¿Cómo navegar el camino hacia la ética en IA?

CLAUDIA LÓPEZ MONCADA

Doctora en Ciencias de la Información y Tecnología por la Universidad de Pittsburgh. Profesora Asistente del Departamento de Informática de la Universidad Técnica Federico Santa María e investigadora en el Centro Nacional de Inteligencia Artificial (CENIA) y en el Núcleo Milenio Futures of Artificial Intelligence Research (FAIR). Líneas de investigación: computación centrada en personas.

claudia.lopez@usm.cl

GABRIELA ARRIAGADA BRUNEAU

Magíster en Filosofía por la Universidad de Edinburgh y candidata a doctora por la Universidad de Leeds. Profesora Asistente en el Instituto de Éticas Aplicadas y el Instituto de Ingeniería Matemática y Computacional, Pontificia Universidad Católica de Chile. Líneas de investigación: ética de la IA y datos, filosofía de la tecnología.

gcarriagada@uc.cl

ALEXANDRA DAVIDOFF

Socióloga por la Pontificia Universidad Católica de Chile. Asistente de investigación en el Núcleo Milenio Futures of Artificial Intelligence (FAIR). Líneas de investigación: ética en IA e infancia.

eadavidoff@uc.cl

RESUMEN. En este artículo relatamos cómo el Centro Nacional de Inteligencia Artificial (CENIA) busca abordar las consideraciones éticas en sus proyectos de investigación de IA. Hemos conformado un Grupo de Trabajo en Ética (GTE) con representantes de cada línea de investigación del Centro y, en conjunto, hemos explorado diferentes desafíos que aquí resumimos.

Primero, discutimos si el enfoque ético en IA requiere un análisis adicional al de los comités de ética de investigación. Luego, revisamos los principios que se han propuesto para resolver los problemas éticos que emergen alrededor de la IA. Finalmente, presentamos la iniciativa de CENIA para priorizar principios éticos propios, que involucra el desarrollo de un enfoque metodológico para evaluarlos. Como conclusión, destacamos la importancia de la ética como una herramienta para mejorar el desarrollo de la IA y subrayamos la necesidad de enfoques éticos específicos para la región latinoamericana.

Lo que inició con preocupaciones por privacidad de datos personales, imparcialidad/justicia (*fairness*) de los resultados y opacidad de los modelos de inteligencia artificial (IA) [1, 2, 3], hoy se ha desarrollado hacia múltiples aristas éticas, que van de la mano del despliegue masivo de la IA en nuestra vida cotidiana, como su impacto en el medio ambiente, su relación con los derechos humanos, la falta de mecanismos de rendición de cuentas, entre otras.

En el Centro Nacional de Inteligencia Artificial (CENIA)[1], que busca que la IA esté al servicio de las personas, esto ha inspirado una pregunta central en nuestro trabajo: ¿cómo abordamos las diversas consideraciones éticas que están involucradas en un proyecto de investigación en IA?

Con este propósito, hemos reunido a un Grupo de Trabajo en Ética en CENIA, con el objetivo de enriquecer el diálogo sobre la integración de la ética en la IA en Chile. Este artículo presenta un resumen de algunas de las preguntas e incipientes respuestas que hemos logrado desarrollar hasta el momento.

1 https://www.cenia.cl/.

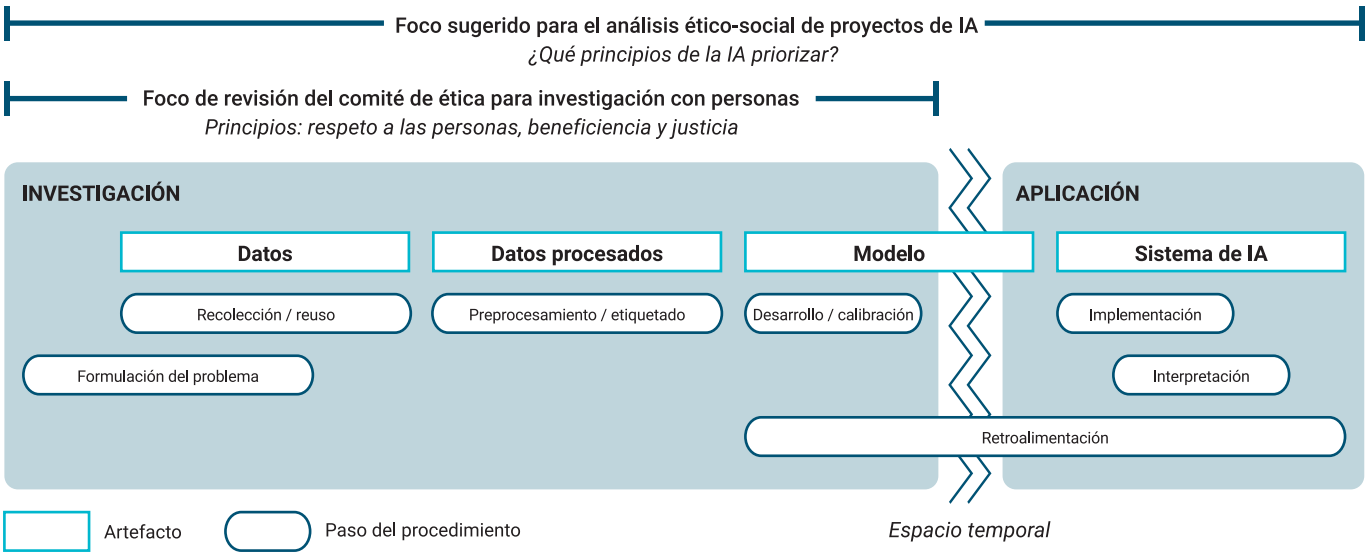

Figura 1. Ilustración del foco de análisis de los comité de ética versus análisis para proyectos de IA.

¿La ética en IA requiere un análisis diferente al de un comité de ética de investigación?

Siguiendo la literatura, hemos llegado a la conclusión que, efectivamente, existen diferencias fundamentales entre las funciones desempeñadas por estos dos enfoques. Si bien son mutuamente complementarios, los comités de ética institucionales están centrados en salvaguardar los derechos de los participantes en la investigación, poniendo un énfasis en la mitigación de los riesgos inherentes durante la ejecución de los estudios (tales como la potencial pérdida de privacidad de datos personales y los posibles efectos secundarios adversos para la salud o la integridad de los sujetos involucrados). No obstante, la evidencia respalda la noción de que los riesgos asociados a la IA a menudo emergen en etapas posteriores a la conclusión de un proyecto de investigación, cuando la tecnología se encuentra en pleno funcionamiento [4].

En la Figura 1, ilustramos el alcance del análisis de riesgos de ambos enfoques, utilizando como ejemplo un flujo estándar en un proyecto de aprendizaje automático basado en datos. El enfoque convencional del análisis adoptado por los comités de ética tiende a centrarse en la fase inicial del proceso. De esta manera, el análisis ético vinculado a la implementación de la inteligencia artificial (IA) experimenta una ruptura temporal, considerándose como una etapa posterior o futura en la formulación de un proyecto.

Diversos casos que exponen los riesgos inherentes a la IA han sido documentados, destacándose la ocurrencia de perjuicios hacia personas en su etapa de aplicación. Estos perjuicios abarcan desde la asignación incorrecta de categorías que afectan el acceso a beneficios gubernamentales, hasta la lesión o exclusión de individuos en procesos de contratación debido a la discriminación fundada en características como género, raza u otros atributos protegidos. A esto se agrega la ausencia de mecanismos que permitan a los usuarios solicitar correcciones o presentar apelaciones en situaciones donde un sistema basado en IA opere de manera errónea, ocasionando daños a nivel individual y grupal en la sociedad.[2]

En el presente contexto, académicos como Bernstein et al. [5] han propuesto poner la atención hacia los riesgos que afectan a la sociedad en fases subsiguientes a la investigación, tales como la implementación o comercialización de la inteligencia artificial, así como su

2 Un reconocido caso ilustrativo es el sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), utilizado en los tribunales de los Estados Unidos para evaluar el riesgo de reincidencia de un individuo. Investigaciones concernientes a esta tecnología revelaron que su índice de falsos positivos era notablemente superior en el caso de personas afroamericanas en comparación con las de origen caucásico. Esta disparidad reproducía las desigualdades raciales presentes en el sistema de justicia de dicho país [6, 7].

aplicación en la formulación de políticas públicas. En esencia, abogan por asumir la responsabilidad de considerar las posibles implicancias de los resultados obtenidos en nuestras investigaciones y buscar formas de atenuar los riesgos a través de la toma de decisiones, proporcionando recomendaciones para aplicaciones futuras. Con este propósito, los autores sugieren abordar tres interrogantes fundamentales en cada proyecto: (1) ¿cuáles son los riesgos involucrados?, (2) ¿qué principios deben seguirse para mitigar estos riesgos?, y (3) ¿cómo se concretan estos principios en el diseño de la investigación?

En la construcción de estas interrogantes, Bernstein et al. [5] siguen una línea de razonamiento similar a la de los comités de ética, al conectar la mitigación de riesgos con principios que tienen como propósito resolver los dilemas éticos que surgen en el transcurso de una investigación, tal como fue propuesto en el Informe Belmont [8]. Dicho informe establece tres principios fundamentales para la investigación con personas: (1) respeto a las personas (autonomía y protección de vulnerabilidad), (2) beneficencia (bienestar y prevención de daños), y (3) justicia (trato justo y no discriminatorio).

El desafío en este momento, por tanto, es reflexionar sobre la ética en IA y poder identificar qué principios nos ayudarán a resolver los problemas establecidos y emergentes, más allá de cuestionamientos metodológicos sobre la formulación del proyecto o las responsabilidades investigativas.

¿Qué principios de la IA son relevantes según el panorama internacional?

En el contexto del rápido avance de la IA, el diálogo a nivel global enfatiza la importancia de orientar su desarrollo en base a principios que favorezcan el bienestar de la sociedad en su conjunto. Esta conversación abarca tanto a actores del ámbito público como del privado, con y sin fines de lucro, los cuales han articulado sus propios principios en consonancia con las prioridades específicas de sus esferas de influencia.

Hemos llegado a la conclusión de que [...] existen diferencias fundamentales entre las funciones desempeñadas por [...] los comité de ética de investigación [y] el análisis para proyectos de IA.

Hace tres años el Centro Berkman Klein de Harvard, publicó una taxonomía de los principios de la IA [9] fundamentada en un análisis de 36 documentos que enuncian principios relacionados con la IA, a nivel nacional, multilateral y organizacional. Esta taxonomía identifica ocho ejes temáticos para categorizar los principios de la IA: privacidad, rendición de cuentas, seguridad y protección, transparencia y explicabilidad, imparcialidad/justicia (*fairness*) y no discriminación, control humano de la tecnología, responsabilidad profesional y promoción de valores humanos. Asimismo, otros organismos como la UNESCO [10] y la ONU [11] también han propuesto recomendaciones, añadiendo aspectos como la alfabetización, sustentabilidad, inclusión (enfatizando género), y proporcionalidad del uso de la IA para alcanzar un objetivo legítimo.

Por otra parte, grandes empresas como Google [12], IBM [13] y Microsoft [14] han formulado sus propios protocolos y principios. Aunque en términos generales están en línea con los mencionados previamente, también ponen énfasis en la aplicación y los posibles usos de sus productos. Además, destacan la relevancia del profesionalismo y la excelencia científica en el proceso de desarrollo tecnológico. Sin embargo, es esencial notar que estos lineamientos han sido ampliamente criticados dada su limitada influencia en la práctica. De hecho, la literatura alerta sobre el uso de los lineamientos éticos como un mecanismo de lavado de imagen (*ethics washing*, en inglés). Más específicamente, las críticas enfatizan cómo las tendencias de operacionalización de los principios éticos sin una reflexión profunda y prudente, ha terminado por subordinar a la ética como una herramienta para los intereses económicos de las grandes corporaciones tecnológicas [15].

Esto nos deja con una escenario de amplios intereses, motivaciones, y principios, haciendo dificultosa la identificación de directrices éticas a seguir.

¿Qué principios de la IA son relevantes desde el punto de vista local?

Por esas dificultades es importante enfocarse en el punto de vista local. En 2021, el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación de Chile, desarrolló una Política Nacional de IA [16], que presenta cuatro principios transversales: (1) IA con centro en el bienestar de las personas, respeto a los derechos humanos y la seguridad; (2) IA para el desarrollo sostenible, (3) IA inclusiva e (4) IA globalizada y en evolución. Además, propone tres ejes estructurales: (1) factores habilitantes, (2) desarrollo y adopción, y un último eje (3) ética, aspectos normativos, e impactos socioeconómicos. Hasta la fecha, siguen trabajando en profundizar este último eje de manera participativa en conjunto con la UNESCO, con el objetivo de presentar una actualización de la Política en octubre de 2023. Además, desde la normatividad legislativa, existe una propuesta de ley para “regular los sistemas

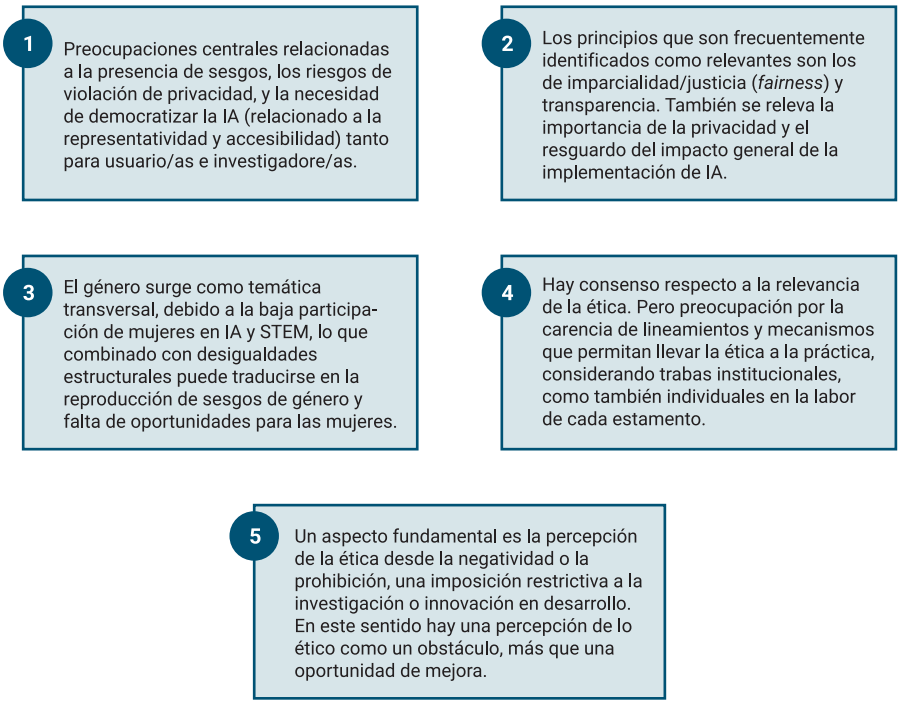

Figura 2. Tendencias centrales para la indagación exploratoria.

de inteligencia artificial, la robótica y las tecnologías conexas, en sus distintos ámbitos de aplicación" [17], la cual está siendo discutida en la Comisión de Desafíos del Futuro, Ciencia, Tecnología e Innovación de la Cámara de Diputadas y Diputados, y la Mesa "Legislando sobre IA" convocada por la Comisión de Desafíos del Futuro, Ciencia, Tecnología e Innovación del Senado de Chile.

Estos pasos a nivel nacional son sólo una de las realidades regionales. Vale la pena destacar las considerables diferencias en las percepciones públicas sobre los avances de IA en la región latinoamericana, según la caracterización hecha por el reciente Índice Latinoamericano de IA [18]. En el índice se enfatiza que es importante tener en cuenta las diferencias culturales y socioeconómicas al analizar la percepción de la IA en cada país, así como considerar la falta de perspectiva crítica asociada al discurso público sobre su desarrollo.

Teniendo en consideración las tendencias regionales, así como la transversal falta de entrenamiento formal en ética en IA en la formación profesional de áreas STEM [19, 20], a fines del 2022 decidimos realizar una investigación exploratoria sobre el conocimiento y las percepciones sobre ética de IA de los distintos integrantes de CENIA. A partir de esta indagación, se corroboró la escasez de conocimiento formal en el campo, aunque también se pudo identificar una gran heterogeneidad en términos de conocimiento informal, determinada por el propio interés o la relevancia del tema de acuerdo con las demandas del rol laboral de cada participante.

Para esta investigación exploratoria, se utilizó una metodología cualitativa, que constó con un total de siete entrevistas semiestructuradas individuales y una entrevista grupal, así como tres *focus groups*, a integrantes de distintos cargos como investigadores, cargos administrativos, desarrolladores y estudiantes. A partir de una metodología de análisis de discurso por medio de rejilla, las respuestas de los participantes se codificaron de acuerdo a los objetivos de la entrevista en lo que refiere a conocimiento, valoraciones, percepciones y prácticas de los participantes. En la Figura 2 presentamos cinco aspectos esenciales que surgieron de estas entrevistas.

Es importante tener en cuenta las diferencias culturales y socioeconómicas al analizar la percepción de la IA en cada país.

En relación particularmente a la percepción de la ética desde la negatividad, es que nos parece necesario reenfocar la percepción de la ética, no como una limitante, sino que como una oportunidad metodológica para mejorar el desarrollo de la IA. Para esto, nos propusimos formular una guía de trabajo para el análisis ético-social. En esta guía, el propósito es robustecer proyectos de investigación identificando nuevas o mejores preguntas base para la formulación de problemas, y el análisis de riesgo.

Ante esta serie de preocupaciones y principios, ¿por dónde empezar?

Como una forma de hacer frente a la multiplicidad de jerarquías posibles asociadas al uso de los principios y preocupaciones alrededor de la IA, el Grupo de Trabajo de Ética (GTE) de CENIA ha buscado sintetizar principios transversales que representen los intereses de nuestro centro. El GTE, constituído por solicitud del Comité Científico de CENIA, está conformado por un representante de cada

Documento	Principios
Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled AI: A Map of Ethical and Rights-Based Approaches to Principles for AI.	Privacidad, rendición de cuentas, imparcialidad/justicia (*fairness*), seguridad, responsabilidad profesional, promoción de valores humanos, transparencia, control humano de la tecnología.
Khan, A. A., Badshah, S., Liang, P., Khan, B., Waseem, M., Niazi, M., & Akbar, M. A. (2021). Ethics of AI: A Systematic Literature Review of Principles and Challenges.	Transparencia, privacidad, responsabilidad, equidad, autonomía, explicabilidad, imparcialidad/justicia (*fairness*), no maleficencia, dignidad humana, beneficencia, responsabilidad, seguridad, seguridad de datos, sostenibilidad, libertad, solidaridad, prosperidad, efectividad, precisión, previsibilidad, interpretabilidad.
Independent High-Level Expert Group on Artificial Intelligence (European Commission). (2019). Ethics Guidelines for Trustworthy AI.	Respeto por la autonomía humana, prevención del daño, equidad y explicabilidad. Agencia humana y supervisión, robustez técnica y seguridad, privacidad y gobernanza de datos, transparencia, diversidad y no discriminación, equidad, bienestar social y ambiental, rendición de cuentas.
Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines.	Transparencia, justicia y equidad, no maleficencia, responsabilidad, privacidad, beneficencia, libertad y autonomía, confianza, sustentabilidad, dignidad y solidaridad.
Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking Artificial Intelligence Principles.	Humanidad, colaboración, compartir (equidad), justicia, transparencia, privacidad, seguridad, protección, rendición de cuentas, AGI/ASI.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.	Beneficencia, no maleficencia, autonomía, justicia, explicabilidad.
Smit, K., Zoet, M., & Van Merten, J. (2018). A Review of AI Principles in Practice.	Mejora (*augmentation*) humana, beneficencia, confiable, centrado en el ser humano. Autonomía, igualdad (diseño y ejecución), trazabilidad, dignidad humana, derechos humanos, transparencia, democratización, privacidad, seguridad, seguridad (diseño y ejecución), colaboración, responsabilidad, comprensibilidad, uso responsable de los datos, precisión y educación y promoción.

Figura 3. Revisión de literatura sobre principios de la IA.

una de las 5 líneas de investigación del centro: (RL1) Aprendizaje profundo para visión y lenguaje, (RL2) IA neuro-simbólica, (RL3) IA inspirada en el cerebro, (RL4) Aprendizaje automático basado en la física, y (RL5) IA centrada en las personas. Además, el GTE ha contado con la asesoría de una socióloga (coautora de este artículo) y dos abogadas.

La síntesis comenzó con una revisión de siete documentos recopilatorios y meta-revisiones de principios de la IA al 2022 (ver Figura 3). Una propuesta inicial de principios para CENIA, fue socializada y discutida en un workshop interno que congregó a la mayor parte de quienes integran el Centro. Con esta retroalimentación, el GTE propuso una

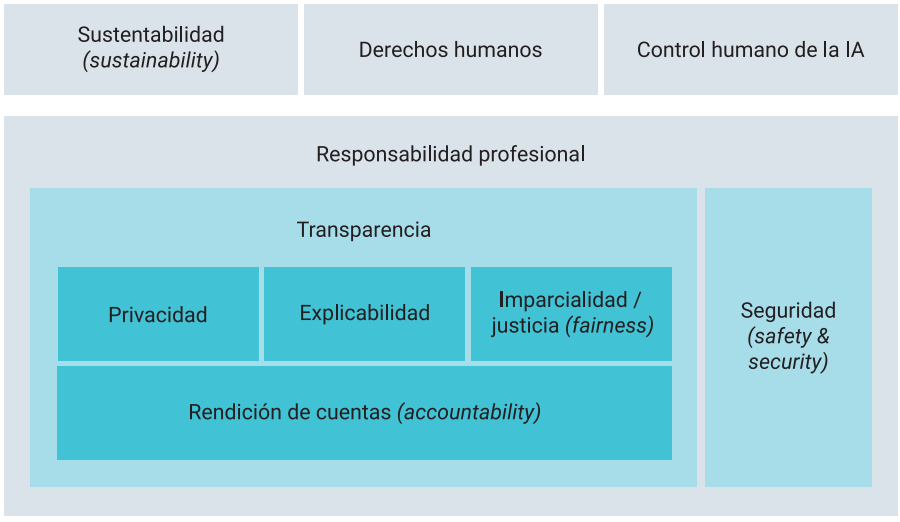

Figura 4. Síntesis de principios de la guía de trabajo para el análisis ético-social de CENIA.

Derechos humanos	Access Now [21] enfatiza la universalidad y el poder vinculante que estos tienen: "En los casos donde no existe legislación nacional, la legitimidad moral de los derechos humanos conlleva un **importante poder normativo**" (p.17).
Control humano de IA	En las directrices para una IA confiable de la Unión Europea, se establece que "la supervisión humana ayuda a garantizar que un sistema de IA **no socave la autonomía humana** o cause otros efectos adversos" [22, p. 16].
Sustentabilidad	Para Floridi et al. [23] se trata de garantizar las condiciones básicas para la vida como una **"prosperidad continuada de la humanidad"** (p. 697) preservando el medio ambiente para generaciones futuras.
Responsabilidad profesional	Según Fjeld et al. [9] esto pasa en parte por "asegurarse de que los involucrados y afectados sean consultados y que los **efectos a largo plazo sean parte de la planificación**" (p.5).
Seguridad	Los estándares de seguridad, según Smith et al. [24] son transversales: "La IA debe diseñarse y ejecutarse para **proporcionar la máxima seguridad contra amenazas internas y externas**, maliciosas o accidentales" (p.8).

Figura 5. Descriptores de principios en la literatura I.

Transparencia	Definida por Leslie [25], hace referencia a la habilidad de "saber **cómo y por qué un modelo funciona de determinada manera**, en un contexto específico" (p.35), haciendo referencia a visibilizar la lógica detrás de su comportamiento.
Explicabilidad	Este principio apunta a tener criterios para diseñar una IA "de modo que **los humanos sean capaces de entender** (lenguaje, presentación) la forma de trabajar de la IA" [24, p. 8].
Privacidad	Privacidad implica respetar la intimidad de las personas y sus datos, pero también su "**capacidad de decisión sobre sus datos** y las decisiones que se tomen con ellos" [9, p. 4].
Imparcialidad	Se busca prevenir la discriminación de individuos y grupos, priorizando el "**evitar sesgos injustos en los sistemas de IA**, que pueden propiciar injusticia social o privar a las personas de su autonomía" [26, p. 11-12].
Rendición de cuentas	Se enfatiza en la importancia de que existan mecanismos disponibles para garantizar que "**la responsabilidad por las repercusiones de los sistemas de IA** se distribuya adecuadamente y que se ofrezcan soluciones adecuadas" [9, p. 4].

Figura 6. Descriptores de principios en la literatura II.

Nos parece necesario reenfocar la percepción de la ética, no como una limitante, sino que como una oportunidad metodológica para mejorar el desarrollo de la IA.

síntesis de principios para CENIA, la cual ilustramos en la Figura 4, y explicamos a continuación.

En esta síntesis hay tres principios generales: sostenibilidad, derechos humanos y control humano de la IA. Estos principios generales intersectan los demás principios, a los cuales englobamos dentro del principio de responsabilidad profesional. Esta decisión busca significar que entendemos que la investigación y el desarrollo de la IA siempre ocurre en el marco de decisiones humanas que deben incorporar estándares técnicos y éticos para garantizar la concreción de los demás principios.

Bajo responsabilidad profesional, situamos a la transparencia como otro principio que ayuda a articular otros principios de la IA. La transparencia incorpora la idea de la explicabilidad, es decir, entender las decisiones puntuales del sistema. Además, es un requisito para garantizar la privacidad de las personas, donde ellas tienen control sobre cómo son usados sus datos, así como para imparcialidad/justicia (*fairness*), que requiere que las personas puedan comprender el funcionamiento e impacto de la IA, y cómo esta puede implicar sesgos y perjuicios. La transparencia, además, habilita la rendición de cuentas (*accountability*), que nos lleva de vuelta a las decisiones humanas involucradas en el proceso de la IA y cómo se deben distribuir las responsabilidades cuando se produce daño debido a la IA, y a la necesidad de desarrollar mecanismos de prevención y reparación asociados a tales daños.

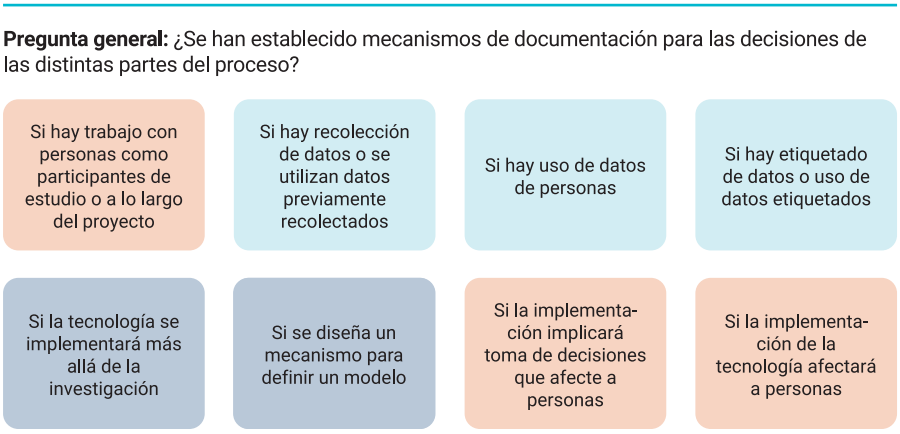

Figura 7. Categorización de preguntas de la guía de trabajo para el análisis ético-social de CENIA.

Por último, la seguridad, si bien no es necesariamente habilitada por la transparencia, sí es parte de la responsabilidad profesional que debe atravesar el diseño de la IA, al sostener los aspectos técnicos y las defensas ante posibles ataques que atenten contra los sistemas de IA o sus principios. En las Figuras 5 y 6 se incluyen descripciones de estos principios según la revisión de literatura que realizamos.

No cabe duda que ante esta lista de principios cabe preguntarse si todo proyecto de IA debe hacerse cargo de absolutamente todos esos principios. La respuesta simple es no, y justamente el desarrollo de nuestros siguientes pasos busca poder generar un instancia metodológica que fomente la priorización e identificación de principios a las contextualidades propias de cada proyecto investigativo. La metodología propuesta consta de una serie de preguntas sobre principios, riesgos y estrategias de mitigación organizados en categorías (ver Figura 7). Por ejemplo, hay preguntas que se deben contestar si hay trabajo con personas, si hay uso de datos, o si existe etiquetado de datos, así como otras preguntas relacionadas al impacto esperado en caso de implementación de la IA en contextos operacionales.

En el segundo semestre del 2023, estamos implementando un piloto con proyectos de CENIA para ir refinando principios y las preguntas que motivan la reflexión ético-social, así como herramientas que permitan articular estrategias de mitigación de los riesgos identificados. Esperamos poder hacer públicos esos resultados a la comunidad de IA y computación en el futuro cercano.

Es clave resaltar que aún quedan muchas preguntas sin abordar y a las cuales deberemos poner atención desde nuestras narrativas latinoamericanas. Por ejemplo, ¿cómo hacemos esto desde América Latina donde tenemos mucha menos evidencia de los impactos de la IA? Esperamos que nuestro trabajo fomente la integración de la ética no como un accesorio o un requisito vacío, sino como un aliado en la innovación responsable de la IA centrada en las personas, apoyando los esfuerzos legislativos y normativos que están surgiendo en nuestra región. ▪

Agradecimientos: *Queremos agradecer a las y los investigadores que integran el Grupo de Trabajo de Ética de CENIA: Margarita Castro, Cristóbal Moenne, Cristóbal Rojas y Federico Fuentes, así como a las abogadas que nos asesoran voluntariamente, Catherine Muñoz y Evelyn López.*

BIBLIOGRAFÍA

[1] O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.

[2] Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

[3] Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St Martin's Press.

[4] Arnold, T., & Scheutz, M. (2018). The "big red button" is too late: an alternative model for the ethical evaluation of AI systems. Ethics and Information Technology, 20, 59-69.

[5] Bernstein, M. S., Levi, M., Magnus, D., Rajala, B., Satz, D., & Waeiss, C. (2021). ESR: Ethics and Society Review of Artificial Intelligence Research (Version 2). arXiv.

[6] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks.

[7] Larson, J., Mattu, S.,Kirchner, L., & Angwin, J. (2016). How we analyzed the COMPAS recidivism algorithm.

[8] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. U.S. Department of Health and Human Services.

[9] Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication, 1.

[10] United Nations Educational Scientific and Cultural Organization. (2022). Recommendation on the Ethics of Artificial Intelligence. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa.

[11] Hogenhout, L. (2021). A Framework for Ethical AI at the United Nations (Version 1). arXiv.

[12] Google AI. (2023). Our Principles. Google. https://ai.google/principles/.

[13] World Economic Forum. (2021). Responsible Use of Technology: The IBM Case Study (White Paper). https://www.weforum.org/whitepapers/responsible-use-of-technology-the-ibm-case-study.

[14] Microsoft. (2022). Microsoft Responsible AI Standard, General Requirements for External Release. https://aka.ms/ResponsibleAI-Questions.

[15] Steinhoff, J. (2023). AI ethics as subordinated innovation network. In AI & SOCIETY. Springer Science and Business Media LLC.

[16] Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. (2021). Política Nacional de Inteligencia Artificial. https://minciencia.gob.cl/areas/inteligencia-artificial/politica-nacional-de-inteligencia-artificial/.

[17] Cámara de Diputadas y Diputados de Chile (2023) Ley 15869-19: Regula los sistemas de inteligencia artificial, la robótica y las tecnologías conexas, en sus distintos ámbitos de aplicación. https://www.camara.cl/legislacion/ProyectosDeLey/tramitacion.aspx?prmID=16416&prmBOLETIN=15869-19.

[18] Centro Nacional de Inteligencia Artificial (2023) Índice Latinoamericano de Inteligencia Artificial. /https://indicelatam.cl/wp-content/uploads/2023/08/ILIA-2023.pdf.

[19] Gómez Mont, C., Del Pozo, C. M., Martínez Pinto, C., & Martín del Campo Alcocer, A. V. (2020). La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países. https://doi.org/10.18235/0002393.

[20] Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: a systematic review. In International Journal of Educational Technology in Higher Education (Vol. 19, Issue 1). Springer Science and Business Media LLC.

[21] AccessNow (2018) Human RIghts in the Age of Artificial Intelligence. https://www.accessnow.org/wp-content/uploads/2018/11/AI-and-Human-Rights.pdf.

[22] Independent High Level Expert Group on Artificial Intelligence. (2019). Ethic Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[23] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. In Minds and Machines (Vol. 28, Issue 4, pp. 689–707). Springer Science and Business Media LLC.

[24] Smit, K., Zoet, M., & Meerten, J.V. (2020). A Review of AI Principles in Practice. Pacific Asia Conference on Information Systems.

[25] Leslie, D. (2019). Understanding artificial intelligence ethics and safety. arXiv.

[26] Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar, M. A. (2022). Ethics of AI: A Systematic Literature Review of Principles and Challenges. ACM International Conference Proceeding Series, 383–392.