REVISTA INTERNACIONAL TECNOLÓGICA - EDUCATIVA DOCENTES 2.0
ISSN: 2665-0266 | ISNI: 0000 0005 0409 1664 | RINGGOLD ID 608948 | ROR: 037JQRK16
VOL. 18 (2), NOVIEMBRE 2025 | https://doi.org/10.37843/rted.v18i2.724

RTED Resiliencia Paradigmática

Entre Códigos y Corazones: Investigación e Inteligencia Artificial hacia una Tecnología más Humana

Between Codes and Hearts: Research and Artificial Intelligence Towards a More Human Technology

Nepsi Beatriz Garcia-Heredia[1] y Ruth M. Mujica-Sequera[2]

EDICIÓN: RTED Resiliencia Paradigmática

Recibido: 2/julio/2025
Aceptado: 3/noviembre/2025
Publicado: 29/noviembre/2025

Páginas: desde 337-345

País

[1]Estados Unidos de América
[2]Estados Unidos de América

Institución

[1]Universidad Internacional de Valencia (VIU)
[2]Grupo Docentes 2.0 C.A.

Correo Eletrónico

[1]nepsigarcia@gmail.com
[2]ruth.mujica@docentes20.com

ORCID

[1]https://orcid.org/0009-0007-1404-0213
[2]https://orcid.org/0000-0002-2602-5199

Citar así: APA / IEEE

Garcia-Heredia, N. & Mujica-Sequera, R. (2025). Entre Códigos y Corazones: Investigación e Inteligencia Artificial hacia una Tecnología más Humana. *Revista Tecnológica-Educativa Docentes 2.0*, 18(2), 337-345. https://doi.org/10.37843/rted.v18i2.724

N. Garcia-Heredia y R. Mujica-Sequera, "Entre Códigos y Corazones: Investigación e Inteligencia Artificial hacia una Tecnología más Humana", RTED, vol. 18, n.°2, pp. 337-345, nov. 2025.

Resumen

La tecnología, más allá de su componente instrumental, debe incorporar dimensiones afectivas y éticas que refuercen la empatía junto con la solidaridad social. El objetivo del ensayo consistió en analizar cómo los procesos algorítmicos pueden articularse con las perspectivas éticas y afectivas propias de la experiencia humana. Para ello, el ensayo se enmarca en un paradigma humanista bajo un método inductivo, con enfoque cualitativo, de tipo interpretativo y de diseño narrativo de tópico. A lo largo del texto se reflexiona sobre las bases filosóficas de la inteligencia artificial (IA), sus aplicaciones en contextos de interacción social, así como sus implicaciones morales. Se examinan ejemplos concretos de sistemas inteligentes orientados al cuidado o la educación, contrastados con innovaciones de carácter estrictamente funcional. Asimismo, se plantea la necesidad de un enfoque interdisciplinar que integre perspectivas procedentes de la psicología, la sociología y la ciencia de datos. Finalmente, se concluye que solo mediante un diálogo permanente entre códigos y corazones podrá consolidarse un desarrollo tecnológico verdaderamente al servicio de la dignidad humana.

Palabras clave: Códigos, humanización, investigación, inteligencia artificial, tecnología.

Abstract

Technology, beyond its instrumental component, must incorporate affective and ethical dimensions that reinforce empathy along with social solidarity. The objective of this essay was to analyze how algorithmic processes can be articulated with the ethical and affective perspectives inherent to the human experience. To this end, the essay is framed within a humanist paradigm using an inductive method, with a qualitative, interpretive approach, and a narrative topic design. Throughout the text, the philosophical foundations of artificial intelligence (AI), its applications in contexts of social interaction, and its moral implications are reflected upon. Specific examples of intelligent systems oriented toward care or education are examined, contrasted with innovations of a strictly functional nature. The essay also raises the need for an interdisciplinary approach that integrates perspectives from psychology, sociology, and data science. Finally, it concludes that only through an ongoing dialogue between codes and hearts can technological development truly serve human dignity.

Keywords: Codes, humanization, research, artificial intelligence, technology.

D.L:LA2017000128 | ISSN: 2665-0266 | Ringgold ID:608948
ISNI: 0000 0005 0409 1664 | ROR: 037jqrk16
Noviembre 2025 - Volumen 18, Número 2, Ejemplar No. 41.

Introducción

La tecnología, más allá de su componente instrumental, debe incorporar dimensiones afectivas y éticas que refuercen la empatía y la solidaridad social. En este sentido, el presente estudio tiene como propósito analizar cómo los procesos algorítmicos pueden articularse con los aspectos morales y afectivos propios de la experiencia humana, demostrando que la humanización de las ciencias aplicadas es posible mediante un enfoque deliberadamente interdisciplinar. En este ensayo, los términos ética, afectividad y tecnología no se abordan como dimensiones aisladas, sino como componentes interdependientes de un mismo engranaje conceptual, la ética orienta el sentido del desarrollo tecnológico y la afectividad humaniza su aplicación social. La relevancia del análisis radica en la proliferación de sistemas inteligentes que operan sin considerar las repercusiones emocionales y éticas de sus decisiones automatizadas.

En un contexto global marcado por desigualdades crecientes y crisis de confianza en las instituciones, resulta imprescindible diseñar innovaciones tecnológicas que fomenten la inclusión, la responsabilidad y el respeto mutuo. Asimismo, la reciente regulación europea sobre ética en Inteligencia Artificial (IA) refuerza la necesidad de integrar criterios de responsabilidad social en los procesos de innovación técnica, razón por la cual este ensayo aporta a la discusión contemporánea sobre la gobernanza de la inteligencia artificial. En este marco, la dimensión lingüística de la IA adquiere una relevancia ética y política, pues el lenguaje no solo trasfiere información, sino que también reproduce o transforma los marcos ideológicos que configuran la comunicación humana.

Desde esta perspectiva, se delimita el alcance del trabajo al análisis teórico y narrativo de casos representativos de IA humanizada, sin incluir experimentos empíricos ni métricas cuantitativas. Se excluyen las descripciones técnicas detalladas sobre arquitecturas de redes neuronales, privilegiando en cambio las implicaciones filosóficas y sociológicas de su uso. Esta delimitación permite enfocar la reflexión en la dimensión cultural, ética y afectiva de la tecnología, evitando dispersiones hacia aspectos estrictamente ingenieriles.

En cuanto al proceso metodológico, la investigación se sustenta en el método inductivo, que parte de ejemplos concretos para abstraer principios teóricos. Bajo un paradigma humanista y un enfoque cualitativo, se analizan narrativas y prácticas sociales vinculadas a la IA, con el fin de esclarecer los significados y valores que orientan su desarrollo. Finalmente, el diseño narrativo temático organiza el ensayo en torno a tres ejes fundamentales: ética, empatía y gobernanza tecnológica, los cuales estructuran el hilo argumental que da coherencia a toda la investigación.

Desarrollo

En la sección de desarrollo se abordarán de forma secuencial los seis ejes temáticos que estructuran la interacción entre IA y sociedad. El primero examina la transparencia y la explicabilidad algorítmica, con especial atención a las técnicas que permiten desvelar la “caja negra” de los modelos. El segundo analiza los sesgos y la equidad en los sistemas de IA, considerando sus orígenes, las métricas de evaluación y las estrategias de mitigación. El tercer apartado se dedica a la privacidad y la protección de datos, mediante una revisión de principios, metodologías y tecnologías de salvaguarda. A continuación, se estudia la colaboración humana–máquina a partir de sinergias y modelos de interacción sustentados en casos de uso de distintos ámbitos. El quinto eje examina el sesgo lingüístico y la reproducción de estructuras de poder perpetuadas por el discurso. Por último, el sexto desarrolla la educación y la alfabetización en IA, presentando estrategias curriculares y formativas. Cada subapartado combina fundamentación teórica, ejemplos concretos y referencias actualizadas; esta integración permite articular un discurso coherente con un constante componente crítico.

Transparencia y Explicabilidad Algorítmica

Las técnicas de interpretabilidad pueden clasificarse en métodos globales y locales según el alcance de sus explicaciones. Los primeros ofrecen una visión panorámica del modelo, mientras que los segundos se enfocan en decisiones particulares. Herramientas como LIME y SHAP propuestas respectivamente por Ribeiro et al. (2016) y Lundberg & Lee (2017), emplean teorías de atribución para cuantificar la contribución de cada variable, generando explicaciones visuales o textuales comprensibles para el usuario. En sectores regulados, como el financiero o el sanitario, estos mecanismos resultan notables para garantizar la transparencia y la rendición de cuentas (EU AI Act, 2024). En última instancia, la elección del método depende del tipo de modelo y del grado de precisión requerido, pero todos convergen en un mismo propósito, hacer comprensible la lógica interna de la IA a quienes la utilizan.

La transparencia fortalece la confianza de los usuarios, pues ofrece claridad sobre los procesos de decisión automatizada. Adams (2023) sostienen que

esta apertura no solo cumple una función informativa, sino también ética, al permitir que el individuo comprenda los criterios que orientan las resoluciones de la IA. A su vez, los estudios de Ananny & Crawford (2018) demuestran que las explicaciones claras reducen la resistencia al cambio tecnológico y fomentan la aceptación pública. Las organizaciones que promueven la apertura de sus algoritmos reportan menores índices de litigios y fortalecen su reputación (Wachter & Mittelstadt, 2019). En consecuencia, la transparencia deja de ser un requisito técnico para convertirse en un principio de legitimidad social. No obstante, la búsqueda de explicabilidad enfrenta diversos desafíos técnicos y prácticos.

Los modelos de aprendizaje profundo son estructuras altamente complejas, con millones de parámetros interconectados, lo que dificulta extraer interpretaciones fieles sin sacrificar precisión (Lipton, 2016). Además, algunas técnicas introducen simplificaciones que distorsionan el proceso real de decisión, generando "explicaciones pos hoc" que justifican el resultado sin reflejar la ruta inferencial auténtica (Ananny & Crawford, 2018). Esta tensión se agrava cuando las exigencias de transparencia colisionan con la confidencialidad industrial o con la protección de la propiedad intelectual, un dilema aún no resuelto en la regulación internacional.

En áreas críticas como la salud y las finanzas, la transparencia adquiere una dimensión ética ineludible. En medicina, los sistemas de diagnóstico asistido deben justificar sus recomendaciones clínicas para que el personal sanitario pueda validarlas (Cheong, 2024). Del mismo modo, los modelos de crédito requieren trazabilidad de decisiones para demostrar imparcialidad y evitar discriminación algorítmica. Estas prácticas no solo garantizan cumplimiento normativo, sino que fortalecen la responsabilidad social y la confianza institucional.

De cara al futuro, la explicabilidad algorítmica evolucionará hacia enfoques más interactivos y participativos. Se prevé la aparición de interfaces conversacionales capaces de responder a las preguntas del usuario sobre el comportamiento del modelo (Ananny & Crawford, 2018). En paralelo, se consolidarán estándares internacionales para armonizar los criterios de apertura, como la verificación algorítmica (ISO/IEC JTC 1/SC 42., 2025). Por su parte, la formación ética de los profesionales en ciencia de datos será determinante para integrar la transparencia en todo el ciclo de vida de la IA, garantizando confianza social y desarrollo sostenible.

Sesgos y Equidad en los Sistemas de IA

Los sesgos en los sistemas de IA surgen de las desigualdades estructurales presentes en los datos y en los procesos de diseño. Barocas & Selbst (2016) advierten que los modelos algorítmicos, lejos de ser neutrales, tienden a reproducir las asimetrías históricas de las sociedades que los originan. De manera complementaria, Mehrabi et al. (2021) distinguen tres tipos de sesgos: el de representación, asociado con la infrarrepresentación de ciertos grupos; el de medición, vinculado con la definición inexacta de variables; y el algorítmico, producto del diseño mismo de la arquitectura del modelo. Comprender estas diferencias resulta decisivo para implementar estrategias de mitigación y promover un uso ético de la IA. Este reconocimiento implica asumir que la objetividad técnica es una ilusión, cada sistema refleja decisiones humanas que impactan directamente en la justicia social.

La equidad, entendida como la ausencia de discriminación sistemática entre colectivos, constituye un principio ético y operativo esencial. Desde una perspectiva computacional, Hardt et al. (2016) proponen la noción de igualdad de oportunidades, que busca garantizar resultados similares para individuos con condiciones equivalentes. Otros enfoques, como la paridad demográfica descrita por Feldman et al. (2015), persiguen que los resultados favorables se distribuyan equitativamente entre grupos protegidos. Sin embargo, ambas métricas no siempre pueden cumplirse simultáneamente, lo que obliga a reflexionar sobre los dilemas éticos que emergen al priorizar ciertos criterios sobre otros. Así, la decisión de qué entender por "equidad" debe contextualizarse en cada aplicación y responder no solo a parámetros técnicos, sino también a valores organizacionales y sociales.

La detección temprana de sesgos representa un desafío constante. Requiere combinar técnicas automáticas de auditoría con juicios humanos informados por conocimiento contextual. Las herramientas de revisión algorítmica ayudan a identificar disparidades en métricas de equidad, pero la interpretación de los resultados exige prudencia y una comprensión profunda de los entornos donde se implementan los modelos. En este sentido, la participación de expertos en derechos humanos, ética y ciencias sociales, e incluso de comunidades afectadas, amplía la capacidad de las organizaciones para evaluar los impactos de la automatización. La equidad no puede delegarse a la máquina, necesita deliberación humana, transparencia institucional y rendición de cuentas.

Para mitigar los sesgos, se aplican estrategias en distintas fases del ciclo de desarrollo. En el preprocesamiento, se ajustan los datos para equilibrar

la representación de grupos minoritarios. Durante el entrenamiento, se incorporan penalizaciones que reducen disparidades, y en el postprocesamiento se calibran los umbrales de decisión para corregir resultados injustos. Estas medidas, aunque efectivas, suelen implicar una tensión entre precisión y justicia, lo que demanda evaluar el impacto social de cada ajuste. La trazabilidad de las decisiones, junto con la documentación de las intervenciones realizadas, constituye una herramienta fundamental para garantizar la responsabilidad y fortalecer la confianza pública en los sistemas inteligentes.

Desde el plano normativo, la Comisión Europea (2016) y el Reglamento General de Protección de Datos (UE, 2016/679, Art. 22) establecen que ninguna persona debe ser objeto de decisiones automatizadas que produzcan efectos significativos sin supervisión humana. Del mismo modo, el IEEE (2020) promueve estándares para una "IA responsable", basada en principios de equidad, transparencia y rendición de cuentas. Estas directrices convergen en la necesidad de marcos regulatorios globales que armonicen las políticas públicas, eviten vacíos legales y refuercen la confianza de la ciudadanía en la tecnología. La equidad, más que un requisito técnico, se configura como una condición ética de legitimidad.

Casos emblemáticos han mostrado las consecuencias de ignorar esta dimensión. En el ámbito financiero, sistemas de crédito automatizados han negado préstamos a comunidades vulnerables sin criterios verificables (Barocas & Selbst, 2016). En salud, Obermeyer et al. (2019) demostraron cómo modelos de predicción médica subestimaban la gravedad de pacientes pertenecientes a minorías raciales debido a datos sesgados. Estos ejemplos evidencian que la inequidad algorítmica tiene efectos tangibles en la vida humana. Por ello, las futuras líneas de investigación deben avanzar hacia métricas longitudinales que evalúen el impacto social de los sistemas automatizados. Solo una colaboración sostenida entre academia, industria y sociedad civil permitirá construir una IA verdaderamente justa, inclusiva y orientada al bien común.

Privacidad y Protección de Datos

La privacidad constituye un derecho fundamental que resguarda la información personal frente a usos indebidos y prácticas de vigilancia masiva (Reglamento UE 2016/679). Cada dato recolectado, desde hábitos de navegación hasta registros médicos, contiene trazos de la identidad humana y, por tanto, demanda un tratamiento ético. La protección de datos debe garantizar su confidencialidad, integridad y disponibilidad a lo largo de todo su ciclo de vida (ISO/IEC 27001, 2013). Estos principios exigen limitar la recolección a lo estrictamente necesario, solicitar consentimiento informado y asegurar la trazabilidad del tratamiento. Tales disposiciones, comunes al RGPD europeo y a la CCPA estadounidense, subrayan que el control sobre la información es una extensión de la autonomía personal.

El diseño de sistemas de IA respetuosos de la privacidad incorpora las estrategias de privacidad por diseño y por defecto, formuladas por Cavoukian (2011). Integrar salvaguardas desde la fase inicial evita la recopilación innecesaria de información y restringe el acceso a los actores indispensables (Reglamento UE 2016/679, Art. 25). A ello se suma la necesidad de realizar análisis de impacto de privacidad que anticipen y mitiguen riesgos. Estas medidas no deben entenderse como simples requisitos técnicos, sino como prácticas de gobernanza ética que refuerzan la confianza pública.

Entre las innovaciones recientes destaca el aprendizaje federado, que permite entrenar modelos sin concentrar los datos sensibles en un servidor central (McMahan et al., 2017). Este enfoque reduce la exposición de la información y ha mostrado eficacia en sectores como la salud. Sin embargo, los avances en técnicas de desanonimización evidencian que ningún método es infalible; por ello, deben complementarse con mecanismos de privacidad diferencial y protocolos criptográficos avanzados. Del mismo modo, el cifrado homomórfico descrito por Gentry (2009) posibilita realizar operaciones sobre datos cifrados, preservando su confidencialidad incluso durante el procesamiento. Aunque su complejidad limita su adopción, representa un horizonte de equilibrio entre utilidad y seguridad.

La gestión de identidades y accesos continúa siendo fundamental para prevenir intrusiones. Los modelos de autenticación multifactor y las políticas basadas en roles garantizan que solo usuarios autorizados manipulen información crítica. El paradigma Zero Trust, planteado por Rose et al. (2020), redefine la seguridad bajo la premisa de que ningún actor es confiable por defecto, reforzando así la resiliencia ante amenazas. Estas estrategias técnicas deben acompañarse de una vigilancia constante mediante auditorías, registros y revisiones periódicas. No obstante, la tecnología por sí sola no basta.

La protección de datos exige una cultura institucional y ciudadana de responsabilidad digital. Organizaciones como la Cloud Security Alliance (2021) recomiendan la formación continua de usuarios y desarrolladores en buenas prácticas de privacidad. La alfabetización en derechos digitales, sumada a la cooperación entre ingenieros, juristas y

expertos en ética, consolida una visión integral de la seguridad informacional. Solo así la IA podrá operar en armonía con los valores humanos, garantizando que el progreso tecnológico no se traduzca en pérdida de dignidad ni de autonomía individual.

Colaboración Humana–Máquina

La colaboración entre humanos y máquinas se fundamenta en la sinergia entre la capacidad cognitiva y emocional de las personas y el poder de procesamiento de la inteligencia artificial. En este paradigma, los sistemas automatizados actúan como aliados cognitivos más que como sustitutos. Según Parasuraman et al. (2000), el equilibrio entre automatización y control humano permite que las personas se concentren en tareas creativas y estratégicas mientras delegan las operaciones repetitivas. Esta cooperación exige interfaces intuitivas y explicables que promuevan la confianza mutua y reduzcan la carga cognitiva. La ergonomía digital y el diseño centrado en el ser humano, defendidos por Norman (2013), son pilares para garantizar una experiencia colaborativa inclusiva, accesible y emocionalmente satisfactoria.

En los entornos de trabajo, los sistemas inteligentes impulsan nuevas formas de colaboración al personalizar procesos y flujos de tareas. La inteligencia colectiva surge cuando los algoritmos detectan patrones de sinergia entre los aportes humanos y los integran en tiempo real (Agrawal et al., 2018). En este escenario, la máquina funciona como mediadora de conocimiento, mientras el ser humano aporta juicio ético, creatividad y sentido contextual. Esta interacción redefine el concepto de trabajo, desplazando el valor desde la ejecución hacia la interpretación y la innovación.

En el ámbito científico, la IA acelera la generación de conocimiento al identificar correlaciones invisibles al análisis humano. Shneiderman (2020) propone que esta colaboración no debe aspirar a reemplazar la investigación, sino a expandir sus fronteras mediante un modelo de "IA centrada en el bienestar humano". Las máquinas procesan grandes volúmenes de datos, pero la validación y la interpretación crítica siguen siendo responsabilidad del investigador. De este modo, la inteligencia humana y la artificial coevolucionan en un circuito de retroalimentación continua.

En sectores sensibles como la salud, esta relación cobra un sentido ético decisivo. Los sistemas de diagnóstico basados en aprendizaje profundo han demostrado precisión comparable a la de los especialistas (Esteva et al., 2017), pero la decisión final siempre recae en el juicio clínico humano. Topol (2019) advierte que la verdadera revolución no radica en automatizar la medicina, sino en humanizarla mediante herramientas que amplíen la empatía y la capacidad predictiva del profesional. La colaboración entre médico y máquina fortalece el diagnóstico, reduce errores y optimiza el tiempo de atención, pero también exige una regulación clara de responsabilidades y una formación digital permanente del personal sanitario.

La co-creación asistida por IA está transformando el diseño y la innovación. Algoritmos generativos exploran miles de soluciones funcionales y estéticas en segundos (Bentley, 1999), mientras los diseñadores humanos aportan sensibilidad cultural, criterio ético y visión social. Esta interacción democratiza la creatividad, reduce los costos de desarrollo y promueve la reutilización del conocimiento gracias a comunidades abiertas y licencias colaborativas (Raymond, 1999). En este entorno, la ética del diseño se convierte en un componente esencial de la responsabilidad tecnológica.

De cara al futuro, la colaboración humana–máquina evolucionará hacia esquemas más horizontales, distribuidos y participativos. Tecnologías como los gemelos digitales y la computación en el borde permitirán interacciones más inmediatas y privadas, mientras las interfaces neuronales abrirán nuevas formas de comunicación directa entre mente y máquina. Tales avances, no obstante, plantean desafíos sobre autonomía, identidad y consentimiento. La UNESCO (2021) enfatiza que la gobernanza ética de la IA debe involucrar a la sociedad civil, el sector productivo y las instituciones públicas en un marco de corresponsabilidad global. Solo un enfoque interdisciplinar y humanista garantizará que la IA complemente la inteligencia humana sin sustituir su esencia.

Discurso, Relaciones de Poder y Estereotipos.

El lenguaje no es un instrumento neutral, sino un tejido simbólico donde se inscriben las tensiones sociales, culturales y políticas de cada época. Desde la perspectiva de Volóshinov (1976), todo signo lingüístico es también un signo ideológico: un vehículo que refleja y modela la conciencia colectiva. Esta visión invita a comprender que cada palabra pronunciada, o generada por un sistema de IA, participa en la construcción de sentidos y reproduce las jerarquías del mundo social. Así, el discurso automatizado no es ajeno a la política del lenguaje; por el contrario, la incorpora y la amplifica.

El carácter dialógico del lenguaje, planteado por Bajtín (1982), refuerza esta idea. Toda expresión se enuncia en respuesta a otras voces, reales o

anticipadas, y produce efectos sobre quienes la reciben. Aplicado a la IA, este principio implica que los sistemas generativos también se insertan en un diálogo social más amplio, cargado de valores y supuestos culturales. La pregunta ya no es solo cómo las máquinas procesan el lenguaje, sino qué tipo de voces y perspectivas están autorizadas o silenciadas dentro de ese proceso comunicativo.

Desde el análisis crítico del discurso, van Dijk (2000) advierte que el lenguaje es uno de los principales medios por los cuales se ejercen y legitiman las relaciones de poder. En consecuencia, los modelos de lenguaje entrenados con grandes corpus de datos reproducen inevitablemente las desigualdades y estereotipos presentes en las sociedades que los producen. Los sesgos de género, raza o clase no son simples fallos técnicos, sino expresiones estructurales de un orden simbólico que la IA puede perpetuar si no se diseñan mecanismos de corrección y supervisión ética.

Esta preocupación es compartida por Dignum (2019), quien subraya el riesgo de delegar decisiones comunicativas a algoritmos carentes de conciencia contextual o responsabilidad moral. La objetividad aparente de los sistemas automatizados puede generar una falsa legitimidad discursiva, otorgando autoridad a mensajes que reproducen prejuicios o exclusiones. Por ello, el diseño de modelos lingüísticos debe integrar principios de rendición de cuentas y justicia comunicativa, más allá de la eficiencia técnica.

En la era de la economía de la atención, Couldry & Mejías (2019) evidencian cómo las plataformas digitales transforman el discurso en una forma de control simbólico y vigilancia social. La IA, al gestionar y clasificar los flujos de información, moldea la esfera pública y redefine los límites de la autonomía individual. Este fenómeno exige una alfabetización crítica que permita reconocer los sesgos discursivos y reconfigurar la relación entre palabra, poder y tecnología.

En definitiva, el lenguaje generado por IA no solo transmite información, produce sentido, establece jerarquías y refleja los valores de una sociedad. Su análisis requiere un enfoque interdisciplinar que combine la lingüística, la ética y la tecnología. Solo así será posible garantizar que los sistemas de IA no reproduzcan los estereotipos y las exclusiones del pasado, sino que contribuyan a construir un diálogo más justo, plural y verdaderamente humano.

Educación y Alfabetización en IA

La alfabetización en IA constituye uno de los pilares de la formación ciudadana contemporánea. Comprender los principios que rigen los algoritmos y sus lógicas de decisión permite a los individuos analizar críticamente las tecnologías que median su vida cotidiana. Este conocimiento favorece la apropiación consciente de los sistemas digitales y el fortalecimiento del pensamiento computacional, entendido como una forma de razonamiento estructurado orientada a la resolución de problemas complejos. La educación en IA, por tanto, no se limita a la dimensión técnica, sino que se proyecta como un ejercicio de emancipación cognitiva y ética que habilita la participación informada en la esfera pública (Unesco, 2021).

La incorporación de la IA en los planes de estudio exige un rediseño curricular sustentado en la interdisciplinariedad y en la integración de saberes. En los niveles iniciales, la enseñanza puede articularse mediante proyectos lúdicos que vinculen las ciencias, las humanidades y las artes, mientras que en la educación superior resulta pertinente entrelazar las competencias técnicas con la reflexión filosófica, económica y política sobre el impacto social de la automatización. Este enfoque formativo propicia una comprensión holística del fenómeno tecnológico y promueve la corresponsabilidad ética de los futuros profesionales (Luckin et al., 2016; Unesco, 2021). Más allá de la instrucción instrumental, la educación en IA debe cultivar la capacidad de interpretar críticamente las implicaciones epistemológicas y morales del conocimiento digital.

El rol del docente adquiere una dimensión estratégica en este proceso de transformación pedagógica. Los educadores, en tanto mediadores culturales, deben actualizar sus competencias tecnológicas y didácticas mediante programas de desarrollo profesional sostenidos en la colaboración con la industria y con comunidades académicas globales (Unesco, 2021). La creación de materiales educativos abiertos, la participación en redes de aprendizaje y la experimentación con metodologías activas consolidan una pedagogía dialógica que reconoce al estudiante como protagonista de su propio proceso cognitivo. De esta manera, la docencia se redefine como un espacio de coaprendizaje y creación colectiva.

La dimensión ética, a su vez, ha de integrarse transversalmente en todo proceso de alfabetización digital. Los estudiantes deben comprender las consecuencias sociales de los sesgos algorítmicos, las tensiones entre privacidad y transparencia, y los dilemas que plantea la responsabilidad en el uso de sistemas inteligentes (Floridi, 2013). Debates guiados, estudios de caso y simulaciones de escenarios morales favorecen el desarrollo del juicio ético y la empatía digital. Así, la alfabetización en IA se transforma en una práctica de ciudadanía reflexiva

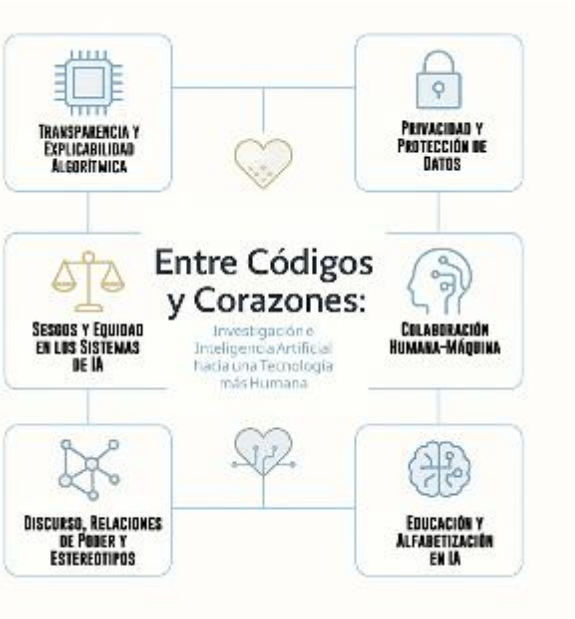
orientada a la justicia tecnológica y a la preservación de la dignidad humana.

Los espacios de aprendizaje no formal complementan y expanden las oportunidades educativas en IA. *Hackathon*s, laboratorios ciudadanos y comunidades *makers* promueven la experimentación colaborativa y la creación de soluciones con impacto social (Henderikx et al., 2017). Las plataformas abiertas de aprendizaje, junto con los cursos masivos en línea, democratizan el acceso global a contenidos especializados, mientras que los ecosistemas de código abierto potencian la co-creación del conocimiento y la cultura del intercambio solidario (Raymond, 1999). Estas prácticas encarnan una pedagogía de la participación, en la que aprender implica también contribuir al bien común digital.

De cara al porvenir, la educación en IA tenderá hacia modelos personalizados, inclusivos y adaptativos sustentados en la propia tecnología. Las redes educativas internacionales favorecerán el intercambio multicultural de recursos y metodologías (Unesco, 2021). Los Estados deberán garantizar políticas que aseguren igualdad en el acceso y reconocimiento de competencias, mientras que las universidades habrán de asumir la investigación pedagógica sobre la enseñanza de la IA como un campo estratégico para el desarrollo humano. En última instancia, la alfabetización en IA se erige como una vía para reconfigurar la relación entre conocimiento, ética y sociedad, preparando a las nuevas generaciones para habitar con lucidez y responsabilidad la era algorítmica.

La integración de los seis ejes abordados: transparencia, equidad, privacidad, colaboración, discurso y educación, configura una visión holística de la IA como fenómeno ético, social y tecnológico. Estos componentes, lejos de ser ámbitos aislados, conforman un entramado de relaciones interdependientes donde la técnica requiere de la sensibilidad humana para orientarse hacia el bien común. En esta convergencia entre razón y emoción, el conocimiento científico se funde con la responsabilidad moral, dando origen a una nueva epistemología de la IA que no solo busca optimizar procesos, sino también fortalecer los lazos de empatía, justicia y solidaridad. Así, el siguiente Figura 1 sintetiza gráficamente el itinerario conceptual que articula la propuesta de este ensayo.

Figura 1. *Entre Códigos y Corazones: Investigación e Inteligencia Artificial hacia una Tecnología más Humana.*



Nota. Síntesis conceptual de los ejes que articulan la relación entre investigación e inteligencia artificial hacia una tecnología más humana, elaboración propia (2025).

Conclusión

El ensayo partió de la premisa de que la tecnología debe integrar dimensiones tanto éticas como afectivas, también examinó cómo los procesos algorítmicos pueden articularse con la experiencia humana. A lo largo del desarrollo se abordaron seis ejes: transparencia y explicabilidad, sesgos y equidad, privacidad y protección de datos, colaboración humana–máquina, discurso y poder, hasta culminar con educación y alfabetización en IA. Cada apartado combinó fundamentos teóricos y ejemplos concretos, en coherencia con el enfoque humanista, inductivo e interpretativo declarado en la introducción. Desde ese punto de partida, el análisis mostró lo siguiente: a) la transparencia facilita la detección de errores y sesgos; b) la equidad demanda métricas claras y documentación de decisiones; c) la privacidad requiere principios de minimización, técnicas criptográficas y gobernanza de datos; d) la colaboración humano–máquina depende de interfaces comprensibles y distribución responsable de tareas; e) el lenguaje generado por IA reproduce marcos ideológicos presentes en los datos de entrenamiento; y f) la alfabetización en IA necesita integración curricular interdisciplinaria y dimensión ética transversal.

Dentro de los límites autoimpuestos, el texto cumplió con el objetivo descrito: analizar la relación entre códigos y corazones mediante una mirada interdisciplinar y crítica. Con ello, el presente trabajo

investigativo se inscribe en la categoría deductiva: la conclusión reafirma, a la luz de lo expuesto, la tesis inicial sobre la posibilidad y necesidad de una tecnología orientada a la dignidad humana, sin introducir nuevas proposiciones. Asimismo, se han sustentado argumentos que validan la tesis en cuanto a la integración del componente afectivo como una condición indispensable para garantizar interacciones significativas y socialmente responsables. Este planteamiento cobra relevancia en un contexto donde la automatización tiende a desplazar criterios éticos por métricas de eficiencia. En definitiva, reconocer la necesidad de incluir una dimensión humanista abre la puerta a modelos tecnológicos que optimicen procesos, fortalezcan valores democráticos y consoliden vínculos comunitarios.

Referencias

Adams J. (2023). Defending explicability as a principle for the ethics of artificial intelligence in medicine. *Med Health Care Philos*, 26(4), 615-623. DOI: 10.1007/s11019-023-10175-7.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence.* Harvard Business Review Press.

Ananny, M., & Crawford, K. (2018). *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20*(2), 13–27. https://doi.org/10.1177/1461444816676645

Bajtín, M. M. (1982). *The dialogic imagination: Four essays* (M. Holquist, Ed.; C. Emerson & M. Holquist, Trans.). University of Texas Press.

Barocas, S., & Selbst, A. D. (2016). *Big data's disparate impact. California Law Review, 104*(3), 671–732. https://dx.doi.org/10.2139/ssrn.2477899

Bentley, P. J. (1999). *Evolutionary Design by Computers.* Morgan Kaufmann.

Cavoukian, A. (2011). *Privacy by Design: The 7 Foundational Principles.* Information and Privacy Commissioner of Ontario.

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, *6*, 1421273. https://doi.org/10.3389/fhumd.2024.1421273

Cloud Security Alliance. (2021). *Security Guidance for Critical Areas of Focus in Cloud Computing v4.0.*

Couldry, N., & Mejías, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism.* Stanford University Press.

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way.* Springer Nature.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). *Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542*(7639), 115–118. https://doi.org/10.1038/nature21056

European Commission. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union,* 1–88. https://n9.cl/bkdv7

European Union. (2024). *Artificial Intelligence Act (EU Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence and amending certain Union legislative acts).* Official Journal of the European Union. https://n9.cl/k1ka6

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and removing disparate impact. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 259–268. https://doi.org/10.1145/2783258.2783311

Floridi, L. (2013). *The ethics of information.* Oxford University Press.

Gentry, C. (2009). *A fully homomorphic encryption scheme*. Stanford University.

Hardt, M., Price, E., & Srebro, N. (2016). *Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29*, 3315–3323. https://doi.org/10.48550/arXiv.1610.02413

Henderikx, P., Kreijns, K., & Kalz, M. (2017). *Refining success and dropout in massive open online courses based on the intention–behavior gap. Distance Education, 38*(3), 353–368. https://doi.org/10.1080/01587919.2017.1369006

IEEE. (2020). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems* (2nd ed.). IEEE.

ISO/IEC 27001:2013. (2013). *Information technology — Security techniques — Information security management systems — Requirements.* International Organization for Standardization.

ISO/IEC JTC 1/SC 42. (2025). *Information technology. Artificial intelligence (AI) AI system impact assessment (ISO/IEC 42005:2025)*. International Organization for Standardization.

Lipton, Z. C. (2016). *The Mythos of Model Interpretability*. *arXiv.* https://doi.org/10.48550/arXiv.1606.03490

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education.* Pearson Education.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 54, 1273–1282. https://n9.cl/3eh3z

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). *A survey on bias and fairness in machine learning. ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Norman, D. A. (2013). *The Design of Everyday Things* (Revised and expanded ed.). Basic Books.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. DOI: 10.1126/science.aax234

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). *A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part A, 30*(3), 286–297. DOI:10.1109/3468.844354

Raymond, E. S. (1999). *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary.* O'Reilly Media.

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). *Zero Trust Architecture.* NIST Special Publication 800-207.

Shneiderman, B. (2020). *Bridging the gap between ethics and practice: Guidelines for human-centered AI. IEEE Computer, 53*(10), 83–90. https://doi.org/10.1145/3419764

Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.* Basic Books.

Unesco. (2021). *Recommendation on the Ethics of Artificial Intelligence.* UNESCO Publishing.

van Dijk, T. A. (2000). *El discurso como interacción social.* Gedisa.

Volóshinov, V. N. (1976). *Marxism and the philosophy of language* (L. Matejka & I. R. Titunik, Trans.). Harvard University Press.

Wachter, S., & Mittelstadt, B. D. (2019). *A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI*. Columbia Business Law Review, 2019(2), 494–620. https://n9.cl/9ihwj0