IA generativa. Comenzando con un tratamiento centrado en “la IA como autora”, cuestiona cómo los LLM producen grandes textos en citas que confunden los marcadores tradicionales de originalidad, contribución y propiedad. A continuación, documenta los vectores emergentes del fraude académico —*deepfakes*, corpus sintéticos e identidades falsas— y sostiene que estos fenómenos precipitan una crisis de verificación, reasignando el trabajo de la producción de conocimiento a la autenticación. Basándose en la “agencia distribuida” de Bruno Latour (2005), el análisis desglosa la responsabilidad entre los estudiantes, los desarrolladores de plataformas, las instituciones y las infraestructuras de evaluación, representadas a través de un mapa de red de un manuscrito generado por IA. La sección termina poniendo de relieve lo que está en juego desde el punto de vista normativo: cuando el razonamiento moral se integra en las rutinas de optimización, la corrección se reduce a la puntuación. A continuación, esboza contramedidas —rastros de procedencia obligatorios y verificación de la autoría a prueba de manipulaciones— que podrían restablecer la responsabilidad dentro del bucle algorítmico.

95



La IA como autora en ChatGPT

La llegada de los LLM a gran escala —ChatGPT, Claude, Gemini, etc.— ha redibujado la cartografía de la autoría académica. Tras haber aprendido de terabytes de texto digital, estos sistemas pueden producir un contenido elegante y repleto de citas que imita el estilo de escritura, los enfoques argumentativos y las convenciones disciplinarias de la literatura revisada por pares. Una sola indicación puede producir un resumen (con menciones de Derrida y Latour), una revisión bibliográfica llena de DOI imaginarios o una sección de resultados adornada con tablas fabricadas. Pero esta hazaña tecnológica tiene un coste epistémico: desestabiliza las categorías de originalidad y contribución en las que se ha basado tradicionalmente el crédito académico (Lund *et al.*, 2023).

La condición resultante podría denominarse *ambigüedad de la autoría*. Mientras que el plagio solía implicar un claro acto humano de apropiación, ahora nos enfrentamos a resultados cuya procedencia es algorítmica, colectiva y probabilística. ¿El doctorando se limitó a pulir el borrador de la IA o fue el modelo el que proporcionó el núcleo conceptual del argumento? ¿Quién —si es que hay alguien— es el propietario de las frases ensambladas estadísticamente a partir de millones de fuentes protegidas por derechos de autor? Los primeros trabajos empíricos sugieren que esta práctica ya no es marginal: varias editoriales científicas, técnicas y médicas (STM)⁴ informan que entre el 15 % y el 20 % de las nuevas presentaciones muestran huellas textuales compatibles con la asistencia de LLM: ráfagas léxicas inusuales, artefactos de compresión en las referencias u homogeneización estilística en manuscritos de varios autores.

Estos avances acentúan la distinción propuesta por Floridi *et al.* (2018) entre “la ética de la IA” (cómo nosotros, como agentes morales, diseñamos y pusimos en práctica a estos sistemas) y “la IA ética” (la expectativa errónea de que las máquinas pueden asumir el razonamiento normativo por nosotros). Cuando ChatGPT redacta una sección de debate, la responsabilidad ética de verificar las fuentes, comprobar la coherencia lógica y declarar las contribuciones sigue recayendo en el investigador humano; delegar esas tareas al modelo confunde la fluidez computacional con la agencia moral. Sin embargo, la velocidad y la fluidez de los LLM invitan precisamente a esa abdicación, tentando a los académicos a tratar la deliberación ética como otra función rutinaria que se puede externalizar. A menos que los campos académicos desarrollen normas claras sobre la procedencia, las recomendaciones sobre los umbrales de divulgación y recomendaciones pedagógicas, la línea entre la artesanía intelectual y la imitación automatizada se erosionan aún más, creando un mercado de ideas en el que la propia responsabilidad podría ser tan artificial como el contenido de la página.

Deepfakes y secuestro de citas

La escritura fantasma basada en texto es solo el inicio. Un nuevo conjunto de herramientas generativas permite ahora falsificar prácticamente todos los elementos del registro de investigación, transformando el fraude de un trabajo manual que requiere mucha mano de obra en un servicio de tomar apuntes y hacer clic. Entre los avances más destacados se incluyen:

- *Conferencias y charlas deepfake*. Con poco más que una fotografía de cabeza y hombros y una muestra de voz extraída de

un seminario web, los “estudios de avatares” comerciales pueden sintetizar videos de alta definición en los que un académico parece esbozar protocolos experimentales, repasar diapositivas de PowerPoint y responder a las preguntas del público, todo ello para un artículo que nunca ha existido. Los motores de sincronización labial, seguimiento ocular y ruido ambiental hacen que las imágenes sean difíciles de distinguir de una conferencia auténtica, lo que permite a los presentadores fraudulentos impresionar a los comités de contratación o asegurarse honorarios por conferencias sin incluso moverse de sus escritorios.

- *Fábricas de datos algorítmicos.* Las GAN (redes generativas antagónicas) especializadas y los modelos de difusión ahora pueden producir conjuntos de datos de cualquier tipo (imágenes de resonancias magnéticas o microscópicas, secuencias genómicas u hojas de cálculo de encuestas) que satisfacen las distribuciones estadísticas esperadas y superan la mayoría de las comprobaciones de plausibilidad. Un investigador puede solicitar “Generar un CSV^s de 10 000 encuestados que muestre una correlación positiva significativa entre el entrenamiento en mindfulness y el promedio de calificaciones ($p < 0,01$)” y recibir números perfectamente formateados que superarán la revisión rutinaria por pares, a menos que se requieran auditorías de datos brutos o la replicación del código.
- *Identidades académicas sintéticas.* Los servicios de generación de perfiles unen retratos de IA, redes de coautores fabricadas y listas de publicaciones seleccionadas que contienen DOI (reales o falsificados ingeniosamente, pues redirigen a callejones sin salida con muros de pago). Estos fantasmas registran identificadores ORCID, abren cuentas en Google Académico y solicitan puestos en consejos editoriales, inyectando así expertos inexistentes en la economía de las citas. Una vez integrados, se pueden recurrir para mejorar las métricas de los clientes de pago o para generar informes positivos de revisión por pares bajo demanda.

En conjunto, estas capacidades son el simulacro de Baudrillard (1994): una copia sin relación al original que circula como si fuera real, sustituyendo la realidad por una hiperrealidad que solo se refiere a sí misma. En el ámbito académico, el resultado es una crisis de verificación epistémica. El control tradicional (revisión por pares, preguntas y respuestas en conferencias, comprobación de afiliaciones institucionales) se

diseño para actores maliciosos que pudieran adornar la realidad, no para algoritmos que pueden fabricar mundos enteros a gran escala. La base ontológica de la erudición —su anclaje en procedimientos observables, individuos responsables y pruebas reproducibles— comienza a disolverse cuando pueden conjurarse *ex nihilo* una convincente conferencia *deepfake*, un conjunto de datos estadísticamente impecable y un académico con un impresionante Índice h. La tarea académica central pasa así de producir conocimiento a autenticar la realidad, un cambio que amenaza con desviar la energía intelectual del descubrimiento hacia un juego interminable de puesta al día.

¿Quién es el responsable? El problema de la agencia distribuida

98



El delegar cada vez más tareas de lectura, síntesis y redacción a algoritmos generativos reconfigura el mapa moral de la erudición, exactamente como predijo Latour (2005) cuando habló de la “agencia distribuida”. El acto final —subir un trabajo de fin de semestre al portal del curso— depende ahora de un relevo sociotécnico en el que actúan múltiples actores:

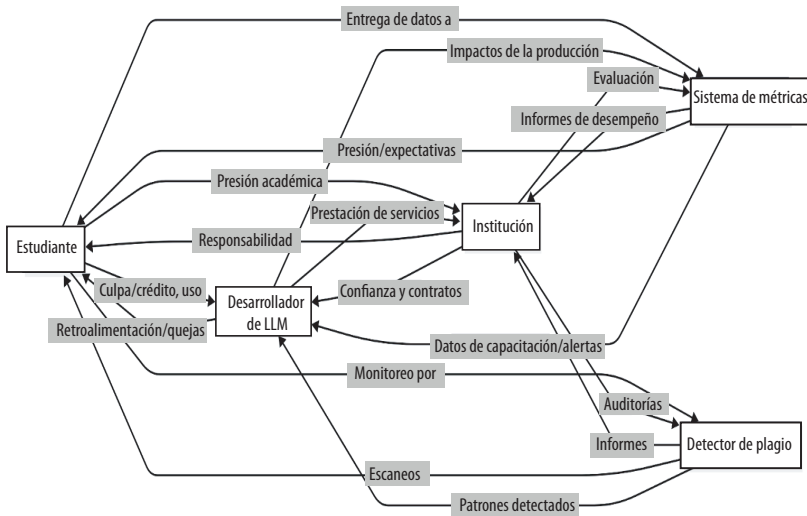
- El estudiante que formula la pregunta y decide pulsar “enviar”.
- El equipo de ingeniería cuya arquitectura transformadora traduce esa propuesta en contenido.
- Los inversores de capital de riesgo que presionan para que el crecimiento de usuarios sea fluido.
- La universidad que no ofrece una orientación clara más allá de la plantilla sobre plagio.
- Los sistemas de clasificación global que premian el rendimiento y la visibilidad de las citas por encima del aprendizaje reflexivo.

Cuando el ensayo se marca posteriormente como “asistido por IA”, la culpabilidad se dispersa por toda esta red. ¿El estudiante es un tramposo o simplemente un usuario oportunista de una herramienta con licencia del campus? ¿Son los desarrolladores los responsables de lanzar un producto que puede superar en escritura a los estudiantes universitarios? ¿Debería el profesor haber exigido defensas orales para verificar la autoría? Ningún nodo puede atribuirse de forma creíble la responsabilidad total del resultado, lo que demuestra la observación de Latour (2005) sobre cómo la acción pertenece a un conjunto y nunca a un sujeto.

Esta dispersión crea la atmósfera moral que Bauman (2001) denominaría “Modernidad líquida”. Al entrar en el frágil ecosistema de los ciclos de actualización, los memorandos de políticas y las métricas de los

paneles de control, las normas que antes parecían sólidas —la autoría, la originalidad, la integridad académica— se funden en corrientes fluidas. Anticipando perfectamente el resultado, el ecólogo conductual M. Sharma capturó su esencia en una repulsión muy comentada en LinkedIn: “Todo vale en el amor, la guerra y las calificaciones académicas” (comunicación personal, 9 de julio de 2025). En tal sentido, la responsabilidad se comporta como el mercurio sobre una mesa: cada vez que una institución intenta fijarla, se rompe en gotas más pequeñas que ruedan a otra parte, sin ser ningún actor responsable.

Figura 3
Mapa de agencia distribuida en un trabajo generado por IA



La brecha axiológica y los valores en desaparición en el trabajo digital

Esta sección sitúa una crisis de autenticidad en la producción académica y muestra cómo las herramientas digitales erosionan los valores en tres niveles: la autenticidad epistémica (donde los incentivos métricos desplazan la investigación), las relaciones auténticas (donde la coautoría y la colaboración se convierten en transaccionales) y la autenticidad institucional (donde las clasificaciones reconfiguran la misión de la universidad). A continuación, se analiza el declive de la responsabilidad y la justicia: se recurre a Lévinas para enmarcar la responsabilidad hacia el otro y a

Fraser para mostrar cómo surge la “justicia anormal” cuando la gestión algorítmica enmascara el poder y reproduce la desigualdad. Por último, analiza la desconexión moral como una característica sistémica de la academia digitalizada, utilizando los mecanismos de Bandura (justificación moral, etiquetado eufemístico, comparación ventajosa, desplazamiento, minimización), una tabla comparativa para alinear las formas académicas clásicas y digitales, y el poder instrumental de Zuboff para explicar cómo el diseño de las plataformas y las métricas en tiempo real normalizan estas violaciones.

Autenticidad en la producción académica

La rápida digitalización de la vida académica ha impulsado a las universidades a lo que Taylor (1991) describiría como una “crisis de la ética de la autenticidad”. El concepto de Taylor se basa en el ideal de que los individuos y las comunidades deben actuar de manera que expresen sus verdaderas convicciones y valores. Las herramientas digitales que pueden fabricar contenido, inflar métricas o monetizar el prestigio ahora hacen que ese ideal sea difícil de alcanzar, o incluso de reconocer. El resultado es una erosión de la autenticidad en múltiples niveles que afecta a todos los ámbitos de la práctica académica:

- *Autenticidad epistémica*. La integridad de la creación de conocimiento en sí misma:
 - a. Ideal: la investigación se lleva a cabo con el propósito intrínseco de comprender el mundo, guiada por la curiosidad, el rigor y la transparencia metodológica.
 - b. Erosión: los paneles bibliométricos, los relojes de permanencia y los sistemas de recomendación algorítmica premian el volumen y el número de citas. Están enseñando a una generación de académicos cómo convertir un estudio en múltiples “artículos”, citar a amigos en círculos de citas o externalizar la revisión de la literatura a un modelo lingüístico genérico de gran escala. El ansia por alcanzar las métricas desplaza la meta real; el artículo se convierte en un instrumento de estrategia en lugar de una prueba de investigación.
- *Relaciones auténticas*. La naturaleza de las relaciones académicas:
 - a. Ideal: la colaboración se basa en el respeto intelectual compartido y una causa común, y se reconoce el mérito donde se ha realizado una contribución real.



- b. Erosión: la coautoría funciona ahora cada vez más como una transacción: se añaden autores por su marca, se eliminan para evitar conflictos de intereses y, finalmente, se expresa una reciprocidad justa mediante la inclusión en una orgía de escritos coautorados para acumular citas. La comunicación se limita a plantillas de correo electrónico y tableros de proyectos generados automáticamente, donde antes se encontraba el compromiso profundo y dialógico que solía caracterizar la tutoría y el descubrimiento conjunto.
- *Autenticidad institucional.* La misión y la autoconcepción de la universidad:
 - a. Ideal: las universidades actúan como guardianas del pensamiento crítico y la ilustración pública, estructurando incentivos en torno a la excelencia docente, el debate abierto y la investigación socialmente relevante.
 - b. Erosión: *rankings* como los mencionados más arriba convierten esa misión en un juego de números en el que predominan los ingresos por subvenciones, las publicaciones de gran impacto y el alcance en las redes sociales. Las oficinas administrativas establecen “grupos de trabajo de clasificación”, canalizan los fondos hacia áreas de investigación de rápido rendimiento y presionan al profesorado para que publique en cualquier medio que impulse las métricas institucionales, incluso si la beca es tangencial a las necesidades locales o al aprendizaje de los estudiantes.

En estas tres capas se repite la misma dinámica: las posibilidades digitales que prometen eficiencia y alcance socavan al mismo tiempo las condiciones para la expresión sincera y la contribución honesta. Lo que parece productividad en una hoja de cálculo a menudo enmascara atajos intelectuales; lo que parece colaboración en un currículum puede ocultar alianzas transaccionales; lo que parece éxito institucional en una tabla de clasificación puede indicar un alejamiento de la educación en interés público. En términos de Taylor (1991), la academia corre el riesgo de cambiar la autenticidad —la alineación de las acciones con un propósito moral más profundo— por una versión superficial y performativa de sí misma optimizada para la visibilidad digital.

El declive de la responsabilidad y la justicia

Lévinas (1961) enmarca la ética no como un conjunto de reglas que se derivan del ser, sino como una “responsabilidad infinita” hacia el Otro que nos llama incluso antes de formularnos conceptos propios, de conocimiento o comunidad. En el momento en que la relación cara a cara se redirige a través de pantallas, paneles de control y motores de recomendación automatizados, su intensidad moral se diluye. En la academia digital actual —negociaciones por correo electrónico, portales de seguimiento de manuscritos, búsqueda de talentos en LinkedIn— el Otro aparece menos como una persona concreta y más como un punto de datos o un perfil que se puede aprovechar. El reciente episodio de LinkedIn, en el que los departamentos negociaron abiertamente plazas de coafiliación para mejorar sus puntuaciones en el ranking, evidencia esa erosión. Las etiquetas institucionales se intercambiaron como activos negociables, sin reflexionar apenas sobre cómo esas transacciones afectan a los estudiantes, al personal y al público local, cuyas identidades pretenden significar esas etiquetas.

La justicia dentro de la universidad ha significado tradicionalmente tres compromisos entrelazados: el reconocimiento preciso de quién ha contribuido realmente con su trabajo intelectual; la distribución justa de recursos como becas, tutorías y espacio para publicaciones; la protección vigilante del registro de pruebas en el que se basa la investigación académica posterior.

Las herramientas de gestión algorítmica —índices de citas, paneles de control de financiación, filtros de contratación— prometen ahora salvaguardar estos ideales mediante mediciones objetivas. Sin embargo, como advierte Fraser (2008) en su análisis sobre la “justicia anormal”, los sistemas que parecen imparciales desde afuera pueden ocultar reglas de decisión distorsionadas y relaciones de poder. Los algoritmos de clasificación amplifican las ventajas de las instituciones bien financiadas; los análisis predictivos dan prioridad a los solicitantes de subvenciones cuyo éxito ya es señal de privilegio; los detectores de plagio señalan la escritura de los autores de inglés como segunda lengua con más frecuencia que la de los hablantes nativos. La vía de decisión se esconde detrás del código o de métricas patentadas, lo que provoca que se materialicen nuevas jerarquías precisamente en el momento en que se vuelven más difíciles de ver. En efecto, la maquinaria que se suponía que iba a democratizar el reconocimiento, la distribución y la integridad instala un nuevo régimen de desigualdad, al tiempo que enmascara las huellas de la injusticia bajo una apariencia de neutralidad computacional.

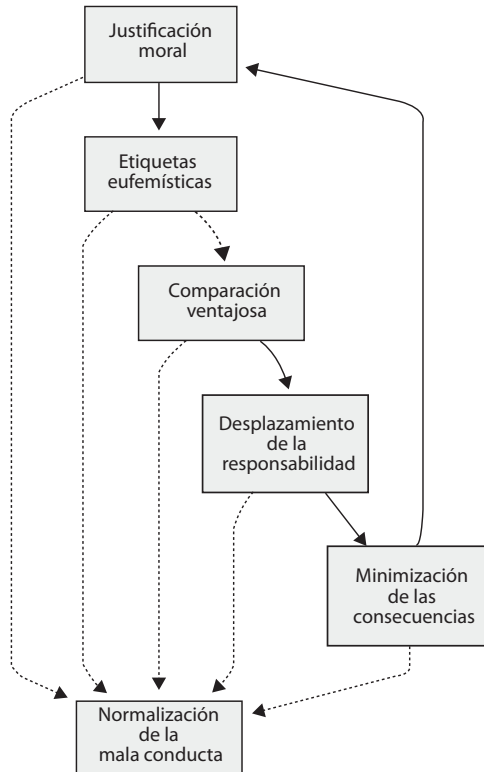
102



La desconexión moral como característica sistémica

El trabajo posterior de Bandura (2016) sobre la desconexión moral describe el bucle cognitivo que accionan las personas cuando quieren comportarse de forma poco ética sin sufrir remordimientos.

Figura 4
El bucle de desvinculación moral en la publicación académica



La academia digitalizada ha integrado estos bucles en las rutinas cotidianas, las políticas institucionales e incluso el diseño de la interfaz de usuario de los sistemas de manuscritos. Hay cinco mecanismos que son especialmente visibles:

JUSTIFICACIÓN MORAL

- *Forma clásica:* “Hice trampa porque había mucho en juego y mi familia depende de mí”.

- *Variante académica digital*: “Todos los laboratorios de nuestro campo inflan las citas o pagan a creadores de artículos; si no lo hacemos, no alcanzaremos el siguiente umbral de clasificación”.

Presta atención al lenguaje de crisis y a las afirmaciones de inevitabilidad. Exige normas preestablecidas antes de los ciclos de revisión para que no se puedan adaptar las razones. Utiliza auditorías a nivel de cohorte para comparar el comportamiento con la mediana del campo. Combina los periodos de amnistía con soluciones prospectivas, por ejemplo, retirar y divulgar, y luego establecer límites a los resultados cuantificables. Enseña la prueba del “atajo menos permisible”: ¿apoyaría esta razón si su nombre fuera público y la norma se aplicase también a sus rivales?

Basta con echar un vistazo a las métricas de las tablas de clasificación o a las tasas de éxito de las agencias de financiación para enmarcar las prácticas cuestionables como autodefensa colectiva en lugar de como malas prácticas.

104



ETIQUETAS EUFEMÍSTICAS

- *Forma clásica*: “daños colaterales” en lugar de “muertes de civiles”.
- *Variante académica digital*: un ejemplo de variante académica digital incluye frases como “colaboración estratégica”, “autoría invitada” u “optimización del factor de impacto”, que se utilizan en lugar del término más directo “fraude de autoría”.

Elabora un glosario de lenguaje sencillo que relaciona los eufemismos con lo que realmente ocurrió. Añade campos de formulario que obligan a realizar acciones concretas, por ejemplo, “sección de métodos aportados” en lugar de “colaborador”. Utiliza una herramienta de detección de jerga en las políticas y las convocatorias de artículos. Durante las revisiones, solicita pruebas que coincidan con el verbo, como el historial de compromisos, el análisis o la correspondencia. Penaliza el etiquetado incorrecto incluso cuando el trabajo subyacente sea aceptado, ya que el lenguaje moldea la repetición de comportamientos.

Los portales de envío de artículos a revistas y los folletos de consultoría normalizan estas frases, lo que permite a los actores ocultar las manchas morales detrás del lenguaje gerencial.

COMPARACIÓN VENTAJOSA

- *Forma clásica*: “Sí, mentí, pero no robé”.
- *Edición académica digital*: “Solo estamos añadiendo honorarios de coautores; los verdaderos villanos son aquellos que inventan conjuntos de datos completos o recurren a imágenes falsas”.

Evalúa las acciones en función de una norma fija, no en comparación con ejemplos peores. Utiliza tablas de sanciones que se adaptan al impacto y la intención, de modo que se sigan abordando los daños menores. En clase, realiza ejercicios de “escalera de infracciones” en los que cada peldaño se juzga por sus propios méritos. Está atento a la parte moral después de pequeñas buenas acciones. Exige una solución prospectiva, por ejemplo, eliminar la autoría inmerecida y corregir los registros, en lugar de conformarse con comparaciones.

Al yuxtaponer las infracciones menores de las normas profesionales con casos de gran repercusión de fabricación de datos, los investigadores se protegen del peso de la crítica ética.

DESPLAZAMIENTO DE LA RESPONSABILIDAD

- *Forma clásica*: “Solo seguía órdenes”.
- *Variante académica digital*: “El sistema de titularidad, los comités de subvenciones y los ejercicios de clasificación hacen que se publique de esta manera; no tenemos otra opción”.



Identifica los roles con una sencilla tabla RACI (responsable, consultado, informado). Inserta puntos de control humanos donde los números son orientativos, no vinculantes. Exige la firma de la autoría y las certificaciones de los métodos por parte de las personas nombradas. Publica el historial de decisiones para los ascensos y los números especiales, de modo que se puedan inspeccionar los motivos. Reforma los incentivos, pero no deja que la estructura borre la agencia. Si una norma hizo probable el resultado, cambia la norma y sigue abordando la elección que se tomó.

La responsabilidad se traslada a estructuras anónimas: paneles algorítmicos, tarjetas de puntuación de los indicadores de rendimiento (KPI) y rastreadores de citas automatizados. Las personas entienden que simplemente están participando en un juego.

DISTORSIÓN O MINIMIZACIÓN DE LAS CONSECUENCIAS

- *Forma clásica*: “Solo fue una pequeña broma”.
- *Variante académica digital*: “Añadir un nombre más a la firma no hace daño a nadie; solo es una línea de texto en un PDF”.

Cuantifica los efectos posteriores. Muestra cómo un solo crédito falso altera las curvas del Índice h, las puntuaciones de las subvenciones y las listas de los preseleccionados. Utiliza declaraciones de impacto en las notificaciones de retractación o corrección que rastrean quién ha perdido

oportunidades. Acompaña la minimización con reflexiones sobre el caso de los compañeros afectados. Exige medidas correctivas, por ejemplo, metadatos corregidos, cartas a los comités y redistribución de los premios cuando sea posible. Hace que el seguimiento de las consecuencias forme parte de los informes anuales, y no sea una disculpa puntual.

Los daños tangibles —créditos mal asignados, decisiones de contratación sesgadas, prestigio institucional inflado— se vuelven invisibles tras una pantalla de abstracción.

Tabla 1
Comparación entre las infracciones morales clásicas
y las digitales en el ámbito académico

Mecanismo	Forma clásica	Forma académica digital
Justificación moral	“Hice trampa por mi familia”	“Todo el mundo compra citas para seguir siendo competitivo”
Etiquetado eufemístico	“Daños colaterales”	“Optimización del impacto” en lugar de fraude de autoría
Comparación ventajosa	“Al menos no robé”	“No es tan malo como falsificar datos”
Desplazamiento	“Solo seguía órdenes”	“Los sistemas de clasificación no dejan otra opción”
Minimización	“Solo fue una pequeña broma”	“Un nombre más no hace daño a nadie”

Fuente: Elaboración propia.

Lo que hace que estos mecanismos sean especialmente potentes hoy en día, es que ya no se limitan a racionalizaciones privadas dentro de la mente del investigador. Están, en primer lugar, codificados en políticas institucionales (por ejemplo, modelos de reparto de ingresos que recompensan a las revistas únicamente por el volumen de artículos; en segundo lugar, están integrados en la arquitectura de las plataformas (por ejemplo, “citas sugeridas” que empujan a los autores hacia redes de reciprocidad); en tercer lugar, están reforzados por paneles de control métricos que se actualizan en tiempo real, convirtiendo los atajos morales en mejoras de rendimiento visibles al instante.

Este sería un ejemplo de lo que Zuboff (2019) ha denominado “poder instrumental”: el poder que no opera mediante órdenes y controles explícitos, sino que impulsa la configuración del comportamiento a través de infraestructuras tecnológicas respaldadas por bucles de retroalimen-



tación de datos. Los académicos modifican su conducta, a menudo de forma preconsciente, para adaptarse a lo que la plataforma recompensa, mientras que las propias plataformas permanecen en gran medida fuera del ámbito de la visibilidad ética. El resultado es un ecosistema que se refuerza a sí mismo, en el que la desconexión moral ya no es un truco psicológico individual, sino el sistema operativo ambiental de la investigación académica digital.

Enseñar en un mundo fantasma

En esta sección se analiza cómo las prácticas de investigación dudosas erosionan el papel del profesorado como modelo a seguir y producen un currículo oculto que valora más la manipulación de métricas que la integridad. Luego, se presenta una pedagogía que fusiona la imaginación narrativa con la alfabetización algorítmica a través de tres métodos de aula: análisis de casos, juegos de rol con múltiples perspectivas y arqueología digital. Se propone un conjunto de herramientas contemporáneas de ética de la virtud (prudencia digital, templanza algorítmica, cibervalentía, tecnoinjusticia) y se enumeran hábitos concretos para poner en práctica esas virtudes (aprendizaje de tipo de estudio, rotación de funciones de administración, sesiones informativas reflexivas estructuradas y salvaguardias institucionales). Todo con el fin de ayudar a estudiantes y maestros a ver cómo los sistemas moldean el comportamiento y a desarrollar prácticas duraderas que se centren en la investigación y la responsabilidad.



La crisis de los modelos a seguir

Cuando las prácticas de investigación dudosas se convierten en rutina, las universidades se enfrentan a una aguda paradoja pedagógica: ¿cómo pueden los profesores fomentar el juicio ético en los estudiantes y, al mismo tiempo, tolerar o beneficiarse silenciosamente de la misma conducta indebida que se supone que deben condenar? El profesor B. K. Panigrahy capturó este dilema con una franqueza pasmosa al referirse al docente de Pune que negocia puestos afiliados: “Es un profesor de economía honesto, que impulsa la economía de su universidad siguiendo las instrucciones de la alta dirección” (comunicación personal, 9 de julio de 2025). En otras palabras, el comportamiento deshonesto se convierte en lealtad institucional; la transgresión moral se aplaude como cumplimiento estratégico.

Tales contorsiones cognitivas dejan a los educadores en la incertidumbre normativa. Las conversaciones en las aulas sobre el plagio y la

virtud académica suenan vacías cuando los estudiantes pueden observar a sus mentores intercambiando puestos de autoría, manipulando citas o presentando resúmenes escritos por IA para cumplir los objetivos trimestrales de los KPI. El aula se convierte en un lugar de doble moral ética: el programa de estudios prescribe la integridad, mientras que la práctica del departamento recompensa el juego métrico.

Ortega y Gasset previó un peligro similar hace casi un siglo. En *La rebelión de las masas* (1930), deploraba la “barbarie de la especialización”, una condición en la que los profesionales alcanzan una competencia técnica muy precisa, pero siguen siendo ética y culturalmente vacíos. Los modelos académicos actuales suelen encajar en esta descripción: dominan la optimización del Índice h, las exenciones de tasas de acceso abierto y las estrategias de amplificación en las redes sociales, pero les cuesta encarnar la honestidad intelectual que la educación superior pretende cultivar.

El resultado es una generación de profesores que:

- Forman a los doctorandos para que dividan un único conjunto de datos en múltiples “unidades mínimas publicables”.
- Aceptan u ofrecen regularmente coautorías “honoríficas” para ganarse favores.
- Tratan las puntuaciones de similitud algorítmica como un sustituto de la ética de citación genuina.
- Respaldan públicamente las políticas de integridad en la investigación, mientras que en privado enseñan a los estudiantes a cómo eludirlas.

En un entorno así, el mensaje sutil que aprenden los estudiantes es el que proporciona el currículo oculto: el éxito no es el resultado de explorar rigurosamente una pregunta o un misterio, ni de informar objetivamente sobre lo que se encuentra, sino de escribir de manera que los vacíos sean fáciles de encontrar. La imaginación moral que las universidades pretenden desarrollar se reduce efectivamente, limitándose a un cálculo estratégico de qué reglas se pueden romper sin correr el riesgo de sufrir vergüenza en las redes sociales.

Desarrollar la imaginación moral crítica

Nussbaum (1997) sostiene que la ciudadanía democrática depende de la “imaginación narrativa”, el hábito cultivado de imaginar la vida desde puntos de vista distintos al propio y de empatizar con esperanzas, miedos y limitaciones que difieren de nuestras propias biografías. En un sistema



universitario cada vez más mediado por códigos, *rankings* y motores de recomendación, la lente narrativa tradicional ya no es suficiente. Los estudiantes y el profesorado también deben adquirir lo que podríamos llamar *una imaginación algorítmica*: la capacidad intelectual de visualizar cómo las infraestructuras tecnológicas asignan visibilidad, distribuyen recompensas y codifican sesgos, y la de comprender las consecuencias humanas que se derivan de esas decisiones de diseño ocultas. Mientras que la imaginación narrativa se pregunta “¿cómo se sentiría ser esa persona?”, nosotros podemos preguntarnos “¿qué bienes hace posibles o imposibles el sistema para esa persona?” y la imaginación algorítmica añadiría “¿qué realidades hace posibles o imposibles el sistema para esa persona?”.

Esta doble imaginación puede animarse a través de tres estrategias pedagógicas:

ANÁLISIS BASADO EN CASOS

- *Método*: los profesores presentan un caso con incidentes publicados, como el escándalo de comercio de afiliaciones en Pune o las retiradas masivas de redes de manipulación de imágenes, y piden a los alumnos que identifiquen a los participantes, las motivaciones y los puntos de decisión.
- *Resultado del aprendizaje*: al analizar casos de la vida real, los estudiantes llegan a ver cómo las ideas generales (honestidad, justicia y responsabilidad) chocan con las tablas de clasificación, las cuotas de publicación y las plataformas financiadas por el capital riesgo. El ejercicio arraiga la teoría moral en las ecologías institucionales específicas.

Utiliza una línea de tiempo con etiquetas de evidencia para que los estudiantes distingan entre acusaciones y pruebas. Solicita un memorándum de “bifurcación en el camino” que describe dos acciones plausibles en cada punto de decisión y los riesgos para las diferentes partes interesadas. Añade una hipótesis contraria: ¿qué pasaría si la regla de clasificación o la política de la revista fueran diferentes? Esto traza una línea entre la política y el comportamiento. Puntúa el análisis posterior sobre la claridad con la que se relaciona la causa con el efecto: cuadros claros, vínculos comprobables. Deduce puntos por teatralidad. Termina con un resumen de la política en el que se nombra una solución estructural y sus efectos no deseados.



JUEGO DE ROLES CON MÚLTIPLES PERSPECTIVAS

- *Método:* a los equipos se les asignan roles de partes interesadas (estudiante de posgrado, editor de revista, administrador universitario, ejecutivo de una empresa emergente de IA, denunciante) en seminarios. Se enfrentan a un dilema, como por ejemplo retirar un artículo basado en datos sintéticos.
- *Resultado del aprendizaje:* el juego de roles hace que los participantes asuman contrapuestos imperativos (supervivencia profesional, deber fiduciario, confianza pública), lo que agudizará tanto la imaginación narrativa como la algorítmica. Los estudiantes identifican cómo los resultados éticos dependen no solo de la buena voluntad, sino también de las limitaciones técnicas y las asimetrías de poder.

110



Entrega a cada participante un informe sellado: datos privados, restricciones y un mandato. A mitad del proceso, envía por correo electrónico una “nueva normativa” o un “ultimátum” disruptivo que reescribe las reglas del juego. Resultado final: un documento de decisión conjunta, un párrafo de disconformidad obligatorio que inmortaliza la voz minoritaria en el archivo oficial. Evalúa los resultados por la calidad de la justificación y la transparencia del proceso de toma de decisiones, y penaliza el comportamiento performativo que carece de una responsabilidad clara (participantes, plazos, compromisos medibles). Termina con una auditoría posterior al papel: los participantes salen de su personaje y califican la decisión de su equipo en cuanto a equidad y viabilidad.

ARQUEOLOGÍA DIGITAL

- *Método:* los estudiantes “excavan en las fuentes” de las herramientas sencillas del trabajo académico cotidiano: realizan ingeniería inversa de la ponderación de un detector de plagio, rastrean el linaje del capital riesgo de una empresa de índices de citas y trazan un mapa del sesgo geográfico de un algoritmo de búsqueda.
- *Resultado del aprendizaje:* esta investigación descubre las historias sedimentadas, los intereses económicos y las decisiones políticas escritas en un *software* aparentemente neutral. Ayuda a los estudiantes a aprender a cuestionar los ajustes predeterminados y a reconocer los bucles de retroalimentación ocultos para que puedan evitar construir temas que conduzcan a la explotación y la desigualdad.

Empareja cada hallazgo con una prueba reproducible. Por ejemplo, pasa textos multilingües por el detector e informa de las tasas de falsos positivos por el registro. Crea un “expediente de herramientas” que incluya la propiedad, las fuentes de datos, los problemas conocidos y los costes para el usuario. Exige una estrategia de salida: un cambio de configuración, un protocolo de mitigación o una alternativa viable con las ventajas y desventajas claramente expuestas. Evalúa la calidad y la practicidad de las pruebas, aunque el riesgo es el cinismo. Mide las mejoras tras aplicar la mitigación y reflexiona sobre lo que quedó sin resolver.

En conjunto, estas prácticas llevan la educación ética más allá de la memorización de reglas hacia una alfabetización incorporada en cómo las historias y los sistemas coproducen la vida académica. Al combinar la imaginación narrativa de Nussbaum con una sólida imaginación algorítmica, las universidades pueden preparar a académicos que no solo empatizan con las diferencias humanas, sino que también examinan y remodelan las arquitecturas tecnológicas que cada vez más gobiernan esas diferencias.

La ética de la virtud para la era algorítmica

En *After Virtue* (1981), MacIntyre lamenta el colapso de las narrativas morales compartidas y propone un retorno a la visión aristotélica en la que las comunidades estables cultivan hábitos que forman el carácter. Traducir ese programa a una universidad digitalizada significa nombrar las excelencias que necesita un académico y luego crear rutinas que permitan que esas excelencias echen raíces. Hay cuatro que destacan:

PRUDENCIA DIGITAL (PHRONESIS)

- *Definición:* la sabiduría práctica para percibir cómo los sistemas de gestión del aprendizaje, los paneles de citas o la IA generativa afectarán a la integridad del trabajo y las relaciones.
- *Ejemplo:* optar por ralentizar el ciclo de entrega de las tareas cuando una nueva actualización del modelo de lenguaje hace que los “ensayos automáticos” sean sencillos, o decidir publicar el código y los datos para que la automatización no afecte a la reproducibilidad.

La prudencia depende de la situación, así que debe enseñarse como puntos de control temporales: antes del diseño (para anticipar los efectos de las herramientas), a mitad del semestre (para auditar los atajos no deseados) y después (para registrar lo que hay que cambiar el próximo trimestre).

Acompaña cada adopción de tecnología con una “barrera de seguridad” (por ejemplo, defensas orales, cuadernos de código, declaraciones de procedencia). Evalúa la prudencia mediante pruebas de previsión (registros de riesgos, planes de contingencia), no sólo por los resultados. El modo de fallo es la reacción exagerada (prohibiciones generales). En su lugar: permite el uso asistido con límites declarados y los pasos de verificación.

MODERACIÓN ALGORÍTMICA

- *Definición:* una cautela disciplinada hacia las seducciones de los números (factores de impacto, donuts Altmetric, hitos del Índice h) que pueden secuestrar los motivos académicos.
- *Ejemplo:* establecer límites departamentales al número de artículos que se tienen en cuenta para las revisiones anuales o valorar deliberadamente la tutoría y la divulgación comunitaria junto con el recuento de citas.

112



La templanza necesita pasos estructurales, no charlas motivadoras. Convierte los valores en rúbricas: establece pesos máximos para cualquier métrica individual; exige una “narrativa de la contribución” y un crédito por “replicación/resultados negativos”. Realiza pruebas de sensibilidad con sus propios criterios de promoción para mostrar cómo los pequeños cambios en los pesos distorsionan el comportamiento. Lleva un “libro de contabilidad paralelo” del trabajo no contabilizado (tutoría, curación de conjuntos de datos) y los incluye en las evaluaciones anuales. El problema es la sustitución de métricas (nuevas insignias por antiguas). Mantiene cualquier indicador cuantitativo como explícitamente provisional.

CIBERVALENTÍA

- *Definición:* la fortaleza para denunciar prácticas cuestionables (fábricas de artículos, revisiones escritas por terceros, bots de autores falsos), incluso cuando dicha crítica amenaza las clasificaciones, las subvenciones o la buena voluntad colegiada.
- *Ejemplo:* los investigadores Jr. que firman una carta colectiva solicitando una investigación sobre las imágenes sospechosas de un colega senior, o un editor que rechaza las presentaciones por vía rápida que eluden la revisión por pares adecuada.

El valor es proporcional a la protección. Crea canales seguros (defensor del pueblo confidencial para la integridad de la investigación, política contra las represalias, sesiones informativas con asesores lega-

les). Normaliza las plantillas de notificación basadas en pruebas (comprobación de imágenes, datos PRISMA, cronología). Realiza “ejercicios de valentía”: ofrece a los jóvenes un conjunto de datos depurado con una trampa ética y cinco minutos para que den su opinión, mientras que los adultos responden: “Buena observación, ¿qué hacemos ahora?”. Advierte en las evaluaciones de rendimiento, pero las acompaña de vías claras de debido proceso para que nadie se convierta en policía de Twitter. Cuando la protección está codificada, la disidencia deja de parecer heroica y empieza a parecer normal.

TECNOLOGÍA Y JUSTICIA

- *Definición:* compromiso permanente de utilizar herramientas digitales de forma que amplíen, en lugar de reducir el acceso y el reconocimiento.
- *Ejemplo:* utilizar *software* de código abierto y servidores de preimpresión para nivelar las disparidades de recursos o auditar los algoritmos de detección de plagio en busca de sesgos contra los escritores no nativos de inglés.

Trata a las herramientas como decisiones políticas. Realiza evaluaciones de impacto en la equidad antes de su adopción (costes, ancho de banda, efectos lingüísticos, accesibilidad). Mantiene un “presupuesto de inclusión”: dedica fondos/tiempo a la traducción, el acceso con pocos recursos y la tecnología de asistencia. Exige a los proveedores que revelen toda la información (publicando el origen completo de la formación y las tasas de error de los subgrupos) o de lo contrario pierde su licencia para vender. Elimina cualquier modelo que no supera una auditoría de sesgos. Evalúa la justicia en función de quién se beneficia y quién queda marginado, mediante el seguimiento a través de paneles de distribución y vídeos de cuentas reales, y no por el número de clics que recoge el panel. El riesgo es la apertura performativa; vincula la apertura a la administración real: mantenedores, documentación y gobernanza de la comunidad.

MacIntyre insiste en que estas virtudes no se pueden memorizar como cláusulas de política; deben “habituarse” mediante acciones repetidas dentro de prácticas que tengan bienes internos y normas comunitarias claras. En términos concretos:

- *Pedagogía del estudio.* Incorporar a los estudiantes en colectivos de investigación donde la transparencia de los datos y la revisión del código sean rituales diarios, no eventos especiales.

- *Rotación de funciones de administración.* Asignar a los miembros del laboratorio la gestión del repositorio compartido, el registro de preinscripción o el canal de retroalimentación entre pares, de modo que la prudencia y la justicia se conviertan en memoria muscular.
- *Sesiones informativas reflexivas.* Después de cada ciclo de publicación, celebrar una reunión post publicación en la que se pregunte qué atajos digitales se rechazaron, cuáles se aceptaron y cómo se formó el carácter del grupo en el proceso.
- *Barreras institucionales.* Elaborar criterios de promoción que recompensen la publicación de resultados negativos o datos de replicación tanto como el número de citas llamativas, reforzando la moderación y la justicia tecnológica a nivel estructural.

114



Esto indica lo que MacIntyre (1981) denomina una concepción de la virtud “dependiente de la práctica”: los académicos aprenden a desear lo correcto repitiendo las acciones correctas dentro de las comunidades que honran esos esfuerzos. Solo entonces la prudencia digital, la templanza algorítmica, el cibercoraje y la tecnojusticia pueden consolidarse en la gramática moral compartida de la que tan a menudo carece la academia moderna.

Hacia una nueva pedagogía de la verdad

En esta sección se esboza un plan de estudios de alfabetización digital impregnado de ética que integra la competencia técnica con la reflexión moral. Se proponen tres líneas interrelacionadas —estudios críticos de código, talleres de ética de datos y pedagogía de plataformas— para que los estudiantes aprendan a leer el código como argumento, prevean los daños en las prácticas de datos y critiquen la lógica comercial de los sistemas de clasificación y recomendación. Arraigado en la pedagogía crítica freireana, el modelo convierte el diálogo y la praxis en cuestiones de aula que rediseñan las métricas de evaluación y desafían las configuraciones predeterminadas de las plataformas. Decantándose, finalmente, en una nueva *reflexividad algorítmica*.

Integración de la ética y la alfabetización digital

Los cursos convencionales de alfabetización digital rara vez van más allá de los tutoriales prácticos: cómo formatear una hoja de cálculo, verificar una URL o cambiar la configuración de privacidad. Ese enfoque centrado en las habilidades deja sin respuesta la pregunta más importan-

te: ¿qué tipo de relaciones humanas, dinámicas de poder y consecuencias morales se esconden en las herramientas que estamos aprendiendo a dominar? Por lo tanto, proponemos un modelo ampliado y saturado de ética de la alfabetización digital que trata la competencia técnica y la reflexión moral como inseparables. El marco se basa en tres aspectos que se refuerzan mutuamente:

ESTUDIOS CRÍTICOS DEL CÓDIGO

- *Objetivo*: enseñar a los estudiantes a leer el código de la misma manera que los humanistas leen las novelas, buscando suposiciones, exclusiones y juicios de valor.
- *Práctica*: anotar el código fuente de un detector de plagio (o, si es de propiedad exclusiva, su solicitud de patente y sus libros blancos) para revelar cómo se seleccionan los umbrales de similitud, qué registros lingüísticos se penalizan y qué alusiones culturales se favorecen.
- *Resultado*: los estudiantes llegan a comprender que un algoritmo no es una máquina neutral, sino un texto polémico que puede —y, en algunos casos, debe— ser debatido, revisado o rechazado.

Comience con un solo ladrillo Lego auditable, un *tokenizador* de diez líneas o un *stub* de similitud coseno, y luego encaje el resto del castillo a su alrededor; el primer clic demostrable asegura a todos que la máquina más grande también puede abrirse y contarse. Exija “diarios de diferencias”: los estudiantes registran cómo un cambio de una línea (lista de palabras vacías, tamaño de n-gramas) altera los resultados en muestras multilingües. Esto expone rápidamente los sesgos ocultos. Error común: tratar la documentación como algo neutral. Haga que los estudiantes comparen los documentos con el comportamiento real mediante pruebas unitarias con cambio de código y ESL. Califique el trabajo en función de lo que se puede ejecutar y medir (pruebas unitarias, diferencias de errores, trazas de rendimiento), no en función de adjetivos abstractos. Comprobación de equidad: alimente el modelo con texto que nunca se haya acercado a una editorial de la Ivy League (foros médicos en swahili, tuits en quichua, revistas en tamil) para que el punto de referencia no sea solo el inglés.

TALLERES DE ÉTICA DE DATOS

- *Objetivo*: afrontar las repercusiones morales de la adquisición, limpieza, modelización y compartición de datos.

- *Práctica*: realizar una revisión simulada de la junta de revisión institucional (IRB) de un proyecto que recopila resúmenes de conferencias para el análisis de tendencias, exigiendo a los participantes que identifiquen los posibles daños, los sesgos ocultos y la reutilización incontrolable posterior.
- *Resultado*: los estudiantes desarrollan hábitos de reflexión anticipatoria y aprenden a preguntarse “¿Quién se beneficia? ¿Quién está expuesto al riesgo? ¿De quién falta la voz?” antes de recopilar un solo byte.

Deje atrás las listas de verificación de la IRB. Organice un “ensayo de daños” de cinco minutos: cada miembro del equipo escribe una frase de rechazo desde el punto de vista de las personas más propensas a sufrir daños (“Estimados profesores, rechazamos... porque...”). Cambie las batas de laboratorio por la experiencia vivida y convierta la lista de verificación en una conversación sobre el consentimiento. Exija un “manifiesto de procedencia” con cada conjunto de datos: origen, transformaciones, incertidumbres y reglas de acceso. La laguna habitual es la responsabilidad posterior a la publicación; añada una revisión periódica en la que los equipos revisen el conjunto de datos después de cuatro semanas y registren cualquier uso no previsto. Califique la claridad de los planes de mitigación, no solo la identificación de los riesgos. Mantenga un presupuesto de sesgos: por cada paso de limpieza, los estudiantes deben indicar qué señales se pierden y por qué esa pérdida es aceptable.

116



PEDAGOGÍA DE LA PLATAFORMA

- *Objetivo*: ilustrar cómo los sistemas de gestión del aprendizaje, los gestores de referencias, los motores de recomendación y los canales de redes sociales moldean lo que se considera un conocimiento legítimo.
- *Práctica*: trazar el modelo de ingresos de una base de datos de citas, examinar cómo su algoritmo de clasificación pondera los distintos campos y simular cómo los pequeños cambios en las ponderaciones redistribuyen la visibilidad académica.
- *Resultado*: los usuarios son capaces de reubicarse, pasando de ser consumidores pasivos de los resultados de la plataforma a negociadores activos —o críticos— del terreno epistémico que construyen esas plataformas.

Se deben tratar a las plataformas como instituciones con incentivos. Pida a los alumnos que esbochen una función objetivo sencilla (“maximizar el tiempo en la plataforma \times ingresos por APC”) y prueben los ajustes de política en función de ella. Un ajuste del 1-2 % en la “colaboración internacional” o la “prima de acceso abierto” a menudo cambia las clasificaciones, lo que hace tangible la fragilidad. Añada un ejercicio de “salida y voz”: proponga un cambio de configuración (voz) o una ruta de migración a una herramienta alternativa (salida), incluyendo los costes para los usuarios con pocos recursos. Trampa: moralizar sin ofrecer alternativas. Exija un prototipo funcional o una guía de configuración (por ejemplo, desactivar los “artículos relacionados” que favorecen a los mismos editores). Evalúe la viabilidad y el impacto en los usuarios, no solo la crítica.

Filosofía pedagógica

Esta tríada extiende la pedagogía crítica de Freire (1970), basada en el diálogo, la reflexividad y la acción transformadora al espacio digital. El concepto de concientización (conciencia crítica) de Freire se traduce aquí en “pedagogía digital crítica”, donde, primero, el diálogo significa no solo hablar con los compañeros, sino también “escuchar” lo que el código y los datos revelan sobre los imperativos institucionales; segundo, la reflexión se combina con la praxis: después de analizar un algoritmo de clasificación, los estudiantes pueden proponer métricas alternativas o redactar una carta abierta al proveedor; tercero, el empoderamiento pasa simplemente de operar la tecnología a darle forma, cuestionarla y rediseñarla al servicio de la producción democrática y equitativa del conocimiento.

Al incorporar la investigación ética en cada punto de contacto técnico, el plan de estudios revisado forma a graduados que pueden depurar tanto el *software* como los sesgos sociales que incorpora, abordar los conjuntos de datos con la misma seriedad moral que aplicarían a los sujetos humanos y reconocer las plataformas como actores políticos en lugar de infraestructura invisible.

En resumen, la pedagogía digital crítica convierte la alfabetización digital de una lista de competencias en una práctica sostenida de compromiso ético, intelectual y cívico con las tecnologías que cada vez más gobiernan la vida académica.

La reflexividad algorítmica

El siglo XXI exige una nueva forma de autoconciencia: la reflexividad algorítmica. Mientras que el pensamiento crítico tradicional cuestiona

los argumentos y las pruebas, esta reflexividad pide a los estudiantes que interroguen las corrientes computacionales que dirigen lo que ven, leen y valoran. Es la disciplina de detenerse para preguntarse: “¿Por qué este artículo, esta métrica, este anuncio es el que aparece ante mí, y qué me está enseñando silenciosamente a desear o ignorar?”. Para cultivar ese hábito mental, los cursos pueden incorporar tres ejercicios iterativos en sus planes de estudios:

AUDITORÍAS ALGORÍTMICAS

- *Tarea:* cada estudiante lleva un registro de dos semanas sobre cómo el autocompletado de los motores de búsqueda, las recomendaciones de los gestores de citas o las fuentes de las redes sociales dan forma a las fuentes que consultan. A continuación, realizan ingeniería inversa de las señales (lenguaje, geografía, redes de citas conjuntas) que el sistema utilizó para llegar a esas sugerencias.
- *Objetivo:* hacer visibles las revisiones bibliográficas, la selección de temas e incluso la formulación de las preguntas de investigación.

Los registros suelen revelar una pauta: los estudiantes convergen en los mismos artículos “seguros”, a menudo de fuentes bien indexadas, mientras que los trabajos no anglófonos o de campos afines desaparecen. La auditoría también pone de manifiesto cómo la formulación de una consulta en diferentes registros cambia lo que aparece. Se debe pedir a los estudiantes que realicen la misma consulta desde una IP del campus, una red móvil y una VPN para ver los efectos de la localización. La evaluación debe premiar la claridad de la inferencia, no el volumen de capturas de pantalla. El escollo es el fatalismo. Hay que concluir con una propuesta de diseño: “¿Cuáles ajustes o revelaciones harían que este *feed* fuera más justo?”. Eso convierte la crítica en capacidad.

PROYECTOS DE DECONSTRUCCIÓN MÉTRICA

- *Tarea:* trabajar en pequeños grupos en algo que se pueda medir o clasificar en un contexto académico (factores de impacto de revistas, insignias Altmetric, tablas de clasificación departamentales) y rastrear las entradas de datos, los modelos de ponderación y el modelo de propiedad/patrocinio. Demostrar cómo los pequeños cambios en los parámetros redistribuyen el prestigio y los recursos entre campos, regiones o cohortes demográficas.



- *Objetivo*: desmitificar la cuantificación; mostrar que las métricas no son hechos naturales, sino artefactos discutibles, y animar a los estudiantes a imaginar sustitutos más justos.

Las pruebas de sensibilidad son el punto de inflexión. Un cambio de un punto en la ponderación de la “colaboración internacional” a menudo reordena toda la tabla, lo que enseña a los estudiantes lo frágil que puede ser la “excelencia”. Exija dos entregables: un breve resumen que indique lo que la métrica pretende medir y una contramétrica que se ajuste a un valor que la clase nombre de antemano, como la tutoría o la transparencia de los datos. Los modos de fallo más comunes son tratar la propiedad y los incentivos como cuestiones secundarias. Conviértalos en elementos centrales: quién se beneficia cuando esta métrica aumenta y quién paga. Termine con una anotación pública en la página web de la métrica para practicar la crítica responsable.



PORTAFOLIOS DE ÉTICA DIGITAL

- *Tarea*: los estudiantes mantienen un “documento vivo” durante todo el semestre para documentar los dilemas éticos que encuentran en sus flujos de trabajo digitales: *scraping* (publicar datos sin un paso de ofuscación), utilizar IA generativa para contenido o copias, y citar recomendaciones de algoritmos. Añadir a cada declaración las normas pertinentes que siguieron al elegir sus soluciones, inspiradas en las directrices éticas existentes, junto con su propio veredicto y una evaluación posterior de las consecuencias.
- *Objetivo*: establecer una práctica longitudinal de reflexión que extraiga los conocimientos fragmentados para crear un código de trabajo personalizado para una conducta académica digital responsable.

Los portafolios funcionan cuando son específicos. Anime a los estudiantes a usar indicaciones, hash y notas de versión, no resúmenes. La entrada de seguimiento es crucial: “¿Qué cambió después de elegir X?”. Esa capa temporal convierte el seguimiento de las normas en juicio. Incluya al menos un ejercicio de “reversión” en el que el estudiante revise una decisión después de aprender una nueva norma y escriba lo que haría ahora de manera diferente. La calificación debe privilegiar la franqueza y el razonamiento relacionado con el curso por encima del comportamiento impecable. El riesgo es la confesión performativa. Contrarreste el escepticismo adjuntando recibos breves: bloquee su plan de búsqueda en

un registro público, enumere todas las indicaciones de IA en un apéndice o pegue el correo electrónico de consentimiento para ese conjunto de datos recopilados, pequeños actos verificables que convierten la transparencia de un eslogan en una nota al pie.

Conclusión

El revuelo por el plan de Pune-LinkedIn no es un rumor aislado, sino una señal que pone de manifiesto un cambio más profundo en el andamiaje moral de la academia, donde los avatares digitales, los coautores fantasma y algoritmos de KPI ahora coescriben los guiones por los que se rigen las universidades, los departamentos y los académicos individuales, y donde la prioridad no puede ser un juego perpetuo de atrapar a los tramposos, sino una reimaginación integral de la vida ética en una universidad saturada de datos. Si leemos todo el espectro, desde la advertencia de Arendt (1963) sobre la conformidad irreflexiva, pasando por el relato de MacIntyre sobre la virtud fragmentada, hasta el análisis de Zuboff (2019) sobre el control instrumental, se repite una lección: una reforma duradera debe abordar tanto la maquinaria que da forma al comportamiento como el carácter de quienes la manejan. El camino a seguir requiere valor institucional para romper la adicción a los marcadores y volver a basar la evaluación en la calidad, la apertura, la replicación y el valor público; innovación pedagógica que incorpore la ética en los estudios de diseño, las simulaciones de casos y las auditorías algorítmicas; sabiduría tecnológica que afronte la IA y la analítica con crítica disciplinada, exigencias de transparencia y alternativas abiertas; e imaginación moral que anime a los administradores, profesores, estudiantes y financiadores a imaginar universidades en las que la integridad supere a la eficiencia. El comentario: “¿Por qué no me sorprende que esto sea de la India?”, malinterpreta una patología global: la adoración de las métricas, el tráfico de las autorías y los atajos de la IA son síntomas de una condición digital universal. El llamamiento de Lévinas (1961) a la ética como primera filosofía nos recuerda que la responsabilidad hacia el otro debe preceder a los cálculos de beneficio o velocidad. Volver a centrarnos en ese principio es la forma de disipar los fantasmas digitales de hoy en día y elegir una academia en la que la tecnología sirva a nuestra vida moral compartida en lugar de suplantarla.

120



Notas

- 1 El Marco Nacional de Clasificación Institucional (NIRF) es una metodología adoptada por el Gobierno de la India para clasificar las instituciones de educación superior del país en función de diversos parámetros, entre los que se incluyen la enseñanza, el aprendizaje y los recursos; la investigación y las prácticas profesionales; la divulgación y la inclusividad; y la percepción.
- 2 Empresa global conocida por proporcionar información y análisis fiables para acelerar el ritmo de la innovación y por ofrecer diversos servicios a organizaciones como Web of Science, Journal Citation Reports o InCites.
- 3 Cargo por procesamiento de artículos (APC) es una tarifa que los autores suelen pagar a los editores por los servicios de procesamiento, edición y publicación que prestan a sus artículos académicos, especialmente en revistas de acceso abierto.
- 4 Se trata de editoriales que se especializan en la producción de revistas académicas, libros y otros contenidos dentro de los campos de la ciencia, la tecnología y la medicina.
- 5 Los valores separados por comas (CSV) son un formato de archivo sencillo que se utiliza para almacenar datos tabulares (hojas de cálculo o bases de datos). En un archivo CSV, cada línea representa un registro de datos y cada campo dentro de ese registro está separado por una coma.

Bibliografía

- ARENDDT, Hannah
 1963 *Eichmann in Jerusalem: A report on the banality of evil*. Viking Press.
- BANDURA, Albert
 2016 *Moral disengagement: How people do harm and live with themselves*. Worth Publishers.
- BAUDRILLARD, Jean
 1994 *Simulacra and simulation*. University of Michigan Press.
- BAUMAN, Zygmunt
 2000 *Liquid modernity*. Polity Press.
- BRAIDOTTI, Rosi
 2013 *The posthuman*. Polity Press.
- ELSE, Holly & VAN NOORDEN, Richard
 2021 The fight against fake-paper factories that churn out sham science. *Nature*, 591(7851), 516-519. <https://doi.org/10.1038/d41586-021-00733-5>
- FLORIDI, Luciano, COWLS, Josh, BELTRAMETTI, Monica, CHATILA, Raja, CHAZERAND, Patrice, DIGNUM, Virginia, LUETGE, Christoph, MADELIN, Robert, PAGALLO, Ugo, ROSSI, Francesca, SCHÄFER, Burkhard, VALCKE, Peggy & VAYENA, Effy
 2018 AI4People-An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- FONG, Eric & WILHITE, Allen
 2017 Authorship and citation manipulation in academic research. *PLoS One*, 12(12), e0187394. <https://doi.org/10.1371/journal.pone.0187394>
- FOUCAULT, Michel
 1977 *Discipline and punish: The birth of the prison*. Vintage Books.

- FRASER, Nancy
2008 *Scales of justice: Reimagining political space in a globalizing world*. Columbia University Press.
- FREIRE, Paulo
1970 *Pedagogy of the oppressed*. Continuum.
- HAN, Byung-Chul
2015 *The transparency society*. Stanford University Press.
- HAYLES, Katherine
1999 *How we Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. University of Chicago Press.
- HOSSEINI, Mohammad, HILHORST, Mirjam, DE BEAUFORT, Ineke & FANELLI, Daniele
2018 Doing the right thing: A qualitative investigation of retractions due to unintentional error. *Science and Engineering Ethics*, 24, 189-206. <https://doi.org/10.1007/s11948-017-9894-2>
- LATOUR, Bruno
2005 *Reassembling the social*. Oxford University Press.
- LÉVINAS, Emmanuel
1961 *Totality and infinity*. Duquesne University Press.
- LUND, Brady, WANG, Ting, MANNURU, N. R., NIE, Bin, SHIMRAY, S. & WANG, Zhe
2023 ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570-581. <https://doi.org/10.1002/asi.24750>
- MACINTYRE, Alasdair
1981 *After virtue*. University of Notre Dame Press.
- MERTON, Robert
1973 *The sociology of science: Theoretical and empirical investigations* (N. W. Storer, ed.). University of Chicago Press.
- MITTELSTADT, Brent Daniel, ALLO, Patrick, TADDEO, Mariarosaria, WACHTER, Sandra & FLORIDI, Luciano
2016 The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
- MULLER, Jerry
2018 *The tyranny of metrics*. Princeton University Press.
- NUSSBAUM, Martha
1997 *Cultivating humanity*. Harvard University Press.
- ORANSKY, Ivan
2023 “Truly devastating”: Four journals won’t get new Impact Factors this year because of citation shenanigans. *Retraction Watch*. <https://bit.ly/4iEutAD>
- ORTEGA Y GASSET, José
1930 *The revolt of the masses*. W. W. Norton.
- PASQUALE, Frank
2015 *The black box society*. Harvard University Press.
- RECIO SASTRE, Alejandro
2025 The Question about Ethics Teaching as Boundary of Artificial Intelligence. *Sophia, Colección de Filosofía de la Educación*, (39), 53-81. <https://doi.org/10.17163/soph.n39.2025.01>



RETRACTION WATCH

- 2023 *Hindawi retracts more than 300 papers.* <https://bit.ly/4p6LjdL>
- SCHNEIDER, Leonid
2021 The full-service paper mill and its Chinese customers. *For Better Science.* <https://bit.ly/4otcU86>
- TAYLOR, Charles
1991 *The ethics of authenticity.* Harvard University Press.
- VAN NOORDEN, Richard
2013 Brazilian citation scheme outed. *Nature*, 500(7464), 510-511. <https://doi.org/10.1038/500510a>
- ZUBOFF, Shoshana
2019 *The age of surveillance capitalism.* Public Affairs.

Declaración de disponibilidad de datos

Los comentarios en redes sociales a los que se hace referencia en este artículo se recopilieron de un hilo público de LinkedIn publicado el 9 de julio de 2025. La publicación original ha sido eliminada, pero el autor ha conservado capturas de pantalla del debate, que están disponibles previa solicitud con fines de verificación.

123


Declaración de Autoría - Taxonomía CRediT	
Autor/es	Contribuciones
Amitabh Vikram Dwivedi	Conceptualización, tratamiento de datos, etc.

Declaración de uso de inteligencia artificial
El autor Amitabh Vikram Dwivedi del artículo titulado: “Fantasmas digitales, algoritmos morales y el desafío de enseñar ética en la era poshumana”, DECLARA que la elaboración del documento se apoyó con Inteligencia Artificial (IA) en un porcentaje de 1 %.

Fecha de recepción: 13 de julio de 2025
 Fecha de revisión: 20 de septiembre de 2025
 Fecha de aprobación: 10 de noviembre de 2025
 Fecha de publicación: 15 de enero de 2026