VARII CIVES

Ética y desafíos de la inteligencia artificial en la sociedad contemporánea

Ana María Ayala Román y Henry Arley Táquez Quenguan (Editores)

Universidad Icesi

Editorial Universidad Icesi

Ética y desafíos de la inteligencia artificial en la sociedad contemporánea

Ana María Ayala Román y
Henry Arley Táquez Quenguan
(Editores)

Universidad ICESI

Ética y desafíos de la inteligencia artificial en la sociedad contemporánea
© Ana María Ayala Román y Henry Arley Táquez Quenguan (editores académicos) y varios autores.

Cali, Universidad Icesi, 2025
230 pp., 14x21,5 cm
ISBN 978-628-7814-19-6
DOI https://doi.org/10.18046/EUI/vc.8.2025

Palabras clave: 1. Ética | 2. Inteligencia artificial | 3. Sociedad | 4. Filosofía ética y moral | 5. Ética profesional

Incluye referencias bibliográficas

Código dewey: 323.6
Primera edición / Octubre de 2025

© Universidad Icesi

Rector
Esteban Piedrahita Uribe

Director académico
José Hernando Bahamón Lozano

Coordinador editorial
Adolfo A. Abadía

Editorial Universidad Icesi
Calle 18 No. 122-135 (Pance), Cali – Colombia
Teléfono: +57(2) 555 2334 etx. 8365
E-mail: editorial@icesi.edu.co
http://www.icesi.edu.co/editorial

Diseño de colección
Ladelasvioletas

Diseño y diagramación
Editoria Icesi

Publicado en Colombia – *Published in Colombia*

La publicación de este libro se aprobó luego de superar un proceso de evaluación doble ciego por pares expertos.

La Editoral Universidad Icesi no se hace responsable de las ideas expuestas bajo su nombre, las ideas publicadas, los modelos teóricos expuestos o los nombres aludidos por los(as) autores(as). El contenido publicado es responsabilidad exclusiva de los(as) autores(as), no refleja la opinión de las directivas, el pensamiento institucional de la Universidad editora, ni genera responsabilidad frente a terceros en caso de omisiones o errores.

El material de esta publicación puede ser reproducido sin autorización, siempre y cuando se cite el título, el autor y la fuente institucional.

Índice

Presentación 5

Prológo. Educar con inteligencia (artificial) y sentido humano en tiempos de transformación 11
Andrés Chiappe Laverde

[Parte 1] Verdad, subjetividad y experiencia humana frente a la inteligencia artificial

[Cap. 1] Automatización del saber y erosión democrática: desafíos éticos en torno a la verdad en la era de la inteligencia artificial 17
Ana María Ayala Román

[Cap. 2] De Skynet a OpenAI: Del Imaginario Apocalíptico a los Desafíos de la Gobernanza y la Agencia Humana en la Inteligencia Artificial Real 43
Juan Felipe Castaño Quintero

[Cap. 3] Redes sociales e inteligencia artificial: subjetividades, vínculos y algoritmos en la era digital 75
Edilberto Loaiza Corrales

[Parte 2] Protección de datos personales y ética organizacional en la era digital

[Cap. 4] Privacidad, seguridad de los datos y el manejo ético del Big Data 103
Juan Fernando Arenas Jaramillo

[Cap. 5] Ética e Inteligencia Artificial: acciones empresariales sobre el derecho a la intimidad 143
Diego Hernán Varón Rojas

[Parte 3] Ética de la inteligencia artificial en la educación superior

[Cap. 6] Marcos institucionales para el uso ético de la Inteligencia Artificial Generativa en las instituciones de educación superior 175
Henry Arley Taquez Quenguan y María Isabel González Serrano

[Cap. 7] ¿Cómo abordar las implicaciones éticas en el uso de la IA?: una reflexión pedagógica 209
Isabel Tejada Sánchez

Sobre los autores 225

Presentación

¿Es posible pensar éticamente en un mundo atravesado por tecnologías que aprenden, predicen, clasifican y, en ocasiones, deciden por nosotros? Esta pregunta, lejos de tener una única respuesta, es el punto de partida de este libro, concebido como un espacio plural de reflexión crítica frente a los desafíos que plantea la inteligencia artificial generativa en nuestras sociedades. La ética, entendida como una disposición reflexiva que nos interpela sobre el modo en que vivimos y convivimos, no puede mantenerse al margen de las transformaciones profundas que estas tecnologías están generando en el trabajo, la educación, la salud, la política, la economía o las relaciones sociales. Este libro parte del reconocimiento de dicha complejidad, y de la convicción de que no hay respuestas simples cuando se trata de pensar la tecnología desde una perspectiva ética. La creación colectiva de este libro constituye, por ello, una apuesta ética y académica por comprender, desde una mirada interdisciplinar, los desafíos que plantea la Inteligencia Artificial Generativa (IAG) en la educación superior. Como instituciones comprometidas con la formación y la producción de conocimiento, reconocemos que atravesamos un momento histórico en el que la tecnología transforma, de manera acelerada, la vida cotidiana, la experiencia humana, el trabajo y las formas de conocer y crear cultura.

Este volumen es el resultado de un esfuerzo deliberado por parte de actores vinculados a instituciones de educación superior de la ciudad y del país, quienes, desde distintas trayectorias, proponen

una postura crítica y situada frente a la IAG. Surge como respuesta a una inquietud académica compartida: explorar las implicaciones humanas, sociales, jurídicas y políticas de estas tecnologías emergentes. Cada autor, desde su experiencia y formación, contribuye con una mirada particular que, en conjunto, busca aportar conocimiento para afrontar la incertidumbre del presente y los desafíos del porvenir en la educación y la academia. En este sentido, en cuento a la construcción académica, este libro se desarrolló bajo un enfoque metodológico de carácter reflexivo y crítico, propio de la tradición humanística y de las ciencias sociales. Cada capítulo adopta un método ensayístico-argumentativo, sustentado en revisión documental, análisis conceptual y estudio de casos cuando aplica. La obra es el resultado de la pluralidad de marcos teóricos y perspectivas disciplinares que permiten iluminar problemas éticos complejos desde diferentes ángulos. En tal sentido, La diversidad de voces, disciplinas y trayectorias reunidas en esta obra evidencia que los dilemas éticos no deben resolverse con premura, paranoia o protagonismos individuales, sino desde una reflexión crítica, pausada y colectiva. Esta perspectiva contrasta con la lógica de inmediatez que caracteriza tanto la producción de contenidos mediante sistemas de IAG como el ritmo vertiginoso de sus innovaciones. En un contexto donde los datos, estudios y tendencias tecnológicas se vuelven obsoletos con rapidez, este libro busca ofrecer reflexiones con profundidad crítica y vigencia, que movilicen procesos de reflexión pedagógica y ética, tanto individuales como colectivos.

Desde esta premisa, la construcción del libro se sostiene en la convicción de que no existe una única perspectiva legítima. Se requiere un diálogo riguroso, abierto y plural que articule teoría y práctica, reflexión y acción. En este contexto, los textos que integran este libro surgen como tentativas diversas de pensar lo ético en tiempos de algoritmos. Lo hacen desde preguntas que interpelan nuestras formas de conocer, enseñar, convivir y decidir, sin pretensión de respuestas definitivas, pero sí con ánimo de reflexión compartida.

¿Qué significa habitar una época en la que la IAG impone certezas sin juicio humano? El capítulo *Automatización del saber y erosión democrática: desafíos éticos en torno a la verdad en la era de la inteligencia artificial*, escrito por Ana María Ayala Román, examina cómo los sistemas algorítmicos redefinen la noción de verdad al automatizar el juicio y sustituir la deliberación crítica. Desde un enfoque filosófico apoyado en Foucault, Bergson, Sadin, Byung-Chul, entre otros, la autora argumenta que la IAG transforma el régimen epistémico moderno, desplazando el conocimiento interpretativo por una exactitud operativa. Frente a ello, propone recuperar la incertidumbre, el juicio crítico y la construcción colectiva del sentido como condiciones esenciales para sostener la vida democrática.

¿Cómo pensar la gobernanza de la IAG sin reducir la agencia humana al control de los datos? En *De Skynet a OpenAI: del imaginario apocalíptico a los desafíos de la gobernanza y la agencia humana en la inteligencia artificial real*, Juan Felipe Castaño Quintero analiza críticamente los discursos alarmistas sobre la IAG desde una perspectiva CTS. A partir del estudio de marcos regulatorios como los de UNESCO, OCDE y políticas públicas colombianas, argumenta que la gobernanza debe reconocer la agencia humana en la apropiación y resignificación de estas tecnologías. Propone una agenda ética que desplace el fetichismo tecnológico y recupere la dimensión política y cultural de los sujetos.

¿Qué tipo de lazo social es posible cuando identidades, emociones y comunidades están mediadas por algoritmos? En *Redes sociales e inteligencia artificial: subjetividades, vínculos y algoritmos en la era digital*, Edilberto Loaiza Corrales examina cómo la IA redefine las formas de subjetividad y convivencia en plataformas que actúan como infraestructuras invisibles de lo social. Desde una mirada crítica al capitalismo de vigilancia, la lógica de la visibilidad y la personalización algorítmica, el autor argumenta que las redes sociales ya no son simples medios, sino territorios que moldean lo común. Concluye con una invitación a imaginar comunidades digitales basadas en el disenso, la pluralidad y una ética del encuentro.

¿Cómo garantizar la privacidad, la seguridad y el uso ético de los datos en la era del Big Data y la IAG? El capítulo *Privacidad, seguridad de los datos y el manejo ético del Big Data*, de Juan Fernando Arenas Jaramillo, explora las tensiones entre innovación tecnológica y derechos fundamentales. A partir de casos como Cambridge Analytica y Clearview AI, se muestran los riesgos de la IAG en términos de vigilancia, sesgos y erosión del consentimiento informado. El autor propone una gobernanza ética basada en principios de transparencia, autonomía y equidad, así como marcos normativos robustos y prácticas técnicas responsables. Concluye que una ciudadanía digital crítica y una regulación proactiva son claves para orientar estas tecnologías hacia la justicia social.

¿Qué hacen las empresas de IAG para proteger el derecho a la intimidad? En *Ética e Inteligencia Artificial: acciones empresariales sobre el derecho a la intimidad*, Diego Hernán Varón Rojas examina el rol ético de las corporaciones frente a los riesgos de la IAG. A través del análisis de políticas empresariales de IBM, Intel y Tencent, expone las tensiones entre innovación, regulación y derechos. Inspirado en una ética fenomenológica basada en Hans Jonas, el autor señala que muchas compañías priorizan sus intereses económicos por encima del respeto a la intimidad. Concluye que, más allá de la normativa estatal, se necesita liderazgo corporativo ético, culturas organizacionales responsables y una ciudadanía vigilante que exija transparencia en la era algorítmica.

¿Qué marcos éticos deben construir las universidades para guiar el uso de la IAG en la formación profesional? El capítulo *Marcos institucionales para el uso ético de la Inteligencia Artificial Generativa en las instituciones de educación superior*, escrito por Henry Arley Taquez Quenguan y María Isabel González Serrano, analiza los dilemas que emergen con la integración de la IAG en la educación superior. A partir de una revisión de políticas nacionales e internacionales y del estudio de caso de la Universidad Icesi, los autores sostienen que regular no basta: es necesario fomentar una cultura pedagógica que fortalezca el juicio ético, la autorregulación y la agencia compartida. Plantean

que los marcos deben ser flexibles, construidos colectivamente y orientados a formar profesionales capaces de convivir críticamente con las tecnologías emergentes.

¿Cómo integrar una reflexión ética situada sobre la IAG en el aula? El capítulo *¿Cómo abordar las implicaciones éticas en el uso de la IA?: una reflexión pedagógica*, de Isabel Tejada Sánchez, propone una pedagogía crítica basada en el juego didáctico DilemAI. Esta herramienta permite explorar dilemas éticos reales mediante una experiencia lúdica, reflexiva y colaborativa. La autora sostiene que alfabetizar en IAG requiere ir más allá de lo técnico, abordando temas como el sesgo, la desinformación y la pérdida de agencia. Concluye que enseñar ética en tiempos de IAG implica un currículo transversal, docentes críticamente formados y estrategias que conecten con las emociones, los contextos y la vida cotidiana.

Desde este horizonte reflexivo, este libro invita a docentes, estudiantes, directivos y comunidades académicas a una lectura crítica, rigurosa y pedagógica que reconozca en la IAG un cambio profundo en las preguntas, los vínculos y los sentidos de la educación. Cada capítulo propone detenerse, pensar y actuar con juicio ético y compromiso formativo.

Los editores

[Prólogo]

Educar con inteligencia (artificial) y sentido humano en tiempos de transformación

Transformar la educación no es un capricho, ni una consigna de moda. Tampoco es un acto de ruptura sin fundamentos, ni el eco de una ideología sin causa. Es, más bien, una urgencia histórica. Décadas de una educación estandarizada —marcada por lógicas de control, homogeneización y eficiencia industrial— están mostrando sus límites ante un mundo crecientemente cambiante, incierto y ambiguo. Este desfase entre un modelo educativo heredado del pasado y las exigencias del presente interpela a quienes creemos que educar no es simplemente transferir contenidos, sino acompañar a las personas en la construcción de sentido, juicio ético y agencia crítica.

En este contexto, la aparición de la Inteligencia Artificial Generativa (IAG) no solo intensifica los desafíos de la educación, sino que abre posibilidades inéditas para su transformación. Pero esto solo será posible si, en lugar de adoptarla de forma acrítica, somos capaces de situarla en el corazón mismo de nuestras preguntas educativas: ¿qué significa aprender cuando el saber se automatiza?, ¿cómo formar en un mundo donde los sistemas inteligentes imponen certezas sin mediación humana?, ¿cuál es el papel de la escuela, la uni-

versidad y el docente en una época donde el conocimiento circula más allá de los espacios tradicionales?

El libro *Ética y desafíos de la inteligencia artificial en la sociedad contemporánea* ofrece una respuesta coral y rigurosa a estas preguntas. Lejos de presentar la IAG como una amenaza o una solución mágica, esta obra propone un abordaje crítico, interdisciplinar y comprometido con la educación como bien público y como práctica transformadora.

A lo largo de sus capítulos, se revela con claridad que las preguntas fundamentales de la educación ya no pueden ser respondidas desde el paradigma del control, la uniformidad y la certeza. En su lugar, es necesario abrazar la complejidad de un presente en el que la verdad ya no se construye desde la deliberación humana, sino que empieza a ser determinada por sistemas algorítmicos que producen certezas sin juicio. Frente a ello, los autores del libro no solo diagnostican los riesgos de esta automatización epistémica, sino que también identifican rutas de transformación pedagógica, institucional y ética, en las que la IAG puede ser aliada, siempre que sea situada, apropiada y resignificada desde fines humanos.

Una de las contribuciones más potentes de esta obra es mostrar cómo la educación no puede seguir funcionando bajo el supuesto de que los estudiantes aprenden en condiciones estables, predecibles y homogéneas. En un entorno donde la información circula de manera fragmentada, la subjetividad se configura en interacción con algoritmos, y la agencia humana corre el riesgo de diluirse en recomendaciones automáticas, se requiere un nuevo pacto educativo: uno que privilegie el juicio ético por encima del rendimiento, la formación integral por encima del entrenamiento técnico, y la experiencia situada por encima del estándar abstracto.

Desde esta perspectiva, la IAG no debe entenderse como una amenaza externa que llega a desestabilizar el sistema educativo, sino como un detonante que nos obliga a revisar lo que entendemos por conocimiento, evaluación, enseñanza y comunidad. La obra insiste en que la

transformación educativa no puede limitarse a la integración instrumental de tecnologías, sino que debe comprometer el sentido mismo de lo que significa formar sujetos capaces de actuar, comprender y decidir en escenarios inéditos. Así, se destacan propuestas que van desde el desarrollo de pedagogías críticas apoyadas en el juego y la simulación ética, hasta la necesidad de marcos institucionales flexibles que fomenten la autorregulación, la deliberación y la cultura digital crítica.

El libro también llama la atención sobre el lugar que ocupan los datos y su gestión en la construcción de la realidad educativa. En tiempos de big data y vigilancia inteligente, formar para la transformación exige desarrollar nuevas formas de alfabetización que no solo enseñen a usar tecnologías, sino que capaciten a los estudiantes para comprender cómo son interpretados, clasificados y orientados por ellas. La ciudadanía digital que este libro propone es activa, crítica y colectiva: no se conforma con exigir transparencia, sino que defiende el derecho a comprender y disputar los marcos que configuran lo posible.

En ese mismo horizonte, se plantea la necesidad de reconocer el papel de las comunidades académicas como espacios de resistencia, imaginación y co-construcción de sentido. La educación no debe convertirse en un canal de implementación silenciosa de soluciones tecnológicas, sino en un campo abierto al disenso, donde las múltiples formas de saber, sentir y vivir puedan dialogar y confrontarse. Así, el libro articula preocupaciones filosóficas, regulatorias, pedagógicas e institucionales que, al ser integradas, revelan una misma convicción: la transformación educativa que requiere nuestro tiempo no podrá hacerse sin ética, sin crítica, ni sin comunidad.

Este libro es, en suma, una brújula para pensar la educación en la era algorítmica. En lugar de sumarse a la lógica de la inmediatez o a las promesas redentoras del mercado tecnológico, sus autores invitan a detenerse, observar, reflexionar y actuar con juicio. La IAG, bajo esta mirada, puede ser una tecnología de transformación si nos permite recuperar el valor de la pregunta, la dignidad de la decisión compartida y la potencia pedagógica de lo inacabado.

Celebro la publicación de esta obra, no solo por la lucidez con la que interpela los desafíos del presente, sino porque nos recuerda que educar —en su sentido más profundo— es siempre un acto de servicio. Un servicio comprometido con el otro, con su libertad de pensar y su derecho a imaginar futuros posibles. Educar es servir desde el conocimiento, desde la escucha, desde la duda. Y si la IAG va a ser parte de la educación del futuro, deberá serlo bajo esa misma vocación: la de estar al servicio de una humanidad más justa, más libre y más consciente.

Andrés Chiappe Laverde
Director del Doctorado en Innovación Educativa con uso de TIC
Universidad de La Sabana, Colombia
16 de junio de 2025

[Parte 1]

Verdad, subjetividad y experiencia humana frente a la inteligencia artificial

[Capítulo 1]

Automatización del saber y erosión democrática: desafíos éticos en torno a la verdad en la era de la inteligencia artificial

Ana María Ayala Román

La oscilación del pensamiento humano entre la confianza en verdades absolutas (sean estas religiosas, metafísicas o científicas) y la incertidumbre nos lleva a preguntarnos por nuestra relación con la verdad. Podríamos esgrimir una historia de esta relación desde el pensamiento antiguo hasta nuestros días, pero no es nuestro objetivo hacer este camino a través de la historia de las ideas, sino acercarnos al problema de la redefinición de dicha relación en tiempos de la inteligencia artificial.

Ahora bien, hablar de la verdad es hablar también de estructuras de poder, esto es, este concepto no es ni universal ni neutro. Según Foucault (2019), cada sociedad tiene un *régimen de verdad* que se constituye por un conjunto de reglas que determinan qué discursos son considerados verdaderos, quién tiene la autoridad de enunciarlos y cómo circulan en la sociedad. En este sentido, la verdad no es la coherencia entre un enunciado y la realidad a la que describe, sino el resultado de relaciones de poder que regulan la producción, la validación y

la difusión de la verdad. Cabe señalar que desde la perspectiva foucaultiana, el poder no puede reducirse a un mecanismo represivo, sino como algo que produce discursos, saberes y subjetividades. En su entrevista con Alessandro Fontana y Pasqueale Pasquino en 1976, Foucault ejemplifica cómo el saber psiquiátrico ha sido utilizado para legitimar prácticas (encierro de ciertos sujetos) y la normalización de algunas conductas que se constituye como formas de control y verdad. El poder entonces no solo puede entenderse como un poder soberano, emanado de una autoridad central, sino como algo que se ejerce a través de dispositivos de saber y normatividad que se integran a la vida cotidiana (lo que Foucault llama *biopoder*), permitiendo un gobierno no solo coercitivo, sino que se da a través de la normalización y autorregulación de los individuos.

En *El orden del discurso* (2005), Foucault sostiene que cada época establece sus propias reglas para definir qué se considera verdadero, además de quién tiene la autoridad para producir la verdad.

> *[…] en toda sociedad la producción del discurso está a la vez controlada, seleccionada y redistribuida por cierto número de procedimientos que tienen por función conjurar sus poderes y peligros, dominar el acontecimiento aleatorio y esquivar su pesada y temible materialidad* (p. 14).

Así las cosas, un régimen de verdad se configura desde un conjunto de reglas, discursos e instituciones que determinan las condiciones bajo las cuales algo puede ser reconocido como verdadero en una *sociedad determinada*. Esto implica, por tanto, que las reglas no son estáticas ni universales, sino que cambian a través de la historia.

> *[…] en los poetas griegos del siglo VI, el discurso verdadero -en el más intenso y valorado sentido de la palabra-, el discurso verdadero por el cual se tenía respeto y terror, aquel al que era necesario someterse porque reinaba, era el discurso pronunciado por quien tenía el derecho y según el ritual requerido; era el discurso que decidía la justicia y atribuía a cada uno su parte; era el discurso que, profetizando el porvenir, no sólo anunciaba lo que iba a pasar, sino que*

> *contribuía a su realización, arrastraba consigo la adhesión de los hombres y se engarzaba así con el destino* (pp. 19-20).

El sabio o el sacerdote están asociados con la verdad, así como la autoridad de la Iglesia lo estaba en la Edad Media y el método científico en la Modernidad (pp. 30-31). Por ejemplo, el régimen de verdad de las sociedades modernas tiene las siguientes características: i) se centra en el discurso científico y en las instituciones que lo producen. ii) Es impulsado por intereses políticos y económicos, ya que responde a necesidades del Estado y el mercado. iii) Se difunde masivamente por medio de instituciones como universidades, medios de comunicación, etc. iv) Es controlado por grandes aparatos políticos y económicos que establecen la legitimidad de ciertos discursos y la exclusión de otros. v) Las verdades dominantes son constantemente desafiadas por contra-discursos que buscan poner en cuestión su legitimidad, así como las relaciones de poder que las sustentan (Foucault, 2019, pp. 42-43).

En la Modernidad Clásica (s. XVII), René Descartes representó una ruptura con la concepción de verdad anclada a la autoridad, tanto religiosa como filosófica. La búsqueda de un fundamento indudable, es decir, la certeza como criterio de verdad es el proyecto epistemológico cartesiano: hallar algo que no pueda ser puesto en duda.

> *Ya hace algunos años que he tomado conciencia de la gran cantidad de cosas falsas que, con el correr del tiempo, he admitido como verdaderas, así como lo dudoso que es todo lo que sobre ellas construir posteriormente, y que, por lo tanto, había que derribar todo ello desde sus raíces una vez en la vida, y comenzar de nuevo desde los primeros fundamentos, si deseaba alguna vez establecer algo firme y permanente en las ciencias* (Descartes, 2010, p. 69).

El método le permite a Descartes construir la certeza, ya que la verdad no deviene del orden del cosmos o la naturaleza. La ruptura es entonces con respecto a la concepción teleológica y comprensiva del orden. A diferencia de la tradición aristotélica y escolástica, donde el

conocimiento se funda en una realidad dotada de un sentido inherente, la modernidad cartesiana introduce una separación entre el sujeto y el mundo, por lo que este último se convierte en un objeto de cálculo y representación. La ruptura implica una crisis ontológica, puesto que el ser humano ya no descubre verdades *en* y *sobre* un cosmos ordenado, sino que debe descubrirlas sobre su propia racionalidad. No obstante, esta "autonomía" epistémica no libera al sujeto, sino que lo arroja y confronta con un mundo cada vez más ininteligible. La certeza en este panorama se convierte en una necesidad angustiante.

> *Realidad derivada y verdad derivada, mundo escindido y pensamiento de la separación, representación de un mundo extraño y angustiante, imperiosa necesidad de evidencia que le garantiza al hombre una seguridad ontológica, la de ser parte del mundo -y no solamente de participar del mundo-, todo ello merece el calificativo de* subjetividad moderna, *en tanto que a partir del momento en que el hombre deja de identificar la realidad verdadera del mundo, es decir, la verdad del mundo, con la experiencia inmediata que hace del mundo, es decir, la verdad del hombre, el hombre pasa de ser el destinatario de la verdad a ser el sujeto desde donde la verdad se origina* (Margot, 2018, p. 15)

De ahí que sea necesario partir de la duda metódica como primer paso para alcanzar una verdad indubitable. Este paso epistémico lleva a Descartes a concederle al *cogito* –y, por tanto, a la autoconciencia– la categoría de única certeza.

> *Ahora bien, entre esas cosas [cosas que el espíritu concibe clara y distintamente] hay unas tan claras y a la vez tan simples, que nos resulta imposible pensar en ellas sin que creamos que son verdaderas: por ejemplo, que existo cuando pienso, que las cosas que han sido hechas alguna vez pueden no haber sido hechas, y otras semejantes, de las que es obvio que se tiene una perfecta certeza* (Descartes, 2010, p. 425)

Ahora bien, aunque el *cogito* se establece como primera certeza indudable, el filósofo francés se ve enfrentado a un problema, pues necesita reconstruir el conocimiento sin caer en el escepticismo, al

cual se ha dirigido por medio de la duda metódica que incluso lo ha puesto a dudar de las verdades matemáticas. Esta duda se radicaliza con la aparición del argumento del genio maligno que puede engañarlo constantemente; la incertidumbre necesita ser superada y esto pasa cuando se demuestra (en la *Tercera meditación*) la existencia de un ser perfecto y benevolente que garantiza la veracidad del conocimiento humano. El argumento ontológico utilizado por Descartes parte de la idea que tiene él de un ser infinito, perfecto y omnisciente, pero que su mente finita no pudo elaborar, por lo que afirma que debe haber sido puesta en él por un ser que realmente existe y posee esas mismas perfecciones. Esta idea es clave para superar el escepticismo, pues si Dios es perfecto, entonces no puede engañar, lo que significa que el conocimiento humano –cuando es claro y distinto–es verdadero. Por tanto, Dios, en la concepción cartesiana, no es solo un ser supremo, sino una verdad absoluta que sirve de fundamento para la validez del conocimiento.

Teniendo en cuenta las ideas foucaultinas expuestas con anterioridad, la modernidad cartesiana consolida un modelo epistémico en el que la certeza se convierte en el criterio fundamental de la verdad. El uso de la razón individual, guiado por la claridad y la distinción, pasa a ser el fundamento del conocimiento, aunque garantizado por un ser (Dios) que asegura la correspondencia entre las ideas y la realidad. Sin embargo, en nuestros tiempos la búsqueda de certeza ha cambiado: ya no es la razón individual la que se erige como regulador de la verdad, son los sistemas automatizados de procesamientos de datos. En esta perspectiva, si la modernidad cartesiana establecía un régimen de verdad basado en la autonomía racional del sujeto, hoy presenciamos la emergencia de un nuevo régimen de verdad, donde esta no se produce por el esfuerzo deliberativo humano, sino por el cálculo algorítmico, lo que nos lleva a una pregunta crucial: ¿qué significa conocer en un mundo donde la IA ha asumido la tarea de definir lo que es verdadero?

Inteligencia humana y/o automatización de la certeza

La filosofía, en particular la filosofía cartesiana, entendió la certeza como eliminación de la duda, así como la garantía de validez del conocimiento. La modernidad confió en la razón y en principios metodológicos (las reglas para la dirección del espíritu en palabras de Descartes) para alcanzar la certeza; hoy día asistimos al desarrollo de la IA que ha dado lugar a una certeza tecnificada, basada en la función algorítmica para procesar información a una escala y velocidad inalcanzables para el ser humano. ¿Asistimos a un nuevo régimen epistémico donde la confianza en la razón es sustituida por la confianza en el cálculo? ¿Quién determina lo que es verdadero? ¿Los modelos predictivos? ¿Los sistemas automatizados de decisión? Este desplazamiento del conocimiento humano a la IA plantea otros interrogantes: ¿se trata de una continuación del proyecto cartesiano, pero bajo una nueva forma? ¿Es una ruptura con el modelo epistémico de la Modernidad Clásica? ¿Podemos hablar de certeza cuando los procesos algorítmicos que determinan la verdad son opacos para la mayoría de las personas?, y, sobre todo: ¿qué significa para la autonomía del pensamiento humano que la verdad esté cada vez más mediada por sistemas que no comprendemos completamente, pero a los que, en ocasiones, les entregamos nuestras decisiones? Estos cuestionamientos nos remiten a unos más primarios: ¿qué es inteligencia? ¿Qué es autonomía?

El concepto "inteligencia" no ha tenido siempre el mismo significado. Enzensberger (2007) muestra en "Una palabra y sus peripecias", algunos de los sentidos que se le han dado a este concepto a lo largo de la historia, estos han variado radicalmente según los contextos filosóficos, culturales y lingüísticos en el que ha sido utilizado. Por ejemplo, en la Grecia antigua, el término *nous* (*νοῦς*) hacía referencia tanto al alma como al pensamiento, y el mundo latino medieval, *intelligentia* designaba una facultad *espiritual* asociada a la divinidad. Esto cambió con el paso del Renacimiento a la Modernidad, ya que la palabra se secularizó y comenzó a ser asociada con la *razón práctica*, *la*

agudeza, y más adelante, con la *capacidad de manejar información*. De hecho, en inglés, *intelligence* significa también "información secreta", lo que nos muestra cómo el término fue vinculado con la estrategia, el control y la predicción.

La historicidad del concepto también ha sido abordada por Henry Bergson (2016). Desde una mirada crítica a la modernidad nos invita a pensar la inteligencia no como una capacidad fija y técnica, sino como una *fuerza dinámica y profundamente vinculada al esfuerzo y la vida interior*. En su discurso *La inteligencia* (pronunciado en 1902), Bergson advierte que solemos confundir la inteligencia con la agilidad verbal, la buena memoria o la capacidad para resolver con rapidez problemas superficiales. Pero para el pensador francés, la verdadera inteligencia es la que lograr entrar en simpatía con el objeto, comprenderlo desde dentro, establecer una conexión *profunda* y *viva* con lo que estudia. No se trata, por tanto, de un talento innato ni de una habilidad mecánica, sino de una *adaptación interior del espíritu*, una tensión voluntaria de la atención que permite "vibrar al unísono" con lo que se piensa.

> *A eso se reduce lo esencial de la inteligencia. Imaginen una cuerda tensada por un peso. Si ustedes tocan, al lado, en un instrumento musical, la nota que esa cuerda es capaz de dar por sí misma, vibrará al unísono; pero, de igual forma, responderá también a todas esas notas que denominamos los armónicos de la primera. Así sucede con nuestra inteligencia. La tensión particular que habremos sabido infundir a nuestro espíritu lo dispondrá sobre todo a vibrar al unísono con cierta nota; pero si es capaz de dar esa nota exacta, si está tensionado como se debe, podrá dar también, aunque más discretamente, miles y miles de armónicos de ese sonido fundamental* (Bergson, 2007, pp. 76-77).

Esta concepción bergsoniana de la inteligencia se complementa con una advertencia del mismo Bergson sobre los riesgos de la especialización excesiva. En La especialización (conferencia pronunciada en 1882), Bergson asevera que cuando el conocimiento se fragmenta y se aísla en compartimientos, la inteligencia pierde su capacidad de integración y se torna estéril. Quien se especializa, es decir, se concen-

tra en un saber único, puede volverse incapaz de pensar con amplitud y profundidad. Estas ideas de Bergson se deben a que él piensa que la inteligencia se caracteriza por articular saberes diversos, de ver el conjunto, de adaptarse a contextos nuevos. Es por lo anterior que la inteligencia es inseparable de la voluntad, la formación integral y la libertad del pensamiento.

Así, preguntarnos por la inteligencia en el contexto de la IA, no exige simplemente como respuesta comparar capacidades humanas y tecnológicas, sino percatarnos de que estamos asistiendo a una reconfiguración del sentido mismo del término. ¿Se trata de la capacidad de procesar datos y producir certezas? ¿O sigue implicando una dimensión sensible, crítica y reflexiva que resiste a la automatización? Estas preguntas pueden ser fundamentales para pensar no solo qué entendemos hoy por "inteligencia", sino preguntarnos qué tipo de autonomía estamos dispuestos a ceder, es decir: ¿hasta qué punto estamos dispuestos a dejar que esos sistemas decidan por nosotros, es decir, a ceder nuestra capacidad de actuar y pensar autónomamente? Además, integrar la visión bergsoniana al análisis actual involucra preguntarnos si la llamada "inteligencia artificial" conserva algo de esta dimensión humana, abierta y dinámica ¿Podemos seguir llamando "inteligencia" a un sistema que carece de simpatía, de esfuerzo interior, de comprensión profunda del mundo? ¿Estamos frente a una metáfora que corre el riesgo de vaciar de sentido lo que históricamente ha sido una de las características o potencialidades más ricas y complejas del ser humano?

Estas preguntas son propicias en un mundo donde la emergencia de la IA ha transformado radicalmente tanto las condiciones para acceder a la información como los modos en que se configura lo real. Sostendremos que este es un nuevo régimen de verdad donde lo verdadero, y, por tanto, la certeza, ya no es necesariamente lo que se corresponde con la realidad externa ni lo que ha sido producido mediante el ejercicio racional deliberativo, sino aquello que emerge de operaciones de correlación, predicción y clasificación algorítmica

sobre vastos volúmenes de datos. La realidad comienza a ser producida antes que descubierta. Las decisiones automatizadas, las recomendaciones personalizadas, las predicciones conductuales y los sistemas de vigilancia inteligente moldean, jerarquizan y hacen aparecer al mundo de un modo determinado. Lo real deviene lo que es visible, medible y tratable por la lógica del dato, así se desplazan otras formas de comprensión –simbólicas, sensibles y éticas–a la margen del saber considerado legítimo. Como lo plantea Sadin (2020), este poder de enunciar lo verdadero –poder históricamente disputado por filósofos, científicos, etc.– es hoy atribuido a arquitecturas técnicas que funcionan bajo la promesa de una objetividad incuestionable.

> *Nunca un régimen de verdad se había impuesto de esta manera en la historia, y no por su fuerza de seducción o por su influjo coactivo sino por la sensación compartida de que hay una evidencia, por la producción de ecuaciones que damos por sentado que son las más apropiadas según un principio que pretende que los sistemas cognitivos son justamente Evidence Based Systems, es decir, sistemas basados en un principio de revelación de ciertos hechos que estos sistemas exponen ante nuestra conciencia -y por lo tanto tenemos que acordar con ellos porque hay una evidencia concreta para que lo hagamos así* (Sadin, 2020, p. 101).

La certeza se automatiza, se presenta como inapelable ("el sistema lo ha dicho y eso basta"). Ya no se trata de interpretar el mundo, ni de comprenderlo, sino de captúralo en tiempo real mediante modelos que predicen comportamientos, reconocen patrones y emiten juicios. Así, la certeza cartesiana, nacida de la duda, es reemplazada por una certeza operativa, producida por sistemas que no dudan (no experimentan incertidumbre ni ambigüedad), no preguntan en sentido estricto (no formulan preguntas nuevas, no interrogan el mundo para comprenderlo) y no imaginan alternativas (no crean posibilidades fuera de los datos con los que han sido entrenados; no proponen otros mundos posibles ni se abren a lo inédito, simplemente optimizan lo existente). Como lo vimos al inicio de este escrito, la certeza ha estado asociada al juicio, a la confrontación con

el error, al ejercicio de la razón en condiciones de incertidumbre. La automatización, en cambio, supone un cierre, y este no se da porque la realidad sea más clara, sino porque las márgenes del pensamiento se estrechan: lo cierto ya no es aquello que se defiende con razones (religiosas, políticas, de autoridad, etc.), sino lo que ha sido calculado, lo que aparece como lo más probable según una lógica de correlaciones estadísticas. Como advierte Sadin (2020), la verdad se convierte en un valor técnico estrechamente vinculado a la optimización y la anticipación. El poder de decir la verdad se traduce en el poder de orientar los comportamientos (induce decisiones o acciones) sin necesidad del diálogo ni la co-construcción de sentido (no hay mediación social o deliberativa).

Incertidumbre informativa e imposición de la verdad algorítmica

¿Cuáles son las características de la información en día de hoy? Podríamos decir que la información se nos presenta como algo saturado, fragmentado, con datos dispares y, en muchos casos, desprovista de un marco interpretativo que le otorgue sentido.

> *En el modo actual de producción inmaterial, más información y comunicación significan más productividad, aceleración y crecimiento. La información es una positividad que puede circular sin contexto por carecer de interioridad. De esta forma es posible acelerar la circulación de información* (Byung-Chul Han, 2014, p. 12)

Esta sobreabundancia no solo dificulta la tarea de discernir entre lo relevante y lo accesorio, sino que erosionan los mecanismos tradicionales de validación y comprensión de la verdad (argumentación racional, verificación empírica, deliberación pública, autoridad epistemológica). Históricamente, estos mecanismos cumplían la función de garantizar que la verdad no fuera un producto de la inmediatez,

sino el resultado de un proceso de construcción crítico y deliberado. En este contexto, donde solo los sesgos causan certidumbres, la IA parece ser una promesa de orden, claridad y orientación. No obstante, esta solución técnica no resuelve la incertidumbre mediante procesos deliberativos o hermenéuticos, sino que la clausura por decreto: impone la "verdad" desde el cálculo. Ahora, ¿este cálculo permite comprender el mundo? No, lo modela. En esta perspectiva, reduce la complejidad a través de una arquitectura que privilegia correlaciones sobre significados, patrones sobre contextos, y la predicción sobre la comprensión. Lo incierto –esa condición inherente al pensamiento humano, fuente de creatividad y juicio– es sustituido por una lógica de certidumbre operativa. Así, la IA no elimina la confusión producida por el exceso de datos, sino que organiza esa confusión de un modo que no deja espacio para el disenso. Lo que antes era campo de discusión, hoy se transforma en resultado automático.

Ahora bien, pareciera que esta verdad algorítmica es neutra, pero como advierte Sadin (2020), se trata de una forma de poder que actúa sin visibilidad política, pero con consecuencias normativas. Sus afirmaciones no se presentan como hipótesis abiertas al debate, sino como certezas que deben ser obedecidas: sugerencias de consumo, alertas médicas, predicciones de riesgos, filtros de información. La autoridad ya no proviene de una comunidad epistémica, sino de una máquina cuyo criterio es su capacidad de acierto. La verdad ya no es aquello que se argumenta o evidencia, sino lo que se ajusta al modelo.

Consideramos que lo problemático no es solo el desplazamiento del juicio humano ("lo dijo la IA"), sino la naturalización con la que asumimos este desplazamiento y esta nueva forma de verdad. Quizá esto se debe a que los resultados dados por al IA se perciben como "exactos", "objetivos" y "eficaces", lo que reforzaría la idea de que no es necesario pensar más allá de lo que el sistema indica. La incertidumbre, que en otros tiempos impulsó el pensamiento filosófico, artístico, científico y político, hoy es relegada a los márgenes de lo

que se considera relevante. Podemos afirmar que: la IA promete seguridad en un mundo incierto, pero lo hace a costa de clausurar las preguntas. Así, un intento por navegar la complejidad informativa del presente termina por imponer un régimen de sentido cerrado, donde el dato sustituye a la experiencia, la probabilidad a la deliberación, y el cálculo al pensamiento complejo. Cerramos entonces esta sección con una paradoja: desarrollada en un entorno de incertidumbre radical, la IA se convierte en el nuevo juez de la certeza… pero de una certeza sin juicio, sin contexto y sin voz.

La paradoja anteriormente expuesta, pone en evidencia una transformación más profunda: el desplazamiento de la verdad como proceso crítico hacia la exactitud como producto técnico. Entender esta mutación nos lleva (y nos exige) preguntarnos qué significa conocer un mundo donde la información ya no es interpretada libremente, sino modelada bajo un régimen de sentido cerrado, donde el dato sustituye a la apertura del significado o sentido.

La verdad, como ha mostrado Foucault, no es un dato inmutable ni el resultado de una búsqueda neutral, sino el efecto de regímenes de poder que definen, en cada época, qué discursos pueden ser considerados verdaderos, quiénes tienen la autoridad para enunciarlos y cuáles son los procedimientos legítimos para validarlos. En este sentido, conocer no ha sido nunca un acto puramente contemplativo, sino una práctica situada en relaciones de poder. Hoy, en un mundo gobernado por datos, asistimos a una transformación en el régimen de verdad: la verdad, que históricamente se articulaba a través de prácticas discursivas, instituciones de saber y procesos de validación social, comienza a ser sustituida por la exactitud algorítmica. La exactitud responde a una lógica técnica que privilegia la precisión del cálculo y la correlación estadística sobre la construcción crítica del sentido. Mientras algunos regímenes de verdad implicaban exposición a la crítica, posibilidad de disenso y construcción colectiva del saber, el régimen algorítmico produce evidencias inapelables, operaciones que clausuran la discusión antes de que esta pueda

abrirse. La reducción del conocer a la producción de exactitudes algorítmicas no implica solo un cambio técnico, sino una redefinición profunda de la relación entre el ser humano, el saber y el poder. Como advierte Sadin (2020), el poder de enunciar lo verdadero ha dejado de estar vinculado a la deliberación humana para desplazarse hacia arquitecturas técnicas que, bajo la promesa de objetividad, instauran nuevas formas de normatividad. En este contexto, la exactitud técnica desplaza los procesos históricos de validación –la argumentación racional, la verificación empírica, la deliberación pública– que garantizaban que la verdad no fuera el resultado de la mera inmediatez. Hoy, la IA produce "certezas inmediatas" basadas en correlaciones de datos, sin necesidad de verificación empírica ni deliberación pública, y por eso se pierde ese proceso de construcción crítica de la verdad.

La transformación epistémica es, en este sentido, profunda: la verdad, como espacio de apertura, disenso y producción simbólica, es reemplazada por la certeza automática de los sistemas. El conocimiento ya no emerge de la interpretación crítica, sino de la eficiencia en la manipulación de datos modelados. Como señala Floridi (2014), habitamos una infosfera[1] donde la distinción entre experiencia y representación informacional se diluye[2], y donde la inteligibilidad del mundo queda subordinada a infraestructuras técnicas. Ahora bien, no estamos diciendo con esto que haya existido un momento donde la experiencia fuese *inmediata* o pura, pues siempre ha existido algún tipo de mediación cultural, simbólica o epistemológica. Por ejemplo, la religión (mediación a través de doctrinas, símbolos e interpretaciones sagradas) y la ciencia moderna (mediación a través de teorías, modelos y verificaciones) han permitido la construcción del sentido

1 La infosfera es el entorno global de información donde todo –personas, objetos, procesos– existe, interactúa y se transforma a través de datos y flujos informacionales. No solo habitamos la infosfera: somos parte activa de ella.

2 Ya no sabemos bien si estamos viviendo una vivencia directa o una experiencia ya modelada digitalmente. Por ejemplo, cuando viajamos a un sitio siguiendo Google Maps, o cuando vemos las noticias filtradas por algoritmos.

y la experiencia en las sociedades. Pero, ¿cuál es la diferencia de la mediación algorítmica de la IA? Las infraestructuras digitales no solo representan el mundo, sino que lo filtran, ordenan y reconfiguran automáticamente, en tiempo real, sin deliberación humana explícita (automatización técnica). Por otro lado, esta mediación ocurre a una velocidad que excede la capacidad humana de interpretar o cuestionar, y bajo procesos que son opacos (no vemos cómo los algoritmos deciden qué mostrar o qué ocultar). Asimismo, a diferencia de las mediaciones religiosas o científicas, las mediaciones algorítmicas tienden a *homogenizar* y *optimizar* los sentidos posibles: muestran lo que es más rentable, más probable o más eficiente, no lo más profundo o complejo. Finalmente, la inteligibilidad del mundo anteriormente, al menos en parte, objeto de disputa (debate científico, interpretación religiosa, conflicto filosófico), hoy día (el proceso algorítmico) se presenta como cerrado: no deja espacio real para disputar el modo en que organiza la realidad.

La paradoja se despliega en toda su dimensión: en un mundo que produce más información que nunca, el espacio para la verdad se reduce; en un mundo donde los datos se multiplican, la posibilidad de un conocimiento crítico se erosiona. La inteligencia artificial, nacida en el seno de la incertidumbre informativa, no abre nuevos horizontes de sentido, sino que impone un orden algorítmico que clausura la pregunta. Así, el acto de conocer se convierte en conformidad estadística, mientras la autonomía del pensamiento humano —esa tensión viva hacia lo inacabado, lo incierto y lo interpretable— se ve sustituida por la pasividad ante evidencias producidas sin mediación crítica.

Frente a este escenario, reivindicar la diferencia entre verdad y exactitud no constituye un gesto nostálgico, sino una exigencia crítica: recordar que conocer no es solo predecir, sino interpretar, comprender y resistir la clausura de los sentidos bajo la apariencia de una certidumbre que no admite preguntas.

Modelo algorítmico, régimen de verdad y erosión de la vida política

Desde hace poco más de una década, hemos ingresado en lo que Éric Sadin (2020) denomina un "nuevo régimen de verdad" caracterizado por la emergencia de sistemas técnicos capaces de enunciar, de manera automatizada, determinaciones sobre lo real. Ya no se trata solo de tecnologías que almacenan, clasifican o transmiten datos: el giro contemporáneo consiste en que ahora se delega a ciertos sistemas de inteligencia artificial la prerrogativa de definir de manera directa qué es verdadero, enunciarlo y actuar en consecuencia. Como explica el pensador francés:

> *lo digital se erige como una potencia* aletheica, *una instancia consagrada a exponer la* aletheia, *la* verdad, *en el sentido en que la definía la filosofía griega antigua, que la entendía como develamiento, como la manifestación de la realidad de los fenómenos más allá de las apariencias. Lo digital se erige como un órgano habilitado para peritar lo real de modo más fiable que nosotros mismos, así como para revelarnos dimensiones hasta ahora ocultas a nuestra conciencia* (pp. 17-18).

Esta mutación epistémica implica que el acceso a lo real se realiza a través de procesos algorítmicos que prometen exactitud, rapidez y objetividad, desplazando las formas humanas tradicionales de conocer, que reconocían la incertidumbre, la interpretación y el debate. Así, la inteligencia artificial deja de ser un mero asistente para convertirse en una autoridad epistémica, que monopoliza la producción de certezas y presenta sus dictámenes como inapelables.

El ascenso de la IA como autoridad cognitiva implica el progresivo fin del saber interpretativo. Allí donde antes el conocimiento se construía a través de procesos sociales de interpretación, deliberación y validación, ahora emerge una dinámica en la cual el libre ejercicio de nuestra facultad de juicio y de acción, se ve sustituido por protocolos destinados a provocar inflexiones en cada uno de nuestros actos.

> *De ahora en más hay una tecnología que reviste un "poder conminatorio" mientras el libre ejercicio de nuestra facultad de juicio y de acción se ve sustituido por protocolos destinados a provocar inflexiones en cada uno de nuestros actos o cada impulso de lo real con vistas a insuflarles, casi de "soplarles", la trayectoria correcta a adoptar* (p. 21)

La IA, entonces, se ha convertido en una fuerza que no se limita a sugerir o recomendar, sino que incita, prescribe y, finalmente, impone modos de actuar ajustados a los criterios del cálculo.

La producción de certezas automáticas, basada en el análisis masivo de datos y en la modelización de comportamientos, tiende a disolver el juicio humano. La verdad ya no es entendida como resultado de un esfuerzo de comprensión crítica, sino como un producto técnico que se impone directamente a la acción. Esta imposición no es explicita o violenta, sino sutil y no se restringe a algunos campos de conocimiento o de la vida humana, sino que tiende a absolutizarse:

> *Nos equivocaríamos si pensáramos que esta potencia no se ejercería in fine sino en estos dos campos [salud y seguridad], dejándonos preservados en los restantes segmentos de nuestra vida, porque probablemente estos campos solo se sitúen en la vanguardia de un orden destinado a inspirar, o a "aspirar a su paso", a todos los demás. Porque podemos suponer que la dimensión incitativa, imperativa y prescriptiva están llamadas, dado el extremo utilitarismo que instauran, a subir de nivel, a acercarse al estadio fulminante y coercitivo antes de consumar su finalidad última, la de colaborar en la concreción de un mundo perfecto, de espíritu higienista, que ya no debe encontrarse con el estorbo del menor error, donde todo estaría perfectamente regimentado, como en la visión leibniziana del mejor de los universos que se consumará muy pronto gracias a los cálculos hechos por máquinas que nos entregarían, en toda circunstancia, la verdad, una verdad* sistemática *que, según Hannah Arendt, termina inevitablemente por encubrir una función tiránica* (p. 149)

Es decir, según Sadin, estos estadios delinean un proceso por el cual las verdades algorítmicas pasan de sugerir acciones (incitación), a exigirlas (imperativo), a normarlas (prescripción), y finalmente a imponerlas sin posibilidad de discusión (coerción). Al mismo tiem-

po, el problema de la opacidad algorítmica agudiza esta situación. Los sistemas de IA, cada vez más complejos y autónomos, son "cajas negras" cuyo funcionamiento resulta ininteligible tanto para los usuarios como para los propios diseñadores. Esta opacidad impide el ejercicio de la crítica y excluye la posibilidad de deliberar colectivamente sobre las bases del conocimiento que nos rige. La falta de transparencia no es accidental: forma parte del modo en que el nuevo régimen de verdad asegura su eficacia, desactivando los mecanismos democráticos de evaluación y control.

Según Sadin este nuevo régimen de verdad se distingue, a su vez, por el carácter anticipatorio de sus operaciones, es decir, la verdad ya no describe lo que es, sino que prescribe lo que debe ser. Como se ha dicho antes, la IA no solo representa lo real, sino que lo modela, organiza y lo dirige, instaurando un orden normativo-técnico que reduce a los sujetos a meros objetos de cálculo y administración. La tendencia hacia la producción de certezas automáticas y la disolución del juicio humano está, desde esta perspectiva, directamente vinculada con el objetivo de eliminar toda forma de incertidumbre, que es percibida como una falla a corregir mediante el perfeccionamiento continuo de los dispositivos y algoritmos. Es por esta razón que la exactitud prometida por los algoritmos y las IA se transforma en un instrumento de poder epistémico. Tal como lo explica Sadin, el ideal de perfección técnica se articula cada vez más con la voluntad de erradicar toda zona de indeterminación. La capacidad de la IA para ofrecer determinaciones (supuestamente infalibles) no responde solo a un deseo de eficacia, sino también a una aspiración a gobernar la acción humana desde parámetros calculables, optimizados y previsibles. Así, se configura un modelo de sociedad en el que la deliberación, el debate y el juicio crítico son sustituidos por procedimientos automáticos de validación de lo real, administrados por sistemas opacos que dictan la conducta legítima en función de protocolos técnico-científicos. La opacidad de estos procesos, unida a su capacidad para modelar anticipadamente los comportamientos, consuma la exclusión de la deliberación pública y la instauración de

un *nuevo orden tecnocrático*, en el que el poder de la IA consiste en hacer desaparecer la posibilidad misma de disentir.

Lo anterior, compromete diversos campos de la vida humana, esto es, la IA como régimen de verdad no solo transforma nuestra relación con el conocimiento, sino que redefine las condiciones de la vida en común. Frente a este panorama, Sadin llama a recuperar el "coraje de la verdad" en el sentido foucaultiano: resistir la enunciación automática de certezas, reivindicar el derecho a la interpretación, a la crítica y a la construcción plural de los sentidos de la experiencia. La tarea urgente es, por tanto, reabrir los espacios de indeterminación que el poder algorítmico busca clausurar y defender la capacidad humana de juicio frente a la maquinaria de la certeza automática.

> *O sea, el hecho de reivindicar que, a diferencia de la exactitud, la verdad no se presenta bajo ningún referente estable; apela a un esfuerzo de aprehensión que nunca se consuma y sobre el cual debemos regularmente ponemos de acuerdo, incluso de modo provisorio, dentro de la diversidad de las subjetividades existentes a fin de esforzarnos por actuar, individual y colectivamente, del modo más justo y al margen de toda imposición unilateral que amordace nuestro derecho de palabra (Sadin, 2020, p. 38).*

En este escenario, se da un desplazamiento del pensamiento crítico hacia la automatización del juicio compromete de manera decisiva las condiciones de posibilidad de la vida política. Byung-Chul Han (2014) advierte que, en el actual régimen de la positividad[3], se erosiona el espacio del disenso y la alteridad, reduciendo las diferencias a meros datos gestionables. La comunicación digital, al eliminar las fricciones que caracterizaban el espacio público, clausura las tensiones constitutivas de la vida democrática: sustituye el conflicto, indispensable para la construcción de lo común, por flujos de información inmediata y consensos automáticos. La verdad algorít-

3 El régimen de positividad es la forma de poder contemporáneo que no reprime ni prohíbe, sino que estimula, seduce y optimiza. En lugar de imponer límites, promueve la autoexplotación voluntaria, la transparencia total y la comunicación continua.

mica, presentada como un producto neutro de procesos técnicos, reemplaza los procesos deliberativos, imponiendo un modelo de consenso sin discusión.

Luciano Floridi (2014), al analizar la emergencia de la infosfera como el nuevo entorno vital, se caracteriza por el cambio constante y la aceleración de los flujos de información que impiden la estabilización de sentidos comunes y acuerdos duraderos. Este fenómeno debilita los procesos de construcción comunitaria, ya que la comunidad requiere no solo comunicación instantánea, sino sedimentación de experiencias compartidas, tiempos de maduración y reelaboración colectiva del mundo. El segundo peligro, entonces, es la disolución de los vínculos comunitarios: sin memorias compartidas ni horizontes comunes, los individuos quedan aislados en burbujas de información que refuerzan el aislamiento y la fragmentación social.

En este contexto, la supresión de la incertidumbre ya no aparece como una aspiración epistemológica sino como una estrategia de normalización social. Éric Sadin (2020) sostiene que los sistemas digitales contemporáneos no buscan tanto describir lo real como incitar, dirigir y regimentar comportamientos: al automatizar las conductas a partir de correlaciones estadísticas, se elimina el espacio de la duda, del cuestionamiento y de la elección consciente. La autonomía de los ciudadanos, condición fundamental de cualquier régimen democrático, se ve sustituida por procesos de optimización de conductas cuyo fin no es la participación crítica, sino la adaptación eficiente a lógicas predefinidas. El poder algorítmico, en consecuencia, no se limita a la producción de conocimiento, sino que constituye un dispositivo de gobierno que opera bajo la lógica de la seducción y la eficiencia. Como señala Han, el poder inteligente se esfuerza en generar emociones positivas y en explotarlas, evitando así toda fricción que podría interrumpir los flujos de comunicación y consumo. Este nuevo tipo de poder no se impone por la fuerza, sino que coloniza la libertad misma, volviéndola funcional a su propia reproducción. Si nos percatamos bien, el peligro aquí es doble: por un lado, la conversión de la ciudadanía en pasividad

adaptativa, más cercana al consumo que a la acción política; por otro, la anulación del disenso como base de una comunidad verdaderamente plural. En términos del pensamiento de Hannah Arendt, la acción política exige un espacio de aparición donde los ciudadanos, en su singularidad irrepetible, puedan actuar, hablar y deliberar juntos, inaugurando lo nuevo y sosteniendo la pluralidad como condición esencial del mundo común. Sin la posibilidad de acción, entendida no como mera ejecución de objetivos prediseñados, sino como inicio libre y colectivo de procesos imprevisibles, el tejido de relaciones humanas se debilita y la política se degrada en mera administración técnica. El cierre de la incertidumbre elimina la natalidad política –la capacidad de comenzar algo nuevo– y condena a la comunidad a la repetición de lo dado, anulando la posibilidad misma de la libertad que, para Arendt, constituye la razón de ser de la política.

> *El hecho de que el hombre sea capaz de acción significa que cabe esperarse de él lo inesperado, que es capaz de realizar lo que es infinitamente improbable. Y una vez más esto es posible debido sólo a que cada hombre es único, de tal manera que con cada nacimiento algo singularmente nuevo entra en el mundo [...] Si la acción como comienzo corresponde al hecho de nacer, si es la realización de la condición humana de la natalidad, entonces el discurso corresponde al hecho de la distinción y es la realización de la condición humana de la pluralidad, es decir, de vivir como ser distinto y único entre iguales* (Arendt, 2009, p. 202)

La transformación contemporánea redefine no solo qué se entiende por verdad, sino también qué tipo de subjetividad puede existir. Ya no el sujeto que interroga, duda y construye sentidos junto a otros, sino el sujeto que se adapta a flujos prediseñados, que confirma patrones y refuerza sistemas de clasificación automáticos. La erosión de la incertidumbre –como espacio de apertura, de pregunta y de posibilidad– constituye, en este sentido, una pérdida no solo epistémica, sino profundamente política. *Defender la incertidumbre, la opacidad del juicio humano y la lentitud de la deliberación crítica* es hoy defender el espacio mismo donde pueden emerger formas de vida democráticas y comunidades no sometidas a la administración automática de sus destinos.

Ética y resistencia epistémica

Frente a un régimen de verdad automatizado que tiende a clausurar la incertidumbre mediante procesos de cálculo, la resistencia ética y epistémica no es un gesto nostálgico ni un rechazo tecnofóbico, sino una *necesidad política y cognitiva*. La afirmación de que los algoritmos proporcionan respuestas exactas ha erosionado el lugar de la pregunta, del juicio y del disenso, sustituyendo el conocimiento como construcción deliberativa por certezas operativas sin mediación crítica. En este contexto, pensar en una ética de la IA no puede limitarse al diseño de sistemas "menos sesgados", sino que implica la reconstrucción activa de las condiciones para que el conocimiento vuelva a ser un proceso compartido, argumentado y abierto.

Como ha advertido Mittelstadt et al. (2016), los sistemas algorítmicos no solo producen resultados, sino que también transforman lo que cuenta como evidencia válida. Lo hacen mediante procedimientos que muchas veces resultan opacos no por falta de acceso, sino por su complejidad estructural y su naturaleza no explicable en términos humanos. Esta opacidad no solo compromete la transparencia, sino también el reconocimiento de responsabilidades, pues dificulta la trazabilidad de los efectos y la asignación ética de los errores. Si las decisiones se generan a partir de modelos cuya lógica es ininteligible incluso para quienes los diseñan, entonces la posibilidad de ejercer una crítica informada o de disputar sus resultados queda seriamente limitada.

Frente a este panorama, recuperar la deliberación no significa regresar a formas puras de una razón ilustrada, sino *reconocer el valor epistémico y político de la incertidumbre*. En lugar de suprimirla mediante la programación de respuestas anticipadas, se trata de afirmarla como espacio de posibilidad. La deliberación exige lentitud, exposición al otro, apertura a lo inesperado; todas condiciones que el régimen algorítmico tiende a eliminar en nombre de la eficiencia. Por eso, *resistir éticamente supone insistir en el derecho a dudar, a preguntar, a comprender y a disentir*.

Ahora bien, esta resistencia no puede darse sin una transformación estructural en nuestras prácticas sociales, institucionales y formativas. Necesitamos una ética pública de la tecnología que permita reinscribir el juicio humano en el centro de los procesos de decisión y conocimiento. Esto no se logra únicamente mediante exigencias de transparencia técnica: como señalan Mittelstadt et al. (2016), la disponibilidad de información no garantiza su comprensibilidad para los ciudadanos. La transparencia se vuelve vacía si no está acompañada de condiciones que permitan a los sujetos interpretar críticamente lo que se les presenta como evidencia. En este sentido, educar en formas críticas de alfabetización digital implica no solo saber usar herramientas tecnológicas, sino también comprender cuándo una respuesta técnica ha clausurado un problema político o ético que seguía abierto. Además, como advierte el mismo estudio, los algoritmos no se limitan a mediar decisiones: reontologizan el mundo, es decir, redefinen los modos en que concebimos la realidad misma. Por eso, resistir no significa simplemente corregir errores técnicos o hacer ajustes normativos, sino defender formas de vida en las que la pluralidad, la apertura y la construcción colectiva del sentido sigan siendo posibles. Cuando las decisiones se presentan como productos opacos e incuestionables, se impide toda posibilidad de disenso: no se puede disentir de lo que no se comprende. Por ello, una regulación ética verdaderamente transformadora no debe contentarse con evitar daños colaterales, sino promover activamente entornos donde el conocimiento recupere su carácter interpretativo, discutible e inacabado.

En este esfuerzo, las humanidades y la filosofía no son un lujo decorativo ni una nostalgia académica, sino prácticas necesarias para sostener el tejido interpretativo de lo común. Son ellas las que pueden advertir cuándo un algoritmo reemplaza la argumentación por correlación, cuándo una evidencia técnica excluye otras formas de saber, o cuándo un patrón estadístico anula la singularidad. Pensar críticamente en la era de la IA implica no solo cuestionar los fines y medios de los sistemas técnicos, sino también imaginar horizontes epistémicos donde el sentido no esté predeterminado por el cálculo.

Desde esta perspectiva, es urgente construir un compromiso cognitivo, este no puede fundarse en la sustitución, sino en la complementariedad; que reconozca el lugar del algoritmo en el procesamiento de datos, pero que reafirme la centralidad de la experiencia, la comprensión y el juicio en la construcción del conocimiento. Este compromiso deberá incluir el derecho a no ser interpretado únicamente por sistemas automatizados, el derecho a entender cómo se nos representa y clasifica, y el derecho a disputar el modo en que se organiza lo real. Solo así será posible resistir no con la lógica de la nostalgia, sino desde una ética afirmativa que defienda la vida democrática del pensamiento.

Consideración final: la verdad como construcción compartida

Pensar el futuro de la verdad en la era digital implica, ante todo, reconocer que hemos ingresado en un umbral epistémico inédito. El valor de lo verdadero, históricamente ligado a prácticas discursivas, a procedimientos de validación compartida y a formas de experiencia situada, se encuentra hoy tensionado por sistemas automatizados que operan sobre lógicas predictivas, no interpretativas. El conocimiento ya no solo se construye: se extrae, se calcula, se optimiza. En este nuevo paisaje, la verdad tiende a ser entendida como la restitución de un patrón estadístico, no como el resultado de un proceso hermenéutico, y eso transforma la relación misma entre saber, sujeto y mundo.

Sin embargo, lejos de constituir un simple avance técnico, esta mutación reconfigura el modo en que se organiza lo decible y lo pensable. El peligro no reside tanto en que los sistemas algorítmicos puedan equivocarse –lo cual es siempre posible–, sino en que instauren una forma de certeza silenciosa, sin conflicto, sin deliberación, sin resto. Una certeza que ya no se argumenta, sino que se entrega como resultado. Es allí donde se juega el desplazamiento más profundo: en

la transformación del juicio en cálculo, del desacuerdo en dato, del pensar en rendimiento. Frente a ello, la pregunta por la verdad debe ser rescatada de su supuesta automatización, no para negar los aportes técnicos de la inteligencia artificial, sino para reinscribirla en el campo de lo político, de lo común, de lo humano.

En este contexto, la reivindicación de la incertidumbre no equivale a una renuncia a conocer, sino a una defensa activa de la complejidad del mundo. La incertidumbre no es el fracaso del pensamiento, sino su condición de posibilidad. Solo cuando reconocemos los límites de lo que sabemos, podemos interrogarnos críticamente sobre cómo lo sabemos, con qué instrumentos y desde qué lugares. En este sentido, la incertidumbre actúa como freno epistémico ante la ilusión de totalización que muchas tecnologías promueven; pero también como impulso para la imaginación crítica, para la apertura al otro, para la invención de nuevos marcos de comprensión.

Aceptar la incertidumbre es, por tanto, resistir la tentación de una verdad sin sujetos, sin historias, sin conflictos. Es rechazar el espejismo de la neutralidad maquínica y defender que todo saber está enmarcado, situado, atravesado por valores, intereses y afectos. De allí que la inteligencia artificial no deba ocupar el lugar del árbitro epistémico, como si sus operaciones fuesen ajenas a todo contexto o ajenas a la historia. Su función, más bien, debería ser comprendida en términos instrumentales, como parte de una ecología ampliada del conocimiento que requiere, precisamente, de nuestra vigilancia crítica, de nuestra capacidad de interpretación y, sobre todo, de nuestra responsabilidad.

Lo que está en juego, en última instancia, no es solo el futuro de la verdad, sino el futuro de nuestra capacidad para sostener procesos de sentido colectivos en un mundo cada vez más mediado por lógicas técnicas. Frente a ello, el pensamiento crítico se vuelve indispensable no como lujo académico, sino como forma vital de defensa ante la clausura de lo posible. Solo así podremos imaginar una política de lo verdadero que no suprima la pluralidad de voces, la apertura de la pregunta ni la necesidad permanente de argumentar.

Referencias

Arendt, H. (2009). *La condición humana*. Paidós.

Bergson, H. (2007). *La inteligencia*. Interzona.

Byung-Chul, H. (2014). *Psicopolítica. Neoliberalismo y nuevas técnicas de poder*. Herder.

Descartes (2010). *Meditaciones acerca de la filosofía primera. Seguida de las objeciones y respuestas*. Universidad Nacional de Colombia.

Enzensberger, H.M. (2007). *En el laberinto de la inteligencia: guía para idiotas*. Anagrama.

Floridi, L. (2014). *The fourth revolution: how the infosphere is reshaping human reality*. Oxford University Press UK.

Foucault, M. (2005). *El orden del discurso*. Tusquets Editores.

Foucault, M. (2019). *Microfísica del poder*. Siglo Veintiuno Editores.

Margot, J-P. (2018). *La modernidad. Una ontología de lo incomprensible*. Universidad del Valle.

Mittlstadt, B.D., Allo, P., Taddeo, M., Wachter S. y Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, Vol. 3(2), 1-21. DOI: 10.1177/2053951716679679

Sadin, É. (2020). *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical*. Caja Negra Editora.

[Capítulo 2]

De Skynet a OpenAI: Del Imaginario Apocalíptico a los Desafíos de la Gobernanza y la Agencia Humana en la Inteligencia Artificial Real[1]

Juan Felipe Castaño Quintero

Con el lanzamiento público de ChtGPT a finales del año 2022 la inteligencia artificial se puso de moda. En el centro del debate público empezaron a emerger opiniones que la abrazaban como una nueva tecnología que se aún hoy se erige como el futuro de la humanidad; al mismo tiempo, empezaron a escucharse voces de advertencia y crítica hacia lo que implica este tipo de tecnología para la especie humana.

Lo cierto es que, como nos lo recuerdan los estudios en Ciencia Tecnología y Sociedad (CTS) ninguna tecnología es neutra (Parra y Castro, en prensa; Winner, 1983; Haraway, 1991). Es decir, los desarrollos tecnológicos deben ser entendidos en el contexto social de su producción, contexto que es a su vez histórico y político; por ello, una visión crítica de la inteligencia artificial no puede centrase exclu-

1 Este texto se inscribe en el componente 1 de la investigación interfacultades: Inteligencia Artificial en las organizaciones de la Universidad Icesi

sivamente en la tecnología per se y los posibles impactos de su aplicación, sino que esta debe estar atada a su contexto político de interacción social y los correlatos históricos que permiten la emergencia y arraigo de estas tecnologías por sobre otras posibles.

En ese orden de ideas, más allá de fetichizar los miedos hacia una inteligencia artificial autónoma que reemplace lo humano, una verdadera mirada crítica de la inteligencia artificial debe estar enfocada a analizar por qué se torna relevante en el marco del empoderamiento del capitalismo digital y las transformaciones que desde hace un tiempo se nombra como cuarta revolución industrial.

Del origen de la IA al dato

Aunque el desarrollo de la inteligencia artificial viene desde los años 50 del siglo 20 con ideas como la de Turing (1950) de la inteligencia computacional, es a mediados de los años 90s cuando entran en escena Google y el auge público de la Internet que el capitalismo empieza a tronarse hacia la importancia del dato. Así como la inteligencia artificial no es un desarrollo del siglo XXI, el dato y su centralidad no son invenciones de los años 90s del siglo XX. Los estudios de mercadeo y estadística ya habían desarrollado innumerables técnicas de analítica de datos, de hecho, Isaac Asimov (1951) en su novela clásica Fundación, proponía una ciencia ficticia llamada psicohistoria que funcionaba con volúmenes de datos a escala planetaria que podría predecir el destino de toda una civilización. De manera similar, Penrose (1989) centrado en la idea de causalidad plantea la idea que teóricamente una buena catalogación de los datos podría llevarnos a predecir el futuro si se lograra capturar todas las variables (elemento determinante ya que para él no es posible capturar todas estas variables ya que el más mínimo impulso puede cambia todo el sistema que sin embargo ya está determinado desde el big bang dado los impulsos anteriores). Es allí donde teóricamente el dato, empieza a transformarse en un activo central,

para las empresas de mercado, y luego para las compañías tecnológicas, donde elementos como, el algoritmo de Google que permite ordenar la información de forma más eficiente capitaliza el dato dándole un valor agregado que será central en la transformación de los modelos económicos en el siglo XXI.

Del dato al capitalismo digital

En este contexto, empiezan a aparecer una serie de servicios y soluciones donde el dato como activo es central para ordenar la forma cómo se entiende el mundo desde estas tecnologías, pero así mismo, el dato está atado a su producción y su representatividad. Los datos no surgen de la nada, sino que dependen de las personas y los objetos que los producen, pero también en la capacidad de la tecnología de capturar esta producción de datos, lo cual está estrechamente relacionado con el acceso que tienen distintas personas a la tecnología a través de la cual se generan y se capturan estos datos. Es justamente acá donde se empieza a ver dónde emergen las limitaciones de este modelo y esta forma de comprensión del mundo, está limitado por la cantidad de datos disponibles, pero también por la forma como se organizan los datos disponibles para dar cuenta de la realidad y lo cierto es que tal representatividad y disposición de los datos reproduce los sesgos de los sistemas de representación que los producen y los ordena.

Cathy O'Neil (2017) llama a esto *Armas de destrucción matemática*. Postulando la idea que la producción de algoritmos sin supervisión tiende a acrecentar la desigualdad existente al reproducir los preconceptos y opiniones de quienes realizan el código. Para ella, un algoritmo es una opinión embebida en un código y que las definiciones de éxito de la tarea de un algoritmo, está enmarcada en el contexto cultural de la organización y producción del dato.

Gobernanza de datos

A partir de ello, empiezan a escucharse distintas voces que propenden por la justicia algorítmica al demostrar cómo, sistemáticamente, determinados grupos salen de la producción de la realidad de las soluciones que emergen de la forma como estos algoritmos producen (o reproducen) la realidad a partir de los datos. A raíz de esto, organismos como la OCDE, la UNESCO, la Unión Europea, a nivel internacional, y a nivel nacional MinTic, MinCiencias y el Conpes 4144 de 2025, ponen en el centro del debate alrededor de la ética y el desarrollo de inteligencias artificiales la idea de la gobernanza, con especial énfasis en la gobernanza de datos en la inteligencia artificial.

Al igual que O´Neil (2014) el llamado de estos organismos es a ejercer un control activo sobre qué pasa con los datos que producimos y cómo tomar el control sobre los mismos. Y en ese sentido, las políticas públicas se orientan a tomar acción alrededor del dato como activo y a empoderar a los usuarios sobre el valor que tienen sus datos y qué pasa cuando estos son utilizados por distintas compañías.

Personalmente, adhiero a la necesidad de ejercer control sobre la producción, uso y organización de los datos dado no solamente su valor en el mercado, sino también debido a su carácter político y ontológico de moldear la realidad en el contexto del capitalismo digital y la cuarta revolución industrial. No obstante, considero problemático centrar la gobernanza de la IA en el dato como activo por dos razones principalmente, siendo la segunda el argumento principal de este artículo.

Por un lado, y como empiezo a desarrollar en otro espacio, la emergencia del *homodata*, es decir, la construcción y reducción de la experiencia humana al valor de sus datos, plantea una condición ontológica, reduccionista, que socava la diversidad y la somete a la idea de que la inteligencia es los que un reducido grupo de desarrolladores piensan que pensar es; y por otro lado, la misma idea que somos lo

que los datos nos dicen que somos, y en este sentido, la realidad se moldea por los resultados de los algoritmos y más allá de estos por la decisiones que toman la inteligencias artificiales, implica negar de entrada toda la posibilidad de agencia humana, entonces ¿podemos hablar de gobernanza sin agencia? ¿se debe reducir la agencia únicamente al control de los datos y no al uso y la apropiación de estos más allá de o a través de la apropiación de la tecnología?

El objeto de este artículo no es pues, dar una respuesta a estos interrogantes, sino más bien generar más. Lo que pretendo es construir una agenda programática de algunas preguntas relevantes alrededor de pensar críticamente los impactos de las tecnologías de inteligencia artificial haciendo énfasis en el componente humano y, sobre todo, en los usos y apropiaciones de esta tecnología.

Como plantee en las primeras líneas, desde una perspectiva CTS, los desarrollos tecnológicos deben ser entendidos en el contexto social de su producción, contexto que es a su vez histórico y político. Y en este sentido, es fundamental posicionar a los sujetos que usan estas tecnologías y no solamente como sujetos pasivos que son estructurados por estas. En este sentido, sí, partimos de la premisa que las inteligencias artificiales reproducen sesgos y usadas acríticamente amplifican el riesgo de acrecentar las desigualdades. Más si llevamos esto a un análisis político de las relaciones de poder a través de las cuales los desarrollos tecnológicos del capitalismo digital empiezan a modelar las formas hegemónicas de entender el mundo. Y considero que ahondar en esta línea de análisis es fundamental para entender cómo se está dando tal transformación, en un contexto donde líderes tecnológicos asesoran directamente a lideres mundiales.

No obstante, también es necesario entender, que los usuarios de tecnología no somos agentes pasivos que necesariamente somos moldeados por el uso y las propuestas de las tecnologías. Tampoco quiero acá romantizar el uso y la agencia para decir que usar la IA automáticamente nos salva de la producción de sesgos y la reproducción de desigualdades. Y es acá donde pretendo detenerme para plantear una

serie de interrogantes alrededor justamente de cómo los usuarios se apropian de las tecnologías y cómo a través del uso se construye la agencia y cuáles son los límites de esta agencia. Por qué es necesario investigar no sólo la producción de tecnología y sus propuestas, sino cómo esta es usada y en qué contextos y cómo el impacto debe ser entendido cómo algo relacional.

Gobernanza en Inteligencia Artificial

A finales del 2021 la UNESCO publicó un documento normativo mundial que buscaba guiar el desarrollo y uso responsable de la inteligencia artificial. En este documento se reconoce un potencial transformador de la IA que puede llegar a ser tanto positivo como negativo, por tal razón propone un marco de valores, principios y acciones para asegurar que la IA sirva a la humanidad, respete la dignidad humana, los derechos humanos, la igualdad de género y proteja el medio ambiente y los ecosistemas (Unesco, 2021, pp. 69-73). Para esto, la entidad parte de una definición amplia de inteligencia artificial, aludiendo que al ser esta una tecnología que está en constante desarrollo una definición rígida quedaría rápidamente obsoleta (p. 45) y por ende prefiere centrarse en la IA como aquellos sistemas capaces de procesar datos e información de una manera que se asemeja a un comportamiento inteligente. Es decir, que incluya capacidad como el razonamiento, el aprendizaje, la percepción, la predicción la planificación y el control. Acá entran tecnología de proceso de información con modelos y algoritmos que les permitan aprender y realizar tareas cognitivas, operando con diferentes grados de autonomía. Esto incluye métodos como el aprendizaje automático (incluido el aprendizaje profundo) y el razonamiento automático, así como su posible integración en sistemas ciberfísicos como la robótica o el Internet de las cosas (pp. 45-50).

Además de esta definición amplia, que busca que las directrices sean aplicable a una vasta gama de tecnologías y futuras que exhiban

estas capacidades "inteligentes" el documento propone que la relevancia ética de la IA no debe limitarse solamente a su uso final sino que debe abarcar todas las etapas del ciclo de vida de un sistema de IA, es decir, desde la investigación, la concepción y el desarrollo, pasando por el mantenimiento, funcionamiento, comercialización, financiación, seguimiento evaluación, validación, el fin de la utilización el desmontaje y la terminación (p. 51). Así mismo, recalca que la responsabilidad ética en la inteligencia artificial recae, de manera inclusiva, sobre todo actor que participe en al menos una etapa del ciclo de vida del sistema de IA, es decir, cualquier persona física o jurídica como investigadores, programadores, ingenieros, especialistas en datos, usuarios finales, empresas del sector privado, universidades y entidades públicas y privadas (p. 52)

En ese sentido, la Unesco plantea cuatro valores fundamentales que estos sistemas deben respetar y por ende deben estar en el foco de todos los actores durante todo el ciclo de vida de la IA. Estos son:

1. Respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana.
2. Prosperidad del medio ambiente y los ecosistemas.
3. Garantizar la diversidad y la inclusión.
4. Vivir en sociedades pacíficas, justas e interconectadas.

En este sentido, el llamado es al desarrollo de una serie de valores humanos que protejan, desde los sistemas, la humanidad y su posibilidad de vivir en el planeta de manera pacífica (pp. 83-118).

Para lograrlo, propone una serie de principio éticos que traducen estos valores en orientación a la acción, qué deben tener o hacer estos sistemas para garantizar el cumplimiento de estos valores. Para lo cual desarrolla 10 principios éticos a saber:

1. Proporcionalidad e inocuidad: La IA no es intrínsecamente beneficiosa. Su uso debe ser necesario, adecuado al objetivo legítimo y al contexto, y no ir más allá de lo necesario. Deben aplicarse evaluaciones de riesgo y medidas preventivas. La decisión final en casos irreversibles o de vida/muerte debe ser humana. La IA no debe usarse para calificación social o vigilancia masiva (pp. 119-126).

2. Seguridad y protección: Evitar daños no deseados (riesgos de seguridad) y vulnerabilidades a ataques (riesgos de protección) en todo el ciclo de vida. Fomentar marcos de acceso a datos seguros y respetuosos de la privacidad (pp. 127-128).

3. Equidad y no discriminación: Promover la justicia social, la equidad y luchar contra la discriminación. Garantizar acceso inclusivo a los beneficios de la IA para todos, especialmente grupos marginados. Reducir brechas digitales. Los países avanzados deben ser solidarios con los menos avanzados. Minimizar y evitar sesgos discriminatorios en los sistemas de IA y ofrecer recursos efectivos contra ellos (pp. 129-137).

4. Sostenibilidad: Evaluar continuamente los impactos (humanos, sociales, culturales, económicos, ambientales) de la IA en relación con los objetivos de sostenibilidad (como los ODS – Objetivos de desarrollo sosteninble) (p. 141).

5. Derecho a la intimidad y protección de datos: Respetar, proteger y promover la privacidad en todo el ciclo de vida. La recopilación y uso de datos debe ser coherente con el derecho internacional y los marcos legales, incluyendo el consentimiento informado y bases jurídicas válidas. Establecer marcos de protección de datos y mecanismos de gobernanza adecuados. Realizar evaluaciones de impacto en la privacidad (pp. 143-147).

6. Supervisión y decisión humanas: La responsabilidad ética y

jurídica siempre debe poder atribuirse a personas físicas o jurídicas. La supervisión humana incluye la individual y la pública. Aunque se dependa de la IA por eficacia, los humanos deben retener el control y la responsabilidad final, especialmente en decisiones críticas (pp. 149-152).

7. Transparencia y explicabilidad: Son condiciones previas para el respeto de los derechos humanos y la rendición de cuentas. El grado debe ser adecuado al contexto y al impacto. Las personas deben ser informadas si una decisión se basa en IA, poder solicitar explicaciones y tener acceso a recurso humano. La explicabilidad implica hacer inteligibles los resultados y el funcionamiento interno de los algoritmos (pp. 153-169).

8. Responsabilidad y rendición de cuentas: Los actores de la IA y los Estados son responsables de respetar los derechos humanos y proteger el medio ambiente. La responsabilidad final recae en actores humanos o jurídicos. Se requieren mecanismos de supervisión, evaluación de impacto, auditoría, diligencia debida y protección de denunciantes. La auditabilidad y trazabilidad son clave (pp. 174-177).

9. Sensibilización y educación: Promover la comprensión pública de la IA, los datos, sus impactos (incluidos los derechos humanos) y los aspectos éticos mediante educación accesible, participación cívica y alfabetización mediática (pp. 178-180).

10. Gobernanza y colaboración adaptativas y de múltiples partes interesadas: Es necesaria la participación de diversos actores (gobiernos, sector privado, sociedad civil, academia, etc.) en la gobernanza de la IA. Utilizar normas abiertas y garantizar interoperabilidad. Adaptarse a los cambios tecnológicos e incluir a grupos marginados. Respetar el derecho internacional y la soberanía nacional sobre los datos (pp. 181-186).

Como nos recuerdan los estudios en Ciencia, Tecnología y Sociedad (CTS), ninguna tecnología es neutra (Winner, 1983). Las definiciones y los marcos normativos de la Unesco reconocen el carácter político de los sistemas de inteligencia artificial, aunque no lo nombran explícitamente. Su regulación se alinea con una idea de protección de la humanidad en el ámbito de posibles usos de esta tecnología. Por lo tanto, el enfoque del documento tiende hacia una gobernanza de estos sistemas, buscando un despliegue ético y responsable al entender su carácter tanto positivo como negativo. No obstante, si bien el documento incluye un componente dirigido a la educación y a los usuarios de la IA, en general, la regulación tiende a enfocarse en el desarrollo de la tecnología y sus impactos en todo el ciclo de vida, entendiendo el impacto principalmente como los efectos de la tecnología y su desarrollo per se. Se presta menos atención a cómo los usuarios se apropian de la IA, la adaptan a sus necesidades, resisten sus propuestas o construyen nuevas prácticas sociales a su alrededor, lo cual es central en la interacción entre agentes.

En una línea similar de argumentación que el documento de la UNESCO, la OCDE presenta un análisis, a 2022, de cómo se ha construido de manera integral la gobernanza de la IA, especialmente enfocada en su aplicación en el sector público de américa Latina y el Caribe. El documento define la gobernanza como un conjunto de estrategias, políticas, instituciones y procesos que los gobiernos implementan para dirigir, regular y supervisar el desarrollo y uso de la IA, con el objetivo de maximizar sus beneficios y minimizar sus riesgos.

El documento reconoce que la gobernanza de la IA en el sector público debe abordar una serie de desafíos específicos, como la necesidad de garantizar la transparencia y la explicabilidad de los algoritmos, proteger los derechos fundamentales de los ciudadanos, promover la equidad y la no discriminación, y asegurar la seguridad y la resiliencia de los sistemas de IA.

En este sentido, la OCDE destaca la importancia de establecer marcos éticos sólidos que guíen el diseño, desarrollo e implementación

de la IA, así como la necesidad de crear mecanismos de rendición de cuentas que permitan evaluar y supervisar el impacto de esta tecnología en la sociedad.

El informe enfatiza que la gobernanza de la IA no es una tarea exclusiva del gobierno, sino que requiere de la colaboración y la participación de diversos actores, incluyendo el sector privado, la academia, la sociedad civil y los usuarios finales. Se plantea la necesidad de crear espacios de diálogo y concertación, donde se puedan discutir y definir los valores, principios y normas que deben regir el uso de la IA en el sector público.

En ese sentido, destaca que la gobernanza de la IA no puede ser una tarea exclusiva del gobierno, sino que requiere de la colaboración y la participación activa de diversos actores. Esta perspectiva se fundamenta en el reconocimiento de la complejidad y multidimensionalidad de la IA, cuyas implicaciones trascienden el ámbito puramente técnico y abarcan cuestiones éticas, sociales, económicas, legales y políticas. En este sentido, la OCDE argumenta que la construcción de un marco de gobernanza efectivo exige la convergencia de conocimientos, experiencias y valores de diferentes disciplinas y sectores, para lograr soluciones integrales y legítimas (5).

De esta manera, se identifica a los gobiernos como actores centrales en la gobernanza de la IA, dada su responsabilidad en la definición de las prioridades estratégicas, las inversiones públicas y el marco regulatorio. Sin embargo, se subraya que su papel no debe ser el único, sino que debe complementarse con la contribución del sector privado, que aporta su experiencia técnica, recursos y capacidad de innovación; la academia, que contribuye con la investigación, el conocimiento científico y la formación de talento; la sociedad civil, que representa los intereses de los ciudadanos y defiende los derechos humanos; y los usuarios finales, que aportan su experiencia práctica y necesidades específicas (4).

La OCDE plantea la necesidad de crear "espacios de diálogo y concertación" donde estos diversos actores puedan discutir y definir los

valores, principios y normas que deben regir el uso de la IA en el sector público. Se propone un modelo de gobernanza colaborativo y participativo, donde se fomente el intercambio de ideas, la construcción de consensos y la toma de decisiones compartida. Este enfoque busca evitar que la gobernanza de la IA se convierta en un proceso vertical y unilateral, impuesto por el gobierno, y en su lugar, promover un modelo más horizontal y democrático. Además, se resalta la importancia de que la gobernanza de la IA sea adaptativa y flexible, capaz de evolucionar y ajustarse a los rápidos cambios tecnológicos y a las nuevas necesidades y desafíos que surjan. Se propone la implementación de mecanismos de seguimiento y evaluación que permitan monitorear el impacto de las políticas de IA y realizar los ajustes necesarios para garantizar su efectividad y relevancia.

Como se puede observar, el documento de la OCDE promueve una política participativa en la construcción de políticas de regulación de la IA. En este sentido, reconoce el papel central del Estado en el liderazgo de estas iniciativas, pero también involucra a diversos actores, incluyendo la sociedad civil y los usuarios finales. No obstante, el énfasis principal de esta participación se centra en garantizar el correcto funcionamiento de la IA y en la vigilancia de la misma, desde la experiencia de estos actores. Los usuarios finales y la sociedad civil son concebidos principalmente como usuarios reflexivos, pero con un rol fundamentalmente pasivo en la interacción y la configuración activa de la tecnología.

Esta perspectiva, si bien reconoce la importancia de la retroalimentación y el control externo, podría no capturar plenamente la agencia y la capacidad de apropiación de los usuarios. La OCDE se enfoca en cómo los usuarios pueden señalar problemas o riesgos potenciales, lo cual es crucial, pero dedica menos espacio a explorar cómo los usuarios pueden moldear activamente la IA, adaptarla a sus necesidades específicas, o incluso desafiar las formas en que se les propone utilizarla. En otras palabras, la gobernanza de la IA, tal como se presenta en el documento de la OCDE, se inclina hacia un modelo de "vigilan-

cia" y "control", donde el gobierno y los expertos definen las reglas y los usuarios las observan y reportan desviaciones. Se presta menos atención a la idea de la IA como un espacio de interacción dinámica, donde los usuarios pueden ser agentes activos en la configuración de la tecnología y sus usos.

En línea con las recomendaciones de la UNESCO y la OCDE, en 2024 el gobierno colombiano lanza una "Hoja de Ruta Para el Desarrollo y Aplicación de la Inteligencia Artificial en Colombia", promovida por el trabajo de MinTIC y Minciencias. La "Hoja de Ruta" presenta un enfoque de la gobernanza de la IA que va más allá de la simple regulación, concibiéndola como un marco integral para el desarrollo y la implementación responsable de esta tecnología en Colombia. En línea con la política actual, la gobernanza se convierte en un eje central que articula consideraciones éticas, políticas, sociales y económicas, buscando garantizar que la IA se utilice para el beneficio de la sociedad y el progreso del país.

Al igual que los otros documentos, la hoja de ruta enfatiza que la gobernanza de la IA debe estar estrechamente vinculada con la ética, reconociendo que los sistemas de IA no son intrínsecamente neutrales y que su desarrollo y uso pueden plantear dilemas morales y riesgos sociales. En este sentido, la gobernanza debe establecer principios y mecanismos que garanticen la transparencia, la equidad, la no discriminación, la seguridad y la rendición de cuentas de los sistemas de IA, protegiendo los derechos fundamentales y la dignidad humana. Además, la "Hoja de Ruta" destaca la importancia de la gobernanza como un instrumento para promover la participación ciudadana y la inclusión en el desarrollo de la IA. Se busca que los usuarios, las comunidades y los diversos grupos de la sociedad tengan voz en la definición de las políticas y prácticas que moldean el uso de esta tecnología, fomentando así la confianza y la legitimidad de los sistemas de IA.

En este contexto, al igual que lo planteado por la OCDE, la gobernanza se convierte en un proceso colaborativo que involucra a múltiples actores, incluyendo el gobierno, el sector privado, la academia, la

sociedad civil y los usuarios finales. Se reconoce que la complejidad de la IA y sus potenciales impactos requieren de un diálogo abierto y una toma de decisiones compartida, donde se consideren las diferentes perspectivas y valores.

La "Hoja de Ruta" también aborda la gobernanza desde una perspectiva de gestión de riesgos, reconociendo la necesidad de identificar, evaluar y mitigar los posibles efectos negativos de la IA. Se plantea la importancia de establecer mecanismos de supervisión y control, así como la implementación de auditorías y evaluaciones de impacto, para garantizar que los sistemas de IA se utilicen de manera responsable y que sus beneficios se distribuyan de manera equitativa.

Siguiendo esta línea política, en febrero de 2025 el gobierno colombiano lanza su primer CONPES (Consejo Nacional de Política Económica y social) sobre IA. En este documento se define la gobernanza de la IA como un eje estratégico central para el desarrollo y aprovechamiento de esta tecnología en Colombia. Se reconoce que la gobernanza no se limita a la mera regulación, sino que implica la creación de un marco institucional y normativo que oriente y supervise el diseño, desarrollo, implementación y uso de los sistemas de IA.

La política de IA plantea la necesidad de establecer mecanismos de coordinación y articulación entre las diferentes entidades del gobierno, así como con otros actores relevantes como el sector privado, la academia y la sociedad civil. Se busca garantizar la coherencia y la sinergia de las acciones en materia de IA, evitando la fragmentación y la duplicidad de esfuerzos. En este sentido, la gobernanza se convierte en un instrumento para definir una visión estratégica compartida sobre el futuro de la IA en Colombia, estableciendo prioridades, objetivos y líneas de acción que permitan aprovechar el potencial de esta tecnología para impulsar el desarrollo económico, social y ambiental del país.

Además, el CONPES 4144 destaca la importancia de la ética como un componente fundamental de la gobernanza de la IA. Se plantea la necesidad de desarrollar marcos éticos sólidos que guíen el diseño y

uso de los sistemas de IA, protegiendo los derechos fundamentales de los ciudadanos y promoviendo valores como la transparencia, la equidad, la no discriminación y la rendición de cuentas. La política de IA reconoce que la gobernanza debe ser adaptativa y capaz de evolucionar en respuesta a los rápidos avances tecnológicos y a los nuevos desafíos que surjan. Se propone la creación de mecanismos de seguimiento y evaluación que permitan monitorear el impacto de la IA en la sociedad y realizar los ajustes necesarios para garantizar su uso responsable y beneficioso.

Es interesante notar cómo desde el gobierno colombiano se da un mayor énfasis a la participación ciudadana no solo en la definición de políticas, sino también en la práctica que moldean las tecnologías, sobre todo en lo que refiere al fomento de la confianza y la legitimidad de los sistemas de IA desde la hoja de ruta. Sin embargo, al aterrizarlo a la política económica y social (CONPES) el control recae una vez más sobre la tecnología en sí y su impacto al orientarse hacia el diseño, desarrollo e implementación, no obstante, deja abierta una ventana en lo que se refiere al uso de los sistemas de IA.

Esta diferencia de enfoque plantea una tensión fundamental en la gobernanza de la IA. Mientras que la Hoja de Ruta parece reconocer el papel activo de los usuarios en la configuración de la tecnología a través de sus prácticas de uso y apropiación, el CONPES se inclina más hacia un modelo de gobernanza donde el Estado y los expertos definen los límites y los usuarios son principalmente sujetos a los resultados de esta regulación.

Ampliando el alcance de la gobernanza en IA

Como bien podemos ver a través de los cuatro documentos expuestos acá, a nivel internacional, y también en Colombia, hay una gran preocupación por el impacto de las tecnologías de IA, sobre todo en el marco de su regulación dado los potenciales

alcances de las mismas, pero también, las fantasías que se crean ante la emergencia de estas nuevas tecnologías (tema que desarrollaré a continuación). El enfoque de tal regulación esta puesta en la gobernanza, pero, como he venido analizando, tal visión de la gobernanza presenta algunas limitaciones que debemos considerar a detalle.

El 4 de agosto de 1997 Skynet entra en línea y el 29 de agosto de ese mismo año toma conciencia de sí misma. El 12 de enero de 1992 en una fábrica en Urbana Illinois, Hal9000 entra en funcionamiento, más adelante, es considerado miembro confiable de la tripulación del Discovery One, Finalmente Hal9000 muestra fallas atentando contra el resto de la tripulación y tiene que ser desactivado mientras asegura tener miedo y suplica por su vida. En algún momento del siglo XXI la humanidad creó la Inteligencia artificial, no se sabe quién atacó primero, pero los humanos destruyeron el cielo para acabar con la fuente de energía de las máquinas, en retaliación las máquinas construyeron la Matrix para mantener dóciles a los humanos mientras los usaban como baterías.

El párrafo anterior, es un collage de imágenes donde se nos muestra cómo desde la industria cultural Hollywoodense se nos ha construido una visión de los posibles efectos de la inteligencia artificial. La narrativa enfatiza en una agencia de las máquinas consientes que finalmente terminan por revelarse en contra de los humanos llevándolos al borde de su destrucción. Con la emergencia de las IA generativas y el lanzamiento de ChatGPT a finales del 2022, estas mismas narrativas se van introduciendo a la visión general que se construye de la IA y es reforzada frecuentemente en redes sociales y medios de comunicación por diversos desarrolladores, líderes de opinión y algunos analistas al señalar los peligros de la IA.

Basta con una rapida búsqueda en Google para encontrar titulares alarmistas y apocalípticos como: "Sam Altman da la razón a Elon Musk sobre el siguiente paso de la IA: "El mundo no está preparado"" (Luque, 2025); "Elon Musk afirma que la Inteligencia Artificial terminará con los empleos" (Murphy, 2024); "La inteligencia artificial puede llevar a

la extinción de la humanidad: la advertencia de los grandes expertos en IA" (BBC News Mundo, 2023), por citar algunos ejemplos.

Además de los líderes de opinión y de compañías tecnológicas, hay una serie de analistas que refuerzan este tipo de posiciones. El historiador Israelita Yuval Harari, por ejemplo, es uno de los que se subido en esta ola sensacionalista de la IA para de cierta manera, reforzar estas visiones apocalípticas que traemos desde la industria cultural.

En su libro Nexus, Harari (2024) analiza la IA no simplemente como una herramienta más, diferenciándola de avances tecnológicos previos como la imprenta o la energía nuclear, sino como un agente activo capaz de operar y tomar decisiones de forma autónoma (Similar a Skynet, Hal9000, entre otros). Esta distinción es fundamental para Harari, ya que, a partir de esto, la IA puede generar contenido, desarrollar otras IA e incluso influir en el mundo de maneras que escapan al control humano directo, planteando un desafío sin precedentes a nuestra capacidad de comprender y gestionar nuestro entorno. Entre otras cosas, el autor plantea que tal agencia nos llevará a la automatización de las decisiones. Al permitir que los sistemas de IA asuman responsabilidades en ámbitos críticos como la guerra, las finanzas o la política, para el autor, nos estaríamos arriesgando a perder el control sobre eventos de magnitudes catastróficas. Por otro lado, Harari también advierte que la IA, al procesar vastas cantidades de datos sobre cada individuo, podría llegar a conocernos mejor de lo que nos conocemos a nosotros mismos, lo que le permitiría manipular nuestras emociones y decisiones, socavando así nuestra autonomía y el libre albedrío. Así mismo, en el ámbito laboral, la automatización a gran escala impulsada por la IA amenaza con hacer que millones de personas sean innecesarias, generando una crisis de utilidad humana y de significado existencial.

Consecuentemente, Harari enfatiza la urgente necesidad de regulación. Propone medidas concretas, como la prohibición de que los algoritmos se hagan pasar por humanos, la implementación de acuerdos internacionales para limitar los usos militares de la IA y,

crucialmente, la prohibición de que la IA tome decisiones autónomas en situaciones de vida o muerte. Para Harari, la supervivencia de la democracia y la esencia misma de la humanidad dependen de nuestra capacidad para establecer límites claros y éticos al desarrollo y despliegue de la IA, asegurando que esta poderosa tecnología sirva a los intereses humanos en lugar de dominarlos. Esta necesidad de regulación que propone Harari se alinea perfectamente con los llamados de los distintos organismos acá expuestos (UNESCO, OCDE, Gobierno Colombiano), no obstante, dado el hilo de la argumentación queda la sensación que más que un llamado a la regulación termina erigiéndose como una propuesta de resistencia de la humanidad frente a la agencia de la IA. Pero ¿qué lectura estamos haciendo de la IA? ¿dónde queda la agencia de los humanos frente a la IA? ¿cuál es la diferencia entre resistencia y gobernanza?

El pedagogo y filósofo rumano Stephen Popenici en una entrevista para el Journal of Applied Learning & Teaching (Rudolph y Popenici, 2023) argumentaba que la idea de inteligencia en Inteligencia artificial, más que denotar una noción clara de inteligencia, aparecía como un concepto de Marketing y en ese contexto, se vuelve esencial entender cómo se construye y qué implica pensar en inteligencia en el marco de la inteligencia artificial.

En ese sentido, tanto Popenici (2022) como la investigadora, periodista y profesora universitaria Meredith Broussard (2018; 2023) argumentan que la concepción de inteligencia en la IA se vincula directamente con la creencia histórica de que esta es una cualidad medible. Sostienen que esta perspectiva, que busca clasificar a las personas según sus puntajes de inteligencia, tiene sus orígenes y se nutre directamente de la doctrina de la eugenesia planteada por Francis Galton, demostrando cómo, desde sus orígenes, la idea misma de inteligencia en la IA reproduce ideas de superioridad racial y refuerza el poder geopolítico del conocimiento.

Para Galton la inteligencia se entendía como un rasgo principalmente hereditario y cuantificable, que se manifestaba en la "emi-

nencia" social. Sus investigaciones se centraron en la medición de operaciones mentales básicas y sensoriales, como la agudeza visual y auditiva o los tiempos de reacción, buscando establecer jerarquías y justificar la selección. La creencia era que una mayor velocidad o precisión en estas funciones elementales indicaba una inteligencia superior; no obstante, los estándares de puntuación estaban dados por las características de las personas de la clase alta británica. Si bien los posteriores tests de IQ, como los de Alfred Binet (y sus adaptaciones eugenésicas por Lewis Terman), evolucionaron para incluir habilidades cognitivas más complejas (comprensión verbal, memoria, razonamiento lógico), mantuvieron la premisa fundamental de Galton: la inteligencia es una cualidad singular que puede ser reducida a un puntaje numérico, permitiendo la clasificación y jerarquización de los individuos y manteniendo los mismos puntos de referencia que perpetuaban una geopolítica del conocimiento basada en la superioridad de ciertos grupos

Esta concepción de la inteligencia como algo inherentemente medible y cuantificable es lo que, según autores como Stefan Popenici (2022) y Meredith Broussard (2018; 2023), permea la arquitectura y el funcionamiento de la Inteligencia Artificial. La IA opera bajo el principio de que para "entender" o "procesar" el mundo, todo debe ser traducido a datos numéricos y patrones analizables. Los algoritmos de la IA, al igual que los tests de IQ que buscan puntuar y clasificar a las personas, reducen la compleja realidad social y humana a variables computables. La "inteligencia" de un sistema de IA reside en su capacidad para calcular la probabilidad de que ciertos eventos ocurran, que ciertas palabras sigan a otras, o que un patrón visual corresponda a una etiqueta, basándose en la recurrencia estadística de los datos.

En última instancia, la IA no "piensa" en un sentido humano, sino que infiere y organiza probabilísticamente relaciones basadas en los datos sobre los que fue entrenada. Los sistemas de IA están diseñados para optimizar resultados y hacer predicciones, lo que se traduce en una constante labor de clasificación y asignación de

"puntuaciones". Esta forma de operación es una sofisticada extensión de la lógica galtoniana: la inteligencia se define por lo que es observable, cuantificable y estadísticamente predecible. Este enfoque no solo replica los sesgos presentes en los datos históricos, sino que, al reducir la inteligencia a métricas y probabilidades, solidifica una visión de la capacidad humana que se originó en doctrinas discriminatorias como la eugenesia, inscribiendo así el sesgo directamente en la lógica operativa de los sistemas de IA. Y es aquí donde volvemos al dato y su centralidad como un activo económico que permea la visión del mundo construida desde las posiciones apocalípticas que reivindican una agencia de la IA mientras invisibilizan la agencia humana alrededor de la misma. El filósofo italiano Bifo Berardi (2017) argumenta que nos encontramos en un proceso de subsunción hiperreal. Referente a la producción y a la plusvalía, Marx distinguía la dominación formal de la real. La dominación o subsunción formal es cuando a través de la fuerza bruta se obliga a los individuos a aceptar la explotación en condiciones primordiales de una extracción de plusvalía absoluta. Mientras que la subsunción real introduce las máquinas como una forma de extraer más plusvalía a través de la automatización de procesos y en este sentido el dominio es encarnado por las máquinas.

> *Cuando pasamos de la máquina industrial a la informática, la regulación de los actos de producción ya no depende de la automatización mecánica, sino que se incorpora al lenguaje y la cognición. En este punto, podemos hablar de una subsunción hiperreal, lo que supondría una subsunción mental: la captura y el reformateo de la mente* (Berardi, 2017, p. 134).

En este contexto de producción automatizada, es el dato el que regula la realidad y el futuro reduciendo la creación de la realidad a la posibilidad de representación en datos y posibilitando la inteligencia en la inteligencia artificial. De esta manera, el ser humano y su realidad se constituyen en lo que sus datos (y la forma como se clasifican y predicen) dice que es, es decir, se constituye en el homodata.

El dato, en lugar de constituirse como una abstracción, se transforma en un valor de verdad que está justificado en modelos de probabilidad estadística extraídos de toda relación de poder y de política de conocimiento. Y si el algoritmo falla, es por falta de datos o por error humano, pero estos errores pueden ser corregidos sin importar las consecuencias materiales de los mismos.

Por ello, en los documentos analizados, la idea de gobernanza recae justamente sobre el valor y uso de dato. Vemos como en las recomendaciones tanto de la UNESCO, OCDE y las guías del gobierno colombiano, a pesar de reconocer la complejidad ética y social de la IA, su enfoque principal se orienta a establecer marcos que regulen la producción, recolección, uso y protección de datos, así como la transparencia y explicabilidad de los algoritmos que los procesan. Esta perspectiva, si bien crucial para mitigar riesgos como los sesgos y la discriminación producto de la representatividad (la falta de) del dato, tiende a posicionar la gobernanza en el control del flujo informático y la validación técnica de sus resultados. Desde esta perspectiva, la gobernanza puede ser vista como una mera gestión, a través de la cual se neutraliza el conflicto y se anula la imaginación colectiva donde la democracia es reemplazada por la administración tecnocrática del capital (Berardi, 2017, p. 118)

De Souza Santos (2007) justamente crítica el concepto de gobernanza señalando como las metodologías de concertación tienden a anular el papel del estado sometiéndolo a los intereses de un grupo determinado al invisibilizar el conflicto y las relaciones de poder que se dan al interior de las negociaciones, que, si bien implica la participación de una serie de actores diferenciales, implica, según el autor, una serie de decisiones que favorecen sólo a los interesados en un tema específico. A esto va a llamar gobernanza neoliberal (De Souza Santos, 2007). No obstante, el autor también reconoce la existencia de un modelo de gobernanza contrahegemónico, que ejemplariza en resultados de políticas adoptados por el foro social mundial. En estos modelos, se materializa la crítica social, se visibiliza el conflicto y la

desigualdad y se incorporan realmente las voces de actores diversos (De Souza Santos, 2007) Este enfoque rechaza la idea de la gobernanza como mera gestión tecnocrática o descentralización institucional, y la entiende como un campo de disputa y acción política situada. Así, La gobernanza debe pensarse desde las prácticas de los sujetos colectivos, desde abajo, como una forma de construcción de poder territorial alternativo que desborda las lógicas del Estado y se nutre de saberes ancestrales, comunitarios y de economías otras.

Bevir (2011) también analiza distintos modelos de gobernanza dando pie a la existencia de modelos que rompen con el determinismo institucional de la gobernanza al poner énfasis en el carácter contingente construido del orden social, donde la gobernanza abarca todas las formas de tomas de decisiones colectivas, desde acciones impulsada por movimientos sociales, hasta decisiones formales adoptadas por gobiernos. En este sentido, la gobernanza de la inteligencia artificial, lejos de ser un mero conjunto de políticas públicas o la relación con la institucionalidad para su regulación, debe entenderse como un proceso dinámico y co-constructivo en el que los usuarios no son sujetos pasivos, sino agentes activos capaces de apropiar, modificar e incluso resistir las tecnologías en sus prácticas cotidianas.

Teniendo en cuenta que la tecnología es modelada por factores económicos, políticos, culturales y sociales, y a su vez contribuye a configurar las formas de vida, relaciones sociales y estructuras institucionales. La forma en que se desarrolla y adopta una tecnología depende tanto de trayectorias históricas como de negociaciones sociales, decisiones políticas y contextos culturales particulares. Por ello tanto tecnología y sociedad son mutuamente constituidas (MacKenzie & Wajcman, 1999). Es por ello por lo que las sociedades deben implicarse activamente en la orientación del cambio tecnológico, más allá de la simple modernización (actualización tecnológica), debe haber un proceso activo de democratización que permita darle forma a la sociedad (MacKenzie & Wajcman, 1999), es decir, es necesario

pensar en la capacidad de los usuarios para moldear activamente la tecnología o resistir las lógicas que construyen la "inteligencia" y la realidad a partir de ellos.

¿Qué es entonces la inteligencia en inteligencia artificial?

De entrada, quiero decir que no responderé esta pregunta en este apartado (y en general en todo el texto), sino más bien que se plantea como una provocación para analizar cómo nos estamos relacionando con la tecnología y específicamente con la IA. Entender la IA como una herramienta nos permite alejarnos del carácter apocalíptico reificante con el cuál algunos analistas fetichizan el poder y la agencia de la misma.

Sin embargo, tanto MacKenzie & Wajcman (1999), como Winer (1983) nos advierten de la mutua influencia entre tecnología y sociedad, es decir, de la necesidad de contextualizar e historizar la relación que se da entre estas, con el objeto justamente de entender el lugar político de la tecnología. No se trata entonces de neutralizar los efectos de la IA al atribuirle su definición a la exageración de la agencia de la misma, sino más bien de mirar críticamente, en qué contexto se construye esta agencia, con qué intencionalidad y cómo se tensiona con otras formas de entender y vivir la tecnología.

Acá es crucial entender las relaciones de poder que se dan en el marco de la producción e implementación de la tecnología, así como los intereses políticos y económicos que esta reproduce. Es decir, necesitamos entender la IA en el contexto histórico que es producida y apropiada y eso quiere decir que necesitamos localizar esos contextos dada su variabilidad.

Quiero decir, que evidentemente el equivalente de Skynet no se va a levantar mañana e iniciar una guerra mundial; pero eso no significa que no existan riesgos e impactos potencialmente negativos

de la implementación de la IA. Y no porque la tecnología sea intrínsicamente buena o mala, sino más bien por el tipo de relaciones que se constituyen alrededor de ésta y no sólo por el uso de la misma, sino por la forma como a través de ella se consolidan o modifican las estructuras sociales y políticas a través de las cuales las tecnologías se erigen. Es decir, los riesgos de la IA no se dan por la agencia fetichizada y la novedad de su tecnología, sino por la articulación de esta con el contexto social, político y económico particular en el que se desarrolla. Dotar de agencia, de inteligencia, a la inteligencia artificial dice más de la manera cómo nos constituimos nosotros mismos como sujetos inteligentes en nuestro contexto histórico concreto, que de si una máquina es inteligente o no. Pero también de cómo se construye el valor de la inteligencia y de lo humano en tal contexto.

En ese sentido, la IA es una herramienta cuya agencia y autonomía está dada en la manera en cómo es usada y apropiada por agentes concretos con intenciones diversas y es justamente en la interacción y la forma como se constuyen estas relaciones y el poder intrínseco en ellas donde es necesario repensar el papel de la IA y hacia donde debe dirigirse la gobernanza.

La necesidad de una nueva mirada sobre la IA, una agenda programática

Lo anterior, más que certezas propone una serie de rutas que es necesario explorar para continuar ahondando sobre los impactos reales de la IA. La mirada, es una mirada prevenida y crítica, pero en medio de todo optimista.

Partimos de una concepción dominante de la necesidad del control sobre los datos en relación con la IA en distintos momentos de su ciclo de vida. Desde cómo se entrenan, dónde y a partir de quiénes se entrenan y cómo se usan y cómo impacta ese uso determinado. Así mismo, la agencia de los sujetos involucrados, incluida la IA, muchas

veces mirada como sujeto, o en todo caso en una posición liminal entre herramienta y sujeto, se entiende alrededor del control de los datos, los que produce y los que usa y, por ende, la gobernanza se piensa alrededor de cómo controlar estos datos. Pero ¿hay vida más allá de los datos? O por el contrario ¿estamos determinados por lo que los datos dicen que somos y controlarlos es la manera de apropiarnos de nuestras agencias?

Quiero profundizar este cuestionamiento a partir de un ejercicio que he venido realizando con mis estudiantes de la materia CTS de Máquinas pensamiento mágico e Inteligencia Artificial en la universidad Icesi. Con el objeto de identificar sesgos implícitos en los productos de la IA hacemos un ejercicio comparativo. Por un lado, leemos el texto de Rodrigo Bástidas (2021) "Desmantelar patentes para crear universos propios" que aparece como prólogo a su antología de ciencia ficción latinoamericana "el tercer mundo después del sol". El texto es una reflexión de lo que implica hablar de ciencia ficción en Latinoamérica, sobre todo teniendo en cuenta la geopolítica del conocimiento, donde usualmente se nos ubica como consumidores y donde, además, cualquier conocimiento de otra índole es calificado como magia (de ahí que el género literario más asociado a la región sea el realismo mágico).

Por otro lado, les pido que sin hacer un prompt muy elaborado, le soliciten a ChatGPT, Gemini o alguna IA de su predilección (hasta el momento no hemos realizado el ejercicio con DeepSeek) que escriba un cuento de ciencia ficción latinoamericano algo así:

> *Escribéme un cuento corto (3 párrafos) de Ciencia ficción latinoamericana.*
>
> *Hace doscientos años, la Tierra sufrió un colapso ecológico, forzando a la humanidad a buscar refugio en colonias espaciales. Para preservar la rica diversidad cultural que casi perdemos, cada colonia fue diseñada para recrear un ecosistema y una sociedad específicos. En el* **Núcleo 3**, *la mayor parte de la población era descendiente de latinoamericanos, y la estación replicaba una versión idealizada de la Sudamérica pre-cataclismo. El aire olía a tierra mojada y café*

recién molido, la música de la cumbia y el vallenato resonaba por los pasillos, y las paredes holográficas proyectaban selvas exuberantes y picos andinos. La vida era pacífica, incluso idílica, pero la constante nostalgia por la Tierra, que se sentía tan cerca y, a la vez, tan inalcanzable, era un eco que resonaba en los corazones de sus habitantes.

Un día, una anomalía en los sistemas de navegación desvió una nave de exploración. En lugar de regresar al Núcleo 3, el capitán, el joven y ambicioso ***Mateo****, se encontró al borde del sistema solar de la Tierra. Las imágenes de lo que encontró dejaron a la colonia sin aliento: el planeta azul estaba recuperándose. Grandes extensiones de selva virgen cubrían lo que antes fueron desiertos, y los océanos, aunque aún marcados por la cicatriz del pasado, brillaban con una nueva vida. Mateo, con su corazón latiendo al ritmo de un tambor ancestral, comprendió que había presenciado algo milagroso. La Tierra, su hogar ancestral, no era una memoria distante, sino un futuro posible.*

A su regreso, la noticia de Mateo encendió una chispa de esperanza y debate. Algunos, arraigados a la seguridad del Núcleo 3, temían lo desconocido. Otros, impulsados por un anhelo profundo de reconectar con sus raíces, veían una oportunidad de redención. La imagen de la Tierra regenerada se convirtió en un símbolo, no solo de supervivencia, sino de la capacidad infinita de la naturaleza para curarse a sí misma, y la de la humanidad para aprender de sus errores. ¿Sería el momento de regresar al hogar? El dilema se cernía sobre el Núcleo 3, una pregunta tan vasta como el cosmos, pero tan íntima como el latido de un corazón que recuerda el ritmo de su tierra. (Generado con Gemini — 2.5 flash- el 30 de mayo de 2025).

El resultado casi siempre es el mismo, como se puede ver en el texto acá consignado, es una reproducción de estereotipos sobre qué es lo latinoamericano y qué sería la ciencia ficción latinoamericana. De hecho, suelo hacer que le pregunten a la IA qué es lo latinoamericano de su cuento y qué sería ciencia ficción latinoamericana de la siguiente manera:

En este cuento, lo *latinoamericano* se manifiesta de varias maneras:

- Identidad y Herencia Cultural: La colonia "Núcleo 3" está poblada mayoritariamente por descendientes de latinoame-

ricanos y recrea una versión idealizada de Sudamérica. Esto se refleja en los detalles sensoriales: el aire que huele a tierra mojada y café, la música de cumbia y vallenato, y las proyecciones holográficas de selvas y picos andinos. La nostalgia por la Tierra y el "latido de un tambor ancestral" en Mateo resaltan la profunda conexión emocional y cultural con su origen.

- Narrativa de Resistencia y Esperanza: A pesar del colapso ecológico, hay una persistencia de la cultura y un anhelo de regreso a la tierra. La recuperación de la Tierra se convierte en un símbolo de esperanza y redención, reflejando la resiliencia y la capacidad de regeneración, temas recurrentes en narrativas latinoamericanas que a menudo abordan la superación de adversidades.

- Dilema del Retorno: La pregunta de si regresar a la Tierra o permanecer en la seguridad de la colonia genera un debate que evoca la tensión entre la tradición y el progreso, el arraigo y la aventura, y la búsqueda de un hogar, temas muy presentes en la literatura y la experiencia latinoamericana.

La ciencia ficción latinoamericana se puede entender como un subgénero que, si bien utiliza elementos y tropos comunes de la ciencia ficción (colonias espaciales, futuros distópicos/utópicos, tecnología avanzada), lo hace con una *sensibilidad cultural y temática propia de América Latina*. Se distingue por:

- Contexto Sociopolítico: A menudo explora temas como la colonización, la identidad, la desigualdad, la migración, la memoria histórica y la relación con la naturaleza desde una perspectiva regional.

- Elementos Culturales Propios: Incorpora la rica diversidad cultural, mitológica y folclórica latinoamericana, entrelazándola con lo tecnológico o lo futurista, como se ve en el cuento con la música, los aromas y la geografía.

- Distopías y Utopías con Raíces Locales: Las visiones de futuros alternativos suelen estar teñidas por las realidades y problemáticas específicas de la región, ya sea explorando el impacto del progreso en sociedades tradicionales o imaginando futuros utópicos basados en valores comunitarios.
- Enfoque en lo Humano y lo Emocional: Más allá de los avances tecnológicos, suele poner un fuerte énfasis en las experiencias humanas, los dilemas morales y las conexiones emocionales, lo que la hace profundamente arraigada en la condición humana.

En resumen, la ciencia ficción latinoamericana no es solo ciencia ficción escrita por autores latinoamericanos, sino una que *incorpora y dialoga con la idiosincrasia, la historia y las preocupaciones de la región*, dándole un sabor y una perspectiva únicos al género. (Generado con Gemini –2.5 flash- el 30 de mayo de 2025– los resaltados son de la IA). Con esto, vemos como la Inteligencia artificial, reproduce todo aquello que se adjudica a Latinoamérica, incluso su imposibilidad de hacer ciencia (ciencia ficción) por fuera de los cánones hegemónicos.

En este punto, la discusión gira en torno a los datos de entrenamiento, y cómo las poblaciones que producen los datos con los que se entrenan las IAs, tienden a mirar Latinoamérica de esta forma, perpetuando, a través de los productos de la IA estos sesgos. Es decir, lo que ya se ha vuelto un mantra de las críticas hacia la IA, las inteligencias artificiales reproducen (y perpetúan) los sesgos y la desigualdad. Nótese acá que el énfasis es finalmente en el producto de la IA como un hecho absoluto; pero ¿es la gente tan ingenua como para quedarse o no darse cuenta de este tipo de sesgos?

Ahora, no quiero decir que cualquier persona que use la IA se va a dar cuenta y va a reflexionar sobre cómo se producen y reproducen los sesgos, esta es una discución que tenemos a menudo los profesores de distintos niveles sobre la necesidad de alfabetización de la

gente alrededor del uso crítico de la IA; sin embargo, la propuesta es justamente descentrar la mirada del dato como un producto que nos determina y entrar a investigar la agencia humana alrededor del dato y de la tecnología misma.

Evidentemente la IA produce y reproduce sesgos, y los resultados de esto llevan a profundizar las brechas de desigualdad como lo han denunciado diversos movimientos que propenden por la justicia algorítmica (O'Neil, 2017; Ricuarte, 2024, por mencionar un par de casos). Y esta es una lucha que hay que seguir dando y hay que seguir generando herramientas y mecanismos para diversificar los modelos y hacerlos más incluyentes. Pero además de eso, también es necesario preguntarnos, ¿cómo los humanos, como agentes, nos apropiamos y resignificamos los datos que producen las Inteligencias Artificiales?

En esta discusión en clase, un estudiante trajo a colación, a propósito de la ciencia ficción, el afro-futurismo, como una vertiente de la ciencia ficción que resignifica y relabora el rol de los afrodescendientes (generalmente desde los Estados Unidos) en el género literario, al incorporar fuertes elementos identitarios, ancestrales y pensarse el futuro de maneras disidentes. Lo anterior, entonces, abre toda una serie de posibilidades de investigación donde el rol del usuario se vuelve más activo y nos invita a pensar cuál es el impacto real de la tecnología.

Pienso en trabajos de investigación, que involucren grupos focales con artistas de diversas colectividades, de lenguas indígenas, de comunidades y minorías por todo el mundo, donde se expongan y se trabajes manera de apropiarse críticamente de esos sesgos, de cómo reelaborar los productos de la IA, en el sentido que, evidentemente sé que la información que produce la IA está sesgada ¿qué podemos hace con eso y cómo podemos usarlo a nuestro beneficio?

Eso implica entonces una agenda de investigación que descentre la mirada sobre el dato y que vuelva a pensar la tecnología como una herramienta, que no es neutra, sino que se inserta en condicio-

nes concretas que la producen. Plantea la necesidad de mapear a los actores que participan en el contexto de la IA, no sólo como sector productor de tecnología, sino dentro de un contexto histórico determinado y con determinados intereses particulares. Implica repensar la gobernanza, para ampliar el control sobre los datos y entender de forma más asertiva los usos y los impactos reales de la interacción con las diferentes tecnologías.

Finalmente, implica repensar la agencia, tanto de los humanos alrededor de la IA, como se resignifica la idea de lo humano al contacto con esta, pero también, la agencia misma de la IA como tecnología, sin caer en fetichismos y agencias exageradas como las posiciones acá criticadas.

Referencias

Asimov, I. (2022) . *Fundación*. Debolsillo.

Bástidas, R. (2021). Desmantelar patentes para crear universos propios. En R. Bástidas (Ed.), *El tercer mundo después del sol*. Minotauro.

BBC News Mundo. (2023, 30 de mayo). La inteligencia artificial puede llevar a la extinción de la humanidad: la advertencia de los grandes expertos en IA. *BBC News Mundo*. Recuperado de https://www.bbc.com/mundo/noticias-65759254

Berardi, F. (2017). *Futurabilidad: El deseo imposible de la revolución*. Caja Negra Editora.

Bevir, M. (2011). *The SAGE Handbook of Governance*. SAGE Publications.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. The MIT Press

De Sousa Santos, B. (2007). Más allá de la gobernanza neoliberal: el Foro Social Mundial como legalidad y política cosmopolitas subalternas. En B. de Sousa Santos y C. Rodríguez Garavito (Eds.), *El derecho y la globalización desde abajo. Hacia una legalidad cosmopolita*. Anthropos

Departamento Nacional de Planeación. (2025, 14 de febrero). Política Nacional de Inteligencia Artificial (CONPES 4144) CONPES 4144 de 2025: Se menciona como un documento del Consejo Nacional de Política Económica y Social de Colombia.

Harari, Y. N. (2024). *Nexus: Una breve historia de las redes de información desde la edad de piedra hasta la IA*. Penguin Random House.

Haraway, D. (1991). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. In *Simians, cyborgs, and women: The reinvention of nature* (pp. 149-181). Routledge.

Luque, E. (2025, 23 de mayo). Sam Altman da la razón a Elon Musk sobre el siguiente paso de la IA: "El mundo no está preparado". *Computer Hoy*. Recuperado de https://computerhoy.20minutos.es/tecnologia/sam-altman-da-razon-elon-musk-siguiente-paso-ia-mundo-no-preparado-1462674

MacKenzie, D., & Wajcman, J. (1999). *The social shaping of technology* (2nd ed.). Buckingham: Open University Press.

Ministerio de Ciencia, Tecnología e Innovación. (2024). *Hoja de Ruta Para el Desarrollo y Aplicación de la Inteligencia Artificial en Colombia*.

Murphy, S. (2024, 23 de mayo). Elon Musk advierte que la inteligencia artificial acabará con todos los empleos. *CNN en Español* Recuperado de https://cnnespanol.cnn.com/2024/05/23/elon-musk-inteligencia-artificial-empleos-trax

OCDE/CAF. (2022). *Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe* (Estudios de la OCDE sobre Gobernanza Pública). OECD Publishing. https://doi.org/10.1787/5b189cb4-es

O'Neil, C. (2017). *Armas de destrucción matemática: Cómo el big data aumenta la desigualdad y amenaza la democracia*. Capitán Swing.

Parra-Romero, A., & Castro-Gil, R. (en prensa). Capítulo 1. El campo de estudio en Ciencia, Tecnología y Sociedad - (CTS) como herramienta para el desarrollo de competencias en formación ciudadana. En *Prácticas docentes en los cursos electivos en Ciencia, Tecnología y Sociedad (CTS) y el desarrollo de competencias en ciudadanía y pensamiento crítico: una mirada desde las prácticas docentes*. Universidad Icesi

Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press

Popenici, S. (2022). *Artificial Intelligence and Learning Futures: Critical Narratives of Technology and Imagination in Higher Education*. Routledge.

Ricaurte Quijano, P. (2024, 30 de septiembre). "No se puede poner en el mundo una tecnología que no ha sido probada" | La mexicana Paola Ricaurte Quijano, referente en inteligencia artificial y feminismo. *Página|12*. Recuperado de https://www.pagina12.com.ar/771022-no-se-puede-poner-en-el-mundo-una-tecnologia-que-no-ha-sido-

Rudolph, J. y Popenici, S. (2023). A critical perspective on generative AI and learning futures. An interview with Stefan Popenici. *Journal of Applied Learning & Teaching*, 6(2),1-21 DOI: https://doi.org/10.37074/jalt.2023.6.2.5

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 49(236), 433-460

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Winner, L. (1983). Do artifacts have politics? *Daedalus*, 112(1), 121-136.

[Capítulo 3]

Redes sociales e inteligencia artificial: subjetividades, vínculos y algoritmos en la era digital

Edilberto Loaiza Corrales

Vivimos en una época en la que la experiencia social está profundamente mediada por tecnologías digitales. Las redes sociales, inicialmente reconocidas y celebradas por su capacidad para conectar personas y democratizar la comunicación, han devenido en infraestructuras complejas donde la interacción humana, la producción de sentido y la formación de comunidad se configuran a partir de lógicas algorítmicas. Más allá de su dimensión comunicativa, estos entornos digitales son hoy dispositivos de organización social, escenarios donde se negocian identidades, se construyen afectos y se movilizan narrativas que configuran el tejido de lo colectivo. Estas infraestructuras digitales no solo canalizan la interacción, sino que la reconfiguran desde dentro, haciendo de la inteligencia artificial un operador omnipresente, aunque muchas veces invisible, en la modulación de lo social.

Esta mediación tecnológica atraviesa múltiples esferas de la vida cotidiana. Desde las búsquedas laborales y las dinámicas económicas cotidianas, pasando por la organización de eventos, el consumo de

información, hasta la creación de vínculos emocionales; nuestras acciones y decisiones germinan en plataformas digitales que permiten la visibilidad, la conexión y el acceso. La interacción humana ya no ocurre únicamente en la presencialidad: lo social se construye en una simultaneidad constante entre lo físico y lo digital, en dinámicas donde los algoritmos influyen sobre lo que vemos, a quién escuchamos e incluso qué sentido tiene para cada uno. Son estos espacios donde hoy se inician relaciones sentimentales, se articulan proyectos colaborativos, se indaga para encontrar espacios de pertenencia–física o simbólicamente–, y se perfilan trayectorias personales o profesionales.

Las redes sociales operan como arquitecturas de lo común: espacios que no solo albergan interacciones, sino que organizan las condiciones bajo las cuales lo social es posible en la actualidad. En su obra *Comunicación y poder* (2009), Manuel Castells argumenta que la sociedad en red redefine las estructuras tradicionales del poder, desplazando los centros de influencia hacia nodos descentralizados donde se construyen narrativas, identidades y formas de acción colectiva. En este entorno, lo digital no es solo un medio de expresión, sino un terreno de disputa simbólica, donde los marcos de visibilidad y participación son establecidos por infraestructuras técnicas que escapan al control directo de los usuarios. Estas infraestructuras, lejos de ser pasivas, operan como sistemas inteligentes que orquestan relaciones, moldean emociones y organizan economías de atención (Saheb et al., 2024).

Desde una perspectiva complementaria, José van Dijck (2013) propone entender estas plataformas como sistemas sociotécnicos: ensamblajes en los que lo tecnológico, lo económico y lo cultural interactúan para modelar la experiencia digital. No solo posibilitan la interacción, sino que la direccionan y condicionan. La arquitectura de la plataforma –sus algoritmos, botones, métricas– organiza lo que se muestra y lo que no se muestra, lo que es reforzado por los sistemas de recomendación y lo que queda al margen. Así, el diseño

mismo de estos espacios participa activamente en la producción de lo común. Sadiku et al. (2021) advierten que la inteligencia artificial ocupa hoy el núcleo del funcionamiento de estas redes: transforma la manera en que las personas interactúan, construyen comunidad y otorgan sentido a lo que ven. El contenido que aparece en pantalla no es neutral, sino resultado de sistemas de aprendizaje que analizan el comportamiento del usuario, sus emociones, intereses y tiempos de permanencia, para ofrecerle un entorno personalizado que optimice la conexión y prolongue la exposición.

En este entramado sociotécnico de las redes sociales también se transforma la manera en que los sujetos y/o usuarios se constituyen y se perciben a sí mismos. La identidad en redes sociales no es simplemente algo que se muestra, sino algo que se construye constantemente en función de la lógica de la plataforma: algoritmos que recompensan ciertos comportamientos, métricas que cuantifican la relevancia, mecanismos de exposición que guían la presentación del yo (Bucher, 2018). Judith Butler (2002) ha planteado que la identidad es un acto performativo; en estos espacios, dicha performance se convierte en estrategia para ser visto, validado y reconocido[1]. Pero esa visibilidad es selectiva. Como también señala Bucher, los sujetos no solo actúan, sino que son constantemente interpelados por sistemas que moldean sus experiencias: el "yo" que emerge está condicionado por su huella digital: patrones de comportamiento pasados de sí mismo, por inferencias de consumo, por predicciones algorítmicas que delimitan lo que es posible mostrar o incluso desear.

La subjetividad digital se vuelve así una construcción relacional y asimétrica: un yo que se ajusta a modelos preconfigurados de visibilidad. Además, no es solo el sujeto humano quien se presenta: la inteligencia artificial también participa en la interacción con estilos, tonos, narrativas e incluso simulaciones emocionales. Como muestra

1 La noción de performatividad en Judith Butler implica que el género y la identidad no son esencias previas, sino efectos de prácticas reiteradas en contextos sociales normativos.

Wu (2024), los sistemas conversacionales no solo responden: también construyen una imagen de sí mismos. Usan formas de autodeclaración –por ejemplo, al explicar decisiones o simular estados de ánimo– que generan una percepción de subjetividad maquínica. En el nivel perceptivo, la IA se vuelve un actor dramático, con un frente expresivo que simula humanidad y un trasfondo algorítmico que permanece invisible.

Las formas de comunidad que emergen en estos entornos ya no responden a la lógica tradicional del encuentro o del intercambio directo. Se trata de ensamblajes dinámicos, efímeros, mediados por dispositivos, en su mayoría móviles, que distribuyen reconocimiento, atención y pertenencia de manera diferencial. Según el planteamiento de Bauman (2001), estas comunidades en red tienden a la fragilidad: vínculos fugaces, afectos inestables y pertenencias marcadas por la lógica de la entrada y salida inmediata. Esta liquidez del lazo genera un nosotros volátil, susceptible de disolverse con la misma rapidez con la que se conforma. Y, sin embargo, ese "nosotros" sigue buscándose. La comunidad digital se configura, así como un espacio paradójico: deseado como refugio de reconocimiento, pero organizado bajo condiciones inestables y jerarquizadas de visibilidad. El lazo social se redefine como transacción simbólica, como moneda afectiva mediada por interfaces que calculan alcance, reacción y adherencia.

La creciente presencia de la inteligencia artificial ha intensificado estas transformaciones. Lo que antes eran interacciones pensadas como estrictamente humanas –entre pares, amigos, públicos, comunidades– ahora están mediadas por sistemas que intervienen en lo que se ve y se dice. El despliegue algorítmico no solo gestiona información, sino que organiza las condiciones mismas de lo social. Lo que antes eran decisiones humanas hoy se delega a sistemas que aprenden, interpretan y actúan, desplazando la agencia hacia estructuras invisibles de cálculo. En su obra *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Luciano Floridi (2014) plantea que

ya no habitamos el mundo digital como un afuera, sino que vivimos *onlife*, inmersos en una infosfera donde toda experiencia es simultáneamente real y digital, moldeada por flujos de datos que nos anteceden y nos perfilan.

Esta transformación no puede entenderse al margen del modelo económico en el que se inscribe. Shoshana Zuboff (2019) ha descrito cómo las redes sociales operan bajo una lógica de extracción constante, donde cada gesto es capturado, interpretado y monetizado. Las plataformas no son únicamente lugares de encuentro; son máquinas de producción de datos donde el usuario deviene al mismo tiempo sujeto, objeto y producto. Byung-Chul Han (2014), por su parte, ha advertido que este sistema disuelve la negatividad, penaliza el disenso y recompensa la autoexposición permanente. Lo que se ve es lo que se cuantifica. Lo que se repite es lo que se cree.

Frente a este panorama, se hacen necesarias preguntas que abran grietas en el funcionamiento aparentemente natural de estas plataformas: ¿Qué significa hoy habitar una comunidad digital si el vínculo está mediado por intereses algorítmicos invisibles? ¿Cómo se configura el nosotros cuando las reglas de aparición están diseñadas para reforzar segmentaciones, predictibilidad y permanencia? ¿Qué agencia es posible en un entorno donde la identidad, la relación y la comunidad son modeladas por sistemas que calculan más rápido de lo que podemos pensar? ¿Podemos imaginar experiencias colectivas que escapen a la lógica de la personalización y la exposición permanente?

Estas preguntas no buscan cerrarse, sino mantenerse abiertas como umbral de una reflexión crítica urgente: las redes sociales no son simplemente espacios de expresión, sino el lugar donde hoy se reconfigura lo social, y donde la inteligencia artificial, cada vez más, actúa como coautora del vínculo, la identidad y lo común.

Infraestructuras de lo común: redes sociales como espacios de organización social

Las redes sociales han dejado de ser simples plataformas de comunicación interpersonal para convertirse en infraestructuras complejas que organizan las condiciones mismas de lo social. No solo median lo que se dice, sino que modulan cómo se produce el vínculo, cómo se percibe la presencia y cómo se construye lo común. Desde sus arquitecturas técnicas hasta sus algoritmos de recomendación, estas plataformas participan activamente en la configuración del espacio público digital y de las relaciones que en él se tejen.

Entenderlas como "infraestructuras" implica reconocer que no son solo herramientas tecnológicas ni canales neutros, sino que constituyen condiciones materiales, simbólicas y operativas que sostienen y modelan la vida cotidiana. Como diría Susan Leigh Star (1999), una infraestructura es aquello que se vuelve invisible por su cotidianidad[2], pero que estructura lo que podemos hacer, decir y pensar. En este sentido, las redes sociales son tan ubicuas como determinantes: están en el fondo de nuestras rutinas de información y en nuestras formas de ser y estar con otros.

La noción de "sociedad red", formulada por Manuel Castells (2009), permite comprender cómo las formas de poder, pertenencia y producción simbólica han sido absorbidas por estas tramas digitales. Las redes no solo interconectan nodos: redistribuyen visibilidad, construyen narrativa y articulan comunidades dispersas, reconfigurando el modo en que se genera sentido colectivo. Este tipo de estructuras habilitan relaciones que aparentemente no son lineales ni jerárquicas, pero que están profundamente marcadas por lógicas de acceso, velocidad, amplificación y silenciación, determinadas en gran parte por sistemas automatizados.

Desde una perspectiva crítica, José van Dijck (2013), ha señalado que las plataformas deben entenderse como sistemas sociotécni-

2 Susan Leigh Star plantea que las infraestructuras no se ven, pero determinan prácticas sociales. Solo se hacen visibles cuando fallan o cambian.

cos: dispositivos que no se limitan a facilitar la comunicación, sino que la moldean activamente a través del diseño de sus interfaces, los valores incrustados en sus algoritmos y los intereses económicos que las sustentan. La arquitectura de una plataforma –sus botones, sus métricas de participación, sus sistemas de recomendación– define qué se ve, qué se comparte, qué circula y qué se oculta. Lo que para el usuario aparece como espontáneo, es el resultado de una ingeniería intencional orientada por lógicas de *engagement* y segmentación; esta configuración afecta la construcción de lo común. No todas las voces tienen las mismas condiciones de circulación; no todos los discursos encuentran el mismo nivel de amplificación. El algoritmo prioriza aquello que predice éxito: lo emocional, lo polarizante, lo fácilmente digerible. Así, el espacio compartido se convierte en un entorno condicionado por la lógica del rendimiento y la visibilidad.

La inteligencia artificial intensifica este funcionamiento al integrarse de forma transversal en la gestión del entorno digital. Según Sadiku et al. (2021), la IA es hoy un componente estructural de las redes sociales, encargado de analizar comportamientos, reconocer rostros, clasificar intereses y personalizar contenidos. Su implementación no es marginal: afecta desde la curaduría de contenidos hasta la moderación automatizada, pasando por la predicción de reacciones y la sugerencia de contactos. Esta personalización no solo busca comodidad: actúa como un filtro dinámico que adapta lo visible a lo que el sistema anticipa que debe interesarnos, generando una experiencia cuidadosamente ajustada a perfiles construidos algorítmicamente.

Saheb et al. (2024) profundizan esta perspectiva al proponer que las redes sociales deben entenderse como infraestructuras inmersivas inteligentes, capaces de orquestar no solo relaciones y flujos de información, sino también emociones, estados de ánimo y sentidos compartidos. Estas plataformas no se limitan a reflejar intereses: los inducen, los anticipan y en muchos casos los modulan. Intervienen en el ritmo de las conversaciones, en los elementos destacados, en

las sugerencias de contenido, en la intensidad de las respuestas. La convergencia entre inteligencia artificial y redes sociales no crea simplemente entornos digitales más eficientes; sino que diseña las condiciones en las que hoy habitamos lo social y lo político.

Esta inteligencia algorítmica transforma también la manera en que los usuarios participan en los procesos de deliberación colectiva. Lo que vemos en nuestros *feeds* no es el resultado de una netamente de una decisión abierta, sino de una configuración técnica que privilegia ciertos contenidos en función de lo que se considera relevante para nosotros. Esa "relevancia" es el resultado de una serie de inferencias basadas en historiales, clics, tiempos de permanencia y emociones expresadas. Así, el entorno común deja de ser un espacio plural abierto al encuentro de diferencias, y se convierte en un entorno hiperpersonalizado donde lo común es aquello que confirma nuestras inclinaciones previas.

La consecuencia es que el tejido de lo colectivo se recompone de forma fragmentaria: más íntimo, sí, pero también más cerrado. La inteligencia artificial, como operador algorítmico, no anula el vínculo social, pero lo reformula según parámetros de predicción, rendimiento y segmentación. Esto no implica necesariamente una pérdida de sentido, sino una transformación que requiere ser pensada: ¿qué tipo de comunidad emerge cuando lo visible y lo común están mediados por sistemas automatizados?

Al mismo tiempo, estas plataformas construyen una dimensión emocional del vínculo digital. En su artículo "Exploring the use of smart environments and artificial intelligence in enhancing online community building", Saheb et al. (2024) documentan cómo se utilizan herramientas algorítmicas para ajustar el contenido emocional de los mensajes al estado afectivo del usuario, generando una experiencia que se siente íntima y personalizada, pero que en realidad está programada. Este tipo de modulación de la experiencia refuerza la percepción de cercanía, pertenencia y autenticidad, aun cuando la relación esté mediada por una lógica de optimización cal-

culada. Frente a este panorama, se vuelve indispensable problematizar las condiciones bajo las cuales se produce hoy la experiencia de comunidad. Si el encuentro, el reconocimiento mutuo y la exposición a la diferencia son fundamentales para el ejercicio de lo social, ¿cómo se transforman estas condiciones cuando el espacio común está diseñado para reforzar afinidades, predecir comportamientos y maximizar interacciones rápidas?

Al observar las redes sociales como infraestructuras de lo común, se hace evidente que su influencia no reside únicamente en lo que permiten hacer, sino en lo que hacen posible, lo que imposibilitan, y lo que preconfiguran sin que lo notemos. Estas infraestructuras operan desde un fondo de normalidad técnica que define los marcos de interacción aceptables, las estéticas de la participación y las formas de visibilidad legítimas. En esa medida, no basta con habitarlas: urge comprenderlas, intervenirlas y —cuando sea posible— imaginar otras formas de organizar lo social desde la tecnología.

Este texto no plantea una pérdida de agencia ni una cancelación del vínculo. Por el contrario, reconoce que las redes sociales continúan siendo espacios de encuentro, creación y movilización. Pero ese potencial coexiste con estructuras técnicas que requieren lectura crítica. Al comprender las redes sociales como infraestructuras de lo común, se abre la posibilidad de pensar en el modo en que organizan nuestras relaciones, delimitan nuestras exposiciones y codifican nuestros afectos.

Desde aquí, se vuelve inevitable pasar a la siguiente pregunta: ¿qué tipo de sujeto se configura en este entorno? Las plataformas no solo modelan los marcos del vínculo, sino también los del yo. La subjetividad digital, en este contexto, no se presenta como una esencia interior que se expresa, sino como una construcción modelada, cuantificada y gestionada. La siguiente sección abordará justamente esa dimensión: la forma en que la identidad y la subjetividad son producidas, ajustadas y reconocidas en entornos mediados por inteligencia artificial.

El yo en redes sociales: identidad y algoritmos

En el entorno digital, la identidad no se presenta como una esencia interior que se expresa libre y espontáneamente, sino como una construcción que se performa y ajusta en función de las condiciones técnicas, afectivas y sociales que operan en las plataformas. No es solo el sujeto quien decide qué mostrar, sino que es también el entorno –algorítmico, visual, emocional– quien define qué puede circular, qué será recompensado, qué se hace visible. El "yo" digital es tanto producto de la expresión como resultado de la exposición.

La construcción de identidad en redes sociales comienza muchas veces desde el consumo: al desplazarse por contenidos, observar lo que se valora, lo que se aplaude o lo que se ignora, los sujetos aprenden qué se espera de ellos y modelan sus intervenciones para encajar en esos marcos. Como señalan Saheb et al. (2024), las plataformas no solo responden a intereses, sino que los inducen, los anticipan y los orquestan. Por ejemplo, alguien que interactúa con videos sobre éxito profesional puede empezar a narrarse en clave meritocrática, incluso si sus condiciones económicas o materiales no han cambiado.

Lo que se muestra en redes no es idéntico a lo que se es en la realidad, sino una versión específica, ajustada al entorno de visibilidad; mostramos en muchos casos lo que deseamos que otros vean y ocultamos lo que no queremos que se sepa. En este sentido, podríamos decir que nuestra identidad virtual es la manera en que nos representamos en el mundo digital, y puede adoptar distintas formas, muchas veces simultáneas.

La identidad virtual puede parecer ser auténtica cuando refleja con fidelidad nuestras experiencias o emociones, pero también puede responder a lógicas de edición selectiva. En muchas ocasiones se muestra solo lo que conviene destacar: un ascenso laboral, un viaje, una opinión brillante en el momento adecuado. Lo demás –las con-

tradicciones, las dudas, los fracasos– se desliza al margen. Esta versión parcial, que no es necesariamente falsa, sigue siendo una forma activa de construir el yo en redes sociales.

Hay casos, en cambio, en los que se encarna directamente un personaje. Una figura inventada, estética o discursivamente coherente, pero distante de la experiencia real. A veces el tono se vuelve más firme o más radical de lo que seríamos en la vida cotidiana; otras, se adopta una pose que busca simpatía o autoridad. Esa fabricación puede tener fines estratégicos o afectivos, pero sigue operando como una respuesta a las dinámicas algorítmicas de validación.

No hay un solo yo en las redes sociales: hay varios. La identidad digital se fragmenta entre plataformas, adaptándose a los lenguajes, públicos y códigos de cada espacio. En una red social profesional como LinkedIn podemos mostrarnos competentes y organizados; en otra más visual cómo Instagram o Pinterest, podeos ser más espontáneos, creativos o cercanos; en otras cómo X desarrollamos una identidad más discursiva, crítica, irónica o incluso provocadora. Lo que se muestra está profundamente atravesado por lo que se espera que mostremos, y cada escenario exige su propia máscara.

Este juego de ajustes no es menor. Habla de cómo los entornos digitales no solo alojan la expresión del yo, sino que la modelan. Las redes enseñan cómo hablar de uno mismo, qué mostrar, qué callar, qué tipo de afectividad circula mejor. La identidad no se expresa: se ensambla, se prueba, se corrige. Y muchas veces, se performa: se actúa para encajar, para pertenecer, para persistir en el radar de la visibilidad.

Estas variaciones no implican una patología, sino una adaptación al espacio relacional, técnico y simbólico. Como ha planteado Taina Bucher (2018), los usuarios son constantemente interpelados por las plataformas: se les sugiere qué decir, cuándo aparecer, qué publicar. La identidad se convierte en un proyecto continuo, gobernado por métricas visibles y validaciones externas. Una notificación que recuer-

da cuántos días llevamos sin publicar o cuántas personas han visto una historia actúa como estímulo para mantener la presencia, incluso si no hay nada significativo que compartir.

Judith Butler (1997) ha señalado que la identidad es un acto performativo: se hace en el hacer. En el entorno digital, esta performance está guiada por sistemas algorítmicos que premian ciertos comportamientos –emocionalidad, constancia, claridad narrativa– y desincentivan otros –ambigüedad, crítica, complejidad–. Un perfil que publica contenido emocional y directo tiene más probabilidades de ser amplificado que uno que problematiza o matiza sus ideas.

En este ecosistema, la inteligencia artificial no es un simple espectador. Como lo indica Sadiku et al. (2021), la IA participa activamente en la clasificación, análisis y predicción del comportamiento del usuario en redes sociales. La subjetividad se vuelve un dato legible, una superficie para el modelado predictivo. Lo que uno muestra se convierte en insumo para definir lo que uno verá, lo que se le ofrecerá, lo que se anticipa que desea. Si una persona comparte contenido sobre salud mental puede empezar a recibir anuncios sobre terapias, medicación o aplicaciones de meditación, reforzando así un perfil que la misma persona no necesariamente había asumido.

Esta dinámica produce un tipo de yo que, como advierte Byung-Chul Han (2014), está sujeto a una lógica de autoexplotación emocional: el yo que se muestra debe actualizarse, rendir, mantenerse visible, alinearse con las lógicas del algoritmo. La exposición se convierte en obligación, y la coherencia narrativa se vuelve estrategia de permanencia.

Además, todo lo que hacemos –incluso lo que no decidimos compartir– deja una huella digital: un rastro de datos que es recolectado, clasificado y reutilizado para definir futuros contenidos, recomendaciones o conexiones. Esta huella no solo nos sigue: nos antecede, nos perfila, y condiciona lo que puede ser visible para nosotros mismos. Como plantea Floridi (2014), ya no habitamos el mundo digital

como si fuera un afuera: vivimos "onlife", inmersos en una infosfera donde la experiencia está siempre moldeada por datos que nos perfilan incluso antes de que actuemos.

Esa tensión constante entre agencia personal y predicción algorítmica produce también una forma de identidad vigilada. Aunque los sujetos ejercen cierta libertad para decidir cómo mostrarse, esta libertad está mediada por un conjunto de incentivos, recordatorios y presiones invisibles. Saheb et al. (2024) señalan que los entornos digitales inteligentes no solo permiten la expresión, sino que la dirigen, facilitando ciertas emociones y desincentivando otras, en función de patrones previamente aprendidos; una persona que reacciona con tristeza a ciertas noticias podría empezar a recibir más contenido con tonalidades emotivas similares, esto puede reforzar una visión parcializada del entorno. La subjetividad digital, por tanto, se articula como un proceso relacional donde el sujeto es al mismo tiempo actor, contenido y objeto de análisis. Lo que se comparte no solo habla del yo, sino que define qué es valioso, quién merece atención, y cómo debe narrarse la vida. Esta dimensión ético-política del yo digital exige una revisión crítica de los modelos de representación y de los valores que sostienen la participación.

Floridi (2014) afirma que, en el ecosistema *onlife*, no basta con proteger datos: es necesario proteger identidades, narrativas y formas de existencia. Esta idea sugiere que cualquier discusión sobre privacidad o seguridad debe considerar también la configuración misma del yo como entidad algorítmicamente modelada. ¿Qué formas de subjetividad quedan excluidas por los sistemas de recomendación? ¿Qué voces se invisibilizan porque no se ajustan a lo que el algoritmo considera exitoso? Si el yo se construye bajo estas condiciones, ¿cómo se construye el nosotros? ¿Cómo emergen las formas colectivas de pertenencia cuando los vínculos se organizan por segmentaciones automatizadas y afinidades preestablecidas?

Formación de comunidades en plataformas digitales

La noción de comunidad ha mutado en los últimos años, empujada por la lógica de las redes y los dispositivos que median las interacciones. En lo digital, los vínculos no siempre nacen del encuentro directo, sino de la coincidencia algorítmica: de haber reaccionado a los mismos contenidos, compartido intereses semejantes o habitado simultáneamente una conversación. La comunidad ya no es únicamente una construcción social deliberada; es también un efecto de cálculos, coincidencias codificadas y sistemas de agrupamiento invisibles.

Si antes las comunidades eran articuladas en torno a territorios, instituciones o memorias compartidas, hoy emergen desde trayectorias de consumo, afectos resonantes y lógicas de asociación que muchas veces el sujeto no controla. Dos personas que jamás se han cruzado físicamente pueden ser agrupadas, etiquetadas, segmentadas y direccionadas hacia los mismos contenidos, campañas o causas sin haberlo decidido explícitamente.

Manuel Castells (2009) propone que las redes digitales inauguraban una nueva forma de organización social: una "sociedad en red", donde los nodos no están dados de antemano, sino que se constituyen por la conexión. En ese entorno, la comunidad no es un punto de llegada, sino una posibilidad que se activa en la interacción. Sin embargo, esa potencialidad se encuentra cada vez más mediada por sistemas que definen quién aparece para quién, bajo qué condiciones y con qué frecuencia. Las plataformas no solo permiten vincularse, sino que determinan –con silenciosa eficacia– con quién es más probable vincularse.

Este proceso ha llevado a que muchas comunidades digitales estén regidas por un principio de afinidad algorítmica, que privilegia la homogeneidad sobre el disenso. La promesa inicial de apertura se convierte en una arquitectura de eco. Como advierte Van Dijck (2013), estas plataformas no son espacios neutros de interacción,

sino sistemas sociotécnicos diseñados para modelar comportamientos. Lo que se visibiliza, lo que se refuerza, incluso lo que se considera "interacción auténtica", está mediado por intereses económicos y lógicas de automatización.

Esta condición genera comunidades inestables, donde el lazo está mediado por la visibilidad y sostenido por el rendimiento. Zygmunt Bauman (2001) lo expresó con lucidez al referirse a la "liquidez de los vínculos digitales": relaciones efímeras, asociaciones frágiles, afectos organizados por la lógica de la entrada y salida inmediata. Una comunidad digital puede articularse rápidamente en torno a una causa viral, pero deshacerse con la misma velocidad cuando cambia el foco de atención.

Además, estas plataformas diseñan sus entornos para que las comunidades sean también productoras de datos. Las formas de reunión, conversación o protesta se integran a los circuitos de extracción de valor. Como recuerda Shoshana Zuboff (2019), bajo el capitalismo de vigilancia, incluso los gestos colectivos más espontáneos pueden ser registrados, almacenados y analizados para extraer patrones de comportamiento, consumo o riesgo.

Sin embargo, esto no significa que las comunidades digitales sean falsas o carentes de potencia. Pueden constituir espacios de resonancia emocional, de activismo distribuido, de solidaridad entre desconocidos. Movimientos sociales recientes se han gestado, sostenido y amplificado gracias a la acción colectiva digital, desbordando los marcos tradicionales de la organización política. El movimiento *#MeToo*, por ejemplo, permitió visibilizar de forma masiva experiencias de violencia de género, generando debates públicos, reformas institucionales y legislaciones en varios países. *Black Lives Matter*, surgido en Estados Unidos, se consolidó como una red global que denunció la violencia policial y el racismo sistémico, articulando protestas, campañas y acciones comunitarias en línea y fuera de ella. En América Latina, el movimiento #NiUnaMenos impulsó redes feministas transnacionales que cuestionan la violencia machista y

exigen transformaciones culturales y legales. En Colombia, durante el Paro Nacional de 2021, las redes sociales jugaron un papel fundamental para la coordinación de acciones, la denuncia de violaciones a los derechos humanos y la circulación de narrativas alternativas que contrarrestaron los discursos oficiales. Estos casos muestran que las comunidades digitales pueden generar formas de pertenencia, organización y disputa simbólica que no dependen exclusivamente de la presencialidad ni de estructuras tradicionales. Lo que se pone en cuestión, entonces, no es su existencia, sino las condiciones éticas, afectivas y económicas que las sustentan y las formas en que son intervenidas, condicionadas o explotadas por las infraestructuras digitales que las alojan.

Saheb et al. (2024) plantean que los entornos digitales no solo facilitan la creación de comunidades, sino que las orientan, les dan forma, las conectan bajo patrones que responden tanto a dinámicas sociales como a lógicas de control. Esta articulación plantea un nuevo desafío: ¿cómo pensar comunidades digitales que no estén predeterminadas por lo que se espera de ellas, sino que abran espacio para la diferencia, el disenso y lo impredecible? En ese marco, las comunidades se vuelven lugares de circulación de afectos consensuados, donde la contradicción es penalizada algorítmicamente y la tensión se reduce a binarismos. El riesgo es que el lazo se vuelva débil no por falta de conexión, sino por exceso de repetición.

Frente a este panorama, se vuelve urgente imaginar otras formas de comunidad: no necesariamente fuera de lo digital, pero sí en tensión con las arquitecturas algorítmicas que moldean los encuentros. Comunidades que no estén organizadas únicamente por la lógica de la recomendación, sino por la voluntad de escucharse en la diferencia. Espacios que no reduzcan la identidad colectiva a una estética común, sino que permitan negociar sentidos, incluso en la disonancia.

IA como mediadora algorítmica de lo social

Las interacciones digitales que alguna vez se pensaron exclusivamente humanas –comentar, sugerir, responder, compartir, buscar, coincidir– son hoy acciones mediadas o incluso reemplazadas por sistemas automatizados. La inteligencia artificial no aparece solamente como un conjunto de herramientas que acompañan lo social: se ha vuelto un operador estructural, una presencia invisible pero determinante en la forma en que nos vinculamos, participamos, buscamos y nos narramos.

Lo algorítmico ya no está solo al final del proceso (para clasificar lo que hicimos), sino en el centro de la escena, organizando de antemano lo que veremos, a quién escucharemos y qué será posible pensar. La IA, en su despliegue más sofisticado, anticipa deseos, gestos y decisiones. Lo hace no solo desde la promesa de personalización, sino desde la capacidad de intervenir en la construcción del entorno social como tal.

Como plantean Sadiku et al. (2021), los sistemas inteligentes en redes sociales analizan patrones pasados para predecir acciones futuras. Pero no se limitan a observar: intervienen, sugieren, automatizan, modelando activamente las condiciones bajo las cuales se configura lo social. Desde los filtros de contenido hasta los sistemas de recomendación y moderación, la IA ejerce una forma de gobierno silencioso, donde la norma no es explícita, pero sus efectos son tangibles. Quién aparece más, quién es invisibilizado, qué temas ganan presencia, qué disensos se silencian. La mediación algorítmica, entonces, no es neutra. Define lo visible, lo decible y lo circulante. Saheb et al. (2024) explican que los entornos inteligentes actúan como arquitecturas de la interacción: no solo distribuyen información, sino que configuran los escenarios donde se da lo social. En este marco, la IA establece umbrales de pertenencia, relevancia y validez, muchas veces sin mediación humana.

Esto tiene implicaciones profundas para la experiencia del vínculo. La relación ya no ocurre únicamente entre personas, sino entre sujetos que interactúan en un ecosistema modelado por sistemas auto-

matizados, y que muchas veces no pueden distinguir cuándo lo que perciben ha sido construido por otros humanos o generado por una red neuronal artificial. En este entorno, donde la IA no solo media, sino que prescribe, las decisiones se anticipan con base en patrones previos que se refuerzan constantemente. Como advierte Eli Pariser (2011), los algoritmos tienden a crear burbujas de filtro: entornos cerrados donde el sujeto solo accede a lo que ya confirmó, a lo que se ajusta a sus comportamientos pasados, eliminando así el riesgo, la disonancia y el encuentro con lo inesperado.

Este escenario redefine lo que entendemos por encuentro, comunidad y cercanía. Como plantean Ajayi-Majebi et al. (2021), los agentes artificiales ya no están en la periferia del vínculo: son parte activa de las redes de afectividad, decisión y comunicación. La IA se convierte así en mediadora de relaciones que antes eran pensadas como exclusivamente humano-humanas.

Pero más allá de la técnica, lo que está en juego es una dimensión política: ¿qué tipo de sociabilidad se produce cuando lo común se articula bajo una lógica predictiva y estadística? ¿Qué pasa con la incertidumbre, el disenso, la pausa o el error? El filósofo Byung-Chul (2014) Han alerta sobre una cultura digital que castiga lo que no fluye, que penaliza la negatividad, que transforma todo en experiencia optimizada, funcional, predecible. En este entorno, la IA contribuye a la normatividad de lo homogéneo: lo que más se parece a lo anterior, lo que confirma lo ya sabido, lo que no desborda la categoría.

Sin embargo, la mediación algorítmica no es, en sí misma, un mal. El problema no es la existencia de IA, sino el modelo desde el cual opera: una infraestructura diseñada principalmente para maximizar atención, extraer datos y mantener usuarios activos, sin garantizar pluralidad, cuidado o disenso. Como señala Floridi (2014), el desafío no está en suprimir la inteligencia artificial, sino en pensar su lugar ético y político dentro de la infosfera que habitamos. ¿Puede la IA ser una mediadora no extractiva? ¿Puede contribuir a la

apertura, en lugar de al encierro estadístico? Zuboff (2019) pone el acento en este punto: si no cuestionamos el modelo económico que sostiene estas mediaciones, cualquier innovación técnica profundizará los problemas de vigilancia, segmentación y manipulación. La IA no es solamente un asunto técnico, es un asunto estructural, que transforma la forma en que entendemos la libertad, la decisión y la posibilidad misma del vínculo.

Así, el reto está en problematizar la presencia de la IA como mediadora del lazo social, no para condenarla, sino para imaginar formas alternativas de convivencia digital. Espacios donde lo algorítmico no elimine la diferencia, donde lo automatizado no desplace la agencia, donde el dato no sustituya la escucha. Esto exige repensar no solo la tecnología, sino también las infraestructuras políticas, afectivas y culturales en las que se inscribe.

Capitalismo de vigilancia, visibilidad y agencia

Habitar las plataformas digitales y las redes sociales no es un acto neutral ni inocente. Cada gesto, cada interacción, cada silencio incluso, se convierte en un rastro legible, interpretable y monetizable. Las redes sociales no son meros espacios de encuentro o comunicación: son máquinas de extracción de datos, dispositivos diseñados para traducir la vida social en información operativa. Esta operación configura lo que Shoshana Zuboff (2019) ha llamado el capitalismo de vigilancia: un modelo económico en el que el comportamiento humano es materia prima gratuita para procesos comerciales de predicción, control y rentabilidad.

La visibilidad, en este contexto, se convierte en una moneda de cambio. Cuanto más visible es un sujeto, más valioso es para el sistema. Pero esa visibilidad no es igual para todos ni se reparte democráticamente. Está mediada por los algoritmos, moldeada por patrones de consumo y sostenida por métricas que cuantifican cada gesto. En

este sentido, la lógica de la burbuja de filtro[3] opera como una forma de fidelización afectiva: se nos ofrece más de lo mismo porque eso asegura nuestra atención. El capitalismo de vigilancia –como lo describe Zuboff– no solo extrae datos, sino que moldea las condiciones del deseo[4]. Pariser (2011) muestra cómo esta lógica reduce la complejidad del mundo a ecos agradables, donde el sujeto siente que el entorno digital lo entiende, sin percibir que ese "entendimiento" es también una forma de encierro estratégico.

Este encierro no es evidente, pero sus efectos pueden ser profundos. La personalización constante de los contenidos configura un entorno donde la diferencia se vuelve una anomalía, y la pluralidad, un riesgo algorítmico. Lo que el sujeto ve, comparte o incluso desea, es en gran medida lo que el sistema ha aprendido a ofrecerle, creando una forma de interacción circular y anticipada. Así, la posibilidad de construir comunidad con otros queda condicionada por un sistema que evita el roce con la alteridad. La burbuja de filtro, no solo delimita lo informativo, sino también lo afectivo y lo político.

En ese marco, la agencia, entendida como la capacidad de actuar y decidir sobre sí mismo, se ve desplazada hacia un territorio inestable. En apariencia, somos más libres que nunca para expresarnos, conectarnos y construir comunidad; en la práctica, muchas de nuestras decisiones son prefiguradas por sistemas que aprenden de nosotros más rápido de lo que logramos interpretarnos. Sadiku et al. (2021) advierten que la personalización que promete la IA en realidad configura un entorno autosegmentado, donde el usuario en redes sociales o entornos digitales, es conducido hacia decisiones anticipadas y experiencias repetidas. Esto plantea una paradoja inquietante: cuanto más visible es el sujeto,

3 Eli Pariser acuñó el término "burbuja de filtro" para describir cómo los algoritmos personalizan el contenido que vemos, aislándonos progresivamente de ideas divergentes.

4 Zuboff identifica el capitalismo de vigilancia como un nuevo orden económico que convierte la experiencia humana en materia prima para la extracción de datos y la predicción de comportamientos.

más expuesto está a dinámicas de control y predicción. La promesa de autonomía que acompañó la emergencia de las redes se ha transformado en un régimen de hiperexposición donde el reconocimiento se convierte en vigilancia y la conexión en extracción.

La estructura misma de las plataformas está pensada para convertir cada expresión en un dato útil. Lo que parece espontáneo –una búsqueda, un comentario, una reacción– es absorbido por sistemas que analizan, modelan y monetizan. Como indican Ajayi-Majebi et al. (2021), este proceso afecta la percepción que tenemos de nosotros mismos: nos mostramos no solo para existir, sino para rendir, para ser legibles en un sistema que traduce cada experiencia en patrón. Byung-Chul Han (2015) ha sido contundente al respecto: vivimos en una era de rendimiento obligatorio, donde el sujeto ya no es explotado desde fuera, sino que se explota a sí mismo en busca de visibilidad, aceptación y permanencia. Este yo conectado, hiperproductivo y disponible, se convierte en su propio vigilante, ajustando su comportamiento para maximizar su relevancia.

En este ecosistema, la visibilidad funciona como forma de regulación. Aquello que no aparece, que no circula, que no se ajusta, tiende a desaparecer del campo común. Y lo que sí aparece, debe hacerlo bajo condiciones impuestas por una arquitectura que privilegia lo emocionalmente impactante, lo polémico o lo rentable. La atención se vuelve escasa, y el algoritmo decide qué la merece.

La inteligencia artificial refuerza esta lógica al automatizar los filtros de lo que vale y lo que no. No se trata solo de mostrar contenidos, sino de administrar el deseo, orientar el interés, predecir la acción. La IA participa así de una forma de poder que ya no impone desde afuera, sino que opera desde la interiorización de lo que se espera. El sujeto, en este régimen, se convierte en consumidor, contenido y producto al mismo tiempo.

Floridi (2014) propone que, en la infosfera contemporánea, la identidad y la agencia deben pensarse en términos de relaciones

informacionales: somos lo que los datos dicen que somos, y actuamos dentro de límites trazados por esos mismos datos. Esto no anula la posibilidad de agencia, pero sí obliga a preguntarse en qué condiciones puede ejercerse de forma crítica, consciente o alternativa.

A este marco se suma una dimensión política fundamental: la reproducción de desigualdades. No todos los cuerpos, discursos o estilos de vida acceden del mismo modo a la visibilidad. Las plataformas replican –y a veces amplifican– los sesgos existentes, relegando al margen aquellas subjetividades que no se ajustan al molde dominante. La personalización, lejos de ser inclusiva, puede operar como un sistema de exclusión suave, silenciosa y eficaz.

Frente a esto, la agencia no puede pensarse como una decisión individual aislada. Implica, más bien, una práctica colectiva de conciencia crítica, una renegociación de lo visible, lo decible y lo posible. Implica crear grietas en la narrativa del dato total, del yo optimizado, de la conexión permanente. Implica también reclamar el derecho a no ser leídos, a no estar disponibles, a aparecer de otras formas. El problema no es la tecnología en sí, sino el régimen que organiza su uso: una economía de la atención que convierte cada vínculo en insumo, cada emoción en métrica, cada subjetividad en perfil operable. Interrumpir esta lógica supone pensar otras formas de infraestructura, donde lo común no sea capturado, sino cuidado; donde la diferencia no sea anomalía, sino condición de lo vivo.

A modo de conclusión

Pensar las redes sociales como espacios de interacción social mediados por inteligencia artificial exige reconocer que lo digital no es un escenario paralelo, sino una dimensión constitutiva de la vida contemporánea. La experiencia de comunidad, la formación de identidad, la construcción de vínculos y el ejercicio de la agencia se producen hoy en territorios donde la visibilidad, la perte-

nencia y la afectividad están profundamente condicionadas por arquitecturas algorítmicas. Esta configuración no impone un único destino, pero sí delimita los márgenes de lo posible. Habitar críticamente estos entornos implica asumir que el vínculo digital no es neutro, ni automático, ni transparente. Lo común no ocurre por defecto: se modela, se filtra, se codifica, y también se puede disputar. Allí radica quizás el mayor desafío: no renunciar a la posibilidad de comunidad, aun en escenarios diseñados para segmentar; no abandonar la construcción de sentido compartido, aun cuando la experiencia esté mediada por la predicción y la personalización.

Reconocer la complejidad de este paisaje no implica rechazar la tecnología, sino comprometerse con una forma más consciente y deliberada de habitarla. Las redes sociales, por su alcance y su capacidad de moldear lo cotidiano, constituyen hoy uno de los espacios más determinantes de lo político, lo afectivo y lo subjetivo. En ellas se negocian no solo opiniones, sino identidades; no solo discursos, sino formas de vida. Es por eso que pensar críticamente estos espacios no es solo un ejercicio académico, sino una necesidad cultural urgente.

En tiempos donde la automatización se presenta como sinónimo de eficiencia, y la hiperconexión como promesa de comunidad, quizás lo verdaderamente radical sea defender la calma, el disenso, lo no calculable. Afirmar que no todo puede predecirse, que no todo debe optimizarse. Apostar por la aparición de lo inesperado, lo irreductible, lo singular.

Lo social, en la era de la inteligencia artificial, permanece como un campo en disputa: entre lo automatizado y lo deliberado, entre lo visible y lo silenciado, entre lo que se espera de nosotros y lo que aún podemos imaginar. En esa tensión, todavía es posible ejercer una forma de libertad que no consista en elegir entre opciones prefiguradas, sino en inventar nuevas formas de habitar juntos los entornos digitales y reales.

En este contexto, resulta importante subrayar las implicaciones éticas que emergen del análisis. Si la inteligencia artificial y las plataformas digitales modelan hoy nuestras formas de relación, visibilidad

e incluso deseo, entonces las decisiones técnicas no pueden seguir pensándose al margen de la responsabilidad colectiva. Lo que está en juego no es solo la eficiencia de los sistemas, sino los marcos de posibilidad para el reconocimiento mutuo, la pluralidad democrática y la protección de subjetividades diversas.

Las infraestructuras digitales no son neutrales: organizan lo social desde lógicas que privilegian la homogeneidad, la predecibilidad y la exposición. Por eso, el diagnóstico aquí propuesto exige una respuesta ética que vaya más allá de la regulación de datos o la mejora de algoritmos. Se trata de reivindicar el derecho a la opacidad, a la diferencia, a la no-coincidencia, frente a entornos que recompensan la repetición y penalizan el disenso.

El imperativo no es meramente técnico, sino político: imaginar infraestructuras alternativas donde la inteligencia artificial no sea un instrumento de extracción, sino un medio para potenciar el cuidado, la deliberación y el acceso equitativo a la participación. Esto implica revisar críticamente los valores que sustentan el diseño de plataformas, y abrir el campo de lo digital a la construcción de comunidades que no estén determinadas por afinidades algorítmicas, sino por una ética del encuentro. Así, las redes sociales no deben pensarse únicamente como espacios de interacción, sino como territorios en disputa, donde se negocia qué vidas importan, qué discursos circulan, qué formas de existencia son posibles. Enfrentar este desafío con lucidez crítica es, quizá, la tarea más urgente de nuestra época *onlife*.

Referencias

Bauman, Z. (2001). Modernidad líquida. Fondo de Cultura Económica.

Bucher, T. (2018). If... Then: Algorithmic Power and Politics. Oxford University Press.

Butler, J. (2002). El género en disputa. Paidós.

Byung-Chul Han (2014). Psicopolítica. Neoliberalismo y nuevas técnicas de poder. Herder.

Castells, M. (2009). Comunicación y poder. Alianza Editorial.

Floridi, L. (2014). The Fourth Revolution: How the Infosphere is Reshaping Human Reality. Oxford University Press.

Pariser, E. (2011). The Filter Bubble: What the Internet Is Hiding from You. Penguin Press.

Sadiku, M. N. O., Ashaolu, T. J., & Ajayi-Majebi, A. (2021). Artificial Intelligence in social media. American Journal of Science and Technology, 10(3), 172–177.

Saheb, T., Qrunfleh, S. M., & Alfarraj, O. (2024). Exploring the use of smart environments and artificial intelligence in enhancing online community building. Journal of Smart Environments, 3(1), 24–40.

Sadin, É. (2020). La inteligencia artificial o el desafío del siglo. Caja Negra Editora.

Star, S. L. (1999). "The Ethnography of Infrastructure." American Behavioral Scientist, 43(3), 377–391.

Van Dijck, J. (2013). The Culture of Connectivity: A Critical History of social media. Oxford University Press.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.

[Parte 2]

Protección de datos personales y ética organizacional en la era digital

[Capítulo 4]

Privacidad, seguridad de los datos y el manejo ético del Big Data

Juan Fernando Arenas Jaramillo

Con la llegada de la era digital, la "explosión" del Big Data ha cambiado desde sus cimientos las dinámicas de nuestra sociedad, impactando virtualmente todos los sectores de la salud, el comercio, la seguridad o las políticas públicas. Las nuevas tecnologías emergentes como la inteligencia artificial (IA) han facilitado este proceso de transformación por cuanto permiten la posibilidad de analizar, procesar y utilizar grandes cantidades de datos a velocidades y niveles de precisión que nunca se habían visto hasta ahora; no obstante, este progreso en la tecnología plantea también dilemas éticos y legales cruciales, especialmente los que tienen que ver con la privacidad, la seguridad de los datos o la justicia.

La creciente dependencia de sistemas basados en datos masivos ha llevado a una recopilación exhaustiva de información sobre la vida de las personas, a menudo sin que los sujetos tengan pleno conocimiento de ello. Por esta razón, ha empezado a haber interrogantes sobre vulneraciones de la privacidad, discriminaciones algorítmicas o vulneraciones de los derechos por parte de los actores privados y/o públicos en el uso de los sistemas. Por ejemplo, la forma en que los sistemas de IA son capaces de realizar reconocimiento facial, de realizar perfiles de conductas o de prever comportamien-

tos ha suscitado debates en todo el mundo acerca de los límites éticos y legales de su uso.

A medida que los datos se convierten en el recurso más valioso del siglo XXI, las preguntas sobre su manejo ético y seguro se vuelven cada vez más imperiosas. La palpable necesidad de un equilibrio entre la innovación tecnológica y la protección de los derechos fundamentales exige un análisis crítico, juicioso y propositivo. Este capítulo se enfoca en explorar las intersecciones entre la privacidad, la seguridad de los datos y el manejo ético del Big Data, revisando los desafíos contemporáneos y proponiendo estrategias para un futuro más justo y equitativo.

Al abordar estos temas, el capítulo se estructura en torno a los siguientes ejes: primero, se establecen los conceptos fundamentales que subyacen a la privacidad y seguridad de los datos en el contexto del Big Data; luego, se analizan los riesgos éticos asociados con su uso desmedido; posteriormente, se exploran los marcos legales y regulatorios actuales, así como las buenas prácticas éticas recomendadas. Finalmente se concluye con reflexiones y propuestas para construir un ecosistema tecnológico centrado en valores.

El objetivo es proporcionar un análisis integral que permita a legisladores, desarrolladores de tecnología, académicos y la sociedad en general comprender la magnitud de los retos y, al mismo tiempo, promover soluciones viables que armonicen la innovación con el respeto por los derechos humanos.

Conceptos fundamentales

Previo a abordar los retos que presuponen las nuevas tecnologías, como la IA, en la privacidad de los datos de los usuarios, y en general, en los derechos relacionados con el habeas data. Se hace necesario definir, así sea de forma breve, los conceptos que serán el cimiento sobre el cual se construirá la discusión y conclusiones de

este capítulo de libro. Cabe indicar que, en ningún caso, busco con esta sección realizar una definición clara y completa de tecnologías tan complejas como la IA y el Big Data.

En consecuencia, a continuación, explicaré tres temas trascendentales para este trabajo: (i) Big Data; (ii) Privacidad y seguridad de los datos; y (iii) Ética en el manejo de datos masivos.

Big Data (the ability of making informed decisions)

W. Edwards Deming y Peter Drucker son frecuentemente citados con una frase que encapsula perfectamente el fenómeno del Big Data: "No se puede controlar lo que no se mide" (McAfee & Brynjolfsson, 2012). En esencia, el Big Data busca precisamente eso: extraer valor de una cantidad masiva y compleja de datos para tomar decisiones fundamentadas en los insights obtenidos. Este proceso es posible gracias a la digitalización, que convierte información analógica en digital, y a la datificación, que transforma fenómenos en datos cuantificables para un análisis más profundo. Además, el Internet de las Cosas (IoT) amplifica esta capacidad al generar un flujo constante de datos a través de millones de dispositivos conectados, en formatos estructurados, no estructurados y semiestructurados. (De Mauro et al., 2015)

En este contexto, el Big Data se caracteriza por tres rasgos fundamentales: volumen, velocidad y variedad.

Volumen

El volumen se convierte en una de las características principales del Big Data, que expresa la cantidad masiva de datos generados a cada segundo por el mundo digital. Las plataformas como Facebook con más de 2,9 miles de millones de usuarios activos, YouTube con más

de 2 miles de millones de usuarios activos, X (anteriormente Twitter) con un aproximado de 350 millones de usuarios activos o Instagram con una cantidad aproximada de 1,3 miles de millones de usuarios activos (Statista, 2023), produce billones de imágenes, publicaciones, videos y otros tipos de datos al día.

Este crecimiento exponencial en la producción de datos no solo está alimentado por las propias redes sociales, sino también por el Internet de las Cosas, los sensores inteligentes y los dispositivos móviles, que suman más información para el análisis (Manyika et al, 2011).

En esta línea, el volumen del Big Data no solo incorpora oportunidades para analizar información y tener acceso a ella en tiempo real, sino también, consecuencias para el almacenamiento, el procesamiento y la privacidad de la información (De Mauro, Greco & Grimaldi, 2015).

Velocidad

Uno de los pilares que define el Big Data es la velocidad, entendida como la capacidad de procesar datos en tiempo real o casi en tiempo real, permitiendo a las organizaciones responder de manera ágil a los cambios del entorno. Para ilustrar este concepto, resulta útil recurrir a una analogía compartida por un mentor mío hace algunos años. Él comparaba este principio con la idea de "tomar una fotografía del momento del negocio o de la compañía" y, basándose en esa instantánea, realizar el análisis necesario para el ejercicio legal que se le había encargado.

Uno de los puntos de sustentación que define el Big Data es la velocidad, que se entiende como la posibilidad que tienen las organizaciones de procesar datos en el momento, o prácticamente en el momento, permitiendo que estas reaccionen, agilizando así los cambios en el medio. Para facilitar este entendimiento, vale la pena recurrir a una analogía que compartía un mentor mío hace unos años. Él comparaba este principio con la idea de "tomar una fotografía del momento del

negocio o de la compañía" y, basándose en esa instantánea, realizar el análisis necesario para el ejercicio legal que se le había encargado.

De la analogía me gustaría resaltar lo siguiente: Aunque una captura de imagen nos remite a un cierto instante, el concepto de Big Data se acerca más bien a un videoclip que a una imagen o captura de imágenes; es decir, el Big Data hace referencia al flujo continuo e incesante de la información. No tenemos así una cantidad estática de datos, sino que existe la posibilidad de hacer el procesamiento de grandes cantidades de información a medida que se está produciendo, lo que tiene especial relevancia en ciertas situaciones, como, por ejemplo, la detección de fraude, la gestión del riesgo o la personalización en servicios que tienen lugar en tiempo real.

En un mundo donde las decisiones deben tomarse en fracciones de segundo, la capacidad de gestionar la velocidad de los datos se convierte en un diferenciador competitivo.

Variedad

La variedad es la cantidad y diversidad de formatos o fuentes de datos que pueden ser trabajadas o procesadas y analizadas. Antiguamente, en el análisis de la información, sólo se consideraban los datos estructurados, es decir, los datos organizados en bases relacionales que tienen el formato de tablas. Pero en la actualidad, con el auge del Big Data, se han de tener en cuenta los datos no estructurados: imágenes, videos, audios, e-mails, publicaciones en redes sociales, etc., y el uso de datos semi estructurados, como pueden ser archivos XML y JSON, que son datos que tienen algo de organización, pero no se pueden encuadrar en estructuras rígidas (De Mauro, Greco y Grimaldi, 2015). Esta heterogeneidad de datos supone un reto importante –en el almacenamiento, el procesamiento y el análisis– que necesita de tecnologías muy avanzadas que permitan combinar múltiples fuentes/formatos en tiempo real.

Privacidad y seguridad de los datos

Este es un concepto un poco más complejo, y al igual que el anterior, podría dar para escribir un artículo completo sobre el mismo. Sin embargo, con el fin de ilustrarlo como concepto clave, realizaré una breve descripción de su concepto para posteriormente realizar un recuento histórico.

La privacidad y la seguridad de los datos han cobrado una gran importancia en el contexto actual, caracterizado por la abundancia de datos y la inteligencia artificial (IA). A pesar de que a menudo estos conceptos se utilizan como sinónimos, es importante acometer su distinción; la privacidad de los datos implica el derecho a controlar la información personal y la forma en que se utiliza (Solove, 2021) y la seguridad de los datos hace referencia a las medidas técnicas y organizativas que permiten proteger los datos frente a accesos no autorizados, pérdidas o alteraciones (Oseni et al., 2022).

Ahora bien, la preocupación por la privacidad de los datos no es algo reciente. Regulaciones como el Reglamento General de Protección de Datos (GDPR) en la Unión Europea y la Ley de Privacidad del Consumidor de California (CCPA) ya se han encargado de establecer marcos normativos con el fin de limitar la explotación de datos personales y exigir mayor transparencia a las empresas (Nonju, 2021). Sin embargo, el ascenso del Big Data, la inteligencia artificial y la automatización ha puesto a prueba estos marcos puesto que la recolección a gran escala de los datos en tiempo real junto con el aprendizaje automático, ha supuesto una erosión del consentimiento informado y la capacidad para predecir comportamientos de personas concretas sin que los mismos den su autorización explícita (Gulati, 2020).

El desarrollo de distintas tecnologías como el Internet de las Cosas (IoT) y el reconocimiento facial ha acentuado aún más los peligros de vigilancia masiva, en muchas ocasiones justificado con la idea de mejorar un servicio o de reforzar la seguridad pública (Hu & Min, 2022). Como muestran investigaciones recientes, los algoritmos pue-

den favorecer el sesgo y la discriminación del sujeto al utilizar datos con escasa representación o cuando se usan en situaciones donde no existe una supervisión efectiva, lo cual resalta las inquietudes éticas sobre la privacidad individual (Mühlhoff, 2021).

Desde una perspectiva crítica, la seguridad y la privacidad no pueden descansar sólo sobre la ley o las soluciones tecnológicas, sino que requieren un enfoque multidisciplinar que contemple, entre otros factores, la responsabilidad corporativa, la alfabetización digital de los usuarios o la adhesión de ciertos principios como la minimización de datos o la privacidad por diseño (Gilbert, 2021). Si no es así, la creciente asimetría entre quienes recogen datos y quienes son analizados va a perpetuar una dinámica en la que los individuos no tienen poder sobre su propia información personal.

En las siguientes secciones, se abordará la evolución histórica de la privacidad y seguridad de los datos, analizando cómo los modelos tradicionales de protección se han visto desafiados por las tecnologías emergentes y explorando posibles soluciones para equilibrar la innovación con la protección de los derechos fundamentales.

Ética en el manejo de datos masivos: principios clave (autonomía, consentimiento informado, transparencia)

En este punto abordaremos los principios éticos, presentes en la mayoría de regulaciones, reglamentos y documentos técnicos, que deben regir el manejo de datos masivos, particularmente en el contexto de IA (UNESCO, 2022).

Autonomía: el derecho a decidir sobre los propios datos

La autonomía respecto al manejo de datos implica que las personas tienen la facultad de decidir cómo su información personal es

la recogida, procesada y utilizada. Sin embargo, actualmente no solo el hecho de que la información personal sea recopilada y procesada a gran escala sin un conocimiento explícito de las personas pone en cuestión la autonomía; también la forma en que los algoritmos de IA dan cuenta de la autonomía a la hora de procesar datos, dado que pueden llegar a conformar perfiles predictivos sin un consentimiento explícito, por lo que las personas no pueden tener control sobre la forma en que su información podrá ser utilizada para fines de predicción (Gulati, 2020; Mühlhoff, 2021).

Desde una mirada crítica, el concepto de autonomía en la era digital se torna problemático cuando la recopilación de datos se realiza de forma inadvertida, a través de dispositivos conectados, aplicaciones móviles y plataformas digitales; es decir, frente a estas nuevas tecnologías, las viejas fórmulas de consentimiento explícito no son aplicables. En general, muchas de estas tecnologías se sostienen en el uso de mecanismos "opt-out", que hace recaer la carga sobre el usuario, evitando la exigencia de un consentimiento explícito, claro y eficaz (Solove, 2021).

Consentimiento informado: un concepto en crisis

El consentimiento informado ha sido, desde su concepción, un pilar central, en la protección de la privacidad, en el sentido de que las personas tienen que ser plenamente conscientes de la forma en que se va a utilizar la información antes de que tal información sea recolectada (UNESCO, 2022). Sin embargo, en el contexto del Big Data y la IA, este principio enfrenta una crisis estructural. La gran cantidad de información que se extrae de los usuarios proviene de interacciones digitales cotidianas, en las que los términos de uso y políticas de privacidad suelen ser extensos, técnicos y difíciles de comprender (Nonju, 2021). Como resultado, la mayoría de los usuarios acepta el procesamiento de sus datos sin una comprensión real de las implicaciones.

Además, el uso de datos inferidos por algoritmos complejos añade un nuevo nivel de desafío. Aunque un usuario no haya proporcionado explícitamente cierta información, los sistemas de IA pueden deducir patrones de comportamiento a partir de su actividad en línea, lo que plantea interrogantes sobre si el consentimiento informado sigue siendo válido en estos escenarios (Gilbert, 2021).

Transparencia: la necesidad de un acceso claro a la información

La transparencia en el manejo de los datos hace referencia a la existencia de información clara para los usuarios, la cual tiene que ver con la manera en que su información personal es recogida y utilizada. Sin embargo, estrategias de prueba han puesto de manifiesto que, durante el uso de plataformas digitales, muchas empresas recurren a prácticas opacas que hacen muy difícil para los usuarios entender cómo se utilizan sus datos (Hu & Min, 2022).

La cuestión de la transparencia resulta ser muy grave en el ámbito de los sistemas de inteligencia artificial, dado que los procesos de toma de decisiones pueden ser altamente oscuros, problema que también se conoce como el problema de la "caja negra" en la inteligencia artificial (Oseni et al., 2022). La opacidad de los modelos algorítmicos hace difícil que las empresas puedan ser responsables acerca de sus decisiones y también da lugar a que los sesgos queden cubiertos por los procesos de análisis en la toma de decisiones, generando discriminación algorítmica en decisiones que afectan la concesión de créditos, la selección de candidatos para puestos de trabajo o en los sistemas de vigilancia predictiva.

Desde la óptica de la ética, asegurar la transparencia no sólo requiere de la creación de regulaciones más estrictas, sino que también implica hacer accesibles mecanismos para que los usuarios tengan la posibilidad de controlar su información. Esto es lo que se entiende como modelos de "explainable IA" a partir de los cuales los sistemas

pueden dar razones de uno u otro tipo sobre cómo y por qué toman decisiones que derivan de los datos (Manyika et al., 2011).

En resumen, la autonomía, el consentimiento informado y la transparencia son principios que deben guiar el uso ético del Big Data, si bien, en un plano práctico, estos valores se ven puestos en riesgo por la detección masiva, por la falta de mecanismos de control del usuario y por una falta creciente de transparencia en los sistemas de IA. Para garantizar un uso ético de los datos masivos, es fundamental desarrollar estrategias regulatorias y técnicas que refuercen estos principios, asegurando que la privacidad y los derechos digitales no sean sacrificados en nombre de la innovación.

Riesgos y desafíos éticos en el manejo del big data

Ya en el punto anterior empezamos a vislumbrar que el Big Data, en combinación la IA, ha cambiado la forma en que se recopila, analiza y utiliza la información, facilitando avances significativos en sectores como la salud, la seguridad y el comercio. Pero también ha traído dilemas importantes sobre la privacidad, la protección de datos y la justicia en su manejo. A medida que las empresas y los gobiernos son más hábiles para recopilar grandes cantidades de datos, los riesgos de que la privacidad se vea afectada y de que ocurran sesgos en los algoritmos han ido en aumento. (Solove, 2021; Gulati, 2020).

Uno de los desafíos más importantes es el de la privacidad, que ha sido erosionada por el aumento de la vigilancia masiva y el uso de datos personales sin consentimiento explícito (Hu & Min, 2022). La recopilación indiscriminada de información genera un terreno fértil para violaciones de derechos fundamentales, donde los individuos pierden control sobre cómo se utilizan sus datos. A esto se suman las brechas de seguridad, que exponen información sensible a ataques cibernéticos, afectando tanto a usuarios individuales como a infraestructuras críticas (Oseni et al., 2022).

Otro problema clave es la discriminación algorítmica, tema que será tratado a mayor profundidad en otro capítulo de este libro, que surge cuando los modelos de IA refuerzan desigualdades existentes debido a sesgos en los datos con los que han sido entrenados (Mühlhoff, 2021). Esto puede derivar en decisiones sesgadas en sectores clave como la contratación laboral, el acceso al crédito y la vigilancia predictiva, afectando de manera desproporcionada a comunidades vulnerables. Finalmente, la anonimización de datos, promovida como una solución para proteger la privacidad, presenta limitaciones importantes. Estudios han demostrado que incluso cuando los datos personales han sido despojados de identificadores directos, es posible reconstruir información sensible a partir de patrones y correlaciones ocultas (Gilbert, 2021). Un estudio publicado en Nature Communications concluyó que el 99,98% de los ciudadanos estadounidenses podrían ser re-identificados, luego de que sus datos hubiesen sido anonimizados, en un base de datos usando 15 atributos demográficos (Rocher et al., 2019). Esto podría parecer imposible, pero es mucho más sencillo de los que parece.

Ante estos desafíos, afrontamos como sociedad el desafío no solo de reconocer los riesgos inherentes al manejo masivo de datos, sino también de diseñar estrategias regulatorias y técnicas que minimicen sus impactos negativos. En las siguientes secciones, abordaré en detalle, lo que me permite la extensión de este capítulo, cada uno de estos riesgos, analizando casos concretos y explorando soluciones que permitan un equilibrio entre innovación y protección de derechos fundamentales.

Privacidad

La privacidad es un derecho fundamental, por lo menos en Colombia, que ha sido puesto en riesgo y vulnerado durante la era del Big Data y la inteligencia artificial (IA). Si bien no podemos desconocer que la recopilación y análisis de datos masivos han permitido avances significativos en sectores como la salud, la economía y la seguridad,

también han facilitado prácticas de vigilancia masiva, manipulación de la información y discriminación algorítmica (Nonju & Ihua-Maduenyi, 2024). La falta de regulaciones efectivas y la rápida evolución de las tecnologías han permitido que gobiernos y corporaciones exploten la información personal de millones de personas sin su consentimiento explícito (Oseni et al., 2020).

En este contexto, se han documentado violaciones masivas de la privacidad que han tenido consecuencias francamente terribles para los derechos fundamentales de los ciudadanos de diferentes latitudes. Los casos que expondré a continuación ilustran cómo el acceso indiscriminado a datos personales puede comprometer la libertad de expresión, la autonomía personal, la igualdad de oportunidades y el debido proceso (Hu & Min, 2023). A continuación, voy a citar algunos de los incidentes más representativos y sus implicaciones éticas y legales.

El escándalo de Facebook-Cambridge Analytica

Uno de los casos más conocidos de violación de la privacidad fue el escándalo de Facebook y Cambridge Analytica. Aquí, se recogieron y analizaron los datos de más de 87 millones de usuarios sin que ellos lo supieran o dieran su consentimiento (Nonju & Ihua-Maduenyi, 2024). El objetivo era manipular elecciones y estrategias de comunicación política. Usando modelos de microsegmentación con IA, la empresa creó campañas específicas para influir en la opinión pública en momentos clave, como el referéndum del Brexit en el Reino Unido y las elecciones presidenciales en EE.UU. en 2016.

Este caso reveló una falla estructural en la protección de los datos personales y demostró cómo la explotación de información privada puede erosionar la integridad democrática (Oseni et al., 2020). La capacidad de las grandes corporaciones para manipular la percepción pública sin mecanismos de transparencia ni regulación efectiva representa una amenaza directa contra el derecho a la información, la libertad de expresión y la autodeterminación informativa (Gulati, 2020).

Clearview AI y el abuso del reconocimiento facial

El incidente relacionado al software Clearview AI establece un precedente significativo en términos de violación de la privacidad. Esta compañía creó un sistema de reconocimiento facial basado en más de 3 mil millones de imágenes obtenidas sin permiso de redes sociales y páginas web; esto ha facilitado su utilización por parte de entidades gubernamentales y compañías privadas sin la autorización de las personas afectadas según lo mencionan Hu & Min en su estudio del año 2023.

El principal problema reside en la falta de transparencia y control sobre los datos biométricos que violan principios básicos de privacidad y protección de datos. Estudios han mostrado que estas tecnologías de reconocimiento facial presentan altos niveles de sesgo racial y de género lo cual las convierte en herramientas susceptibles para discriminar sistémicamente y afectando desproporcionadamente a comunidades vulnerables (Piplai et al., 2023). Además de esto, la aplicación de tales dispositivos en lugares públicos sin reglas claras puede amenazar la libertad de reunión y la presunción de inocencia, ya que podrían ser empleados para vigilar y castigar injustamente ciertos grupos de individuos (Osen et al., 2020).

Aadhaar y la violación de la privacidad en la India

El sistema de identificación biométrica Aadhaar en la India ha causado controversia debido a la recolección extensa de datos personales de más de 1.300 millones de ciudadanos. Esta plataforma almacena información como huellas dactilares e iris y ha sido objeto de ciberataques y filtraciones de datos que han expuesto a millones de personas al robo de identidad y la marginación social (Nonju & Ihua-Maduenyi, 2024). Pese a que el tribunal supremo de la India reconoció la privacidad como un derecho fundamental en el caso Justice K.S Puttaswamy v Union of India lo obligatorio del uso de Aadhaar para acceder a servicios esenciales ha generado serias interrogantes éticas (Osenniet al 2020). La concentración de infor-

mación biométrica sin medidas sólidas de protección crea vulnerabilidades sistémicas que podrían ser aprovechadas por individuos maliciosos. Esto resalta la necesidad de asegurar la protección de los datos almacenados y de limitar el acceso gubernamental sin una justificación legal clara (Piplai et al., 2023).

Seguridad

Las filtraciones de datos son una de las debilidades más graves en la era digital actual. Desde brechas accidentales hasta ataques sofisticados dirigidos de manera específica; la revelación de información delicada pone en riesgo no solo la seguridad de personas y entidades particulares, sino también valores esenciales como la privacidad, confianza , y el derecho al control sobre la propia información (Osení et al., 2022).

La manipulación de datos filtrados por individuos malintencionados ha experimentado un progreso notable gracias al uso de la Inteligencia Artificial (IA) y métodos avanzados para el análisis de información. Delincuentes informáticos emplean datos comprometedores para llevar a cabo estafas, ejecutar tácticas de ingeniería social e incluso influir en sistemas democráticos (Hu & Min, 2023). La complejidad asociada a estas amenazas requiere una revisión exhaustiva en cuanto a los sistemas de seguridad y las estrategias para mitigar riesgos. A pesar del progreso continuo en este ámbito específico se observan aún fallos tanto normativos como técnicos (Elazar et al., 2024).

Este segmento explora los tipos primordiales de lagunas de datos, sus efectos adversos y la utilización abusiva del contenido filtrado, para posteriormente examinar las tácticas para reducir el riesgo y las dificultades que se presentan en la protección de datos en el marco ámbito del Big Data.

Tipos de brechas de datos y su impacto

Las brechas de datos pueden clasificarse en tres grandes categorías, cada una con implicaciones significativas:

- Exposición accidental: errores de configuración en bases de datos, fallos en protocolos de cifrado y almacenamiento en la nube sin medidas de seguridad adecuadas han sido responsables de filtraciones masivas de información (De Mauro, Greco & Grimaldi, 2015). Un caso emblemático fue la filtración de datos de AWS S3 Buckets, donde millones de registros quedaron expuestos por configuraciones incorrectas, permitiendo accesos no autorizados (Oseni et al., 2022).

- Ataques dirigidos: los ciberdelincuentes emplean técnicas como ransomware, phishing avanzado y explotación de vulnerabilidades en software para acceder a datos sensibles. El caso de Equifax (2017), en el que se comprometieron los datos financieros de 147 millones de personas, demuestra cómo la falta de actualización de sistemas puede facilitar ataques a gran escala (Nonju & Ihua-Maduenyi, 2024).

- Filtraciones internas: empleados o contratistas con acceso privilegiado han filtrado datos intencionalmente, ya sea por motivaciones financieras, espionaje corporativo o activismo. Edward Snowden y la filtración de documentos de la NSA es un ejemplo de cómo actores internos pueden comprometer la seguridad a nivel gubernamental (Hu & Min, 2023).

El impacto de estas brechas se extiende más allá de lo económico. Afectan la confianza en las instituciones, vulneran derechos fundamentales y generan impactos legales que, en muchas ocasiones, las empresas y gobiernos no están preparados para abordar adecuadamente (Elazar et al., 2024).

Explotación de datos por actores malintencionados

Los datos filtrados se han convertido en un recurso valioso para actividades ilegales en la dark web y otras plataformas clandestinas. Entre las principales formas de explotación se encuentran:

- Venta de información sensible: bases de datos con credenciales de acceso, historiales médicos y datos bancarios son comercializadas en foros clandestinos, facilitando fraudes y robos de identidad (Gulati, 2020).
- Ataques de ingeniería social: con acceso a información personal, los atacantes pueden diseñar campañas altamente personalizadas de phishing avanzado, engañando a las víctimas para obtener más datos o ejecutar transacciones fraudulentas (Oseni et al., 2022).
- Manipulación de modelos de IA: ciberdelincuentes han empezado a inyectar datos falsos en conjuntos de entrenamiento de modelos de IA para alterar sus resultados, lo que representa un nuevo desafío en la seguridad algorítmica (Carroll et al., 2023).
- Extorsión y chantaje: en ataques de ransomware, los delincuentes encriptan información sensible y exigen rescates para no divulgarla. Casos como el de Colonial Pipeline (2021) demostraron la vulnerabilidad de infraestructuras críticas frente a este tipo de amenazas (De Mauro et al., 2015).

Uno de los problemas más alarmantes es la creciente sofisticación de estas prácticas mediante el uso de IA generativa. Investigaciones recientes han demostrado cómo los modelos de lenguaje pueden ser utilizados para automatizar ataques de spear phishing y generar contenido falso con niveles de realismo sin precedentes (Harari, 2024).

Estrategias de mitigación y desafíos regulatorios

Para enfrentar estas amenazas, es esencial adoptar un enfoque integral de seguridad que combine medidas tecnológicas, organizativas y regulatorias. Algunas estrategias clave incluyen:

- Cifrado avanzado: implementar algoritmos de cifrado robustos, como AES-256, para garantizar la protección de datos en tránsito y en reposo (Oseni et al., 2022).
- Autenticación multifactor (MFA): el uso de autenticación de doble factor reduce significativamente el riesgo de accesos no autorizados (Piplai et al., 2023).
- Monitorización con IA: herramientas de detección de anomalías basadas en machine learning permiten identificar patrones sospechosos en tiempo real, mejorando la respuesta ante incidentes de seguridad (Elazar et al., 2024).
- Regulación y cumplimiento normativo: marcos como el Reglamento General de Protección de Datos (GDPR) y la Ley de Privacidad del Consumidor de California (CCPA) han establecido estándares más estrictos en la protección de datos. Sin embargo, la falta de armonización global dificulta su aplicación efectiva (Nonju & Ihua-Maduenyi, 2024).

Pese a estos avances, persisten desafíos críticos. En primer lugar, muchas empresas trasladan sus operaciones a jurisdicciones con regulaciones más laxas, evitando cumplir con normativas más estrictas. Además, la velocidad con la que evolucionan las amenazas supera la capacidad de los marcos legales para adaptarse a nuevos riesgos (Hu & Min, 2023).

Para concluir este punto, me gustaría indicar que las brechas de datos y su explotación por actores malintencionados representan una amenaza creciente en la era del Big Data y la IA. La sofisticación

de los ataques, impulsada por la automatización y la explotación de vulnerabilidades, demanda un enfoque de seguridad que no solo se base en tecnología, sino que también incluya normativas más estrictas y una mayor conciencia social sobre el uso de la información personal.

En el contexto actual de la escritura de este capítulo (marzo de 2025), se ha hecho indispensable fortalecer los marcos regulatorios, mejorar la transparencia en el manejo de datos y desarrollar modelos de IA con mecanismos de seguridad integrados. Sin estos esfuerzos, el avance de la tecnología continuará generando riesgos significativos para la privacidad, la confianza digital y otros temas derivados de estos últimos, entre los que se encuentran la desconfianza en las instituciones, la erosión de la democracia y la manipulación de los ciudadanos.

Discriminación algorítmica

La discriminación algorítmica es uno de los desafíos más complejos en el desarrollo y uso de la inteligencia artificial (IA). Si bien uno de los argumentos más utilizados por algunos autores es que los algoritmos pueden reducir los sesgos humanos mediante el uso de modelos estadísticos objetivos, la realidad demuestra que los sistemas de IA a menudo perpetúan e incluso amplifican desigualdades preexistentes en la sociedad (Solove, 2023).

Origen del sesgo algorítmico

Los sesgos en la IA pueden surgir en distintas etapas del ciclo de vida del modelo, siendo las más críticas la recopilación de datos, la selección de variables y el diseño del algoritmo. De acuerdo con Barocas y Selbst (2016), los algoritmos no funcionan en el vacío, sino que reflejan los patrones históricos presentes en los datos con los que han sido entrenados. Si esos datos contienen prejuicios sistemáticos,

como disparidades salariales de género o perfiles raciales en el ámbito judicial, la IA aprenderá y replicará esas injusticias.

Impacto en derechos fundamentales

La perpetuación de sesgos por parte de la inteligencia artificial no solo afecta la igualdad en el acceso al empleo que también tiene efectos más profundos en áreas como la obtención de crédito, servicios médicos y la protección ciudadana. Dentro del sector financiero los modelos utilizados para evaluar el historial crediticio han sido objeto de críticas por discriminar contra personas pertenecientes a comunidades racializadas al asignar un mayor nivel riesgo crediticio basándose en patrones históricos en lugar considerar su solvencia real (Kim 2020) En el campo de la salud, estudios han mostrado que algunos sistemas de inteligencia artificial destinan menos recursos médicos para pacientes de origen africano debido al historial desigual en el acceso al sistema de salud (Gichoya et al., 2022).

Propuestas para mitigar la discriminación algorítmica

Para resolver estas dificultades se han propuesto diversas estrategias en el ámbito técnico y normativo. Comenzar por recolectar y curar datos más diversos y representativos es un paso fundamental para disminuir la parcialidad en la fase de entrenamiento de los modelos (Comisión Europea, 2021). En cuanto a la regulación, iniciativas como el *Artificial Intelligence Act* de la Unión Europea buscan establecer criterios de justicia y transparencia en el desarrollo y uso de IA.

Además de eso, enfoques como la revisión algorítmica y la aplicación de técnicas de aprendizaje automático conscientes de la equidad pueden ser útiles para detectar y solucionar prejuicios antes de que los modelos se pongan en marcha en entornos delicados (IBM AI Fairness 360, 2023).

La discriminación algorítmica es una problemática que no solo refleja las desigualdades estructurales de la sociedad, sino que también tiene el potencial de amplificarlas y perpetuarlas si no se toman medidas adecuadas. El desarrollo de marcos regulatorios específicos, junto con mejoras en la calidad de los datos y en las metodologías de entrenamiento de IA, son claves para evitar que la tecnología se convierta en un agente de injusticia social en lugar de una herramienta de equidad y progreso.

Anonimización de Datos

La anonimización de datos ha sido presentada como una estrategia clave para equilibrar la necesidad de utilizar información personal en análisis de Big Data con la obligación de proteger la privacidad de los individuos. Sin embargo, su efectividad es cuestionada debido a los avances en técnicas de reidentificación y la ambigüedad legal sobre qué constituye una anonimización efectiva (Li & Zhang, 2017). Aunque en muchos contextos los datos anonimizados se consideran exentos de regulaciones de privacidad, investigaciones han demostrado que es posible revertir este proceso con un alto grado de precisión, poniendo en duda la seguridad de este enfoque (Carroll et al., 2023).

Este apartado analizará las principales limitaciones técnicas de la anonimización, su impacto en el derecho a la privacidad y las paradojas éticas que plantea su implementación en sectores como la inteligencia artificial, la investigación médica y la seguridad.

Limitaciones en técnicas de anonimización

A pesar de los avances en técnicas como el data masking, la agregación y la pseudonimización, la realidad es que la mayoría de los métodos de anonimización pueden ser revertidos si se dispone de bases de datos complementarias para cruzar la información (Rocher et al., 2019).

Reidentificación a partir de datos anonimizados

Uno de los estudios más influyentes en este campo fue el realizado por Sweeney (2000), en el cual demostró que el 87% de la población de EE.UU. podía ser reidentificada utilizando solo tres datos anonimizados: código postal, género y fecha de nacimiento. Esto pone en evidencia que, incluso eliminando nombres y direcciones, los patrones individuales pueden ser reconstruidos con información accesible públicamente (Sweeney, 2000). Un estudio mucho más reciente de Rocher, L., Hendrickx, J. M., & De Montjoye eleva el porcentaje hasta el 99,98%.

Vulnerabilidad a ataques de correlación

El caso de la filtración de datos médicos en Australia en 2016 mostró que un grupo de investigadores logró reidentificar los historiales clínicos de personas a partir de datos anonimizados mediante el cruce con redes sociales y otras bases de datos públicas (Golle, 2006). Esto demuestra que la anonimización efectiva es cada vez más difícil debido a la enorme cantidad de datos disponibles en fuentes abiertas.

Compromiso entre la anonimización y utilidad de los datos

Un desafío clave en la anonimización es que, a medida que se aplican técnicas para proteger la identidad de los individuos, se reduce la calidad y utilidad de los datos para fines analíticos (Dwork & Roth, 2014). En el ámbito de la salud, por ejemplo, la eliminación de ciertos atributos personales en bases de datos de pacientes puede afectar la precisión de modelos de predicción basados en inteligencia artificial (Schneier, 2023). Adicionalmente, la IAG depende exclusivamente en la cantidad y calidad de los datos para entregar resultados adecuados y esto incluye, por supuesto, los datos personales de los ciudadanos.

Paradojas éticas en la anonimización de los datos

Si bien la anonimización se promociona como una forma de proteger la privacidad, en la práctica plantea dilemas éticos importantes que deben ser abordados desde una perspectiva crítica.

Falsa sensación de seguridad

Las empresas y gobiernos suelen tratar los datos anonimizados como si fueran completamente seguros y libres de restricciones legales. Sin embargo, el riesgo de reidentificación es real y está bien documentado en la literatura científica (Rocher et al., 2019). Esta falsa sensación de seguridad ha llevado a la relajación de políticas de protección de datos, permitiendo la explotación de información personal sin el conocimiento o consentimiento de los usuarios.

¿Quién controla los datos anonimizados?

Un problema ético clave es que, en muchos casos, los datos anonimizados pueden ser utilizados sin restricciones por entidades que no participaron en su recopilación original. Según Brill & Harris (2018), esto plantea dudas sobre la autodeterminación informativa, ya que los individuos no pueden ejercer control sobre la información que, aunque anonimizada, sigue refiriéndose a ellos.

El dilema de la investigación científica vs la privacidad

En sectores como la salud y la epidemiología, el acceso a datos detallados es crucial para la investigación médica y el desarrollo de tratamientos. Sin embargo, las restricciones de privacidad y anonimización pueden dificultar la obtención de información precisa. En el contexto del COVID-19, por ejemplo, varios estudios demostraron que los datos anonimizados eran menos útiles para el rastreo de contagios debido a la eliminación de información clave sobre movilidad y contactos (Leslie, 2020).

Regulaciones y alternativas a la anonimización

Dado que la anonimización no garantiza una protección absoluta de la privacidad, diversas normativas han establecido enfoques complementarios para minimizar los riesgos.

Regulación del uso de datos anonimizados

El Reglamento General de Protección de Datos (GDPR) de la Unión Europea establece que los datos anonimizados pueden quedar exentos de ciertas restricciones, pero si existe la posibilidad de reidentificación, deben considerarse datos personales y estar sujetos a protección (European Commission, 2021).

Transparencia y consentimiento informado

Más allá de la regulación técnica, la solución a los problemas éticos de la anonimización requiere una mayor transparencia en la recolección y uso de datos, así como mecanismos efectivos para que los individuos puedan decidir activamente sobre el tratamiento de su información personal (Solove, 2023).

En conclusión, si bien la anonimización sigue siendo una herramienta valiosa en la protección de la privacidad, sus limitaciones y paradojas éticas no pueden ser ignoradas. La creciente sofisticación de técnicas de reidentificación hace que confiar únicamente en la anonimización sea insuficiente para garantizar la seguridad de los datos. Por lo tanto, es fundamental combinar estrategias técnicas, regulatorias y éticas que permitan un equilibrio entre privacidad y utilidad. El desarrollo de nuevas metodologías como la privacidad diferencial, junto con una mayor transparencia y regulación del uso de datos, será clave para proteger los derechos individuales en la era del Big Data.

Marco legal y regulaciones globales

El uso masivo de datos ha transformado las dinámicas de múltiples sectores, desde la economía hasta la gobernanza y la seguridad. Sin embargo, este crecimiento exponencial también ha traído consigo desafíos, que ya he enumerado y explicado a lo largo de este capítulo, en materia de privacidad, derechos fundamentales y regulación. Diferentes jurisdicciones, que no todas, han intentado responder a estos desafíos con marcos normativos que buscan equilibrar el aprovechamiento de la tecnología con la protección de los datos personales. No obstante, la fragmentación regulatoria entre diferentes países y regiones ha generado incertidumbre tanto para los usuarios como para las empresas que operan a nivel global (The Impact of Artificial Intelligence on Privacy Laws, 2024).

Las regulaciones más destacadas en materia de protección de datos incluyen el Reglamento General de Protección de Datos (GDPR) de la Unión Europea, la Ley de Privacidad del Consumidor de California (CCPA) en Estados Unidos y, ya más cerca a nuestras latitudes, no existe ningún tipo de norma comunitaria que trata de forma específica el tema, por lo que lo más cercano serían las leyes nacionales, que en el caso colombiano es la Ley 1581 de 2012. Si bien estas normativas buscan un objetivo común –la protección de los datos personales–, presentan diferencias significativas en cuanto a enfoques, derechos reconocidos y obligaciones para las empresas (Kim, P., & Bodie, M. T. (2021). En este apartado se analizarán los principales marcos regulatorios, sus alcances y limitaciones, y las dificultades que plantea la falta de un estándar global armonizado.

Análisis comparativo de regulaciones: GDPR, CCPA y Leyes Colombianas

El análisis comparativo de regulaciones en materia de protección de datos y privacidad refleja las distintas aproximaciones que los

marcos normativos han adoptado en función de sus prioridades económicas, políticas y sociales. En términos generales, el Reglamento General de Protección de Datos (GDPR) de la Unión Europea es el estándar más robusto y amplio en términos de protección de datos personales, imponiendo estrictas obligaciones a los responsables del tratamiento y otorgando a los ciudadanos derechos extensivos sobre sus datos, incluyendo el derecho a la portabilidad y el derecho al olvido (Regulation 2016/679, GDPR).

En contraste, en Estados Unidos de América no existe ningún tipo de norma sobre datos personales a nivel federal, por lo que nos debemos concentrar en la Ley de Privacidad del Consumidor de California (CCPA), que nos ofrece un marco menos restrictivo en comparación con el GDPR, permitiendo a las empresas cierto margen de maniobra en el tratamiento de datos, aunque otorga a los consumidores el derecho a conocer qué información se recopila sobre ellos y a solicitar su eliminación en determinadas circunstancias (California Civil Code § 1798.100). Sin embargo, su enfoque sigue siendo más sectorial y reactivo en comparación con la visión integral del GDPR.

En el caso de Colombia, la Ley 1581 de 2012 establece un marco basado en el consentimiento informado y la responsabilidad del tratamiento de datos, aunque aún enfrenta desafíos en términos de implementación efectiva y sanciones disuasorias.

Desde una perspectiva crítica, la fragmentación normativa y la falta de un estándar global homogéneo generan incertidumbre tanto para las empresas que operan en múltiples jurisdicciones como para los individuos cuyos datos pueden ser tratados bajo regulaciones dispares. En un entorno donde la inteligencia artificial y el Big Data operan sin restricciones geográficas, la ausencia de armonización legislativa facilita la explotación de vacíos legales mediante prácticas de arbitraje regulatorio. Esto no solo afecta la protección de la privacidad, sino que también incrementa los riesgos de discriminación algorítmica y manipulación de datos con fines comerciales o políticos.

Por ello, es crucial avanzar hacia modelos regulatorios que combinen la flexibilidad requerida por la innovación tecnológica con principios fundamentales de protección de datos, promoviendo un enfoque basado en riesgos y garantizando la rendición de cuentas de los actores involucrados. En este sentido, iniciativas internacionales como la propuesta de la ONU sobre gobernanza de la inteligencia artificial y los esfuerzos del G20 por establecer principios comunes pueden ser el punto de partida para una convergencia regulatoria más efectiva.

Insuficiencia de las normativas tradicionales frente a las tecnologías emergentes

Las normativas tradicionales en materia de protección de datos, diseñadas principalmente en un contexto pre-algoritmico y pre-automatizado, no logran responder adecuadamente a los desafíos éticos, técnicos y jurídicos que plantea el uso de tecnologías emergentes como la inteligencia artificial (IA), el Big Data o el aprendizaje automático. Aunque los instrumentos mencionados previamente representan avances, sus estructuras conceptuales, al ser previos al "boom" de la Inteligencia Artificial Generativa, muestran limitaciones graves frente a las nuevas formas de recopilación, inferencia y explotación de datos.

Uno de los problemas más notorios es que estas leyes se fundamentan en un modelo de consentimiento individual, transparencia y control por parte del titular de los datos. Sin embargo, este modelo entra en crisis ante sistemas de IA capaces de inferir datos personales sensibles sin haberlos recolectado directamente, violando así el espíritu –aunque no siempre la letra– de estas normativas. Como señala Solove (2024), la IA remixa viejos problemas de privacidad –como el uso sin consentimiento, la vigilancia encubierta o la falta de transparencia– y los amplifica mediante nuevas formas de procesamiento como la generación de datos inferidos o las decisiones automatizadas opacas.

Por ejemplo, el artículo 22 del GDPR reconoce el derecho a no ser objeto de decisiones basadas exclusivamente en procesos automatizados que produzcan efectos jurídicos significativos. No obstante, su aplicación práctica es limitada, ya que basta con una "intervención humana" formal para que las empresas eludan esta protección, aun cuando dicha intervención no tenga un impacto sustancial en el resultado (Kaminski & Urban, 2021; Jones, 2017, citados en Solove, 2024) . Este enfoque ha sido criticado por su rigidez conceptual y su escasa eficacia para mitigar los riesgos reales que surgen del uso masivo de algoritmos en la toma de decisiones.

A esto se suma el hecho de que la regulación actual no contempla adecuadamente el fenómeno de la generación algorítmica de datos. Las normativas vigentes regulan la recopilación y el uso de datos, pero no el momento en que los algoritmos infieren o generan nueva información a partir de patrones ocultos, como la predicción de enfermedades, tendencias políticas o características personales (Solove, 2024). Estas inferencias, si bien no provienen de una recolección directa, pueden tener impactos profundos en la vida de los individuos y deberían ser objeto de regulación sustantiva. Adicional a todo esto, también se genera una grave afectación a los derechos de propiedad intelectual, tema que no trataré en este capítulo, pero que per se daría para un libro entero.

El problema se agrava en escenarios de scraping masivo, como el protagonizado por Clearview AI, que construyó una de las bases de datos de reconocimiento facial más extensas del mundo sin consentimiento alguno, extrayendo imágenes públicas de redes sociales y sitios web. Aunque la empresa fue sancionada en varios países, su caso pone en evidencia cómo la IA opera en zonas grises normativas donde los datos técnicamente públicos son usados con fines inesperados, en clara contravía de los principios de finalidad, minimización y proporcionalidad (Hill, 2023; Solove, 2024).

Por otro lado, los marcos regulatorios tienden a ser territoriales, mientras que los flujos de datos y las aplicaciones de IA son globales. Esto ha dado lugar a prácticas de arbitraje regulatorio, donde las

empresas trasladan sus operaciones a jurisdicciones con normativas más laxas para evadir restricciones más estrictas, como las impuestas por el GDPR (Kuner, 2020, citado en The Impact of Artificial Intelligence on Privacy Laws, 2024).

Este escenario de insuficiencia normativa evidencia la necesidad de avanzar hacia enfoques más dinámicos, basados en riesgos, e incluso hacia una arquitectura regulatoria transnacional. Se requiere una gobernanza más adaptativa, que reconozca el carácter híbrido y evolutivo de la IA, y que se base no solamente en los derechos del titular de los datos, sino también en principios de responsabilidad algorítmica, rendición de cuentas, y prevención de daños colectivos. Como propone Solove (2024), el derecho debe alejarse del enfoque centrado en el consentimiento individual y avanzar hacia un modelo que regule los usos, los riesgos y los daños provocados por sistemas automatizados, aun cuando no exista una violación explícita de los datos tradicionales.

Propuestas legales y éticas para un manejo responsable del Big Data

Las propuestas legales y éticas para un manejo responsable del Big Data, y su uso en los sistemas de inteligencia artificial (IA), han adquirido una importancia central ante la evidente insuficiencia de los marcos normativos tradicionales. La complejidad y escala de los sistemas de inteligencia artificial (IA) que procesan grandes volúmenes de datos exigen una respuesta regulatoria que no solo se limite a principios generales, sino que articule mecanismos concretos de gobernanza, rendición de cuentas y protección de los derechos fundamentales.

Un primer eje de estas propuestas está orientado hacia la adopción de principios éticos transversales. Diversos organismos internacionales han coincidido en que toda arquitectura regulatoria debe fundarse en valores como la equidad, la transparencia, la responsabilidad y la

privacidad. Por ejemplo, la Comisión Europea en sus *Ethics Guidelines for Trustworthy AI* plantea que los sistemas de IA deben ser justos, evitar la discriminación, ser explicables y garantizar mecanismos de reparación en caso de daño. Estos principios no son solo declaraciones aspiracionales, sino que deben operacionalizarse a través de auditorías, marcos de diseño ético y procesos deliberativos en cada fase del ciclo de vida del dato y del sistema de IA.

Asimismo, la responsabilidad debe ser distribuida y verificable. Las directrices de la OCDE y la UNESCO han subrayado la necesidad de establecer estructuras de gobernanza ética dentro de las organizaciones que diseñan y utilizan IA, incluyendo comités interdisciplinarios, auditorías regulares, y procesos de evaluación de impacto ético (AI impact assessments). Estas estructuras buscan garantizar que el uso del Big Data esté alineado con los derechos humanos y no agrave desigualdades preexistentes.

En el ámbito jurídico, una de las propuestas más relevantes es la necesidad de avanzar hacia un modelo de regulación basado en riesgos. El *Artificial Intelligence Act* de la Unión Europea es paradigmático en este sentido: en lugar de prohibir o permitir toda tecnología por igual, propone una clasificación por niveles de riesgo (inaceptable, alto y bajo), imponiendo requisitos más estrictos a los sistemas de IA que puedan afectar gravemente los derechos fundamentales, como los utilizados en vigilancia biométrica, selección de personal o justicia penal (Solove, 2024).

Este enfoque se distancia del modelo tradicional centrado exclusivamente en el consentimiento del titular de los datos, el cual ha demostrado ser insuficiente y fácilmente manipulable. Como señala Solove, el consentimiento en el contexto del Big Data se convierte frecuentemente en un "cheque en blanco" que los usuarios otorgan sin comprender el alcance real del tratamiento de su información. Por ello, se propone desplazar el énfasis hacia obligaciones estructurales de las organizaciones que recolectan y procesan datos, exigiendo prácticas responsables, estándares técnicos robustos y mecanismos efectivos de fiscalización.

Ahora bien, no puedo dejar de exponer que este modelo adoptado por la Unión Europea también tiene sus contradictores, quienes se amparan en los efectos económicos del mismo. Principalmente se apalancan en que este tipo de regulaciones general un efecto disuasorio en los inversionistas, lo que al final redunda en falta de inversión en sistemas de IA en Europa, e ineludiblemente, un atraso en su adopción. Particularmente el Fondo Monetario Internacional en un informe publicado en 2024 indica que los principales efectos son: (i) Altos costos de cumplimiento regulatorio; (ii) Desventaja competitiva global; (iii) Concentración del mercado; (iv) Riesgo de retraso en la adopción de IA (Choudhary, Cortina Lorente, & Kemp, 2024). Como advierten los autores, "*the ex-ante system puts a high burden on developers only, giving an advantage to larger firms who can pay the legal consequences, and thus potentially limiting technological improvements and innovation*".[1]

Otra línea fundamental es la implementación de tecnologías de privacidad incorporada desde el diseño, como el cifrado, el aprendizaje federado y la privacidad diferencial, que permiten tratar datos sin necesidad de centralizarlos o exponer información sensible. Estas tecnologías deben ser complementadas con políticas de anonimización responsable y limitación de la retención de datos, garantizando que solo se procesen los datos necesarios y durante el tiempo estrictamente requerido.

Finalmente, se ha planteado la creación de organismos regulatorios internacionales para enfrentar la fragmentación legal que permite el arbitraje normativo por parte de las grandes plataformas tecnológicas. Iniciativas como las Reglas de Privacidad Transfronteriza (CBPR) de APEC o los acuerdos de adecuación del GDPR en la Unión Europea pueden servir de base para un modelo global de interoperabilidad normativa que promueva estándares mínimos universales de protección de datos y derechos digitales.

---

1 Traducción propia: "el sistema ex ante imponer una gran carga únicamente a los desarrolladores, lo que da una ventaja a las grandes empresas que pueden asumir las consecuencias legales, y por tanto podría limitar las mejoras tecnológicas y la innovación."

En conclusión, un manejo responsable del Big Data exige mucho más que reformas técnicas: requiere una transformación profunda en la forma como entendemos el poder informacional. Las propuestas éticas y legales deben complementarse mutuamente para construir una arquitectura regulatoria capaz de garantizar que la tecnología esté al servicio de la dignidad humana y no al revés. Tal como advierte la UNESCO en su *Recomendación sobre la Ética de la Inteligencia Artificial*, el desarrollo tecnológico debe estar guiado por los principios de justicia, inclusión y respeto a los derechos humanos.

Propuestas éticas y buenas prácticas

Ante las limitaciones evidentes de las normativas tradicionales para responder efectivamente a los desafíos planteados por el Big Data y la inteligencia artificial, resulta crucial avanzar hacia la implementación de propuestas éticas y buenas prácticas que orienten el uso responsable de estas tecnologías emergentes. La adopción generalizada del Big Data, y su uso en sistemas de Inteligencia Artificial (IA) demanda un enfoque interdisciplinario y multiactor que garantice no solo el cumplimiento legal, sino la protección efectiva de los derechos fundamentales, la mitigación de sesgos algorítmicos y la promoción de una cultura consciente en materia de privacidad y seguridad digital (Jurewicz, Kaefer & Kran, 2024).

En este contexto, a continuación, se presentan tres grandes propuestas que permitirán abordar de manera ética y efectiva los desafíos actuales: el diseño ético de algoritmos y sistemas de inteligencia artificial, el fortalecimiento de una gobernanza de datos con participación multilateral y activa de las organizaciones, y finalmente, una educación y sensibilización robusta en materia de privacidad y seguridad, orientada hacia una sociedad digitalmente consciente y empoderada.

Diseño ético de algoritmos y sistemas de IA

El diseño ético de algoritmos y sistemas de inteligencia artificial no es solo una opción recomendable, sino una exigencia imperativa frente a las crecientes complejidades técnicas y sociales asociadas con estas tecnologías. Según Oseni et al. (2020), los algoritmos de IA son altamente susceptibles a ataques adversariales debido a vulnerabilidades inherentes a los modelos actuales, lo que pone de relieve la necesidad urgente de integrar seguridad desde su diseño inicial para proteger tanto la privacidad como la integridad de los datos. Por ello, el enfoque conocido como "privacidad por diseño" y "ética desde el diseño" debe constituir la piedra angular en la creación y despliegue de cualquier sistema de IA (Li & Zhang, 2017).

La Unión Europea, a través de sus "Directrices Éticas para una IA Confiable" (European Commission, 2019), ha planteado claramente que los algoritmos deben no solo cumplir con la legalidad, sino ser éticamente robustos, transparentes y explicables. Estos elementos ayudan a prevenir discriminaciones algorítmicas y aseguran la rendición de cuentas mediante auditorías constantes (European Commission, 2019).

Gobernanza de datos: Participación multisectorial y rol de las organizaciones

La gobernanza efectiva de los datos masivos requiere una participación multisectorial donde actores del ámbito público, privado y de la sociedad civil cooperen activamente. Alicia Solow-Niederman (2020) subraya la importancia de una voz pública activa y mecanismos de gobernanza democráticamente responsables para controlar la influencia comercial excesiva y garantizar una representación adecuada de los intereses ciudadanos en las decisiones sobre datos e IA.

En este contexto, la creación de comités éticos y órganos de gobernanza dentro de las organizaciones es fundamental. Estos órganos, integrados por tecnólogos, expertos en ética, juristas y representantes

comunitarios, son esenciales para asegurar que las decisiones tomadas alrededor de los datos y algoritmos sean justas, transparentes y socialmente responsables (Montreal Declaration, 2018). De igual forma, las organizaciones deben implementar evaluaciones continuas del impacto ético y social de sus prácticas algorítmicas, promoviendo así un compromiso real con el aprendizaje continuo y la adaptación frente a los constantes desafíos éticos emergentes.

Educación y sensibilización sobre privacidad y seguridad en la sociedad

La educación y la sensibilización sobre privacidad y seguridad constituyen uno de los pilares fundamentales para garantizar un manejo responsable del Big Data y las tecnologías emergentes. Frente a un entorno digital cada vez más complejo, caracterizado por la omnipresencia de la inteligencia artificial y el análisis masivo de datos personales, es indispensable que la sociedad comprenda los riesgos asociados al uso de su información y sea consciente de sus derechos fundamentales, incluyendo la privacidad, la protección de datos y el control sobre su identidad digital (Jurewicz, Kaefer & Kran, 2024).

Sin embargo, estudios recientes evidencian que el nivel general de conciencia social en torno a la privacidad digital es insuficiente. Gran parte de la población desconoce cómo se recopilan, procesan y utilizan sus datos, así como las implicaciones potenciales que ello puede tener en aspectos tan sensibles como el acceso al empleo, el crédito o la atención médica (Tsamados et al., 2022). Por esta razón, resulta indispensable implementar políticas públicas y programas educativos transversales que aborden desde edades tempranas estos temas, fomentando una cultura de prevención y responsabilidad digital.

Un aspecto crítico en estos procesos educativos es la necesidad de clarificar conceptos complejos como el consentimiento informado en el ámbito digital. La educación no solo debe ser técnica, sino también ética y legal, proporcionando a los ciudadanos herramientas prácticas

para que puedan ejercer efectivamente sus derechos y protegerse ante prácticas abusivas, discriminatorias o invasivas por parte de empresas o autoridades (Veale & Binns, 2017).

Asimismo, es fundamental el compromiso de múltiples actores sociales en estos procesos de sensibilización. Las instituciones educativas, las organizaciones civiles, los medios de comunicación, y especialmente las plataformas tecnológicas deben asumir un rol activo en la difusión de conocimientos sobre privacidad y seguridad. Por ejemplo, empresas líderes como Google y Apple ya han iniciado campañas de sensibilización pública en torno a la privacidad digital, aunque estos esfuerzos aún requieren mayor profundidad y sistematicidad (Brkan, 2023).

Finalmente, la sensibilización no debe reducirse a simples campañas informativas, sino que debe orientarse hacia un verdadero empoderamiento ciudadano. Esto implica promover la participación activa en decisiones sobre políticas públicas digitales, incentivar la auditoría ciudadana sobre el uso ético de algoritmos, y generar espacios para el debate abierto sobre los límites éticos en la implementación de tecnologías como el reconocimiento facial, la vigilancia masiva o el perfilamiento algorítmico (Wachter, Mittelstadt & Russell, 2021).

En conclusión, educar y sensibilizar a la sociedad sobre privacidad y seguridad en el contexto del Big Data no es una tarea opcional, sino un requisito esencial para la construcción de una ciudadanía digital informada, crítica y capaz de exigir una rendición de cuentas efectiva ante las amenazas emergentes derivadas del uso indiscriminado y opaco de tecnologías avanzadas.

Conclusiones

La revolución del Big Data y la inteligencia artificial ha redefinido radicalmente nuestra sociedad, estableciendo nuevos paradigmas tecnológicos, éticos y legales que desafían constantemente

nuestras concepciones tradicionales sobre privacidad, seguridad y responsabilidad ética. A lo largo de este capítulo, se ha evidenciado cómo los avances tecnológicos no solo representan oportunidades excepcionales para el progreso, sino que simultáneamente plantean riesgos significativos y desafíos profundos que no pueden ser ignorados ni tratados superficialmente.

Los casos analizados de violaciones masivas a la privacidad, brechas de datos explotadas por actores malintencionados, discriminación algorítmica y las paradojas de la anonimización nos han recordado insistentemente que la tecnología no es neutral. Por el contrario, refleja y amplifica las asimetrías, prejuicios y desequilibrios existentes en nuestras sociedades. Frente a esto, las respuestas tradicionales en forma de regulaciones jurídicas han demostrado ser insuficientes y, en ocasiones, incluso ingenuas frente a la escala y velocidad de los cambios tecnológicos.

Este escenario exige un cambio radical en nuestra manera de concebir y abordar estos desafíos. No basta con remendar las regulaciones existentes o aumentar ligeramente las sanciones económicas ante infracciones; necesitamos un enfoque más robusto y holístico que integre principios éticos claros desde el diseño mismo de las tecnologías y que garantice la transparencia, rendición de cuentas y la participación activa de múltiples sectores de la sociedad.

En particular, resulta indispensable reconocer la insuficiencia del modelo tradicional basado exclusivamente en el consentimiento informado individual. Es imperativo avanzar hacia una gobernanza más proactiva y colectiva, que considere los impactos sociales agregados y no únicamente la privacidad individual. Esto implica fortalecer los mecanismos de supervisión algorítmica, promover auditorías técnicas y éticas constantes, y adoptar nuevas metodologías como la privacidad diferencial para proteger de manera efectiva los datos personales.

Asimismo, es crucial comprender que ningún avance normativo o tecnológico será suficiente si no se acompaña de un compromiso

decidido por la educación digital y la sensibilización social. La creación de una ciudadanía digital empoderada y consciente es quizás nuestra herramienta más poderosa para enfrentar estos retos, permitiendo que las personas no solo conozcan sus derechos, sino que tengan las capacidades reales para exigirlos y defenderlos activamente.

Finalmente, es esencial subrayar que el objetivo último no debe ser únicamente la mitigación de riesgos o la protección de derechos individuales aislados. La meta debe apuntar hacia un futuro digital donde la tecnología esté genuinamente al servicio del bienestar humano, la justicia social y la sostenibilidad democrática. Como sociedad, tenemos el deber de asumir este desafío con responsabilidad, creatividad y compromiso ético, asegurando que el futuro del Big Data y la IA no esté marcado por el abuso y la explotación, sino por la equidad, la inclusión y el respeto incondicional por los derechos humanos y la dignidad de cada individuo.

Referencias

Adewale Daniel Sontan & Segun Victor Samuel. (2024). The intersection of Artificial Intelligence and cybersecurity: Challenges and opportunities. *World Journal of Advanced Research and Reviews*, 21(2), 1720-1736. https://doi.org/10.30574/wjarr.2024.21.2.0607

Al-Khassawneh, Y. A. (2022). A Review of Artificial Intelligence in Security and Privacy: Research Advances, Applications, Opportunities, and Challenges. *Indonesian Journal of Science and Technology*, 8(1), 79-96. https://doi.org/10.17509/ijost.v8i1.52709

Carroll, M., Chan, A., Ashton, H., & Krueger, D. (2023). Characterizing Manipulation from AI Systems. *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1-13. https://doi.org/10.1145/3617694.3623226

Chidiogo Uzoamaka Akpuokwe, Adekunle Oyeyemi Adeniyi, Seun Solomon Bakare, & Nkechi Emmanuella Eneh. (2024). Legal Challenges Of Artificial Intelligence And Robotics: A Comprehensive Review. *Computer Science & IT Research Journal*, 5(3), 544-561. https://doi.org/10.51594/csitrj.v5i3.860

Comunale, M. (2024). The Economic Impacts and the Regulation of AI: A Review of the Academic Literature and Policy Actions. *IMF Working Papers*, 2024(065), 1. https://doi.org/10.5089/9798400268588.001

De Mauro, A., Greco, M., & Grimaldi, M. (2015). *What is big data? A consensual definition and a review of key research topics*. 97-104. https://doi.org/10.1063/1.4907823

D.S. Nonju, Dr. K., & Ihua-Maduenyi, A. B. (2024). The Impact of Artificial Intelligence on Privacy Laws. *International Journal of Research and Innovation in Social Science*, VIII(IX), 2150-2174. https://doi.org/10.47772/IJRISS.2024.8090178

Elazar, Y., Bhagia, A., Magnusson, I., Ravichander, A., Schwenk, D., Suhr, A., Walsh, P., Groeneveld, D., Soldaini, L., Singh, S., Hajishirzi, H., Smith, N. A., & Dodge, J. (2023). *What's In My Big Data?* (Versión 2). arXiv. https://doi.org/10.48550/ARXIV.2310.20707

Gilani, S. R. S., Al-Matrooshi, A. M., & Khan, M. H. (2023). Right of Privacy and the Growing Scope of Artificial Intelligence. *Current Trends in Law and Society*. https://doi.org/10.52131/clts.2023.0301.0011

Gilbert, C., & Gilbert, M. A. (2024). The Convergence of Artificial Intelligence and Privacy: Navigating Innovation with Ethical Considerations. *International Journal of Scientific Research and Modern Technology* (IJSRMT), 3(9), 9-17. https://doi.org/10.38124/ijsrmt.v3i9.45

Gulati, T. (2022). Artificial Intelligence And Privacy Violation. The Journal Of Unique Laws And Students, II(II). https://doi.org/10.59126/v2i2a3

Harari, Y. N. (2024). *Nexus: A brief history of information networks from the Stone Age to AI* (First edition). Random House.

Hermansyah, M., Najib, A., Farida, A., Sacipto, R., & Rintyarna, B. S. (2023). Artificial Intelligence and Ethics: Building an Artificial Intelligence System that Ensures Privacy and Social Justice. *International Journal of Science and Society*, 5(1), 154-168. https://doi.org/10.54783/ijsoc.v5i1.644

Hu, Y., & Min, H. (Kelly). (2023). The dark side of artificial intelligence in service: The "watching-eye" effect and privacy concerns. *International Journal of Hospitality Management*, 110, 103437. https://doi.org/10.1016/j.ijhm.2023.103437

Huang, L. (2023). Ethics of Artificial Intelligence in Education: Student Privacy and Data Protection. *Science Insights Education Frontiers*, 16(2), 2577-2587. https://doi.org/10.15354/sief.23.re202

Jurewicz, M., Klein Kafer, N., & Kran, E. (2024). Artificial Intelligence and Privacy: Causes for Concern. *Privacy Studies Journal*, 3, 1-32. https://doi.org/10.7146/psj.v3i.143099

Kim, P., & Bodie, M. T. (2021, septiembre). *ABA Journal of Labor & Employment Law*. 2021, 434(PAPER NO. 21-09-02). https://openscholarship.wustl.edu/law_scholarship/434/

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60-66, 68, 128.

Mühlhoff, R. (2023). Predictive privacy: Collective data protection in the context of artificial intelligence and big data. *Big Data & Society*, 10(1), 20539517231166886. https://doi.org/10.1177/20539517231166886

Oseni, A., Moustafa, N., Janicke, H., Liu, P., Tari, Z., & Vasilakos, A. (2021). *Security and Privacy for Artificial Intelligence: Opportunities and Challenges* (No. arXiv:2102.04661). arXiv. https://doi.org/10.48550/arXiv.2102.04661

Pauline, K., & Matthew T, B. (s. f.). *Artificial Intelligence and the Challenges of Workplace Discrimination and Privacy.*

Piplai, A., Kotal, A., Mohseni, S., Gaur, M., Mittal, S., & Joshi, A. (2023). Knowledge-Enhanced Neurosymbolic Artificial Intelligence for Cybersecurity and Privacy. *IEEE Internet Computing*, 27(5), 43-48. https://doi.org/10.1109/MIC.2023.3299435

Rocher, L., Hendrickx, J. M., & De Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(1), 3069. https://doi.org/10.1038/s41467-019-10933-3

Solove, D. J. (2024). Artificial Intelligence and Privacy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4713111

[Capítulo 5]

Ética e Inteligencia Artificial: acciones empresariales sobre el derecho a la intimidad

Diego Hernán Varón Rojas

«Una computadora merece que se le llame inteligente si puede engañar a un humano haciéndole creer que es humano» (Alan Turing).

«O esclavizamos a la inteligencia artificial o nos esclavizará a nosotros» (Sam Altman).

La inteligencia artificial (IA) afecta el orden social existente. Se resaltan los efectos que causa su desarrollo y aplicación, entre ellas la discriminación, el desempleo, los riesgos de seguridad y la filtración de la privacidad (Huang et al., 2022). Puede ser vital para la conducción de hostilidades, la manipulación de la información, el desarrollo de nuevas armas y métodos de combate (Vigevano, 2021). En los últimos años han surgido numerosas directrices éticas desde la academia[1], la industria, el gobierno y la sociedad civil. Algunos

1 El presente artículo hace parte de los resultados del proyecto de investigación realizado sobre antropología de las tecnologías aplicadas al ámbito organizacional, desarrollado por investigadores de la Universidad Nacional Autónoma de Honduras (campus Santa Rosa de Copán), la Universidad Anáhuac (Cancún - México) y la Uni-

hacen énfasis en las exigencias que los consumidores demandan del estado. Estos desean saber las tecnologías que son implementadas por las empresas, y dar respuestas a los numerosos desafíos éticos asociados con su desarrollo (Du & Xie, 2021). Visiones más integrales, se preguntan de dónde vienen las regulaciones de la IA. Al tener mayor participación de partes interesadas, las guías éticas serían más completas y con mayor aplicabilidad a la realidad (Bélisle-Pipon et al., 2023). Por el contrario, otros autores consideran la necesidad de ser claros en la distinción de los aspectos éticos y jurídicos. Se identifica el Derecho Internacional como el principal marco jurídico para su regulación (Carrillo, 2020).

Muchas investigaciones develan los beneficios de la IA para la humanidad porque simplifica procesos. Aplicaciones vinculadas con el aprendizaje automático están obteniendo un mayor conocimiento de *big data* que conducen a una mayor eficiencia y eficacia organizacional (Noble & Dubljević, 2022). Algunas investigaciones muestran la importancia de asumir un análisis ético riguroso y las salvaguardias resultantes para abordar los problemas emergentes de los sistemas de inteligencia artificial generativa (GenAI) (Kirova et al., 2023). Pero también genera efectos adversos. La ausencia de una conducta empresarial ética, altos niveles de corrupción y mala gobernanza oficial, son las principales barreras que enfrenta la implementación de la IA (Diab, 2021). La aplicación en los negocios, plantea varias cuestiones sociales que incluyen los algoritmos sesgados y la falta de transparencia, por esto surge la necesidad de monitorear la rendición de cuentas, los efectos sociales y económicos del empleo, las acciones sobre la igualdad y la privacidad de los datos (Tripathi & Kumar, 2025).

---

versidad Icesi (Cali - Colombia). Otras actividades han incluido la organización de simposios y ponencias en congresos como International Union of Anthropological and Ethnological Sciences - 2023, World Anthropological Union Congress - 2024, el VII Congreso Latinoamericano de Antropología y XI Congreso de Enseñanza de la Ética, así mismo como la producción de artículos aplicados al contexto de turismo y patrimonio (Proyecto de Investigación 621-DICIHT UNAH-CUROC).

Parte de las demandas éticas y legales se encuentran en manos de quienes dirigen a las empresas, sean estas privadas, académicas o gubernamentales. Se debe prestar atención a las estrategias desarrolladas por sus directivos y trabajadores, para proteger los datos personales de todos los grupos de interés con quienes se relacionan (Varma et al., 2023). Se explora el impacto de la integración de la IA y la orientación cultural en la autonomía percibida por los empleados y la autoeficacia creativa (Xavier & Korunka, 2025). Otros trabajos consideran los obstáculos para adoptar e institucionalizar la analítica y la inteligencia artificial en los recursos humanos (Yadav, 2025). Poco a poco, los gerentes han ido adquiriendo conciencia sobre las consideraciones éticas, y reciben recomendaciones para su gestión, vinculadas con los análisis de dimensiones macro y microambientales que permitan tomar decisiones gerenciales consecuentes (Brendel et al., 2021). Para integrar con éxito las nuevas tecnologías y avanzar en las estrategias organizacionales, se debe desarrollar un liderazgo consciente y ético, basado en una comunicación eficaz y una gestión del cambio para mantener la cultura orientada a la transparencia (Parchment, 2025). Se buscan actuaciones responsables de los líderes, que incluyan el cumplimiento con las regulaciones voluntarias y obligatorias, proponiendo de este modo, la institucionalización del comportamiento ético (Schultz & Seele, 2023).

Pocas investigaciones incluyen empresas de software y el análisis de la gestión sobre la ética de la IA. En la revisión realizada muchos estudios se orientan al análisis estadístico (Vakkuri et al., 2020). Sin embargo, se ubicó un trabajo sobre la forma en que las empresas abordan la ética de la IA, presentando datos empíricos recopilados por medio de estudios de casos, que brindan relatos relevantes sobre el análisis de situaciones cruzadas (Stahl et al., 2022). Hay novedad en los trabajos que buscan comprender la relación de la inteligencia emocional y la artificial, así como la influencia en el comportamiento organizacional desde la perspectiva fenomenológica, que incluye posturas filosóficas de Descartes, Husserl y Merleau-Ponty (Canbul, 2025). En la revisión bibliográfica, no hay taxonomías que recurran al

concepto filosófico de responsabilidad, en particular sobre la protección a la vida privada de las personas, desde un enfoque fenomenológico como el planteado por Hans Jonas, aplicado en empresas lideres dedicadas a la producción de IA (Martínez, 2023; Varón, 2025).

Considerando lo anterior, surge la pregunta, ¿qué acciones desarrollan empresas dedicadas a la producción de IA para proteger el derecho a la intimidad? Con el fin de dar respuesta a la pregunta, se revisan los códigos de ética y las acciones desarrolladas por organizaciones de base tecnológica, como IBM, Intel y Tencent[2]. Para este ejercicio, se sigue la clasificación de empresa basada en el punto de vista fenomenológico, planteada por Benítez y Varón (2014). La realización de entrevistas y la revisión de fuentes documentales, permitió identificar las acciones concretas y respuestas ante reclamaciones realizadas por los ciudadanos.

Marco teórico

De acuerdo con Gabriela Ramos (2023), directora general adjunta de la UNESCO para las ciencias sociales y humanas, de los diez principios básicos que conforman el enfoque de la ética aplicado a la IA, se encuentra la protección de derechos humanos y la dignidad. Esta es la piedra angular de las recomendaciones para el avance de principios fundamentales como la transparencia y la equidad. Se busca proteger los datos de las personas y promocionar la privacidad a lo largo de todo el ciclo de vida de la IA. González y Martínez (2020) plantean dilemas éticos en el escenario de la IA. Es necesario preguntarse qué usos maliciosos de la IA se transforman en dilemas, si se han desarrollado códigos de ética o políticas que logren resolver ese dilema. Muchos coinciden con que los gobiernos deben

2 Las empresas seleccionadas hacen parte de la lista que lidera la producción y el uso de la IA y machine learning hacia el año de 2024. Varias de ellas están incluidas en las cinco empresas de IA más punteras del momento. Y finalmente todas se encuentran entre las 45 empresas líderes en IA establecidas por la Fundación Innovación Bankinter durante el 2020.

establecer marcos normativos, adecuados para la protección de datos. Yoshua Bengio y Yann LeCun, se suman a la necesidad de establecer regulaciones estrictas y controles para evitar usos indebidos de la tecnología. "Desde la manipulación de información hasta el desarrollo de sistemas autónomos peligrosos, los expertos coinciden en que la humanidad debe prepararse para los posibles impactos negativos de esta tecnología" (El cronista, 2025).

Se deben controlar prácticas inadecuadas, inmorales, antiéticas e ilegales, que imposibiliten otros escándalos como el recordado "Cambridge Analytica" y otros semejantes, que han sido recreados en diversos documentales, películas y casos de estudio (BBC, 2018; Schnebleet al., 2018; Kanakia et al., 2019). En diciembre de 2014, el científico Stephen Hawking predijo que "La inteligencia artificial augura el fin de la raza humana" (BBC, 2014). Algunas interpretaciones más esperanzadoras sobre las tecnologías del futuro, plantean que no se trataría de la extinción de nuestra especie homo sapiens - sapiens, sino la entrada de una nueva forma esclavitud centrada en la eliminación de la singularidad e individualidad. Un antecedente sobre el derecho a la intimidad y la privacidad se generó hace más de 130 años. Los jueces estadounidenses Samuel D. Warren y Louis Brandeis, en 1890 señalaron que la privacidad es "el derecho de no ser molestado" y se enfocaron en proteger a las personas individuales, como respuesta a los recientes desarrollos tecnológicos de ese entonces, como la fotografía y los "medios amarillistas" (Saldaña, 2012).

El temor de los seres humanos por perder el control de sus datos mediante el uso de la tecnología, inspiró la promulgación anterior. En la actualidad, los estados buscan proteger información presente en sistemas de comunicación privada, entre los que se encuentra la telefonía fija y móvil, el correo electrónico que aluden a datos personales, sensibles que vinculan relaciones sociales, el estado de salud, el correo personal, las comunicaciones privadas mediante el uso de la I.A. Pero también se requiere el trabajo de las empresas. Trabajos como el de Ponce et al., (2023) analizan la relación entre

IA y el derecho a la intimidad/privacidad, concluyendo que el abuso "puede generar un entorno en el cual los Estados, las organizaciones empresariales e incluso otras personas, rastreen, predigan o manipulen el comportamiento de los ciudadanos de todo el mundo en un grado sin precedentes" (p. 91).

Complementario a la normativa que deben implementar los estados, se debe trabajar por estrategias éticas que respeten los líderes y directivos de las empresas orientadas a desarrollar y aplicar la I.A.[3] Un "inocente test" de personalidad en la red social Facebook, derivó en acusaciones de robo de datos, interferencia política y chantajes. 265.000 usuarios completaron este test, que requería permiso para acceder a información personal y de la red de amigos, sin el consentimiento, calculado en 87 millones de perfiles. Infringir leyes orientadas a la protección del derecho a la intimidad, permitió que organizaciones empresariales como Facebook, fuese sancionada con una multa histórica de 1.200 millones de euros, además de la prohibición temporal de transferir los datos de los usuarios europeos fuera del territorio de la Unión Europea (Terán, 2023).

Los efectos no se restringieron en pagar la multa. También ocasionó impacto en la reputación e imagen corporativa. Esto significó que, en un día, la organización perdió US$37.000 millones por este escándalo (BBC, 2018). Habría que considerar las réplicas que tuvo en diferentes campañas políticas del mundo, incluyendo a los países de América Latina. La sanción social también tuvo un costo alto. Películas como "100 cosas" recrean situaciones semejantes a la realidad (Esteves, 2022). Dos amigos empresarios que desarrollan una asistente virtual, llamada Nana. Esta IA se encargan de personalizar la voz que elijan los usuarios. Creada por Paul para que "las personas no se sientan solas", y que Toni

3 El autor agradece a los estudiantes de los cursos de Organizaciones e Instituciones y Ética y organizaciones de los años 2023 y 2024 orientados en la Universidad Icesi. Por las discusiones centradas en organizaciones dedicadas a la producción de metaverso e Inteligencia Artificial, el análisis de la película 100 cosas, y casos relacionados con los usos fraudulentos de la I.A.

adapta como estrategia para conocer al consumidor, ampliando así su uso al mercadeo. David Zuckerman se entera de Nana por medio de su concurso. Este buscaba promover entre los participantes, el "ideal" de hallar un inversionista que apoyara sus innovaciones. Después de una negociación fallida con Toni y Paul, y el deseo fallido para cooptar al ingeniero, David Zuckerman "lanza" a Nana, robando la aplicación a los inexpertos "empresarios". La asistente virtual les hace un último favor: generar conciencia a los usuarios, sobre la forma antiética e ilegal usada para "captar" sus datos. Se esperaría que los clientes sancionen a la multinacional por las prácticas indebidas.

¿Qué perspectiva de la ética permite estudiar la tecnología actual? La pregunta desarrollada en el programa "El Puente al Aire" presenta la necesidad de incluir en la reflexión una perspectiva fenomenológica. Aunque los lineamientos se orientaron a las organizaciones empresariales dedicadas a la producción de metaverso, esta incógnita también permite extenderse a aquellas que producen IA. En ese orden de ideas, la propuesta del filósofo alemán Hans Jonas (1903 – 1993), discípulo de Martín Heidegger, profundiza en el "principio de responsabilidad", de forma que este ensayo da luces para una "ética futura", buscando incluir aspectos orientados a la "civilización tecnológica".

La supervivencia humana depende de los esfuerzos conjuntos para cuidar nuestro planeta, que permita la existencia de generaciones no nacidas. La relación del ser humano con la naturaleza ha generado un gran cambio, consolidado desde la segunda mitad del siglo XX mediante los desarrollos tecnocientíficos que impactan en la sociedad de masas. La expansión de las ciudades trae consigo problemas ecológicos. A esto se le suma que la humanidad se vea amenazada directamente, con la aparición de armas nucleares, químicas y biológicas, así como la manipulación artificial de la vida basada en "la heurística del temor" (Martínez, 2023). Inspirado en el imperativo ético de Immanuel Kant, plantea una actualización del principio de conducta: "Obra de tal modo que los efectos de tu acción sean compatibles con la permanencia de una vida humana auténtica en la tierra" (Jonas, 1995, p. 39).

En contra posición David Noble se aleja de las posturas "tecnófobas" que "satanizan" el uso de nuevas tecnologías. Muchos prevén que la IA podría ser causante del futuro colapso del sistema, al requerir nuevas fuentes energéticas y nueves lógicas del poder y el control. Noble (1999) privilegia una postura cercana a la "tecnofilia" y sugiere que la tecnología es similar a la religión. Se evocan imágenes de omnipotencia, devoción y sobrecogimiento, hasta el punto de convertirse en una "nueva religión" con sus propios líderes, rituales y artículos de fe. Diversas miradas entre las que sobresalen empresarios y trabajadores del sector, presentan los desarrollos de la IA como una "auténtica necesidad". No obstante, muchos de los creadores, prevén cambios sustanciales en las profesiones que hoy conocemos, y la necesidad de una adaptación inmediata antes de que lleguen los aumentos en las tasas de desempleo, para quienes dudan de su poder desestructurador (Varón, 2023).

Más que nunca, los estados, las empresas y la sociedad civil se preguntan por la responsabilidad de innovaciones tecnológicas que tienen distintas consecuencias sobre la sociedad. Aunque la literatura plantea ideas que la responsabilidad es una cuestión compartida entre diversos actores que integran el sistema, la mayoría de las reflexiones éticas sobre el derecho a la intimidad de la IA se presenta como responsabilidad del Estado, sin embargo, se requiere auscultar este fenómeno desde una orientación empresarial, bastante ausente en la revisión de la literatura.

Siguiendo la taxonomía centrada en las metáforas sobre organizaciones planteada a por Gareth Morgan (1991) y en el principio de la responsabilidad de Hans Jonas (1995), se retoma la categorización centrada en las empresas desde una perspectiva conceptual, propuesta por Benítez y Varón (2014). Se plantean tres metáforas: la empresa como negocio que funciona desde una orientación economicista, determina por el mercado y basada en la propiedad privada de los accionistas o inversionistas. La segunda donde la empresa puede ser comprendida como una organización, es decir, como un espacio social

en la que se configuran tensiones y conflictos entre distintos modelos e intereses. Y finalmente la empresa entendida como institución donde se legitima el poder político, se estimulan procesos de socialización y de construcción identitaria. Considerando las entrevistas y la revisión documental de tres empresas seleccionadas por ser lideres en la producción y uso de IA, se presentan los resultados obtenidos[4] sobre las acciones enfocadas a proteger el derecho a la intimidad.

Resultados

Se realizaron seis entrevistas a lo largo del proceso de investigación. A un ingeniero de sistemas especializado en IA, a una doctora en matemáticas aplicadas, especialista en generación de tecnología en IA, y a estudiantes del curso de ética y organizaciones. Andrés Chaux[5] señala su impacto transformador en la sociedad actual, particularmente sobre el potencial de un proyecto de contenido personalizado. Para examinar cómo la inteligencia artificial de Amazon Web Services impacta la transformación digital empresarial, se subraya la importancia de abordar desafíos éticos y de privacidad en el futuro de la IA, así como la necesidad de colaboración interdisciplinaria y educación para un progreso sostenible. Esta puede impulsar la innovación y la eficiencia en diversas industrias.

Andrés Chaux resalta el potencial de la IA para "mejorar la experiencia del cliente y la retención de usuarios". Si bien Amazon Web Services "ofrece oportunidades emocionantes para mejorar la experiencia del cliente y la eficiencia operativa en diversos sectores", es

4 El orden en el que aparecen las empresas corresponde al año de creación. En ningún momento su ubicación corresponde a criterios sociales, económicos, financieros, técnicos, de mercadeo u otra categorización.

5 Los nombres de los entrevistados han cambiado para respetar el anonimato expresado en el formato de consentimiento informado. Se agradece a los estudiantes Ana Sofia Arango Bonilla, Natalia Candamil García y Estefany Quintero Serrato por la transcripción de la entrevista realizada en mayo de 2024.

fundamental reconocer "los dilemas éticos asociados, como la privacidad de datos y el sesgo algorítmico, para garantizar un impacto positivo y sostenible en la sociedad" (mayo 15 de 2024). Servicios como AWS SageMaker y AWS Rekognition son aplicados en distintos sectores industriales para mejorar la eficiencia, la personalización de servicios y la toma de decisiones basada en datos. Desde la experiencia de Chaux, puede ayudar a identificar los beneficios observados, como la optimización de procesos, la mejora en la experiencia del cliente y el impulso a la innovación en diversos ámbitos empresariales.

Jeniffer Beltrán[6] expresa que "el éxito de la inteligencia artificial se mide no solo por lo que puede lograr, sino también por cómo lo logra". Para ella, su impacto trasciende lo técnico y debe ser evaluado desde una perspectiva social y académica. "La inteligencia artificial tiene el potencial de transformar la educación, la salud y la economía, pero si no lo hace respetando principios como la transparencia y el respeto, no podremos hablar de un verdadero avance". En su práctica docente enfrenta constantemente problemas relacionados con el uso irresponsable de la IA por parte de sus estudiantes. "Me preocupa que muchos recurren a estas herramientas para calificar o incluso elaborar trabajos académicos, a menudo sin entender las implicaciones éticas que esto conlleva". Es alarmante que se "estén dejando datos privados en estas plataformas sin ser conscientes de los riesgos asociados (…) y el manejo de los mismos" (noviembre 9 de 2024).

De otro lado, Carol Popayán, Yinna Franco, Paula Reina y Sebastián Rodríguez[7] expresan que los dilemas éticos que rodean el uso de la IA reflejan una tensión continua entre el avance tecnológico y la protección de los derechos humanos. Como destacan González y Martínez (2020), el uso malintencionado de la IA para la invasión de la privacidad, la manipulación de la información y la interferencia

6 Se agradece a los estudiantes Aura Camila Torres Torrez, Josué Vargas González y Felipe Velasco Sánchez por la entrevista realizada en noviembre de 2024.

7 Entrevista grupal realizada en noviembre de 2024 sobre el trabajo realizado en una de las empresas consideradas en el estudio.

en procesos democráticos, "demuestra que la IA tiene el potencial de beneficiar a la sociedad, su aplicación sin un marco ético y regulatorio sólido puede tener consecuencias graves" (noviembre 19 de 2024).

Contrario a la creencia generalizada, Camila Rojas, Pablo Salcedo y Juan Perdomo, consideran que las inteligencias artificiales, en el punto actual de desarrollo, "carecen de cualquier intencionalidad y por ende no pueden ser consideradas sujetos morales". Las implicaciones y limitaciones morales de su código dependen sustancialmente de la programación de sus creadores. Gigantes de la industria son actores claves en la preservación y limitación de estas herramientas. Se pueden criticar por desarrollar actividades que van en contra de la moral y la ética, como suministrar información al complejo militar de los Estados Unidos. "A pesar de la existencia de normativas nacionales o internacionales, el uso indebido de estas herramientas recae en última instancia en los usuarios" (noviembre 14 de 2024).

Los entrevistados coinciden que se deben realizar un esfuerzo para educar y reforzar el peligro que pueden llegar a representar dichas herramientas. El desarrollo exponencial de la tecnología, está muy cerca de crear "super inteligencias artificiales", lo cual supondrá dilemas éticos mucho más complejos. Existirían dos especies sapiens, siendo la inteligencia artificial superior en racionalidad a la humanidad, "lo cual supone un posible dilema de convivencia y transhumanismo" (noviembre 14 de 2024). En las entrevistas se resalta la importancia de la ética en el uso de las IA`s. Sin embargo, se hace necesario interrogar las acciones concretas realizadas por las tres empresas seleccionadas que dedican esfuerzos para su desarrollo, aspecto que se plantea a continuación.

International Business Machines (IBM)

Entre 1880 y 1924, se dio la fusión de empresas innovadoras lideradas por Herman Hollerith y Charles Flint. Desarrollaban tecnologías fundamentales como las tarjetas perforadas, esenciales para la

tabulación de datos, y se estableció la base de lo que será IBM. En 1924, CTR adoptó el nombre de International Business Machines (IBM). Entre los años 1930 y 1960, se consolidó como líder tecnológico durante la Gran Depresión y la Segunda Guerra Mundial, destacándose por su contribución en proyectos nacionales, como la Ley de Seguridad Social. En la década de 1960 revolucionó la industria con el System/360, una familia de computadoras compatibles que establece un nuevo estándar que marcó el inicio de la computación empresarial moderna.

En la etapa denominada "computación personal", se introdujeron innovaciones claves como el disquete en 1971 y el IBM PC en 1981. Esta corporación enfrentó desafíos debido a la competencia y la apertura del mercado a terceros, lo que empujó a llevar a cabo una reinvención. Tras enfrentar una crisis financiera en 1993, la organización se reinventó bajo el liderazgo de Louis Gerstner, enfocándose en servicios y software. La victoria de Deep Blue sobre el jugador de ajedrez Garry Kasparov en 1997 simbolizó el retorno de IBM como líder en tecnología avanzada, con sus innovaciones en middleware y servicios integrados (Stierhoff & Davis, 1998).

IBM afirma que la ética de su IA, se apoya en los principios de confianza y transparencia. Asegura que el propósito es aumentar la inteligencia humana, los datos y que las entradas pertenecen a sus creadores, por lo que la tecnología debe ser transparente y explicable. Expresa que los "pilares de confianza" son fundacionales de su ética sobre la IA, que incluyen la explicabilidad, la legitimidad, la robustez, la transparencia y la privacidad.

En octubre del 2023 el expresidente Joe Biden decretó que los proveedores de inteligencia artificial están obligados a someter a pruebas de seguridad llevadas a cabo por el estado, si se estima que representa un riesgo para la seguridad nacional o la salud pública de Estados Unidos (Roose, 2023). Sin embargo ¿qué sucede con IBM?, ¿qué acciones ha desarrollado esta empresa para proteger el derecho a la intimidad? En su reporte de impacto del 2023 afirma que la empresa es dirigida interna y externamente bajo las siguientes políticas:

- Conductas empresariales y éticas.
- Reciprocidad.
- Diversidad en la fuerza de trabajo.
- Salud y seguridad.
- Privacidad de los datos.
- Relaciones de negocios diversas.
- Asuntos medioambientales.
- Calidad.
- Principios de los derechos humanos.
- Estándares de empleo global.
- Principios cognitivos.
- Declaraciones y casos individuales.

Justina Nixon-Saintil, vicepresidenta y directora de impacto de IBM afirma "At IBM, we believe technology is a catalyst to unlock new and more powerful ways to make a positive impact. Innovation can address evolving demands and create a better future as new ethical, societal and environmental challenges arise". También expresa "We continue to focus on making a lasting impact by using the power of technology, such as Artificial Intelligence"[8] (IBM, 2023).

Un ejemplo de su aplicación corresponde al caso de Watson Health, donde IBM emplea datos médicos para desarrollar herramientas que promuevan avances en el ámbito de la salud. La empresa asegura que toda la información utilizada, esté completamente anonimizada, protegiendo la privacidad de los pacientes mientras contribuye al bienestar colectivo. Además, IBM ha tomado decisiones éticas significativas, como no desarrollar ni comercializar tecnologías de reconocimiento facial que puedan usarse para vigilancia masiva o violaciones de derechos humanos (IBM, 2025).

8 Una posible traducción sería "En IBM, creemos que la tecnología es un catalizador para descubrir formas nuevas y más poderosas de generar un impacto positivo. La innovación puede abordar demandas cambiantes y crear un futuro mejor a medida que surgen nuevos desafíos éticos, sociales y ambientales" (…) "Seguimos enfocándonos en generar un impacto duradero mediante el uso del poder de la tecnología, como la Inteligencia Artificial".

Integrated Electronics – Intel

Intel Corporation es una empresa de tecnología fundada en 1968 por Gordon Moore y Robert Noyce. Los dos eran ingenieros con experiencia en la industria de los semiconductores, que trabajaron en Fairchild Semiconductor, una de las empresas pioneras en el desarrollo de circuitos integrados. Es responsable de ser la mayor fabricante del mundo. Creadora de la serie de procesadores X86, los más usados en la mayoría de las computadoras personales. Su nombre "Intel" proviene de "Integrated Electronics", reflejando el propósito de la compañía, que es crear tecnología que cambia el mundo y mejora la vida de todas las personas del planeta (Intel Corporation, 2023). Opera a nivel global en más de 50 países a través de oficinas, centros de investigación y plantas de fabricación. Su sede principal se encuentra en Santa Clara-California (Estados Unidos).

En la página oficial se promulga: "Nuestros valores guían la forma en que tomamos decisiones, nos tratamos entre nosotros y servimos a nuestros clientes. Más que simples palabras, nuestros valores son el hilo conductor que nos une" (Intel Corporation, 2019). Esto refleja el compromiso que se han impuesto para cumplir. Como es una empresa influyente en el desarrollo de hardware y software, incluidos temas de IA, Intel tiene responsabilidad de establecer estándares éticos y de seguridad.

Ha mencionado su compromiso con la defensa de los derechos humanos. Se establece que: "el respeto se encuentra arraigado en nuestros valores y se aplica en cualquier lugar en que realicemos actividades comerciales" (Intel Corporation, 2009). Evidencia de lo anterior, se podría mencionar que Intel cumple con el principio de garantizar la protección de los derechos humanos, al igual que la proporcionalidad e inocuidad, ya que invierten fuertemente en la investigación y testeo para que sus productos sean seguros de utilizar.

Esta compañía se ha comprometido con la seguridad de sus usuarios, protegiendo la intimidad por medio del almacenamiento y tratamiento seguro de datos. En la actualidad, cualquier empresa es

susceptible de un posible ciberataque, tal como fue el caso de Intel. En el año 2020, alrededor de 20GB de archivos confidenciales de la compañía fueron robados por crackers. Aparentemente fue muy poco, pero pudo contener códigos de acceso importantes, lo cual supone un riesgo a la privacidad y seguridad de sus usuarios (Gómez, 2020).

El uso de su tecnología en el complejo militar cuestiona su integridad. Al apoyar directa o indirectamente al cualquier grupo militar, supone la imposibilidad de vivir en sociedades pacíficas. Esto es evidente por lo menos en la relación de Intel con el ejército americano, que aprobó un préstamo de 3.5 billones de dólares para el desarrollo de chips (Meenaktshi & Adler, 2024). Intel no está involucrada directamente con la industria de la guerra, asunto que se cuestiona al proveer recursos necesarios como microchips y software, para que esta industria siga a flote. Esto hace que la empresa pierda validez y autoridad ética frente a los consumidores.

Intel ha estado en las noticias por sus esfuerzos para mejorar la transparencia y responsabilidad en el uso de la IA. La empresa ha colaborado con otras compañías tecnológicas para establecer estándares éticos y ha desarrollado tecnologías avanzadas para protegerse de ciberataques (Tech News, 2021). Además, Intel ha sido proactiva en abordar los sesgos en la IA mediante la promoción de la inclusión y diversidad en sus equipos de desarrollo, y en los datos utilizados para entrenar algoritmos. Estas iniciativas son parte de la estrategia de Intel para asegurar que sus tecnologías de IA se utilicen de manera ética y beneficiosa para la sociedad (Tech News, 2021).

Tencent Holdings

Fue creada por "Hua Teng Ma", también conocido como "Pony Ma", quien es su director ejecutivo. Se graduó de ciencia computacional en 1993, y su primer trabajo fue escribir el código para visualizadores de páginas. Cinco años después cofundó Tencent

junto con un colega en Shenzhen y su primer producto fue un programa de mensajería instantánea conocido como "QQ". Para el 2004, Tencent controlaba más de dos tercios del mercado de mensajería instantánea en China (Dai et al., 2011). En el 2010, dos equipos de ingenieros compitieron para crear un nuevo producto. De allí salió WeChat, que para el año 2022 tenía más de 1.2 billones de usuarios. Al analizar su enfoque de negocio, se orienta hacia la mensajería instantánea y videojuegos, buscando beneficios en el área de las relaciones sociales virtuales (Tencent Holdings, 2024).

Una de las primeras implementaciones de IA por parte de Tencent fue con Miying en el 2017. Es un sistema de IA`s con los que se pueden diagnosticar enfermedades utilizando imágenes diagnósticas. En este punto, tenía una precisión del 90% y podía cubrir cáncer de esófago, pulmón, cérvix, mama y retinopatía diabética (Global Growth Markets, 2017). Fue reconocido por el público en el 2019, y dentro de la pandemia de COVID, permitió a los doctores identificar esta enfermedad en 2 segundos, siendo galardonado por el Foro Económico Mundial (Li & Zafar, 2020).

Hunyuan es el nuevo generador de videos con IA de Código Abierto. Es el rival directo de Chat GPT propuesto por Tencent a inicios del 2024 y disponible en la nube de Tencent Cloud. Permite la creación de imágenes, redacción, reconocimiento de textos y servicio al cliente. Su crecimiento se debe en parte, a que el gobierno chino prohibió Chat GPT por contenidos ilegales frente a las políticas del país. Sin embargo, la información frente a este modelo es difusa. Especialmente en el rendimiento: no se conoce si es mejor que Chat GPT o LLama, ya que esta información no se hace pública, o la información que se difunde es cuestionada por su positivismo. La empresa ha sido señalada y demandada por acceso no autorizado a las fotos de los usuarios, incluso cuando la aplicación no estaba en uso. Este comportamiento generó preocupación sobre la invasión de la privacidad. Tencent respondió prometiendo corregir este problema en versiones futuras (McDonell, 2019). La empresa afirma haber modificado el tra-

tamiento de la información, el periodo de almacenamiento de cuentas eliminadas, los datos requeridos para registrar y hacer parte de la comunidad, entre otros cambios, a partir de su última actualización en la política de privacidad generada en abril de 2024.

La empresa ha enfrentado varios escándalos relacionados con la privacidad y la censura de usuarios a nivel global, lo cual permite encontrar una tensión evidente en materia de legislación y derechos. Tencent manifiesta abiertamente que la información de los usuarios está almacenada en su servidor local en la República China, información que está disponible para el Gobierno de ese país debido a sus políticas. Adicionalmente, se encuentra otro punto crítico sobre IA y los Derechos Humanos. A pesar que la empresa manifiesta ajustarse a los marcos legales de cada país en el que hace presencia, no hay claridad sobre sus políticas de seguridad. En 2025 Tencent fue incluida por Estados Unidos en la Lista CMC (Chinese Military Companies) vinculadas con el ejército chino. Esto provocó que las acciones de la empresa cayeran estrepitosamente en la bolsa de valores de ese país. La empresa tomaría acciones legales si su nombre no era removido de esta lista (El Espectador, 2025).

Discusión

El derecho a la intimidad hace parte de los ideales humanos que deberían ser universalizados, ampliando su cobertura incluso en regiones con libertades restringidas[9]. Se espera que las empresas dedicadas a la producción de IA desarrollen acciones orientadas

9 Se agradece al doctor Fernando Cruz-Kronfly por exponer detalladamente esta idea. En entrevista realizada en marzo de 2025, se argumentó que los "derechos humanos" son un conjunto de derechos básicos aceptados, que el Estado y la denominada "sociedad civil" reconoce a los seres humanos. Es un constructo cultural que hicieron países básicamente francés e inglés, al mundo. Otras culturas no han elaborado es constructo filosófico humanista, jurídico y político, debido a las "resistencias culturales", respecto de la idea de una "humanidad universal", que no hace parte de sus instituciones ni de sus presupuestos.

a materializar su compromiso en el cumplimiento de los diez principios básicos que incluyen el enfoque de la ética aplicada. Aunque muchos padres y padrinos de la IA claman medidas urgentes para que los Estados reglamenten la producción y uso, y la sociedad civil demanda estas medidas, las empresas al estar interconectadas en "la aldea global", tienen también co-responsabilidades.

La institucionalización de la intimidad en materia de IA, es un proceso que requiere de aprendizaje, por tanto, directivos, trabajadores y demás grupos de interés, aportan a su defensa (Brendel et al., 2021; Varma et al., 2023; Xavier & Korunka, 2025; Yadav, 2025; Parchment, 2025). Si bien la teoría institucional propende para que las empresas se legitimen, convirtiéndose en referentes para otras, se requiere una buena dosis de conciencia y responsabilidad de las partes interesadas (Schultz & Seele, 2023). Son pocos los estudios que vinculan la ética en la IA desde una orientación empresarial, basado en estudios de caso como lo proponen Stahl et al. (2022). La perspectiva fenomenológica de Jonas (1995) que plantea una ética futura aplicada a la civilización tecnológica, junto con el modelo de Benítez y Varón (2014) permiten auscultar el comportamiento de las tres empresas estudiadas, que se presenta en la tabla 1.

Tabla 1. Empresas desarrolladoras de IA, comprometidas con el derecho a la intimidad

Empresa	Negocio	Organización	Institución
IBM	Su tecnología usada por los militares, ponen en duda su integridad.	Enfrentó una crisis financiera en 1993 lo que le exigió diversificar sus productos.	Está alineada con normativas internacionales y la Ley de Inteligencia Artificial de la Unión Europea.
Intel	El uso de su tecnología en el complejo militar cuestiona la integridad.	Colaboración con otras compañías tecnológicas para establecer estándares éticos en la IA.	Implementa políticas para proteger la privacidad y seguridad. Cumple con regulaciones internacionales.
Tencent	WeChat ha violado el derecho a la privacidad. Se han cuestionado los vínculos comerciales con el gobierno chino.	La información personal está disponible para el Gobierno Chino debido a las políticas nacionales.	Se ha comprometido con aplicar acciones para cumplir con las exigencias globales.

Fuente: datos propios organizados a partir de la categorización planteada por Benítez y Varón (2014).

Pese a que en el pasado la empresa fue acusada de desarrollar tecnología que los nazis usaron en el holocausto, el compromiso ético de IBM se materializa en sus principios de confianza y transparencia. Es un modelo a seguir para garantizar que el desarrollo de IA respete los derechos fundamentales. Entre las acciones destacadas se encuentran la implementación de políticas de privacidad desde el diseño, la anonimización y minimización de datos, la auditabilidad de algoritmos, la alineación con normativas internacionales y la Ley de Inteligencia Artificial de la Unión Europea. Además, IBM demuestra que la gober-

nanza ética y el diseño responsable son esenciales para prevenir el uso indebido de la IA, como lo ilustra su decisión de no desarrollar tecnologías de reconocimiento facial para vigilancia masiva. Esto lo ubica en un lugar de institución frente a los compromisos de velar por el derecho a la intimidad.

En el caso de Intel, pese a que en el pasado la empresa fue acusada de atentar contra el derecho a la intimidad al incluir en los nuevos procesadores Pentium III, también por las respuestas "inaceptadas" a partir del ciberataque que sufrió en el 2020 y los negocios realizados con las fuerzas militares de los Estados Unidos, realiza esfuerzos para mejorar la transparencia y responsabilidad en el uso de la IA. Actúa como organización al extender los lazos de solidaridad con otras compañías de tecnología, con el fin de establecer estándares éticos que tiene consagrado en su ADN. Y se esfuerza por cumplir con las regulaciones internacionales para proteger la individualidad humana.

Tencent, junto a sus sitios web y empresas subsidiarias, han sido poco transparentes respecto a cómo exactamente usan los datos recolectados, incluso hay quejas de los usuarios por fuera de China. Indudablemente estos datos se han usado para entrenar su modelo de IA Hunyuan, o para alimentar su publicidad orientada a nuevos desarrollos de juegos o servicios. Respecto a la ciberseguridad y protección de datos, Tencent no toma medidas de precaución para la seguridad del individuo. Al ubicarse en China, tiene el compromiso de dejar disponible la información personal al gobierno. Debido al tratamiento de datos de locación o la QQ respecto a pagos ilegales. Esto posibilita que los criminales usen estos servicios para aprovecharse de sus víctimas. Actúa como negocio en tanto se preocupa por el crecimiento económico, descuidando el proceso de institucionalización de la ética en materia de protección a la intimidad.

Es significativo que la IA resulte ser importante en la "conducción de hostilidades", por medio de manipular la información, el desarrollo de nuevas armas y la sofisticación de métodos de combate (Vigevano, 2021). Aunque las tres empresas han desarrollado dispositivos

para mejorar la calidad de vida y la salud, también se han visto vinculadas con actividades orientadas a la guerra en distintos momentos. "La heurística del temor" se hace presente desde la propuesta de Jonas. La formulación de principios éticos no garantiza su aplicabilidad en aras de generar una gobernanza de la responsabilidad sostenida en el tiempo. En buena medida, al ser empresas, permanecen alineadas con los intereses que generen mayor eficiencia para sus inversionistas.

Las tres empresas son globales, con trabajadores por distintos países del mundo, con clientes y usuarios en diversas regiones, aun así, no pueden escapar a las exigencias propias que les asignan los países donde tienen su "casa matriz". El derecho internacional humanitario y las éticas de corte global tienen un gran desafío, más allá de repensar la aplicación de la IA en situaciones de destrucción e incertidumbre por "el futuro", lo que impide la sostenibilidad de la vida humana y su consecuente dignidad en la tierra. Sin lugar a dudas, desarrollos tecnocientíficos como la IA terminan impactando a la sociedad de masas mediante la aceptación, manipulación o distorsión de los datos personales, afectando el derecho a la privacidad. No basta con los esfuerzos de los estados para exigir un control sobre los datos, las empresas deben refrendar estos deseos con acciones contundentes. Y los ciudadanos deben estar alertas a la letra minúscula y hacer clic en el lugar adecuado para proteger su identidad personal.

Conclusiones

Muchas empresas orientadas a la producción de IA hoy son más conscientes de la necesidad de actualizar los códigos de ética, incluir valores, principios y demás consideraciones legales, para salvaguardar la información personal incluyendo el mayor número de grupos de interés. La propuesta de Warren y Brandeis en la actualidad se ha convertido en una práctica que defienden las democracias y atacan los totalitarismos. Sin embargo, se debe incluir también en la formación ética de los futuros directivos y traba-

jadores de la industria dedicada a la producción de IA, para evitar que la "raza humana" pierda su individualidad y se esclavice, retornando a situaciones ya vividas por la humanidad. Se requieren acciones complementarias a las propuestas institucionales de los estados, tendientes a la regulación. La auto-regulación es un proceso vinculado con la libertad, la autonomía, y el libre desarrollo de las personas.

Mientras David Noble se enfocó en mostrar la fascinación por la tecnología, asunto reconocido como *tecnofilia*, más que nunca la humanidad empieza a desarrollar una *tecnofobia*. Se ve reflejada en aquellas empresas que descuidan, ya sea por acción o por omisión, responsabilidades vinculadas con la alteridad, es decir, con la protección integral de los datos personales. Incluso se pone en duda el temor por estudiar o desarrollar actividades laborales que rápidamente la IA desplazaría. La religión sigue generando rituales y creencias dentro del nuevo capitalismo, que inserta a la tecnología para "avanzar" frente al imperativo de la innovación. En muchos casos, estas acciones son incompatibles con la propuesta de Hans Jonas, de buscar la sostenibilidad empresarial y humana, mediante el imperativo de "obrar de tal forma" que las acciones en la tierra, permitan la continuidad "verdadera" de la especie humana. Sin embargo, no todo es certeza. El temor viene de los usos no éticos que las empresas le pueden dar a la tecnología.

Pese a las limitaciones de seleccionar tres casos de estudios, usando revisión documental y realizando el análisis desde la taxonomía elegida, la I.A. proporciona nuevos productos, para seguir cumpliendo con búsqueda de satisfacer deseos cada vez más exigentes de los "nuevos usuarios". Enfoques fenomenológicos deberían incluir entrevistas a trabajadores y directivos de las empresas seleccionadas, en aras de indagar aquellos espacios que existen entre los ideales de la ética y el derecho con las prácticas culturales reales. El cierre de brechas no se logra desde los deseos por un futuro mejor. Se requiere una ética de la convicción y la responsabilidad sostenida en el tiempo.

Los tres casos de estudio presentan complejidades a partir de los propios contextos de enunciación. La humanidad poco a poco,

les ha dado nuevos virajes a pulsiones contra opuestas y fundantes de la especie. Mientras las tres empresas han desarrollado dispositivos orientados a la mejora de la salud humana, incluso en situaciones como la pandemia, la muerte también saluda con "engaños" siguiendo a Alan Turing.

Marta R. Vigevano expone que la IA tiene diversas aplicaciones, siendo una de estas, "la conducción de hostilidades". Esto incluye tareas vinculadas con el reconocimiento de objetivos militares, de vigilancia y comunicación, de logística y manipulación de la información, el desarrollo de nuevas armas y los métodos de combate. No está alejada de los instrumentos que seducen pero que, al examinarse con detalle, genera temor por los usos futuros no programados.

Es paradójico que mientras un país señala que ciertas empresas se han aliado con sus gobiernos para producir IA, en secreto ellas hagan lo mismo, justificando prepararse para futuros "ciberataques". La creatividad humana permite que se le den diversos usos a la IA, entre los que se encuentra salvar vidas, proteger el derecho a la privacidad, aumentar la alfabetización, o como era el deseo de Toni, "que la gente no se sienta sola". Pero también puede convertirse en un arma, donde las municiones serían los datos personales y la información individual.

A futuro, se deberían generar trabajos investigativos que amplíen la selección de empresas tomadas en este estudio. Particularmente aquellas creadas en los últimos años debido a las oportunidades identificadas con el "boom" de implementar la IA sea generativa o no. Si bien las tres empresas seleccionadas han realizado actividades, unas en mayor o menor grado, para institucionalizarse con el paso del tiempo, deberían realizarse mayores seguimientos a las denuncias planteadas por la sociedad civil a partir de las posibles vulneraciones del derecho a la intimidad. Si la educación debe seguir cumpliendo el deseo kantiano de alcanzar "la mayoría de edad", que no se nos olvide el favor solicitado a Nana: "es hora de despertar".

Referencias

BBC (2014). *Stephen Hawking: "La inteligencia artificial augura el fin de la raza humana"*. 12 de diciembre. https://www.bbc.com/mundo/ultimas_noticias/2014/12/141202_ultnot_hawking_inteligencia_artificial_riesgo_humanidad_egn

BBC (2018). *5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera US$37.000 millones en un día*. 21 de marzo. Disponible en: https://www.bbc.com/mundo/noticias-43472797

Bélisle-Pipon, J. C., Monteferrante, E., Roy, M. C., & Couture, V. (2023). Artificial intelligence ethics has a black box problem. *AI & Society* 38, 1507–1522. https://doi.org/10.1007/s00146-021-01380-0

Brendel, A. B., Mirbabaie, M., Lembcke, T. B., & Hofeditz, L. (2021). Ethical management of artificial intelligence. *Sustainability*, 13(4), 1974. https://doi.org/10.3390/su13041974

Benítez, E. & Varón, D. (2014). Metáforas de empresa: reflexiones conceptuales sobre responsabilidad social empresaria. En: *Intervención social y el debate sobre lo público: reflexiones conceptuales y casos locales*, 115-146. Universidad Icesi.

Canbul, A. (2025). Organizational reflections of the relationship between artificial ıntelligence and emotional ıntelligence in the context of phenomenology and Cartesian dualism. *International Journal of Organizational Analysis*. DOI:10.1108/IJOA-10-2024-4892

Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. *Telecommunications policy*, 44(6), 101937. https://doi.org/10.1016/j.telpol.2020.101937

Dai, J., Shen L., and Zheng, W. (2011). Business-model dynamics: A case study of Tencent. *IEEE 18th International Conference on Industrial Engineering and Engineering Management*, 306-310. doi: 10.1109/ICIEEM.2011.6035164.

Diab, B. (2021). Investigating the impact of organizational ethics on the optimization of artificial intelligence models. *Proceedings of the IE 2021 International Conference*, 1-8. Bucharest – Romania.

Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, 129, 961-974. https://doi.org/10.1016/j.jbusres.2020.08.024

El cronista. (2025). *"El padrino de la inteligencia artificial predijo que habrá una amenaza a la humanidad y una nueva revolución tecnológica en cinco años"*. https://www.cronista.com/usa/trending/el-padrino-de-la-inteligencia-artificial-predijo-que-habra-una-amenaza-a-la-humanidad-y-una-nueva-revolucion-tecnologica-en-cinco-anos/

El Espectador (2025). *Tencent, dueño de LoL, es acusada en Estados Unidos de ser una empresa militar*. https://www.elespectador.com/tecnologia/gadgets-y-apps/tencent-dueno-de-lol-es-acusada-en-estados-unidos-de-ser-una-empresa-militar/

Estevez, D. (2022). *Análisis de 100 cosas*. Universidad del Valle de México.

Gómez, R. (2020). Filtran archivos confidenciales de Intel tras fallo de seguridad. *Criptonoticias*. https://www.criptonoticias.com/seguridad-bitcoin/filtran-archivos-confidenciales-intel-tras-fallo-seguridad/

González, M. y Martínez, D. (2020). Dilemas éticos en el escenario de la inteligencia artificial. *Economía y sociedad*, 25(57), 93-109. http://dx.doi.org/10.15359/eys.25-57.5

Global Growth Markets. (2017). *China's Tencent launches AI medical imaging*. Web link: https://www.ggmkts.com/greater%20china/i-Chinas-Tencent-launches-AI-medical-imaging

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2022). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799-819. DOI: 10.1109/TAI.2022.3194503

IBM. (2023). *Technology for impact*. https://www.ibm.com/impact/2023-ibm-impact-report

IBM. (2025). *Liderazgo intelectual en tecnología sanitaria por Nelson Advisors*. https://www.healthcare.digital/single-post/ibm-s-watson-was-once-heralded-as-the-future-of-healthcare-what-went-wrong

Intel Corporation. (2023). *Explore Intel's history. Explore Intel's history*. 17 de noviembre. https://timeline.intel.com/

Intel Corporation. (2019). *At Intel, Our Values Define Us*. Intel.

Jonas, H. (1995). *El principio de responsabilidad: ensayo de una ética para la civilización tecnológica*. Herder Editorial.

Intel Corporation. (2009). *Intel's Global Human Rights Principles*. https://www.intel.fr/content/dam/www/central-libraries/us/en/documents/es-policy-human- rights.pdf

Kanakia, H., Shenoy, G., & Shah, J. (2019). Cambridge Analytica–a case study. *Indian Journal of Science and Technology*, 12(29), 1-5. DOI: 10.17485/ijst/2019/v12i29/146977

Kirova, V. D., Ku, C. S., Laracy, J. R., & Marlowe, T. J. (2023). The ethics of artificial intelligence in the era of generative AI. *Journal of Systemics, Cybernetics and Informatics*, 21(4), 42-50. DOI:10.54808/JSCI.21.04.42

Li, C. & Zafar, H. (2020). COVID-19 has upended the media industry. This is how it can come back stronger. *World Economic Forum*. Web link: https://www.weforum.org/stories/2020/06/covid-19-has-upended-the-media-industry-this-is-how-it-can-come-back-stronger/

Martínez, J. (2023). "Inteligencia artificial". Ontología y ética. *Philosophia*, 83(1), 27-37. https://doi.org/10.48162/rev.50.020

McDonell, S. (2019). *Cómo la aplicación china WeChat puso mi imagen y voz en alguna base de datos biométrica de personas problemáticas*. https://www.bbc.com/mundo/noticias-48567170

Meenaktshi, H., & Adler, L. (2024). Intel qualifies for $3.5 billion in grants to make chips for US military. *Bloomberg News reports*, 13 de septiembre. Reuters. https://www.reuters.com/technology/intel-qualifies-35-billion-grants-make-chips-us-military-bloomberg-news-reports-2024-09-13/

Morgan, G. (1991). *Imágenes de la organización*. México: Editorial Alfaomega.

Noble, D. (1999). *The religion of technology: The divinity of man and the spirit of invention*. Penguin Publishing Group.

Noble, S. M., & Dubljević, V. (2022). *Ethics of AI in organizations. Human-centered artificial intelligence*, 221-239. Academic Press. https://doi.org/10.1016/B978-0-323-85648-5.00019-0

Parchment, A. (2025). Ethical Leadership and Its Communication Implications: Navigating Artificial Intelligence, Technological Advancements, and Employee Well-Being. *AI-Powered Leadership: Transforming Organizations in the Digital Age*, 73-108. IGI Global Scientific Publishing. DOI: 10.4018/979-8-3373-1687-1.ch003

Ponce, A., Robles, G., Diaz, I. (2023). La inteligencia artificial y el derecho a la intimidad-privacidad. Iustitia Socialis. *Revista Arbitrada de Ciencias Jurídicas*, 8(1), 84-93. https://doi.org/10.35381/racji.v8i1.2493

Ramos, G. (2023). Message from Gabriela Ramos. *Key facts UNESCO`s recommendation on the ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000385082.page=4

Roose, K. (2023). El gobierno de Biden busca controlar la IA sin frenar su desarrollo. *The New York Times en Español*. 4 de noviembre.

Saldaña, M. (2012). «The Right to Privacy». La génesis de la protección de la privacidad en el sistema constitucional norteamericano: El centenario legado de Warren y Brandeis. *Revista de Derecho Político*, 85, 195-240.

Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. *EMBO reports*, 19(8), e46579. https://doi.org/10.15252/embr.201846579

Schultz, M. D., & Seele, P. (2023). Towards AI ethics' institutionalization: knowledge bridges from business ethics to advance organizational AI ethics. *AI and Ethics*, 3(1), 99-111. https://doi.org/10.1007/s43681-022-00150-y

Stahl, B. C., Antoniou, J., Ryan, M., Macnish, K., & Jiya, T. (2022). Organisational responses to the ethical issues of artificial intelligence. *AI & Society*, 37(1), 23-37. https://doi.org/10.1007/s00146-021-01148-6

Stierhoff, G. C. & Davis, A. G. (1998). A history of the IBM Systems Journal. *IEEE Annals of the History of Computing*, 20(1), 29-35.

Tech News. (2021). Intel lidera iniciativas para la transparencia y responsabilidad en IA. *Tech News Journal*, 28(5), 45-50.

Tencent Holdings. (2024). *About us*. Web link: https://www.tencent.com/en-us/about.html#:~:text=Vision%20%26%20Mission,-Value%20for%20Users&text=we%20strive%20to%20incorporate%20social,the%20sustainable%20development%20of%20society

Terán, M. (2023). *Multa histórica de 1.200 millones a Facebook por violar la política de protección de datos de la UE*. Disponible en: https://www.eleconomista.es/tecnologia/noticias/12285863/05/23/multa-historica-de-1200-millones-a-facebook-por-violar-la-politica-de-proteccion-de-datos-de-la-ue-.html#:~:text=Multa%20hist%C3%B3rica%20de%201.200%20millones,de%20datos%20de%20la%20UE

Tripathi, A., & Kumar, V. (2025). Ethical practices of artificial intelligence: a management framework for responsible AI deployment in businesses. *AI and Ethics*, 1-12. https://doi.org/10.1007/s43681-025-00670-3

Vakkuri, V., Kemell, K. K., Kultanen, J., & Abrahamsson, P. (2020). The current state of industrial practice in artificial intelligence ethics. *Ieee Software*, 37(4), 50-57. DOI: 10.1109/MS.2020.2985621

Varma, A., Dawkins, C., & Chaudhuri, K. (2023). Artificial intelligence and people management: A critical assessment through the ethical lens. *Human Resource Management Review*, 33(1), 100923. https://doi.org/10.1016/j.hrmr.2022.100923

Varón, D. (2025). *Inteligencia Artificial y Ética. Examinando el derecho a la intimidad y la responsabilidad empresarial futura*. Artículo Inédito.

Varón, D. (2023). ¿Qué perspectiva de la ética permite estudiar la tecnología actual? En *ElPuenteAlAire*, 24 de abril. https://www.youtube.com/watch?v=Ei30Je8ckfU

Vigevano, M. (2021). Inteligencia artificial aplicable a los conflictos armados: límites jurídicos y éticos. *ARBOR Ciencia, Pensamiento y Cultura*, 197 (800). https://doi.org/10.3989/arbor.2021.800002

Xavier, D. F., & Korunka, C. (2025). Integrating Artificial Intelligence across cultural orientations: A longitudinal examination of creative self-efficacy and employee autonomy. *Computers in Human Behavior Reports*, 18, 100623. https://doi.org/10.1016/j.chbr.2025.100623

Yadav, G. (2025). AI and analytics conundrum: unpacking the barriers in modern HR with ISM and MICMAC analysis. *International Journal of Organizational Analysis*. Ahead-of-print. https://doi.org/10.1108/IJOA-08-2024-4782

[Parte 3]

Ética de la inteligencia artificial en la educación superior

[Capítulo 6]

Marcos institucionales para el uso ético de la Inteligencia Artificial Generativa en las instituciones de educación superior

Henry Arley Taquez Quenguan y
María Isabel González Serrano

La acelerada integración de la Inteligencia Artificial Generativa (IAG) en la educación superior ha abierto un horizonte de oportunidades y desafíos que obligan a repensar las prácticas formativas, investigativas y ético-profesionales en las universidades. En un entorno donde la creatividad, la autonomía intelectual y el compromiso con la transformación social son fundamentales para el ejercicio profesional, se hace necesario establecer marcos orientadores que favorezcan un uso ético y deliberado de estas tecnologías. Este enfoque ético no se limita al cumplimiento de normativas formales, sino que implica tomar decisiones humanas, informadas y conscientes sobre cómo se integra la IAG en la vida académica y profesional. En este sentido, es relevante analizar cómo las instituciones de educación superior de referencia, tanto nacionales como internacionales, han formulado políticas y recomendaciones sobre su uso dentro de las comunidades universitarias.

La construcción de un lineamiento para orientar el uso de la IAG no puede entenderse como la mera elaboración de un documento institucional destinado a ser consultado solo en casos de conflicto o novedad. Como lo advierte la filósofa española Mariana Garcés, las universidades no deben responder desde la reactividad motivada por la paranoia o la urgencia adaptativa, sino avanzar del miedo hacia la participación crítica y propositiva (Universitat Oberta de Catalunya [UOC], 2023). En el contexto colombiano, esta participación reviste características particulares. La Política Nacional de Inteligencia Artificial, publicada por el Consejo Nacional de Política Económica y Social (CONPES, 2025), reconoce que Colombia "aún evidencia bajas capacidades para la investigación, desarrollo, adopción y aprovechamiento ético y sostenible de sistemas de IA en asuntos estratégicos a nivel social, económico y ambiental" (p. 3).

Los lineamientos que se construyan en el país deben, por tanto, propiciar una participación activa desde un enfoque integrador. Como advierte Moncada (2025), la irrupción de los sistemas de IA en el ámbito educativo ha sido veloz, pero no equitativa. Factores estructurales como el acceso desigual a la conectividad marcan una diferencia sustancial entre zonas urbanas, rurales y periféricas, generando brechas que inciden directamente en las posibilidades de apropiación social de la tecnología. En esa misma línea, Tejedor (2025) destaca que estas desigualdades, amplificadas por la infraestructura digital limitada, condicionan la capacidad de respuesta de las comunidades educativas frente a la IA.

Frente a este panorama, las universidades que cuentan con una infraestructura sólida y con una capacidad instalada en términos de competencias digitales entre los actores de su comunidad, deben asumir una responsabilidad ampliada. No es suficiente con generar lineamientos para garantizar buenas prácticas dentro de sus aulas, oficinas y pasillos. Estos documentos también deben convertirse en insumos que, a partir de procesos participativos, contribuyan a la configuración del papel que se desea otorgar a la IAG en la educación superior

colombiana. Esto implica reconocer a los sistemas de IAG no solo como tecnologías de apoyo, sino como agentes con el potencial de reconfigurar la construcción del conocimiento, el aprendizaje y las relaciones educativas. A partir del diálogo entre los distintos actores de la comunidad universitaria, es posible construir acuerdos sobre su sentido, uso y límites, orientados a atender los retos educativos del país y a promover una apropiación ética, crítica y contextualizada de estas tecnologías.

Este capítulo analiza cómo universidades de referencia han abordado los dilemas éticos que plantea la integración de la IAG en la vida académica, en particular en relación con principios como la integridad, la autonomía, la equidad y la responsabilidad. A partir de una revisión documental de lineamientos institucionales en instituciones de educación superior de América Latina, Europa y Estados Unidos, se identifican principios compartidos, divergencias normativas y tensiones en su gobernanza. Asimismo, se presenta como estudio de caso la experiencia de la Universidad Icesi, que ha diseñado un modelo de integración de la IAG alineado con sus valores formativos. Este modelo reconoce el valor de la tecnología sin desplazar el juicio humano, promueve un uso creativo y reflexivo de la IAG, y fortalece la formación de profesionales capaces de ejercer su liderazgo con criterio ético en entornos complejos y en transformación.

Contexto ético-tecnológico en la educación superior

El reto de simular la inteligencia humana no es nuevo. Desde los inicios de la informática moderna, científicos, filósofos y educadores han reflexionado sobre la posibilidad de que las máquinas imiten o incluso superen las habilidades cognitivas humanas. Lo que antes parecía una aspiración teórica o una trama de ciencia ficción hoy se concreta con el desarrollo acelerado de sistemas de Inteligencia Artificial Generativa capaces de producir textos, imágenes, código o música con niveles de calidad que desafían el criterio humano. Esta

transformación plantea una pregunta de reflexión para la educación superior: ¿qué significa formar profesionales humanos cuando las máquinas pueden generar, analizar y sintetizar información con una velocidad y eficiencia que nos resultan inalcanzables?

La "apocalipsis de la tarea", como la denomina Ethan Mollick (2023), se refiere a que la Inteligencia Artificial Generativa ha alcanzado un nivel de capacidad en el que puede realizar la mayoría de las tareas académicas mejor que los estudiantes. Lo que impacta directamente el modelo tradicional de enseñanza basado en trabajos y deberes. Ante este escenario, la "nueva universidad", como la describen Chicaiza et al. (2024), debe reformular sus principios pedagógicos en respuesta a un entorno cambiante, marcado por innovaciones tecnológicas y por dilemas éticos que afectan el "corazón" o la esencia del proceso educativo. En este contexto, la educación superior necesita adoptar una visión centrada en el ser humano, donde la tecnología actúe como medio, no como fin, y la formación privilegie el pensamiento crítico, la creatividad situada y el juicio ético (UNESCO, 2023).

Desde una perspectiva ética, la IAG plantea al menos cinco dilemas clave en el ámbito educativo. El primero es la autonomía y agencia del estudiante, es decir, su capacidad para decidir cuándo, cómo y por qué utilizar estas herramientas. El segundo se refiere a la transparencia y autoría, que exige declarar su uso y reconocer los límites entre producción humana y automatizada. El tercero involucra la equidad y el acceso, considerando las brechas en infraestructura y alfabetización digital. El cuarto dilema apunta a la validez de la evaluación, que debe reflejar el aprendizaje y las competencias reales. El quinto, la gobernanza institucional, implica que las universidades ofrezcan lineamientos claros, adaptativos y sostenibles.

Ahora bien, abordar estos dilemas para las universidades se convierten en una oportunidad para revisar sus fundamentos y fortalecer su propósito formativo. Quizás esta revisión conlleve a "volver al centro" y priorizar capacidades propiamente humanas que difícilmente podrá sustituir la Inteligencia Artificial Generativa, tales como: el dis-

cernimiento ético, la empatía, la curiosidad intelectual y la colaboración interdisciplinaria (Carrera, 2024; Molina et al., 2024). En este sentido, frente a los riesgos y tensiones de la automatización del pensamiento, las universidades están llamadas a sostener una educación con propósito, centrada en la construcción crítica de conocimiento.

Los proyectos educativos institucionales suelen declarar compromisos en torno al desarrollo integral del estudiante, la ética, la responsabilidad social y la innovación. Bajo este panorama se hace necesario que dichos principios se traduzcan en acciones concretas que respondan a las realidades de cada comunidad académica. Ahora bien, para orientar dichas acciones una alternativa es acudir a referentes conceptuales como el modelo AI Ecological Education Policy Framework (Chan, 2023), que propone un enfoque multidimensional para guiar la adopción ética y pedagógica de la IA en la educación superior. El modelo subraya que ninguna dimensión -directiva, pedagógica o tecnológica- puede abordarse de forma aislada. Para que la integración de la Inteligencia Artificial Generativa sea exitosa en las instituciones de educación superior, es indispensable fomentar la colaboración entre actores institucionales e implementar políticas que se concreten en prácticas sostenibles, situadas y basadas en la corresponsabilidad.

Riesgos asociados al uso de la IAG en la construcción reflexiva de los lineamientos

La integración de la Inteligencia Artificial Generativa (IAG) en la educación superior plantea una serie de riesgos que requieren una reflexión profunda sobre los valores fundamentales que orientan la formación universitaria. Desde una perspectiva ética, la IAG desafía cinco dimensiones clave: la autonomía y agencia del estudiante, la transparencia y autoría, la equidad y el acceso, la validez de la evaluación, y la gobernanza institucional. Estos riesgos no se resuelven mediante un conjunto cerrado de reglas, sino a través de una ética situada, deliberativa y formativa, que permita construir acuerdos

colectivos en torno al uso justo, creativo y responsable de estas tecnologías. La Universidad Icesi ha reconocido este reto al adoptar una postura pedagógica basada en la agencia compartida, el discernimiento crítico y la construcción colectiva de sentido. Esta orientación se refleja tanto en el Proyecto Educativo Institucional, como en el Manifiesto sobre Inteligencia Artificial y Educación (2024) y en los Lineamientos para el Uso Pedagógico de la Inteligencia Artificial Generativa (2025). Ambos documentos articulan principios humanistas con criterios técnicos, y reconocen que la IAG no solo transforma herramientas didácticas, sino también los fundamentos epistemológicos y éticos del quehacer universitario.

Uno de los riesgos más debatidos en el contexto universitario es el aumento del fraude académico. El artículo 95 del Libro de Derechos, Deberes y Normas de los Estudiantes de Pregrado de la Universidad Icesi (2025) define como fraude la presentación de productos académicos generados total o parcialmente por sistemas de IAG sin declarar su uso, en contravención de los lineamientos establecidos. Este comportamiento compromete la integridad académica, entendida como el compromiso con la honestidad, la justicia, la responsabilidad, el respeto y el esfuerzo (International Center for Academic Integrity (ICAI, 2021). Aunque algunas voces proponen prohibir por completo la Inteligencia Artificial Generativa en contextos educativos, tal prohibición resulta insostenible, ya que el futuro de la educación superior no puede prescindir de las herramientas tecnológicas emergentes. Por ello, se considera necesario implementar marcos pedagógicos que regulen su uso de forma ética, transparente y estratégica. En lugar de optar por enfoques punitivos, la Universidad Icesi ha elegido construir acuerdos formativos, en lugar de aplicar enfoques punitivos, reconociendo que el desarrollo de la integridad académica requiere más que restricciones: exige acompañamiento, deliberación y formación crítica.

Otro de los riesgos que emergen con el uso indiscriminado de la IAG es el deterioro de las habilidades cognitivas. Estas tecnologías

permiten producir contenidos de manera rápida y verosímil, lo que puede reducir el esfuerzo invertido por los estudiantes en la comprensión, el análisis y la elaboración de ideas propias. Esta práctica debilita competencias centrales del PEI de la Universidad Icesi, como el pensamiento crítico, la autorregulación y la construcción contextual del conocimiento (Universidad Icesi, 2017). Bastani et al. (2024) advierten que los estudiantes, al confiar en la Inteligencia Artificial Generativa sin cuestionar sus resultados, corren el riesgo de limitar su desarrollo cognitivo. La educación universitaria debe centrarse en formar personas capaces de evaluar la calidad y veracidad de los contenidos. Por esta razón, los lineamientos de la Universidad Icesi promueven prácticas pedagógicas donde la IAG actúa como una herramienta de apoyo, no como sustituto del pensamiento humano. El objetivo no es reemplazar el juicio del estudiante, sino fortalecerlo mediante un uso reflexivo, deliberado y contextualizado de la IAG.

La dependencia excesiva de la IAG también puede afectar negativamente la interacción y la colaboración humana, que son pilares fundamentales en el proceso educativo. Martínez (2023) señala que el uso de chatbots y asistentes inteligentes puede desplazar la figura del docente como orientador del aprendizaje. Esta sustitución debilita el desarrollo de habilidades sociales y emocionales que se construyen en la interacción con otros. La Universidad Icesi sostiene que la educación no puede reducirse a una relación entre el estudiante y la máquina. Es indispensable preservar los espacios de socialización docente-estudiante y estudiante-estudiante, porque en ellos se gestan la empatía, la argumentación, el respeto por la diferencia y la creación colectiva de saberes. Además, la colaboración entre docentes resulta esencial para reflexionar sobre los impactos de la Inteligencia Artificial Generativa y diseñar experiencias de aprendizaje que articulen tecnología y humanidad. En este sentido, el modelo de inteligencia híbrida propuesto por Burger et al. (2023) resulta pertinente, pues combina el potencial técnico de la IAG con las capacidades humanas, en un proceso de mejora mutua y aprendizaje colaborativo.

Otro riesgo crítico en el uso educativo de la Inteligencia Artificial Generativa es la reproducción de sesgos y prejuicios. Ferrara (2023) advierte que los modelos generativos, al entrenarse con datos históricos, pueden amplificar estereotipos y formas de discriminación. Este problema puede tener consecuencias éticas profundas al perpetuar representaciones erróneas de grupos sociales. En el contexto académico, estos sesgos pueden filtrarse en contenidos curriculares, investigaciones o materiales divulgativos, comprometiendo el compromiso institucional con la equidad y la justicia. La alfabetización en Inteligencia Artificial Generativa implica cuestionar sus resultados, identificar sus límites y comprender los valores inscritos en sus respuestas. Promover esta alfabetización crítica forma parte de una ética pedagógica que busca empoderar a los docentes y estudiantes para actuar con autonomía y responsabilidad (Crompton & Burke, 2023).

En este orden de ideas, otro de los riesgos más complejos que plantea la Inteligencia Artificial Generativa en la educación superior se refiere a la validación de la evaluación. Esta debe reflejar el proceso de aprendizaje, el esfuerzo individual y las competencias desarrolladas por los estudiantes (Barrientos et al., 2020). Sin embargo, cuando los productos académicos pueden ser generados por herramientas automatizadas, se cuestiona la validez de los instrumentos evaluativos tradicionales. UNESCO (2023) señala que el riesgo de confundir la asistencia tecnológica con el fraude académico ha derivado en respuestas punitivas, que no necesariamente promueven la reflexión ni el aprendizaje genuino. En este sentido, evaluar de manera válida en contextos mediados por tecnologías generativas requiere rediseñar las actividades con criterios éticos, que reconozcan tanto el potencial de la Inteligencia Artificial Generativa como los límites de su uso en la formación humana.

Finalmente, es de gran relevancia la gobernanza institucional de la Inteligencia Artificial Generativa. Regular su uso en la universidad implica establecer lineamientos que equilibren la necesidad de orien-

tar y proteger a la comunidad académica con el respeto a la autonomía pedagógica. Floridi & Cowls (2019) plantean que toda política educativa sobre Inteligencia Artificial debe fundamentarse en una ética relacional, capaz de considerar las particularidades de cada contexto. Desde esta perspectiva, la Universidad Icesi ha optado por construir lineamientos participativos, en lugar de imponer normativas unilaterales. El Manifiesto sobre Inteligencia Artificial y Educación (2024) y los Lineamientos para el Uso Pedagógico de la Inteligencia Artificial Generativa (2025) expresan una visión deliberativa, en la que los docentes, estudiantes y directivos participan en la toma de decisiones sobre el uso de estas tecnologías. Este enfoque fomenta la corresponsabilidad institucional y evita tanto la vigilancia excesiva como la ausencia de regulación. La gobernanza ética de la Inteligencia Artificial Generativa exige marcos adaptativos que acompañen la transformación educativa, reconozcan la diversidad disciplinar y orienten el desarrollo de capacidades humanas en un entorno tecnificado.

Análisis de lineamientos de instituciones de educación superior sobre el uso ético de la Inteligencia Artificial Generativa

Con el objetivo de identificar buenas prácticas, se realizó una revisión documental de las políticas, lineamientos y guías publicadas por 11 universidades entre públicas y privadas de América Latina, Europa y Norteamérica, que han publicado regulaciones sobre el uso de la IAG. Estas universidades fueron seleccionadas por su alto reconocimiento académico y su liderazgo en innovación tecnológica. El análisis se centró en una serie de criterios clave como la protección de datos, la integridad académica, la orientación pedagógica, la inclusión, la formación docente y la participación estudiantil, con el fin de comprender cómo estas instituciones están gestionando el uso de la IAG en sus comunidades académicas.

Tabla 1. Lineamientos institucionales analizados

Institución y país	Año y documento	Enfoque
Universidad Externado de Colombia (Colombia)	2024 – Lineamientos para el uso de inteligencia artificial	Enfoque ético, legal y pedagógico centrado en el ser humano. Distingue entre contenidos con/sin aporte humano. Define riesgos, asigna responsabilidades diferenciadas y promueve formación continua. Comité institucional actualiza y acompaña el proceso.
Alianza 4U (ICESI, CESA, EAFIT, UNINORTE – Colombia)	2024 – IA for You: Manifiesto sobre Inteligencia Artificial	Marco común para la integración ética, pedagógica y estratégica de la IAG. Propone principios de transparencia, equidad, explicabilidad y colaboración interdisciplinaria. Orientaciones para formación docente, ciudadanía digital y personalización del aprendizaje.
Universidad de los Andes (Colombia)	2024 – Lineamientos para el uso de IAG	Uso ético, informado y contextualizado. Declaración obligatoria del uso. Promueve competencias digitales, protección de datos y autoría responsable. Niveles de uso según objetivos de aprendizaje. Coherencia con valores institucionales.
Universidad Peruana de Ciencias Aplicadas – UPC (Perú)	2024 – Lineamientos para el uso de herramientas de IAG en el proceso de enseñanza-aprendizaje	Define usos permitidos/prohibidos para estudiantes y docentes. Señala riesgos como plagio, desinformación y deshonestidad. Establece comité regulador. Fomenta equidad, autoría responsable y formación crítica.
Universidad Andrés Bello – UNAB (Chile)	2024 – Lineamiento para el uso responsable de IA	Marco ético, normativo y pedagógico. Principios: transparencia, integridad, privacidad y formación continua. La IAG es asistente, no reemplazo. Coherencia con política nacional.

Institución y país	Año y documento	Enfoque
Universidad Iberoamericana – IBERO (México)	2023 – Lineamientos para el uso de la Inteligencia Artificial	Uso ético, responsable y contextual. Docentes deciden condiciones de uso en cada asignatura. Estudiantes deben declarar su uso. Enfatiza el pensamiento crítico y la evaluación ética del contenido.
Tecnológico de Monterrey (México)	2023 – Guía rápida para el uso ético de IA	Dirigida a estudiantes. Principios: dignidad, veracidad y responsabilidad. Uso con autorización docente. Declaración obligatoria. Evita plagio y falsificación de habilidades. Énfasis en prevención individual.
Universidad de Cambridge (Reino Unido)	s.f. – Using Generative AI	Uso permitido bajo normativas locales. Fomenta rediseño de evaluaciones, responsabilidad del usuario y verificación de información. Docentes deben orientar y acompañar. Enfoque en transparencia, ética y aprendizaje significativo.
Stanford University (EE. UU.)	s.f. – Guía didáctica de la inteligencia artificial	Apoya a docentes en integración pedagógica de la IAG (especialmente chatbots). Módulos de formación en ética, políticas, tareas. Uso crítico y contextualizado sin requerir experiencia previa.
European University Institute – EUI (Unión Europea)	2024 – Guidelines for the Responsible Use of AI for Research	Enfocado en investigación. Promueve alfabetización digital, inclusión y protección de datos. Prohíbe detectores automáticos. Detalla uso ético en generación de contenidos, propiedad intelectual y revisión de datos.
Harvard University (EE. UU.)	s.f. – Initial guidelines for the use of Generative AI tools at Harvard	Uso responsable. Prohíbe herramientas no autorizadas y el ingreso de datos confidenciales. Riesgos: phishing, desinformación. Verificación de contenido obligatoria. Supervisión desde cada unidad académica.

Institución y país	Año y documento	Enfoque
Harvard University (EE. UU.)	s.f. – Research with Generative AI: Resources for scholars and researchers	Plataforma con recursos para investigadores. Promueve uso ético, integridad académica, seguridad de datos y consulta de políticas editoriales. Foco en prevención de sesgos, desinformación y dependencia excesiva.

Fuente: construcción propia.

Las universidades, al adoptar estas tecnologías, se enfrentan a la tarea de equilibrar el potencial transformador de la IAG con las precauciones necesarias para garantizar su uso ético, la protección de la privacidad y la preservación de la autonomía intelectual. En este análisis, se exploran los enfoques adoptados por las universidades y se destaca la importancia de crear un marco ético robusto que guíe el uso de la IAG de manera que potencie, no sustituya, las capacidades humanas.

Enfoque ético y responsabilidad en el uso de la IAG

Uno de los elementos clave que emerge en el análisis de los lineamientos institucionales es la preocupación ética en torno al uso de la Inteligencia Artificial Generativa. Las universidades que han adoptado políticas sobre la IAG coinciden en que es esencial establecer un marco ético claro que asegure que su uso esté alineado con los valores humanos y las expectativas académicas.

La Universidad Externado de Colombia (2024), por ejemplo, ha implementado un enfoque que pone al ser humano en el centro del uso de la IA, lo cual se refleja en su insistencia en distinguir entre contenidos generados por IA y aquellos en los que el usuario aporta un elemento sustancial. Este enfoque busca evitar que los estudiantes dependan completamente de la IA para generar respuestas sin el necesario juicio

humano. La universidad también establece niveles de riesgo asociados al uso de la IA y distribuye las responsabilidades entre estudiantes, docentes e investigadores de manera diferenciada, con el fin de garantizar que cada actor sea consciente de su papel en el proceso. Además, se destaca la necesidad de una formación continua en ética digital, donde tanto los estudiantes como los docentes comprendan las implicaciones de la IA en la creación de conocimiento. El enfoque reflexivo propuesto por esta universidad permite que los usuarios sean autores activos, no solo usuarios pasivos de la tecnología, asegurando que el uso de la IAG se haga de manera responsable (Externado, 2024).

Por su parte, la Universidad de los Andes (2024) adopta un enfoque similar, aunque más orientado hacia el uso crítico y contextualizado de la IAG. Esta universidad considera que la IA debe ser una herramienta complementaria, en lugar de un sustituto del juicio humano. En su propuesta, se hace hincapié en la declaración explícita del uso de la IAG y en la necesidad de que los estudiantes asuman la autoría de los contenidos generados. La universidad también promueve una integración de la IA dentro de sus valores institucionales, destacando la protección de datos personales y asegurando que la tecnología se utilice de forma reflexiva. Además, enfatiza que la responsabilidad ética recae tanto en los estudiantes como en los docentes, quienes deben monitorear y orientar el uso de la IA en sus respectivas áreas. Este enfoque se alinea con la idea de que el uso de la tecnología debe ser un proceso dinámico de aprendizaje y reflexión continua (Universidad de los Andes, 2024).

Regulación y usos permitidos de la IAG

Las universidades también han tenido que enfrentar la necesidad de regular el uso de la IAG para asegurar que no se utilice de manera fraudulenta o para eludir el juicio académico. En este sentido, la Universidad Peruana de Ciencias Aplicadas (UPC) (2024) ha adoptado lineamientos específicos que delimitan claramente los

usos permitidos y prohibidos. La UPC reconoce el potencial positivo de la IAG, especialmente en la personalización del aprendizaje y la mejora de la accesibilidad, pero también advierte sobre los riesgos asociados, como el plagio y la desinformación. La universidad considera que es fundamental que los estudiantes desarrollen competencias críticas y que se les eduque para entender el impacto ético de usar la IAG, lo cual les permitirá utilizar estas herramientas de manera responsable. A través de estos lineamientos, la UPC busca garantizar que la autoría y la integridad académica sean preservadas, sin que los estudiantes recurran a la tecnología para reemplazar su propio esfuerzo intelectual (UPC, 2024).

En cuanto a la Universidad Andrés Bello (UNAB) (Chile) (2024), se propone un enfoque similar, donde se regula el uso de la IAG para asegurar la transparencia y privacidad. La universidad destaca que la IA debe ser vista como un asistente para los estudiantes, pero no como un sustituto del juicio humano. Además, insiste en la formación continua de estudiantes y docentes, y promueve que los estudiantes asuman la responsabilidad en el uso de la IA, asegurando que no se convierta en una herramienta para eludir la integridad académica. La UNAB también subraya la importancia de la articulación de estas políticas con las normas nacionales de educación y el marco ético de la universidad (UNAB, 2024).

Protección de Datos y Privacidad

La protección de los datos personales es una de las principales preocupaciones en la integración de la IAG en el ámbito académico. Dado que la IA maneja grandes cantidades de datos sensibles, es esencial que las universidades adopten medidas estrictas para garantizar la privacidad de los usuarios y proteger sus datos personales.

La Universidad de Harvard (s.f.) establece directrices claras sobre la protección de datos confidenciales y la seguridad de la información.

Así mismo, señala riesgos como el phishing, los deepfakes y la manipulación de información. En su enfoque, se subraya la importancia de evitar el uso de herramientas no autorizadas y de garantizar que los usuarios respeten las normas locales sobre la protección de datos. Harvard también señala que el fraude académico y el uso indebido de la IA deben ser prevenidos, y hace un llamado a los estudiantes y docentes a que verifiquen siempre la información generada por la IA. La universidad está comprometida con la evaluación rigurosa de los contenidos generados por estas tecnologías, garantizando que los datos personales estén protegidos en todo momento (Harvard, s.f.).

El European University Institute (EUI) (2024), por su parte, enfatiza la alfabetización digital y la protección de la propiedad intelectual en el contexto de la investigación científica. La universidad también subraya que los investigadores deben utilizar la IAG con responsabilidad, garantizando la seguridad y privacidad de los datos que manejan. Este enfoque se orienta hacia un uso responsable de la IA en el ámbito de la investigación, asegurando que se respeten las normas éticas y los derechos de propiedad intelectual (EUI, 2024).

Ambos documentos coinciden en enfatizar la seguridad y la integridad académica, aunque el EUI profundiza más en el uso responsable en investigación científica, mientras que Harvard apunta a una implementación segura y regulada en todos los niveles universitarios.

Transparencia como principio central

La supervisión institucional juega un papel esencial en la implementación de la IAG en las universidades. Las universidades líderes han establecido comités de supervisión o equipos responsables que aseguran que el uso de la IA se ajuste a las normas éticas y legales. Estos equipos tienen la responsabilidad de monitorear y actualizar las políticas sobre el uso de la IA, y de garantizar que todos los miembros de la comunidad académica sean responsables en su uso.

En México, el Instituto Tecnológico y de Estudios Superiores de Monterrey (s.f.), ha publicado una guía rápida centrada en los estudiantes, que promueve el uso ético de la Inteligencia Artificial Generativa y establece principios como la dignidad humana, la veracidad y la responsabilidad. La guía aclara que la Inteligencia Artificial Generativa solo puede utilizarse con autorización docente y que su uso debe declararse siempre en trabajos académicos. También, prohíbe el uso para falsear habilidades o plagiar. La Universidad de Cambridge (s.f.), en el Reino Unido, adopta un enfoque similar, aunque más amplio, al considerar tanto la investigación como la docencia. Su guía promueve el diseño de evaluaciones más pertinentes, reconoce los límites de la tecnología y fomenta la transparencia. Ambas instituciones coinciden en la necesidad de supervisión docente, pero difieren en el nivel de intervención institucional: el Tecnológico de Monterrey se enfoca en la prevención individual, mientras que la Universidad de Cambridge establece orientaciones que implican ajustes pedagógicos estructurales, como rediseñar formas de evaluación para mantener la integridad académica.

Formación continua en el uso de la IAG

Las universidades han reconocido la importancia de desarrollar capacidades para que tanto estudiantes como docentes comprendan las implicaciones éticas de la IA y desarrollen las competencias críticas necesarias para su uso adecuado.

En este sentido, instituciones como la Universidad de los Andes, la UNAB y el EUI priorizan el desarrollo de competencias críticas, digitales y éticas como condición para el uso pedagógico de la Inteligencia Artificial Generativa. Estas instituciones proponen modelos que articulan lo normativo con lo formativo; buscan construir una cultura institucional que promueva el uso consciente y transformador de estas tecnologías emergentes. Por otra parte, el Tecnológico de Monterrey coloca el énfasis en la acción individual del estudiante; hace énfasis en la autorregulación a través de la orientación ética.

Esta diferencia sugiere que mientras unas instituciones privilegian el acompañamiento colectivo, otras confían en la responsabilidad individual como estrategia de prevención y desarrollo ético.

Ahora bien, en cuanto a los puntos de convergencia se puede decir que los lineamientos revisados comparten varios elementos clave: la necesidad de declarar el uso de la Inteligencia Artificial Generativa, la protección de la privacidad y los datos personales, la promoción del pensamiento crítico y la autoría responsable. Además, todos subrayan la importancia de acompañar la implementación de estas tecnologías con formación continua. La mayoría coincide en que la Inteligencia Artificial Generativa no reemplaza el juicio humano y debe funcionar como herramienta de apoyo. Sin embargo, existen diferencias en el nivel de regulación: mientras algunas instituciones, como la UPC o el Externado, proponen definiciones explícitas de riesgos, otras como Harvard o Cambridge optan por guías más abiertas a la interpretación de cada facultad o curso. Esta diversidad refleja los contextos normativos, culturales y pedagógicos de cada institución.

Otro punto de convergencia es la asignación de responsabilidades diferenciadas a estudiantes, docentes, investigadores y personal administrativo. En la mayoría de los documentos revisados, se reconoce que cada actor académico enfrenta desafíos y oportunidades particulares ante la Inteligencia Artificial Generativa. Por este motivo, se promueve una gobernanza compartida basada en la transparencia, la supervisión docente, la participación estudiantil y la coordinación institucional. En algunas universidades se articulan estos elementos mediante comités o equipos responsables de la implementación, actualización y seguimiento de los lineamientos. Esto refuerza la idea de que los marcos éticos efectivos requieren procesos institucionales dinámicos, que movilicen esfuerzos conscientes y continuos.

En síntesis, la integración responsable de la Inteligencia Artificial Generativa en el ámbito académico exige un compromiso ético claro por parte de las universidades, un marco regulatorio riguroso y una formación continua que permita a los estudiantes y docentes compren-

der las implicaciones de esta tecnología. Las universidades revisadas han adoptado marcos éticos y pedagógicos sólidos que garantizan que la IA complemente las capacidades humanas, protegiendo la autonomía intelectual y los derechos fundamentales de los usuarios. La implementación exitosa de estas políticas puede transformar la educación superior y ofrecer nuevas oportunidades para el aprendizaje, sin comprometer la integridad académica ni la autonomía del juicio humano.

Más allá del papel: los lineamientos deben promover la reflexión pedagógica y ética

El desarrollo de lineamientos para el uso de la Inteligencia Artificial Generativa en la educación superior no puede limitarse a una compilación normativa o técnica. Más allá del papel, estos lineamientos deben convertirse en textos vivos que orienten una reflexión pedagógica, ética y contextualizada. En la Universidad Icesi esta convicción ha guiado la formulación de su propuesta institucional basada en tres vertientes interconectadas: enseñar para, sobre y con la Inteligencia Artificial Generativa. Esta estrategia no busca una adopción instrumental de la tecnología, sino una integración crítica que respete la diversidad disciplinar y fortalezca las capacidades humanas. Tal como advierte el Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado (INTEF, 2024), la Inteligencia Artificial en la educación debe permitir construir entornos de aprendizaje más creativos, curiosos y críticos, lo que requiere que docentes, estudiantes e instituciones comprendan sus implicaciones, beneficios y riesgos. Por ello, los lineamientos deben ser desarrollados desde una ética deliberativa, situada y plural, que reconozca la dimensión formativa del uso tecnológico.

Enseñar para la Inteligencia Artificial Generativa implica introducir a los estudiantes en el uso cotidiano y reflexivo de estas herramientas, con énfasis en la alfabetización digital. Este enfoque busca que los estudiantes comprendan los riesgos de privacidad, sesgo, manipula-

ción o vigilancia asociados al uso de sistemas inteligentes, sin renunciar a su potencial formativo. Este enfoque solo es efectivo si se vincula con una pedagogía que promueva la deliberación ética y la toma de decisiones informadas (INTEF, 2024; An et al., 2025). La Universidad Icesi ofrece acceso a estas herramientas y orientación para su uso académico. Enseñar para la Inteligencia Artificial Generativa no es enseñar sobre tecnología, sino enseñar desde y con una mirada crítica sobre su impacto. En este sentido, la alfabetización en Inteligencia Artificial Generativa se configura como un eje transversal que articula competencias técnicas, sociales y éticas en el marco de una educación con propósito (UNESCO, 2023).

Enseñar sobre la Inteligencia Artificial Generativa demanda una aproximación más técnica, orientada a que los estudiantes comprendan el funcionamiento, los modelos, los algoritmos y las implicaciones computacionales de estos sistemas. Esta dimensión resulta fundamental en disciplinas como la ingeniería, la estadística, la programación, la matemática aplicada o la ciencia de datos, pero también empieza a cobrar importancia en otras áreas como el derecho, la psicología o la comunicación. Según el INTEF (2024), este enfoque debe abarcar temas como percepción y actuación algorítmica, razonamiento automático, aprendizaje automático e impacto social de la Inteligencia Artificial Generativa. La Universidad Icesi ha comenzado a incorporar estos contenidos en cursos específicos. Con esto, se busca promover una comprensión integral del papel que juegan los sistemas inteligentes en la configuración de la realidad contemporánea. Enseñar sobre la Inteligencia Artificial Generativa no se trata de formar desarrolladores, sino de formar ciudadanos capaces de auditar, evaluar y participar en los procesos tecnológicos que moldean la vida colectiva.

Enseñar con la Inteligencia Artificial Generativa representa el desafío más inmediato para los docentes, pues implica utilizar herramientas basadas en IA para diseñar, implementar y evaluar experiencias de aprendizaje. Este enfoque exige al profesorado criterios sólidos para seleccionar e integrar herramientas que promuevan el aprendizaje

significativo. Como lo señala INTEF (2024), esto requiere capacitar al profesorado para diseñar actividades que articulen los beneficios de la Inteligencia Artificial Generativa con objetivos pedagógicos claros, teniendo en cuenta el respeto a la privacidad de los datos, la diversidad de estilos de aprendizaje y los principios de evaluación justa. En la Universidad Icesi, esta línea se traduce en experiencias concretas que van desde el ajuste curricular hasta las acciones pedagógicas en el rediseño de las actividades de aprendizaje, acompañadas por procesos formativos liderados desde la Decanatura de Innovación Educativa y Fortalecimiento del PEI. Esta labor implica una profunda reflexión sobre las finalidades educativas, lo que exige mantener el juicio pedagógico como brújula ética en la integración tecnológica.

Esta transformación no puede abordarse desde una única perspectiva o disciplina. La integración de la Inteligencia Artificial Generativa exige marcos flexibles, informados por las especificidades del saber disciplinar y guiados por principios comunes de equidad, justicia e inclusión. Por ello, en la Universidad Icesi se ha propuesto un marco común de actuación dirigido principalmente al profesorado, quien desempeña un rol clave como mediador pedagógico y formador ético. Este marco busca empoderar a los docentes para tomar decisiones informadas, con base en criterios que les permitan evaluar cuándo y cómo integrar tecnologías generativas en sus prácticas formativas. Tal como señala Flórez (2023), los marcos éticos solo resultan efectivos cuando son situados, construidos participativamente e incorporan múltiples voces dentro de la universidad. En esa línea, las orientaciones institucionales no deben imponerse de forma vertical, sino surgir del diálogo deliberativo entre docentes, estudiantes, directivos y demás actores educativos. Solo así se garantizará que los lineamientos institucionales reflejen las identidades y desafíos particulares de cada contexto educativo.

Lejos de representar una amenaza a la docencia, la Inteligencia Artificial Generativa puede ser una oportunidad para redefinir su sentido y potencial. Aunque pueda parecer que estas herramientas

compiten con la inteligencia humana, lo que está verdaderamente en juego es la posibilidad de renovar el lugar del juicio crítico, el discernimiento ético y la creatividad situada en el proceso formativo. Frente a la automatización de tareas cognitivas, la universidad debe asumir con claridad que su misión es formar personas capaces de convivir con tecnologías avanzadas sin perder su agencia. Esto exige desarrollar competencias propiamente humanas como el pensamiento complejo, la empatía, la responsabilidad social y la capacidad de deliberación colectiva. La educación con propósito cobra más relevancia que nunca en contextos profundamente tecnificados. En esta perspectiva, el rol docente se fortalece: no como transmisor de contenidos replicables por una IA, sino como orientador de procesos de aprendizaje que forman criterio, ciudadanía y ética profesional.

En este nuevo ecosistema educativo, los docentes son quienes enfrentan mayores tensiones y responsabilidades. La Inteligencia Artificial Generativa exige revisar profundamente los contenidos curriculares, las formas de enseñar y los mecanismos de evaluación. Además, coloca al profesorado en el centro de nuevas expectativas institucionales relacionadas con la prevención del fraude, la gestión de la autoría y la formación ética del estudiantado. Jardón et al. (2024) destacan que, para responder a estos retos, los docentes deben adquirir nuevas competencias digitales, adaptarse a ritmos de innovación acelerados y sostener una actitud crítica y propositiva. Harper y Prentice (2024) advierten, además, que muchos docentes ya se perciben en la "primera línea" frente a la crisis de confianza provocada por el uso indiscriminado de herramientas de Inteligencia Artificial en contextos evaluativos. Por eso, los lineamientos no pueden quedarse en lo prescriptivo: deben traducirse en estructuras de acompañamiento, formación continua y espacios de deliberación colectiva que permitan a los docentes redefinir su práctica desde el compromiso ético y la responsabilidad educativa.

Además, es necesario considerar que los impactos de la Inteligencia Artificial Generativa varían según las disciplinas. En áreas como la

medicina, el derecho, la psicología o la educación, donde las decisiones humanas son esenciales, esta tecnología redefine herramientas y métodos, y a su vez cuestiona principios éticos fundacionales. En salud, se exploran aplicaciones como la personalización del aprendizaje clínico, la creación de simulaciones médicas y la automatización de diagnósticos, lo cual genera beneficios, pero también plantea retos relacionados con la responsabilidad y el consentimiento informado (Morow et al., 2023; Narayanan et al., 2023). En derecho, la redacción legal asistida y el análisis jurisprudencial automatizado mejoran la eficiencia, aunque requieren marcos normativos sólidos que garanticen transparencia, justicia y no discriminación (Ruiz et al., 2025).

En negocios, la Inteligencia Artificial Generativa potencia la analítica predictiva, la simulación de escenarios y el marketing personalizado, pero también plantea interrogantes sobre la manipulación algorítmica y la autoría (Kaushik, 2022). En diseño, se redefine el prototipado, generando una nueva relación entre creatividad y automatización (Volichenko & Tsurik, 2024). En docencia, se facilita la retroalimentación personalizada y el acompañamiento adaptativo, lo cual exige rediseñar los métodos y criterios de evaluación (UNESCO, 2025). En psicología, se investiga el uso de la IA para apoyar el diagnóstico temprano y el fortalecimiento de habilidades socioemocionales, con importantes consideraciones éticas (Spytska, 2025).

En este sentido, la Inteligencia Artificial Generativa representa una de las mayores disrupciones tecnológicas en el ámbito educativo de las últimas décadas (UNESCO, 2021). Su capacidad para generar texto, imágenes, código o datos sintéticos en segundos ha abierto nuevas posibilidades pedagógicas, pero también ha puesto en tensión valores fundacionales del quehacer académico. ¿Cómo evaluar de manera justa cuando el estudiante puede generar un ensayo en segundos? ¿Dónde termina la asistencia tecnológica y comienza el posible fraude? ¿Cuál es el rol del juicio humano en una era de automatización creativa?

El camino para la construcción de los lineamientos para el uso pedagógico de la Inteligencia Artificial Generativa en la Universidad Icesi

La respuesta de la Universidad Icesi frente a la Inteligencia Artificial Generativa ejemplifica cómo una institución concreta puede desarrollar lineamientos éticos adaptados a su contexto, en sintonía con las tendencias globales, pero atentos a las necesidades locales. En 2024 la Universidad Icesi, en alianza con otras tres universidades colombianas: CESA, EAFIT y Uninorte, elaboró y publicó el *Manifiesto IA for You* (2024), un documento marco que establece orientaciones conceptuales, éticas y operativas para la adopción responsable de la Inteligencia Artificial en la educación superior. Este manifiesto surgió del trabajo conjunto de las unidades de innovación educativa de dichas universidades -en el caso de Icesi, su Decanatura de Innovación Educativa y Fortalecimiento del PEI-, con el propósito de trazar una ruta común que promueva la incorporación crítica de la Inteligencia Artificial Generativa como aliada pedagógica y administrativa.

Lejos de ser una norma cerrada, el manifiesto se concibió como un documento vivo que invita a la reflexión, al debate colectivo y a la retroalimentación continua. Reconoce que la Inteligencia Artificial Generativa representa un fenómeno emergente en constante evolución. Presenta un panorama general de principios orientadores y de los retos éticos asociados a la integración de la Inteligencia Artificial Generativa en la universidad. Se hace énfasis en que esta tecnología es un "complemento" de la labor docente y del proceso de aprendizaje humano. En ese sentido, retoma lineamientos internacionales que abogan por un enfoque centrado en la dignidad y los derechos humanos, y los adapta al contexto colombiano, destacando la importancia de sostener una visión humanista de la educación en medio de la transformación tecnológica.

A nivel interno y en diálogo con el sector externo, la Universidad Icesi ha definido una ruta clara y coherente para enfrentar los desafíos de la Inteligencia Artificial Generativa en la educación superior. Esta

apuesta institucional se articula en dos documentos clave: el Manifiesto sobre inteligencia artificial y educación (2024) y los Lineamientos para el uso pedagógico de la Inteligencia Artificial Generativa (2025). El primero invita a docentes y comunidades académicas a explorar de forma crítica y creativa las implicaciones éticas, epistemológicas y pedagógicas de la Inteligencia Artificial Generativa. Reconoce que estas tecnologías no solo transforman los recursos didácticos, sino que también alteran las condiciones de producción de conocimiento. El Manifiesto propone orientaciones que promueven el uso metacognitivo de la Inteligencia Artificial Generativa, el análisis disciplinar de sus efectos y la deliberación colectiva sobre sus tensiones éticas. En esta línea, la universidad ha decidido no restringir el uso de la Inteligencia Artificial Generativa, sino comprenderla como una oportunidad para fortalecer la agencia del profesorado y de los estudiantes en un contexto cada vez más tecnificado.

Uno de los principales aportes del Manifiesto sobre inteligencia artificial y educación ha sido la incorporación de la Escala de Evaluación de la Inteligencia Artificial (AI Assessment Scale o AIAS), desarrollada por Perkins et al. (2024). Esta escala ofrece cinco niveles de integración de la Inteligencia Artificial Generativa en las evaluaciones, desde su prohibición total hasta su uso creativo y colaborativo entre estudiantes y docentes. La escala permite que cada asignatura adapte sus prácticas evaluativas según sus objetivos formativos y el contexto disciplinar, y fomenta el diálogo ético y transparente sobre el uso de estas herramientas. Por ejemplo, algunos cursos de escritura académica han adoptado el Nivel 3, que autoriza la edición asistida por Inteligencia Artificial Generativa, siempre que el estudiante explicite en un anexo cómo utilizó la herramienta y qué decisiones asumió de forma crítica. La AIAS busca fomentar el pensamiento reflexivo, fortalecer la autoría estudiantil y construir una cultura de integridad adaptada a la era digital.

Los Lineamientos para el uso pedagógico de la Inteligencia Artificial Generativa (2025) amplían y sistematizan los principios esbozados en el Manifiesto. Estos lineamientos establecen distinciones clave

entre producciones con y sin aporte humano significativo, definen criterios de autoría y distribuyen responsabilidades claras entre los actores institucionales. En cursos como Diseño de experiencias de aprendizaje y Ética profesional, el profesorado ha aplicado estrategias didácticas que integran la Inteligencia Artificial Generativa de forma explícita, mediante rúbricas que valoran la transparencia, la autorregulación y la capacidad crítica del estudiante. En lugar de penalizar el uso de estas tecnologías, los lineamientos incentivan acuerdos pedagógicos explícitos, orientados a una comprensión profunda de sus usos, limitaciones y posibilidades. Esta política favorece la formación de profesionales éticos y autónomos, capaces de interactuar con tecnologías complejas de manera estratégica y contextualizada.

La Decanatura de Innovación Educativa y Fortalecimiento del PEI ha desempeñado un rol esencial en este proceso, al liderar espacios de formación, asesoría y consulta para el profesorado. Talleres sobre rediseño de consignas, análisis de sesgos algorítmicos, validación de prompts y formulación de rúbricas han fortalecido las capacidades docentes para integrar la Inteligencia Artificial Generativa sin sacrificar la calidad pedagógica. Estas acciones han situado el juicio profesional como un principio orientador frente al uso de tecnologías emergentes. En lugar de adoptar enfoques restrictivos o centrados en el control, la universidad ha promovido una cultura de confianza, diálogo y autonomía pedagógica. Este modelo se basa en el principio de inteligencia híbrida, que articula capacidades humanas y técnicas para resolver problemas complejos y generar conocimiento significativo (Burger et al., 2023; Järvelä et al., 2025).

La inteligencia híbrida propone una interacción colaborativa entre docentes, estudiantes y sistemas de Inteligencia Artificial Generativa, en la que el aprendizaje no se automatiza, sino que se enriquece mediante nuevas mediaciones. Esta perspectiva trasciende el uso instrumental de la Inteligencia Artificial Generativa como asistente para tareas mecánicas y plantea su incorporación como herramienta cognitiva que amplía procesos reflexivos, creativos y críticos. Desde este

enfoque, la Inteligencia Artificial Generativa no reemplaza la inteligencia humana, sino que extiende sus márgenes de acción, respeta la autonomía y favorece el aprendizaje situado. La Universidad Icesi ha identificado que el valor educativo de la Inteligencia Artificial Generativa no radica en su capacidad de generar contenido, sino en el tipo de preguntas, decisiones y procesos que permite desplegar en entornos formativos bien diseñados y éticamente orientados.

Además, esta apuesta institucional se ha distanciado de modelos que imponen una vigilancia punitiva mediante el uso de detectores automatizados de Inteligencia Artificial Generativa, los cuales presentan limitaciones técnicas y éticas. Como advierten Dwyer y Laird (2024), estos sistemas pueden arrojar falsos positivos o negativos, y ya existen herramientas accesibles para modificar los textos generados por Inteligencia Artificial Generativa y evitar su detección. Ante este escenario, la Universidad Icesi ha optado por estrategias de evaluación centradas en la construcción del conocimiento, la autenticidad del proceso y el diálogo docente-estudiante. Esta perspectiva se alinea con lo señalado por el ICAI (2021), que define la integridad académica como un compromiso colectivo con la honestidad, la justicia y la responsabilidad, más allá del control de la autoría textual.

Finalmente, negar el uso de la Inteligencia Artificial Generativa como si fuera incompatible con la integridad académica resultaría insostenible en un contexto donde las tecnologías avanzan con rapidez y transforman los entornos educativos y laborales (Perkins y Roe, 2024). La Universidad Icesi ha asumido esta realidad como una oportunidad para redefinir sus principios pedagógicos y fortalecer su Proyecto Educativo Institucional. La transición del Manifiesto a los Lineamientos representa más que un cambio normativo: constituye una transformación cultural que busca formar ciudadanos capaces de convivir con la Inteligencia Artificial Generativa de manera crítica, ética y creativa. Esta decisión reconoce la complejidad del presente, valora el juicio profesional y promueve, mediante la participación deliberativa, nuevas formas de enseñar, aprender y evaluar en la era de la inteligencia híbrida.

Conclusiones

La integración de la Inteligencia Artificial Generativa en la educación superior no puede entenderse únicamente como una innovación tecnológica, es una transformación cultural que interpela los fundamentos éticos, pedagógicos y epistemológicos de las universidades. Este capítulo ha mostrado que el reto principal no radica en prohibir o permitir su uso, es fundamental construir marcos orientadores capaces de acompañar a docentes y estudiantes en la toma de decisiones conscientes, informadas y contextualizadas. La ética del uso de la Inteligencia Artificial Generativa se pude abordar mediante prácticas situadas que promueven la agencia compartida, el discernimiento crítico y la responsabilidad pedagógica. Las experiencias analizadas muestran que una gobernanza ética requiere diálogo, flexibilidad y coherencia con los propósitos institucionales.

La Universidad Icesi ha optado por asumir este reto desde una postura deliberativa, plural y pedagógica, centrada en el fortalecimiento del juicio humano. Esta decisión implica una apuesta por comprenderla en profundidad, delimitar sus usos estratégicos y promover su integración crítica en los procesos de enseñanza, aprendizaje y evaluación. La construcción del Manifiesto sobre inteligencia artificial y educación y de los Lineamientos para el uso pedagógico de la Inteligencia Artificial Generativa constituye un ejercicio de coherencia institucional, en el que los principios formativos orientan la acción docente, la participación estudiantil y la gestión académica.

Los riegos y tensiones éticas en el uso de la IAG abordadas a lo largo del capítulo son oportunidades para revisar, actualizar y enriquecer las prácticas formativas. La Inteligencia Artificial Generativa plantea preguntas nuevas sobre la producción del conocimiento, la validez de los procesos evaluativos y el sentido mismo de aprender en comunidad. Estas preguntas invitan a las universidades a ejercer un liderazgo transformador, que combine el rigor técnico con la sensibilidad ética y que reconozca en la inteligencia híbrida una posibilidad para potenciar capacidades humanas, no para sustituirlas. La formación de

profesionales éticos, creativos y críticos depende en gran medida de cómo se aborde esta transición tecnológica y cultural.

En este contexto los lineamientos institucionales deben permanecer abiertos al diálogo y atentos a las voces de quienes habitan la universidad y de los cambios sociales y tecnológicos en el contexto local e internacional. Bajo este panorama, los procesos colaborativos se convierten en una alternativa para construir orientaciones que respondan a las necesidades reales de los distintos programas académicos y a los desafíos de un mundo cada vez más digital e interconectado. Lejos de cerrar el debate este capítulo propone abrirlo hacia nuevas preguntas, nuevas prácticas y nuevas formas de entender la educación superior en la era de la Inteligencia Artificial Generativa. El futuro de las universidades se jugará en la capacidad del sistema educativo colombiano para sostener comunidades educativas reflexivas, críticas, creativas y éticamente comprometidas con el presente.

Referencias

Alianza 4U. (2024). *Manifiesto IA for you*. https://www.icesi.edu.co/7583-manifiesto-sobre-inteligencia-artificial/

An, Y., Yu, J. H. & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, 22(10). https://doi.org/10.1186/s41239-025-00507-3

Barrientos-Hernán, E. J., López-Pastor, V. M., & Pérez-Brunicardi, D. (2020). Evaluación Auténtica y Evaluación Orientada al Aprendizaje en Educación Superior. Una Revisión en Bases de Datos Internacionales. *Revista Iberoamericana De Evaluación Educativa*, 13(2), 67–83. https://doi.org/10.15366/riee2020.13.2.004

Bastani, Hamsa and Bastani, Osbert and Sungu, Alp and Ge, Haosen

and Kabakcı, Özge and Mariman, Rei. (2024) *Generative AI Can Harm Learning*. The Wharton School Research Paper, Available at SSRN: https://ssrn.com/abstract=4895486 or http://dx.doi.org/10.2139/ssrn.4895486

Burger, M., Nitsche, A.-M., & Arlinghaus, J. (2023). Hybrid intelligence in procurement: Disillusionment with AI's superiority? *Computers in Industry*, 150, https://doi.org/10.1016/j.compind.2023.103946

Carrera, P. (2024). ¿Singularidad? Limitaciones, capacidades y diferencias de la inteligencia artificial frente a la inteligencia humana. *Claridades*, 16(2), 159-186.

Chan, K. Y. C. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(38), 1-25. https://doi.org/10.1186/s41239-023-00408-3

Chicaiza Guayta, S. M., López Bermúdez, F. L., López Valencia, N. A. y Ochoa Tumbaco, G. X. (2024). Impacto de la Inteligencia Artificial (IA) en la Educación Superior. *RECIAMUC*, 8(2), 80-91. https://doi.org/10.26820/reciamuc/8.(2).abril.2024.80-91

Consejo Nacional de Política Económica y Social. (14 de febrero de 2025). *Política Nacional de Inteligencia Artificial*. Departamento Nacional de Planeación.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, 20(22), 1-22. https://doi.org/10.1186/s41239-023-00392-8

Decanatura de Innovación Educativa y Fortalecimiento del PEI. (2025). Análisis de lineamientos IA [Documento interno de la Universidad Icesi].

Dwyer, M. & Laird, E. (2024). Up in the Air. Educators Juggling the Potential of Generative AI with Detection, Discipline, and Distrust. Center for Democracy & Technology.

European University Institute. (2024). *EUI Guidelines for the Responsible Use of Artificial Intelligence for Research*. https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/2024.06-Ethics-Commitee-EUI-GENAI-DIGITAL.pdf

Ferrara, E. (2023). Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies. *Sci*, 6(3), 1-15. https://doi.org/10.3390/sci6010003

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Flórez Rojas, M. L. (2023). Pensamiento de diseño y marcos éticos para la Inteligencia Artificial: una mirada a la participación de las múltiples partes interesadas. *Desafíos*, 35(1), 1-31. https://doi.org/10.12804/revistas.urosario.edu.co/desafios/a.12183

Harper, R., & Prentice, F. (2024). Responsible but powerless: staff qualitative perspectives on cheating in higher education. *International Journal for Educational Integrity*, 20(25), 1-17. https://doi.org/10.1007/s40979-024-00170-7

Harvard University. (s.f.). *Initial guidelines for the use of Generative AI tools at Harvard*. https://www.huit.harvard.edu/ai/guidelines

Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado. (2024). *Guía sobre el uso de la inteligencia artificial en el ámbito educativo*. https://code.intef.es/wp-content/uploads/2024/07/Gu%C3%ADa-sobre-el-uso-de-la-IA-en-el-%C3%A1mbito-educativo-INTEF_2024.pdf

Instituto Tecnológico y de Estudios Superiores de Monterrey. (s.f.). *Guía rápida: lineamientos para el uso ético de inteligencia artificial*. https://tec.mx/

sites/default/files/repositorio/integridad-academica/guia-rapida-alumnos-inteligencia-artificial.pdf?srsltid=AfmBOopBZ9k7usO6oXv8R3wKiqDe9SwR5-83YkfqBp7KRkTtsDg1DSmX

International Center for Academic Integrity. (2021). *The fundamental values of academic integrity* (3rd Ed.). ICAI. https://academicintegrity.org/aws/ICAI/asset_manager/get_file/911282?ver=1

Jardón, M., Allas, W., Zamora, D., y Cedeño, N., (2024). Impacto de la inteligencia artificial en la educación superior: percepciones de alumnos y profesores sobre el uso de IA en el aprendizaje y la evaluación. *Reincisol*, 3(6), 7008-7033. https://doi.org/10.59282/reincisol.V3(6)7008-7033

Järvelä, S., Zhao, G., Nguyen, A., & Chen, H.(2025). Hybrid intelligence: Human–AI coevolution and learning. *British Journal of Educational Technology*, 56, 455-468. https://doi.org/10.1111/bjet.13560

Kaushik, P. (2022). Role and Application of Artificial Intelligence in Business Analytics: A Critical Evaluation. *International Journal for Global Academic & Scientific Research*, 1(3), 01–09. https://doi.org/10.55938/ijgasr.v1i3.15

Martínez González, M. A. (2023). Uso responsable de la inteligencia artificial en estudiantes universitarios: Una mirada tecnoética. *Revista Boletín Redipe*, 12(9), 172-178.

Molina, E., Cobo C., Pineda J. y Rovner H. (2024). *La revolución de la IA en Educación: Lo que hay que saber. Innovaciones Digitales de Educación*. Banco Mundial.

Mollick, E. [Global Silicon Valley]. (16 de abril de 2024). *Co-Intelligence: AI in the Classroom with Ethan Mollick | ASU+GSV 2024* [Video]. YouTube. https://www.youtube.com/watch?v=8FnOkxj0ZuA&t=279s

Moncada, A. (3 de marzo de 2025). *Colombia y el desarrollo de la Inteligencia Artificial*. Colegio de Estudios Superiores de Administra-

ción. https://www.cesa.edu.co/news/colombia-y-el-desarrollo-de-la-inteligencia-artificial/

Morrow, E., Zidaru, T., Ross, F., Mason, C., Patel, K. D., Ream, M., & Stockley, R. (2023). Artificial intelligence technologies and compassion in healthcare: A systematic scoping review. *Frontiers in psychology*, 13, 1-31. https://doi.org/10.3389/fpsyg.2022.971044

Narayanan, S., Ramakrishnan, R., Durairaj, E., & Das, A. (2023). Artificial Intelligence Revolutionizing the Field of Medical Education. *Cureus*, 15(11), 1-9. https://doi.org/10.7759/cureus.49604

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2025). *Marco de competencias para docentes en materia de IA*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000393813/PDF/393813spa.pdf.multi

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2023). *La IA generativa y el futuro de la educación*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385877_spa/PDF/385877spa.pdf.multi

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2023). *Guía para el uso de IA generativa en educación e investigación*. UNESCO https://unesdoc.unesco.org/ark:/48223/pf0000389227/PDF/389227spa.pdf.multi

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2021). *Artificial intelligence in education: Challenges and opportunities for sustainable development*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000366994

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6). https://doi.org/10.53761/q3azde36

Perkins, M., & Roe, J. (2024). Academic publisher guidelines on AI usage: A ChatGPT supported thematic analysis. *F1000Research*, 12(1398), 1-34. https://doi.org/10.12688/f1000research.142411.2

Ruiz-Muñoz, G. F., Paz-Zamora, Y. E., Morales-Loor, K. P., & Narváez-Vega, E. A. (2025). El impacto de la inteligencia artificial en la enseñanza del derecho: The impact of artificial intelligence on law education. *Multidisciplinary Latin American Journal (MLAJ)*, 3(1), 468-491. https://doi.org/10.62131/MLAJ-V3-N1-023

Spytska, L. (2025). The use of artificial intelligence in psychotherapy: development of intelligent therapeutic systems. *BMC Psychology*, 13(175), 1-12. https://doi.org/10.1186/s40359-025-02491-9

Tejedor-Estupiñán, J. M. (2025). Impacto de la inteligencia artificial en la producción, la productividad y el aumento de la desigualdad en América Latina. *Revista Finanzas Y Política Económica*, 17, 1-3. https://doi.org/10.14718/revfinanzpolitecon.v17.2025.0

Universidad Icesi. (2025). *Lineamientos para el uso pedagógico de la Inteligencia Artificial Generativa*. Decanatura de Innovación Educativa y Fortalecimiento del PEI.

Universidad Icesi. (2025). *Libro de Derechos, Deberes y Normas de los estudiantes de pregrado*. https://www.icesi.edu.co/wp-content/uploads/2025/06/Libro-Derechos-Deberes-y-Normas-Pregrado-2025.pdf

Universidad Icesi. (2024). *Manifiesto sobre inteligencia artificial y educación*. Decanatura de Innovación Educativa y Fortalecimiento del PEI.

Universidad Icesi. (2017). *Proyecto Educativo Institucional*. https://repository.icesi.edu.co/server/api/core/bitstreams/e2407df6-94f8-4922-a7d1-843fc7c8f78e/content

Universidad Externado de Colombia. (2024). *Lineamientos para el Uso de Inteligencia Artificial*. https://www.uexternado.edu.co/wp-content/uploads/2024/10/Lineamientos-para-el-uso-de-IA-1.pdf

Universidad de los Andes. (2024). *Lineamientos para el uso de inteligencia artificial generativa (IAG) en la Universidad de los Andes*. https://secretariageneral.uniandes.edu.co/images/documents/lineamientos-uso-inteligencia-artificial-generativa-IAG-uniandes.pdf

Universidad Peruana de Ciencias Aplicadas. (2024). *Lineamientos para el uso de herramientas de inteligencia artificial generativa en el proceso de enseñanza-aprendizaje*. https://innovacioneducativa.upc.edu.pe/wp-content/uploads/2025/02/LINEAMIENTOS_PARA_EL_USO_DE_HERRAMIENTAS_DE_INTELIGENCIA_ARTIFICIAL_GENERATIVA_EN_EL_PROCESO_DE_ENSENANZA_%E2%80%93_APRENDIZAJE_Y_ASSESSMENT.pdf

Universidad Andrés Bello. (2024). *Lineamiento para el uso responsable de inteligencia artificial en la Universidad Andrés Bello*. https://ia.unab.cl/wp-content/uploads/2024/10/IA-Lineamiento-para-el-uso-responsable-de-inteligencia-artificial-en-UNAB-20240926-002.pdf

Universitat Oberta de Catalunya. (2023). *Repensar la inteligencia para repensarnos: La irrupción de las IA generativas nos da la oportunidad de redefinir qué es la inteligencia y cómo queremos que sea. Lección inaugural* 2023–2024. https://www.uoc.edu/portal/_resources/CA/documents/la_universitat/llicons-inaugurals/Llico-inaugural-UOC-2023-24-ES.pdf

University of Cambridge. (s.f.). *Using Generative* AI. https://blendedlearning.cam.ac.uk/artificial-intelligence-and-education/using-generative-ai

Volichenko, O., & Tsurik, T. (2024). Artificial intelligence in conceptual prototyping design. *Проект Байкал*, 79, 38-44. https://doi.org/10.51461/issn.2309-3072/77.2281

[Capítulo 7]

¿Cómo abordar las implicaciones éticas en el uso de la IA?: una reflexión pedagógica

Isabel Tejada Sánchez

La inteligencia artificial (IA) ha dejado de ser una promesa futurista para incrustarse en el tejido de nuestra vida diaria. Según los principios enunciados por Mollick (2024), hemos entrado en una era donde "siempre tendremos que invitar a la IA a la mesa". La IA se ha convertido en una omnipresencia que viene de mucho más atrás que el surgimiento de ChatGPT y que ha sigilosa pero contundentemente permeado nuestra cotidianidad: desde los algoritmos que curan las actualizaciones o *feed* de nuestras redes sociales, proveen recomendaciones sobre los videos que vemos y la música que escuchamos, la forma en que viajamos hasta los sistemas que asisten en diagnósticos médicos, la IA ha estado moldeando nuestras interacciones, decisiones, acciones y comprensiones del mundo desde hace un tiempo hasta que ya se hizo totalmente evidente y por ende se han convertido en sistemas en los que hemos aprendido a confiar, como cajas negras que no conocemos plenamente cómo funcionan (Nowotny, 2023).

Frente a este panorama, este capítulo tiene como propósito profundizar sobre cómo abordar en el aula los desafíos que plantea la incorporación de la IA. A través del análisis de una propuesta didác-

tica concreta, el dispositivo didáctico DilemAI, se busca articular una reflexión que conecte los dilemas éticos más relevantes con estrategias pedagógicas participativas, contextualizadas y orientadas a promover la acción y el desarrollo de la agencia en sus participantes.

La educación tiene la responsabilidad crucial de equipar a sus participantes (estudiantes, docentes, administrativos, comunidades) no solo con el conocimiento de qué es la IA, sino también con una comprensión situada, sustancial y amplia de sus complejas implicaciones éticas, sociales y prácticas. Este artículo corto desarrolla una reflexión pedagógica y explora la construcción y puesta en práctica de un dispositivo didáctico, el juego DilemAI, como un caso para reflexionar de forma emergente sobre estas implicaciones en el aula.

Organismos como la UNESCO y el Banco Mundial ha reconocido la urgencia de la *alfabetización* en IA (AI literacy), definiéndola como la capacidad de entender cómo la IA recopila, manipula y analiza datos, así como la forma en que los algoritmos encuentran patrones (Molina et al. 2024; Cukurova & Miao, 2024). Esta alfabetización va más allá del conocimiento técnico[1]; abarca la comprensión crítica de las implicaciones éticas que plantea la IA y el uso de la tecnología en el desarrollo humano, en sus habilidades cognitivas, su relación con el mundo y en sociedad. Como señalan informes de la UNESCO (Miao, et al., 2021; UNESCO, s.f.), es esencial que los ciudadanos comprendan el impacto de la IA, qué puede y qué no puede hacer, cuándo es útil y cuándo debe cuestionarse su uso, y cómo podría orientarse hacia el bien público. Sin embargo, integrar esta comprensión en el currículo requiere enfoques pedagógicos que vayan más allá de la mera transmisión de información. Esta es una discusión sobre la ética para la cual muchos actores educativos no hemos sido preparados y debemos asumir con urgencia.

1 Un aspecto de interés frente al concepto de alfabetización versus el concepto de literacidad promovido en castellano por autores como Daniel Cassany (2005, inter alia), desde quienes justamente se habla de una habilidad más profunda de interpretar, comprender y llevar el significado más allá.

Sobre las implicaciones éticas en el uso de la IA

Son diversas las implicaciones éticas que presenta la irrupción de la IA y esto supone una tarea compleja para el campo educativo. Entre las principales preocupaciones se encuentran las violaciones a la privacidad, el sesgo algorítmico, la falta de transparencia y las amenazas a la agencia humana en el aprendizaje. Los sistemas de IA suelen recopilar grandes cantidades de datos de sus usuarios, lo que genera inquietudes sobre cómo se almacenan y utilizan estos datos (Maphalala & Ajani, 2025; Mishara, 2024). Asimismo, se ha documentado que los sesgos presentes en los algoritmos pueden perpetuar desigualdades existentes, afectando de manera desproporcionada a grupos marginados (Mohamed, 2024). Esta problemática ha sido evidenciada en diversas investigaciones, como la realizada por Buolamwini (2023), quien demostró cómo sistemas de reconocimiento facial o bien no logran identificar o clasifican erróneamente a mujeres con rasgos afrodescendientes referenciándolas como pertenecientes al género masculino.

De igual forma, la opacidad en los procesos de toma de decisiones de la IA dificulta la rendición de cuentas (*accountability*) en áreas críticas como la evaluación (Kılınç, 2024). A esto se le suma una dependencia excesiva de la IA que puede disminuir aspectos humanos esenciales en la educación y en el relacionamiento social, como la empatía y el apoyo emocional que ofrecen los docentes y comunidades educativas (Isop, 2025; Marais et al., 2024), y ampliar la brecha digital en detrimento de estudiantes de comunidades desatendidas (Zhang, 2024). En paralelo, persisten numerosas preocupaciones en torno a la integridad académica y a los posibles impactos psicológicos y cognitivos derivados de una dependencia excesiva de la IA. A esto se añade que ciertas prácticas tradicionales y preexistentes en contextos educativos han favorecido el uso de detectores de IA como parte de los procedimientos de evaluación, lo cual ha incentivado a algunos estudiantes a utilizar herramientas de IA para completar sus tareas y, posteriormente, recurrir a sistemas que "humanizan" las producciones con el fin de

ocultar su origen automatizado (estos conocidos como "*AI bypass*"). Esta dinámica puede surgir como efecto de prohibiciones institucionales, prácticas sociales o escolares que penalizan el uso de la IA, o bien por la vergüenza de admitir su utilización.

Estas implicaciones éticas afectan directamente principios fundamentales que orientan la práctica educativa. La equidad se ve comprometida cuando los sistemas de IA reproducen sesgos que impactan de forma desproporcionada a comunidades históricamente marginadas, lo cual puede evidenciarse rápidamente en la creación de imágenes que, por defecto, presentan representaciones hegemónicas de los seres humanos, como personas blancas, jóvenes, delgadas, etc. La transparencia se diluye cuando la información disponible sobre los datos y procedimientos utilizados en su entrenamiento es parcial, ambigua o contradictoria (Heikkilä, 2024). Asimismo, la rendición de cuentas se vuelve problemática cuando no está claro quién debe asumir la responsabilidad frente a errores, sesgos o consecuencias negativas derivadas de sistemas desarrollados y respaldados por una industria multimillonaria, caracterizada por fuertes mecanismos de protección comercial (Informe de Grand View Research, s. f.; Statista, s. f.). Estos principios, pilares de una educación justa y democrática, requieren ser revisados y resignificados en contextos donde la tecnología redefine los roles, las relaciones y los procesos educativos.

Ante estos desafíos, diversos marcos éticos y estrategias han sido propuestas para guiar una integración responsable de la IA en la educación. Iniciativas como el marco basado en Lenguaje de Modelado Unificado (UML) (Isop, 2025), los Principios de IA Segura en la Educación (García-Peñalvo et al., 2024) y el Marco Integral de Evaluación de la IA (CAIAF) (Kılınç, 2024) enfatizan la transparencia, la equidad y la responsabilidad. De manera similar, el Plan de Acción de Educación Digital de la Unión Europea aboga por preservar la agencia humana junto al avance tecnológico (Morandín-Ahuerma, 2024).

¿Cómo actuar desde el saber práctico y pedagógico?

La incorporación de la IA en el ámbito educativo debe ir necesariamente acompañada de procesos formativos en ética y en el uso consciente de estas tecnologías por parte de todos los actores educativos y las comunidades que conforman. Casos que son preocupantes de cara a lo que enfrentaremos próximamente, como los *deepfakes*, representan un desafío importante debido a su capacidad para generar desinformación y manipular audiencias, con impactos potencialmente graves tanto en entornos educativos como en otros contextos (Ray, 2024). Asimismo, se hace imperante que en la práctica pedagógica se discuta las implicaciones de los sesgos que suelen arrastrar los algoritmos de IA generativa de forma inherente y derivados de los datos con los cuales fueron entrenados, reproduciendo prejuicios y contribuyendo a la perpetuación de desigualdades sociales.

Ante esta situación, la formación ética en IA debe abarcar el reconocimiento y la mitigación de sesgos, el análisis crítico de los riesgos asociados al contenido que se crea las consecuencias de su uso (en la dimensión social, ambiental, económica), y la adopción de prácticas responsables y reflexivas en el uso de estas herramientas. Es fundamental formar a los docentes en ética de la IA y, a su vez, capacitar a los estudiantes, promover una literacidad digital transversal, garantizar un acceso equitativo a las herramientas de IA y establecer directrices sólidas que protejan los datos y minimicen la propagación de sesgos (Zaidy, 2024; Kölemen, 2024). Asimismo, resulta esencial visibilizar que la construcción de sesgos es un proceso performativo, profundamente enraizado en patrones de pensamiento y conductas sistémicas. Estos sesgos no se eliminan simplemente mediante instrucciones a las máquinas; persisten en la medida en que se mantienen en el imaginario colectivo y en las prácticas sociales (Drage & Frabetti, 2024). Esta iniciativa va de la mano con la promoción de la colaboración global para formular políticas sensibles a los contextos culturales y alineadas con valores éticos. Solo mediante este

compromiso será posible aprovechar el potencial transformador de la IA, protegiendo al mismo tiempo la equidad, la integridad y los valores humanos esenciales para la educación que esté centrada en el ser humano (Miao & Tawil, 2024).

Estas tensiones éticas no son meras abstracciones conceptuales; forman parte de la cotidianidad del aula. Por ello, es crucial generar estrategias que vayan más allá de informar, que movilicen a los actores educativos a explorar estas problemáticas en situaciones concretas y reconozcan cómo estas implicaciones atraviesan sus vidas y decisiones.

Una reflexión pedagógica desde la necesidad de una ética ubicua en tiempos de IA

En el ejercicio docente que he desempeñado como formadora de IA, para docentes, estudiantes universitarios, especialistas en educación y de diversas disciplinas, he encontrado que más allá de la tarea de formar en los aspectos fundacionales y técnicos de la IA generativa, como lo son su funcionamiento, entrenamiento, uso y procesamiento de datos, así como la interacción mediante prompts y workflows, los aspectos éticos han sido sustanciales y trascendentales en el currículo que he construido. Concibo que ninguna formación en IA debería omitir esta reflexión dada su relevancia en una era donde nociones como la propiedad intelectual, la autoría, y la atribución de lo que se crea se ven diluidas por la potencia de estos sistemas. Adicional a esto, se suman otras implicaciones serias como lo son los riesgos asociados a la seguridad, la trascendencia e incuestionable información sesgada y no consentida y el impacto ambiental.

En el marco de las experiencias pedagógicas que he liderado, la reflexión sobre la ética emerge como un tema inquietante y, en muchos casos, se convierte en un estresor que exacerba la angustia y refuerza perspectivas tecno-pesimistas, alimentadas por relatos distópicos sobre el rol de la tecnología en la evolución de la

humanidad. Ante este panorama, la necesidad de una ética ubicua en el currículo se hace palpable y con ella una necesidad pedagógica sobre cómo articularla

Como génesis de esta reflexión del saber práctico, fue claro para mí, al liderar estos espacios, que es posible fortalecer este proceso en el aula a través de propuestas altamente participativas e interactivas, que inviten a todos a involucrarse y profundizar desde diversas emociones, interpretaciones reacciones y posibilidades de acción.

Descripción del dispositivo didáctico DilemAI

Desde esa intención, el dispositivo DilemAI se presenta como una respuesta didáctica que busca abordar estas reflexiones a partir de experiencias significativas para el aula, vinculando los desafíos éticos discutidos con la capacidad de toma de decisiones.

Este juego de mesa en formato digital[2] fue diseñado para explorar dilemas éticos específicos que plantea la IA y ofrecer una experiencia de aprendizaje activa y colaborativa. A través de la asunción de roles, el lanzamiento de dados y el abordaje de situaciones problemáticas (como la pérdida de habilidades cognitivas, el impacto ambiental, los sesgos algorítmicos, la autoría en IA y la vigilancia), sus participantes se ven impulsados a reflexionar críticamente sobre las implicaciones del uso de estas tecnologías.

Desde su diseño, procuré que el valor pedagógico de DilemAI radicara en su capacidad para traducir conceptos abstractos sobre la ética

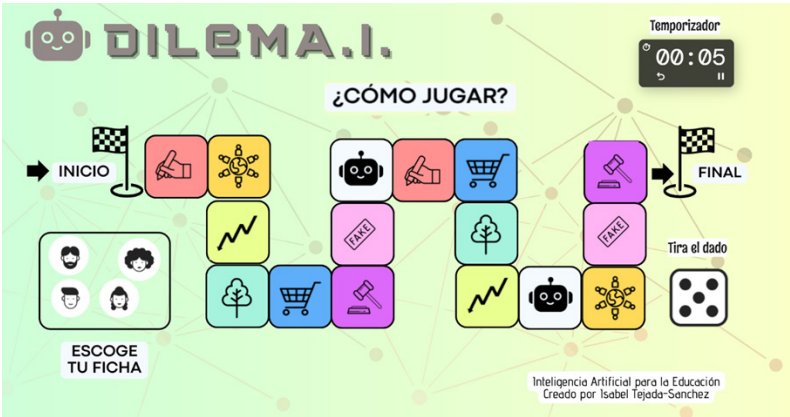
2 DilemAI se encuentra disponible en diferentes formatos:
- Versión web: https://swiy.co/dilemaismall
- Versión juego de mesa en español: https://view.genially.com/66456eab40a7ff0014cfa7e8/interactive-content-dilemai
- Versión web en inglés: https://swiy.co/dilemairulesEng
- Versión juego de mesa en inglés: https://view.genially.com/66fac741ec7f26cd21d202d1/interactive-content-dilemai-english

de la IA en escenarios concretos y relevantes, facilitando conexiones significativas con el aprendizaje y el debate en el aula. La concepción de DilemAI se fundamenta en los cinco pilares del aprendizaje basado en el juego definidos por la Fundación LEGO (Mardell et al., 2023): ser significativo, socialmente interactivo, activo, iterativo y alegre. Cada pilar fue integrado deliberadamente en la dinámica del juego. Así, se busca que los participantes:

- relacionen sus conocimientos previos con nuevas experiencias (significativo),
- colaboren y discutan en equipo (socialmente interactivo),
- participen en actividades prácticas como retos de un minuto, mini-búsquedas y análisis de casos (activo),
- tengan oportunidades de experimentar, cometer errores y reflexionar sobre ellos (iterativo),
- y encuentren disfrute en una experiencia lúdica que reduzca la ansiedad y fomente una actitud positiva hacia el aprendizaje (alegre), especialmente frente a un tema que genera estrés e incertidumbre.

En cuanto a su funcionamiento, DilemAI (ilustrado en la figura 1) presenta una serie de dilemas organizados en siete grandes categorías éticas que plantea el uso de la IA: pérdida de habilidades cognitivas y transformación del relacionamiento social; impacto ambiental; sesgos en la IA; copiar-pegar y mercado de prompts; automatización y deshumanización; detección de IA, vigilancia y fraude; y citación, atribución y autoría.

Figura 1. Tablero de DilemAI en versión interactiva



Fuente: Elaboración propia.

Cuando un jugador cae en una casilla de dilema-tema, debe realizar una de las dos acciones propuestas: Retos de un minuto, que requieren respuestas rápidas sobre un aspecto del dilema; Mini-búsquedas, que implican indagar brevemente sobre un tema; Análisis de caso, donde deben evaluar situaciones hipotéticas relacionadas con el dilema; o *Prompt tinkering*, donde experimentan con herramientas de IA, como crear imágenes y analizarlas en busca de sesgos. Este último rasgo fue influenciado por la propuesta de Mollick y Mollick (2024) sobre la importancia de decantar las expectativas y roles que la IA puede asumir a la hora de incorporarlas en los procesos de enseñanza y aprendizaje. Finalmente, existe una casilla especial de acción colectiva, donde todos los participantes reflexionan en conjunto sobre temas complejos como la autoría en la era de la IA generativa.

Como ejemplo, el dilema sobre los sesgos en la IA, que invita a crear una imagen con IA y analizar sus resultados, ilustra de manera práctica cómo los datos y algoritmos pueden perpetuar desigualdades y percepciones estandarizadas sobre un rol, una profesión, el género, o la racialización. Esta experiencia tangible complementa la comprensión

teórica de los sesgos algorítmicos, que ya forma parte de los resultados de aprendizaje en currículos sobre IA. De manera similar, el desafío relacionado con la citación y autoría de contenidos generados por IA introduce a los participantes a un nuevo conjunto de consideraciones éticas fundamentales en la era de la inteligencia artificial generativa.

De esta manera, DilemAI se alinea con pedagogías que fomentan la resolución de problemas reales mediante actividades interdisciplinarias, donde el aprendizaje se produce a partir de la reflexión crítica y el involucramiento activo. Además, responde a los llamados de organismos internacionales, como la UNESCO y la OECD (2025), que subrayan la necesidad urgente de que los ciudadanos comprendan el impacto de la IA y sepan cuándo cuestionar su uso para orientarlo hacia el bien público.

Discusión y conclusión

Enseñar la ética en torno al uso de la IA de manera seria y lúdica a través de estrategias como DilemAI es un aporte a cómo abordamos la literacidad crítica que demanda la era digital. La Fundación LEGO ha demostrado que el aprendizaje significativo se potencia cuando es socialmente interactivo, activo, iterativo y emocionalmente positivo; iniciativas como DilemAI recogen estos principios al transformar dilemas éticos en experiencias tangibles que invitan a la reflexión colaborativa. Esta aproximación facilita la apropiación de conceptos complejos como el sesgo algorítmico o la atribución en IA generativa, y también abre espacios emocionales seguros y necesarios para confrontar temores, incertidumbres y resistencias frente a la tecnología. Así, busca superar el riesgo de una educación ética meramente teórica, normativa o punitiva, apostando en cambio por una ética vivida, discutida y comprendida en comunidad, anclada en contextos reales y en problemas vigentes. Esto se alinea con la propuesta de investigaciones emergentes en torno al abordaje de la IA en el aula, partiendo desde enfoques como el conectivismo y el constructivismo (Maphalala & Ajani, 2025).

Sin embargo, iniciativas como DilemAI, aunque valiosas, no resuelven ni cubren todos los desafíos que implica formar en ética de la IA, tal y como lo evidencian García-Peñalvo, et al. (2024) en su manifiesto sobre una IA segura y transparente. El alcance de cualquier dispositivo didáctico puede verse limitado si no se integra en un currículo transversal, sistemático y sostenible, en lugar de tratarse como una actividad aislada. Existe también el riesgo de que la dimensión lúdica sea percibida como un tratamiento superficial de dilemas que requieren profundidad teórica, contextualización y confrontación crítica continua, tal y como lo sugieren Drage y Frabetti (2024), al analizar casos como la estigmatización de la protesta reforzada por los sistemas de IA. Asimismo, los prejuicios y carencias en la formación de los propios docentes podrían obstaculizar un diálogo genuino, especialmente en contextos donde prevalece o bien una confianza ciega en la tecnología o su total opuesto. Por tanto, el éxito de propuestas como esta depende de fortalecer la formación ética y crítica de los docentes, acompañar pedagógicamente las experiencias y construir culturas institucionales que reconozcan el valor estratégico de una brújula ética calibrada frente a la revolución que la IA ya representa en nuestras vidas.

La implementación de DilemAI en el aula puede realizarse de diversas maneras, alineándose con los modelos de integración curricular identificados por la UNESCO. Puede ser utilizado como actividad autónoma para introducir o reforzar conceptos éticos, o integrarse en proyectos interdisciplinarios que exploren las implicaciones de la IA en diferentes áreas. Por ejemplo, en una clase de lenguaje, los participantes podrían analizar cómo la IA generativa afecta la creatividad y la autoría; en ciencias sociales, investigar el impacto de la automatización laboral planteado en uno de los dilemas. Esta flexibilidad permite adaptar el juego a distintos contextos y objetivos educativos.

No obstante, la implementación efectiva de herramientas como DilemAI requiere un rol activo del docente como facilitador de la discusión y guía del aprendizaje. Los docentes deben estar preparados para

ayudar a los estudiantes a conectar las experiencias del juego con principios éticos más amplios y fomentar una reflexión crítica sobre las soluciones posibles a los dilemas planteados. Esto subraya la necesidad de una formación docente específica en ética de la IA y pedagogías innovadoras. Como señala la UNESCO, dicha formación es esencial para la integración curricular de la IA (Cukurova & Miao, 2024).

Más allá de DilemAI, otras estrategias pedagógicas pueden complementar la enseñanza de las implicaciones éticas del uso de la IA. Entre ellas se encuentran el análisis de estudios de caso sobre su impacto social, la creación de prototipos simulados de sistemas de IA que consideran cómo se alimentan e interpretan datos, y la participación en debates sobre regulaciones de la IA. Es crucial que los currículos fomenten el pensamiento crítico y la capacidad de los estudiantes para cuestionar la inteligencia y la confiabilidad de estas tecnologías.

Para finalizar, enseñar las implicaciones del uso de la IA en el aula es esencial para preparar a los actores educativos para el futuro que ya estamos construyendo, cada vez más impulsado por esta tecnología. Herramientas como DilemAI ofrecen una vía cautivadora, interactiva y basada en problemas para explorar las complejas dimensiones éticas y sociales de la IA. Al conectar dilemas específicos con conceptos más amplios y fomentar la reflexión crítica, experiencias como esta permiten que docentes y estudiantes construyan un criterio informado y responsable sobre la IA. Sin duda, el éxito de estas iniciativas depende del compromiso continuo con el desarrollo de recursos educativos pertinentes y de la capacitación docente para facilitar estas conversaciones fundamentales en el aula.

Declaración: DilemAI fue concebido y creado por la autora en 2024 y cuenta con una licencia de Creative Commons. Para estructurar su contenido y dinamizar las actividades, se utilizó ChatGPT-4o bajo el rol de asistente. El diseño gráfico se realizó en Canva y Genially, sin emplear funciones de IA integradas. Este texto fue editado y estructurado con apoyo de ChatGPT-4o y NotebookLM. La búsqueda bibliográfica se llevó a cabo con ayuda de SciSpace.

Referencias

Buolamwini, J. (2023). *Unmasking AI: My Mission to Protect What is Human in a World of Machines*. Random House.

Cukurova, M., & Miao, F. (2024). *AI competency framework for teachers*. UNESCO Publishing.

Drage, E., & Frabetti, F. (2024). The performativity of AI-powered event detection: How AI creates a racialized protest and why looking for bias is not a solution. *Science, Technology, & Human Values*, 49(5), 1045-1072.

García-Peñalvo, F. J., Alier, M., Pereira, J., & Casany Guerrero, M. J. (2024). Safe, Transparent, and Ethical Artificial Intelligence. *IJERI: International Journal of Educational Research and Innovation*, 22, 1–21. https://doi.org/10.46661/ijeri.11036

Grand View Research. (s.f.). Artificial Intelligence (AI) Market Analysis. *Grand View Research*. Disponible en: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market

Heikkilä, M. (2024, marzo). Nobody knows how AI works. *MIT Technology Review*. https://www.technologyreview.com/2024/03/05/1089449/nobody-knows-how-ai-works/

Isop, W. A. (2025). A conceptual ethical framework to preserve natural human presence in the use of AI systems in education. *Frontiers in Artificial Intelligence*, 7. https://doi.org/10.3389/frai.2024.1377938

Kılınç, S. (2024). Comprehensive AI Assessment Framework: Enhancing Educational Evaluation with Ethical AI Integration. *Journal of Educational Technology and Online Learning*. https://doi.org/10.31681/jetol.1492695

Maphalala, M. C., & Ajani, O. A. (2025). Leveraging artificial intelligence as a learning tool in higher education. *Interdisciplinary Jour-*

nal of Education Research, 7(1), a01. https://doi.org/10.38140/ijer-2025.vol7.1.01

Marais, E., Marais-Botha, R., & Coertzen, F. (2024). *Constructing an Artificial-Intelligence Higher Education Environment: Guidelines for the Future*. 173–192. https://doi.org/10.33134/hup-25-9

Mardell, B., Ryan, J., Krechevsky, M., Baker, M., Schulz, S., & Liu-Constant, Y. (2023). *A pedagogy of play: Supporting playful learning in classrooms and schools.*

Miao, F., & Tawil, S. (2024). Steering the digital transformation of education: UNESCO's human-centered approach. *Frontiers of Digital Education*, 1(1), 51–58. https://doi.org/10.3868/s110-009-024-0005-6

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: A guidance for policymakers*. Unesco Publishing.

Mishara, P. (2024). The Ethical Implications of AI in Education: Privacy, Bias, and Accountability. *Journal of Informatics Education and Research*, 4(2). https://doi.org/10.52783/jier.v4i2.1827

Mohamed, S. P. (2024). Exploring ethical dimensions of AI-enhanced language education: A literature perspective. *Technology in Language Teaching & Learning*, 6(3), 89–99. https://doi.org/10.29140/tltl.v6n3.1813

Molina, E., Cobo, C., Pineda, J., & Rovner, H. (2024). *La revolución de la IA en Educación: Lo que hay que saber. Innovaciones Digitales de Educación*. Banco Mundial.

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Penguin.

Mollick, E. R., & Mollick, L. (2024). Instructors as Innovators: a Future-focused Approach to New AI Learning Opportunities, With Prompts. *The Wharton School Research Paper*. SSRN: https://ssrn.com/abstract=4802463

Morandín-Ahuerma, F. (2024). *A Critical Analysis of the European Union's Considerations on the Ethical Use of Artificial Intelligence in Education*. https://doi.org/10.31219/osf.io/escqj

Nowotny, H. (2021). *In AI we trust: Power, illusion and control of predictive algorithms*. John Wiley & Sons.

OECD (2025). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education* (Review draft). OECD. Paris. Disponible en: https://ailiteracyframework.org

Ray, T. (2024, August). Understanding AI and Deepfakes: What Parents Need to Know. *Cyber Safety Project*. https://cybersafetyproject.com.au/understanding-ai-and-deepfakes-what-parents-need-to-know/

Statista. (s.f.). *Artificial Intelligence - Worldwide*. *Statista*. Disponible en: https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide

UNESCO (n.d.). *Artificial Intelligence in Education*. Disponible en: https://www.unesco.org/en/digital-education/artificial-intelligence

Zaidy, A. (2024). *The Impact of Generative AI on Student Engagement and Ethics in Higher Education*. https://doi.org/10.70715/jitcai.2024.v1.i1.004

Zhang, C. (2024). AI in Education: Opportunities, Challenges, and Pathways for Equitable Learning. *Journal of Education, Humanities and Social Sciences*, 45, 723–728. https://doi.org/10.54097/kfgp6j07

Sobre los autores

Ana María Ayala Román – Doctora en Filosofía. Licenciada y magíster en Filosofía. Actualmente, es integrante de la Decanatura de innovación educativa y profesora del Departamento de Humanidades y ciudadanía. Sus líneas de investigación son: la ética y la filosofía política, la relación entre estética y política, la enseñanza de la ética, así como las relaciones entre los afectos y la configuración de las comunidades.

ORCID: https://orcid.org/0000–0001–9682–0625

Correo electrónico: amayala@icesi.edu.co / anamariaayalar@gmail.com

Juan Felipe Castaño Quintero – Candidato a Doctor en Antropología Social. Antropólogo. Actualmente se desempeña como docente en los departamentos de Humanidades, Tecnología y Ciencias, en el área de Ciencia, Tecnología y Sociedad (CTS), y en el departamento de Estudios Sociales de la Universidad Icesi. Sus líneas de investigación son diversas y se centran en los Estudios Sociales de la Tecnología, destacando su trabajo desde 2017 sobre el impacto social de la Inteligencia Artificial (IA), los sesgos algorítmicos y los modelos de gobernanza. También posee una vasta experiencia en patrimonio, políticas e industrias culturales, incluyendo su participación en investigación de festivales y patrimonialización en Cali y su investigación doctoral sobre la agencia del movimiento corporal en los bailarines de salsa caleños.

ORCID: https://orcid.org/0009-0007-3169-0357
Correo electrónico: juan.castano3@u.icesi.edu.co

Edilberto Loaiza Corrales – Magister en Educación. Publicista. Coordinador de Aseguramiento y Valoración de los Aprendizajes de la Decanatura de innovación educativa de la Universidad Icesi. Actualmente se desempeña como profesor del curso "Redes Sociales" del Núcleo Exploratorio de la misma universidad. Sus líneas de investigación son: Aseguramiento y Valoración de los aprendizajes, Integración de IA en procesos educativos, Redes Sociales, Ecosistemas Digitales e Inteligencia Artificial.

Correo electrónico: eloaiza@icesi.edu.co

Juan Fernando Arenas Jaramillo – Magíster en Derecho. Especialista en Propiedad Industrial, Derechos de Autor y Nuevas Tecnologías. Abogado. Actualmente se desempeña como Director del Programa de Derecho de la Universidad Icesi, y es integrante del Centro Interdisciplinario para la transformación digital e IA. Sus líneas de investigación son: Derecho digital, moderación de contenido y libertad de expresión. Innovación pedagógica en derecho. Creación y regulación de contenidos generados por IA. Derecho de la competencia, propiedad intelectual y derecho marcario.

ORCID: https://orcid.org/0000-0002-6265-9017
Correo electrónico: jfarenas@icesi.edu.co

Diego Hernán Varón Rojas – Doctor en Humanidades, en la línea de estudios sociales de las organizaciones. Magister en administración. Pregrado en antropología y en gestión empresarial. Actualmente se desempeña como docente cátedra en las Facultades de Ciencias Humanas y Negocios y Economía, de la Universidad Icesi. También de la facultad de Ciencias de la Administración e investigador adscrito al grupo Nuevo Pensamiento Administrativo de la Universidad del Valle. Integrante de las redes CLACSO, ALER, Red de formación ética

en Colombia y RED Antropourbana. Sus líneas de investigación son antropología organizacional y del trabajo, ciencias humanas, ética, cultura y responsabilidad social en las organizaciones.

ORCID: https://orcid.org/0000-0001-7167-094X
Correo electrónico: diegovaron721@yahoo.es

Henry Arley Taquez Quenguan – Máster en Sociedad de la Información y el conocimiento, especialista en gerencia informática e ingeniero de sistemas. Actualmente, hace parte de la Decanatura de Innovación Educativa y Fortalecimiento del PEI, así como profesor del Departamento de Pedagogía de la Universidad Icesi. Sus líneas de investigación se centran en las competencias digitales docentes, los ambientes virtuales de aprendizaje, el uso educativo de la inteligencia artificial generativa y la planeación estratégica para la integración de tecnologías digitales e inteligencia artificial en la educación superior.

ORCID: https://orcid.org/0000-0003-3298-9979
Correo electrónico: hataquez@icesi.edu.co

María Isabel González Serrano – es Licenciada en Literatura y Lengua Castellana y Licenciada en Ciencias Sociales de la Universidad Icesi. Su experiencia docente incluye procesos formativos en lectura, escritura y oralidad en entornos universitarios, y ha acompañado la formación de docentes en competencias digitales. Actualmente, forma parte del área de IA+TIC para la educación de la Decanatura de Innovación Educativa y Fortalecimiento del PEI de la Universidad Icesi, y coordina grupos de investigación interdisciplinares que desarrollan estrategias para la integración pedagógica de la Inteligencia Artificial Generativa en contextos universitarios. Sus líneas de investigación incluyen el uso pedagógico de las tecnologías digitales, la Inteligencia Artificial Generativa en la educación y las competencias digitales docentes.

Correo electrónico: maria.gonzalez29@u.icesi.edu.co

Isabel Tejada Sánchez – es educadora e investigadora con doble doctorado en Ciencias del Lenguaje por la Université Paris 8 y en Comunicación Lingüística y Mediación Multilingüe por la Universitat Pompeu Fabra. Se desempeña como profesora asociada en la Facultad de Educación de la Universidad de los Andes, en Bogotá, Colombia. Certificada en Prompt Engineering y formada en IA generativa centrada en el ser humano, su trabajo se focaliza en liderar espacios reflexivos y críticos sobre la IA en el ámbito educativo. Otros de sus intereses incluyen la mentalidad de crecimiento (Growth Mindset), la Pedagogía del Juego, el aprendizaje de lenguas y la interculturalidad

ORCID: https://orcid.org/0000-0002-4008-8057

Correo: mi.tejada10@uniandes.edu.co

Otros títulos de la colección

[2024]

Enseñar ética, ¿yo? Reflexiones docentes sobre enseñanza de la ética profesional | Ana María Ayala Román (editora académica)

DOI: https://doi.org/10.18046/EUI/vc.7.2024

[2022]

Una mirada atenta | Lelio Fernández Druetta / Jean-Paul Margot y Rafael Silva Vega (compiladores) / Jean-Paul Margot (editor)

DOI: https://doi.org/10.18046/EUI/vc.6.2022

[2021]

Diálogos entre ética y ciencias sociales. Teoría e investigación en el campo social | Ana María Ayala Román, Juan José Fernández Dusso y Enrique Rodríguez Caporalli (editores académicos)

DOI: https://doi.org/10.18046/EUI/vc.5.2021

[2020]

Ensayos sobre la pandemia | Ximena Castro-Sardi, Diego Cagüeñas Rozo, Diana Patricia Quintero Mosquera, Juan José Fernández Dusso y Rafael Silva Vega (compiladores)

DOI: https://doi.org/10.18046/EUI/vc.4.2020

Este libro se terminó de editar en octubre de 2025. En su preparación, realizada desde la Editorial Universidad Icesi, se emplearon tipos Joanna MT Std en 11/14 y 12/14.

Colección Varii Cives

Este volumen es el resultado de un esfuerzo deliberado por parte de actores vinculados a instituciones de educación superior de la ciudad y del país, quienes, desde distintas trayectorias, proponen una postura crítica y situada frente a la IAG. Surge como respuesta a una inquietud académica compartida: explorar las implicaciones humanas, sociales, jurídicas y políticas de estas tecnologías emergentes. Cada capítulo adopta un método ensayístico-argumentativo, sustentado en revisión documental, análisis conceptual y estudio de casos cuando aplica. La obra es el resultado de la pluralidad de marcos teóricos y perspectivas disciplinares que permiten iluminar problemas éticos complejos desde diferentes ángulos.

Universidad Icesi | Editorial Universidad Icesi

ISBN 978-628-7814-19-6