

Documento de análisis

ÉTICA ANTICIPATORIA E INTELIGENCIA ARTIFICIAL

Unidad de Investigación y Vinculación Científica, BCN
Comisión de Ciencia, Tecnología e Innovación Productiva, HCDN



Documento de análisis

Unidad de Investigación y Vinculación Científica, BCN

Comisión de Ciencia, Tecnología e Innovación Productiva, HCDN

ÉTICA ANTICIPATORIA E INTELIGENCIA ARTIFICIAL

El documento analiza los desafíos éticos que surgen con el avance de la inteligencia artificial (IA) y sus tecnologías asociadas. Explora cómo los algoritmos de IA pueden perpetuar sesgos existentes y destaca la importancia de la "justicia algorítmica" para garantizar equidad y transparencia. También se aborda el uso de neurotecnologías y sus implicancias éticas, incluyendo la privacidad mental y los derechos humanos emergentes. Además, se discuten los desafíos de la autonomía y la responsabilidad en sistemas autónomos inteligentes. Se resalta la necesidad de abordar con una perspectiva ético-política los impactos de la IA en el empleo, y pone en marco geopolítico las principales orientaciones en gobernanza de la IA. El informe incluye recomendaciones éticas que permitan anticipar los avances tecnológicos y promover un desarrollo equitativo y sustentable de la IA.

Dra. Lucía Aguerre.

Unidad de Investigación y Vinculación Científica de la Biblioteca del Congreso de la Nación

Instituto de Filosofía. Facultad de Filosofía y Letras. Universidad de Buenos Aires

ÍNDICE

1.Introducción	3
2. El potencial de los algoritmos de la IA para reproducir los sesgos existentes	4
3. Ética y neurotecnologías asociadas a la IA	7
4. Sistemas autónomos inteligentes y responsabilidad	9
5. Impacto de la IA en el trabajo	13
6. Gobernanza de la IA y geopolítica	15
7. IA y democracia	18
9. Referencias	21

1. Introducción

La Inteligencia Artificial agrupa una diversidad de sistemas capaces de procesar datos e información de manera similar a un comportamiento humano inteligente, incluyendo tareas cognitivas de razonamiento, aprendizaje, percepción, predicción, planificación y generación. Estas tecnologías procesan información integrando modelos y algoritmos de aprendizaje automático que detectan patrones en el conjunto de datos aportados, para luego realizar predicciones para la toma de decisiones y la creación de contenidos nuevos en base a dichos datos. Dentro del campo de la IA se desarrollan tecnologías específicas como el aprendizaje automático (*machine learning*), las redes neuronales artificiales (inspiradas en la estructura y función del cerebro humano), el procesamiento de lenguaje natural (que incluye tareas como traducción automática y *chatbox*), el aprendizaje profundo (*deep learning*, que utiliza redes neuronales artificiales para reconocer patrones en datos y proporcionar una predicción), entre otras.

Estas tecnologías tienen el **potencial de transformar radicalmente no sólo los procesos productivos y de investigación científica**, aumentando la eficiencia, reduciendo errores y mejorando los sistemas de análisis, **sino también distintos aspectos de la vida humana que redundan en una mejor calidad de vida.** La IA aplicada a las ciencias médicas, por ejemplo, permite prever el riesgo de desarrollar enfermedades, terapias futuras basadas en la extensión de la vida, entre otras innovaciones potenciales y en curso.

Pero a la vez, **la capacidad de los sistemas de IA para realizar tareas que antes correspondían exclusivamente a los seres humanos plantea nuevos problemas éticos** relacionados con la **autonomía y la atribución de responsabilidad, la disponibilidad de datos y el derecho a la privacidad, la representatividad de los datos que componen los algoritmos, la equidad en el acceso a las tecnologías emergentes, entre otros.** Al mismo tiempo, el uso de IA produce transformaciones radicales en el mundo del trabajo, en el modo de adopción de decisiones y en la interacción social.

Este informe recorre los principales núcleos de abordaje ético en relación con la IA combinando ética anticipatoria, investigaciones que aportan evidencia científica sobre los problemas planteados¹ y recomendaciones para enfocar las políticas y los sistemas de monitoreo y control para garantizar el aprovechamiento máximo de los beneficios de la IA procurando un desarrollo equitativo .

En tiempos de una aceleración sin precedentes de los desarrollos científicos y tecnológicos, **atender los problemas éticos y anticipar los avances es**

¹ Todos los estudios de evidencia empírica referidos en este informe se encuentran referenciados para su acceso digital y profundización en la bibliografía.

fundamental para garantizar el aprovechamiento máximo de los beneficios de la IA. Asimismo se debe procurar un desarrollo de la IA equitativo promoviendo los valores democráticos y la justicia social. Las tecnologías se están desarrollando más velozmente que la capacidad de los Estados y la comunidad internacional para tomar decisiones que involucran procesos de análisis de la información, diálogos multiactorales, consensos y miradas comparadas. **La ética anticipatoria permite abordar los desarrollos que van a transformar las fronteras del conocimiento y tener un impacto en la ciencia, la tecnología y la comunidad, sus impactos sociales, económicos y geopolíticos y así planificar en consecuencia preparando políticas, legislaciones y acciones diversas.** Desarrollar y aplicar una ética de la anticipación no significa centrarse unilateralmente en los riesgos y la mera precaución, sino propiciar su desenvolvimiento hacia una mayor sostenibilidad y una distribución más justa de bienes, capacidades y oportunidades.

Este informe pretende aportar un **análisis situado, atento a las desigualdades interdependientes que determinan los procesos de producción, circulación e impactos de las tecnologías emergentes.** Las asimetrías afectan el desenvolvimiento de los sistemas entrelazando también las dimensiones de las desigualdades con las afectaciones y riesgos. En ese marco, nuestro país tiene el desafío de desarrollar IA para sectores productivos estratégicos nacionales en un marco ético e inclusivo, que procure disminuir los riesgos que conllevan las brechas con los países centrales.

2. El potencial de los algoritmos de la IA para reproducir y reforzar los sesgos existentes

En los sistemas que operan con IA, **los datos funcionan como la base para automatizar el aprendizaje y entrenar la formulación de predicciones.** Estos datos están formados por textos, imágenes, documentos, sonidos, datos de sensores, datos científicos, publicaciones en redes sociales, etc., dependiendo de los resultados esperados del sistema. **En esta información se expresan apreciaciones subjetivas y representaciones sociales que pueden portar o arrastrar prejuicios y estereotipos que circulan en el entramado de interacciones sociales.** De igual manera, cuando los datos utilizados para entrenar un algoritmo no son representativos de la realidad o están desequilibrados en relación a un segmento de la población o a un fenómeno, ya sea por una recolección parcial o incorrecta de ellos, los prejuicios surgen del diseño mismo del sistema y la implementación del algoritmo.

A esto se le llama “**sesgo algorítmico**”, esto es, una **anomalía en los resultados de los sistemas de IA debido a prejuicios acarreados por los datos de entrenamiento o suposiciones erróneas asumidas durante el proceso de desarrollo del sistema,** que arrojan predicciones que benefician

sistemáticamente a un grupo de individuos frente a otro, resultando injustas, desiguales y no generalizables (Estevez Almenzar et al. 2022).

Recientemente, **diversas investigaciones científicas aportaron evidencia sobre la presencia de sesgos algorítmicos en distintos sistemas de IA.** Estudios evaluaron la precisión en los algoritmos de sistemas comerciales de reconocimiento y análisis facial y pusieron en evidencia la existencia de sesgos raciales y de género, con disparidades de precisión, es decir, diferencias en la exactitud del reconocimiento, entre hombres y mujeres y entre los subgrupos de piel más oscura y más clara, debido a sesgos en los datos de entrenamiento (Buolamwini y Gebru 2018, Raji y Buolamwini 2019). También se documentó la aparición de sesgos de género en traductores automáticos como *Google Translate* o *Microsoft Translator* que en variados casos asignaban el género femenino o masculino a palabras neutras en idioma inglés, reproduciendo estereotipos de género relacionados con la distribución de identidades sociales, al traducir, cuando no había una indicación previa, *nurse* por enfermera, *doctor* por médico y *programmer* por programador (Stanovsky, Smith y Zettlemoyer 2019)². En Argentina, un estudio del Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional (sinc(i) CONICET-UNL) mostró la importancia del equilibrio de género en los conjuntos de datos de imágenes médicas utilizados para entrenar sistemas de IA para el diagnóstico asistido por computadora para enfermedades torácicas. Se encontró una disminución constante en el rendimiento para los géneros subrepresentados, lo que aporta una evidencia para las agencias encargadas de regular y aprobar los sistemas de diagnóstico asistido por computadora para incluir recomendaciones explícitas sobre el equilibrio de género y la diversidad (Larrazabal et al. 2020)³. También fue demostrada mediante estudios científicos la existencia de sesgos algorítmicos ligados a los discursos y las representaciones sociales en los Modelos de Lenguaje con Aprendizaje Automático (LLM, por sus siglas en inglés)⁴. Al comparar las respuestas de tres LLM distintos —GPT-3.5 Turbo⁶ (OpenAI), Command-nightly (Cohere) y Bard (Google)— con las opiniones de diversos grupos sociales disponibles en encuestas de opinión pública, el estudio realizado por expertos argentinos demostró la falta de representación equitativa de

² Este estudio, desarrollado por el Instituto Allen para la IA (AI2) representa la primera evidencia cuantitativa multilingüe a gran escala sobre el sesgo de género en la traducción automática. Se seleccionó un conjunto de ocho lenguas con género gramatical mostrando que cuatro sistemas comerciales populares y dos modelos académicos recientes de traducción automática de última generación son significativamente propensos a arrojar traducciones basadas en estereotipos de género.

³ La investigación proporcionó evidencia empírica respaldada por un estudio a gran escala, basado en tres arquitecturas de redes neuronales profundas y dos conjuntos de datos de imágenes de rayos X disponibles públicamente utilizados para diagnosticar varias enfermedades torácicas en diferentes condiciones de desequilibrio de género.

⁴ OpenAI fue la primera empresa en implementar esta tecnología en un producto concreto, gratuito y abierto a todo el público. En noviembre de 2022 la llegada de ChatGPT (generative pre-trained transformer, GPT, por sus siglas en inglés) impulsó una expansión significativa en el uso de estos modelos preentrenados para generar respuestas similares a una conversación en ámbitos diversos de la sociedad, aunque existen otros LLM disponibles en la actualidad como Bard, Claude o Cohere (Ortiz de Zárate et al. 2024).

algunos segmentos de la población⁵ (Ortiz de Zárate et al. 2024). Un estudio similar realizado en EE. UU. evaluó la alineación de las opiniones de LLM con las de 60 grupos demográficos encontrando una desalineación sustancial entre las opiniones reflejadas por los LLM actuales y las de la diversidad de grupos demográficos estadounidenses (Santurkar et al. 2023).

En este escenario, que va desde la necesidad de aplicar mejoras hasta la parcialidad y distorsión digital, **el concepto de “justicia algorítmica” fue acuñado para dar cuenta de la necesidad de implementar estrategias para aumentar la equidad y la transparencia algorítmica**. Entre estas estrategias, las **auditorías algorítmicas** ganan terreno como una herramienta clave para exponer y corregir los sesgos sistemáticos incorporados en las plataformas de software mediante la evaluación de su proceso de desarrollo, incluyendo el diseño y los datos utilizados para entrenar el sistema (Raji y Buolamwini 2019, Aránguiz Villagrán 2022).

A las implicancias éticas de los sesgos señalados en el primer apartado, se añaden **implicancias ético-políticas cuando el uso de LLM sesgados para la toma de decisiones en políticas lleva a la toma de decisiones gubernamentales erróneas**, distorsiona la percepción pública sobre un tema o, incluso, plantea inconvenientes ligados a la privacidad y la seguridad de los datos con los cuales fue entrenado el sistema. Esto abona la importancia de la **transparencia algorítmica en el sector público** mediante repositorios o registros en línea de algoritmos públicos como herramientas para defender los derechos democráticos y la rendición de cuentas de los gobiernos (GPAI 2024).

A partir de estas consideraciones sobre el **riesgo de la reproducción de sesgos en los algoritmos de la IA, se recomienda:**

- Asegurar que los datos utilizados para entrenar los algoritmos sean diversos y representativos de la población completa, ampliando las bases de datos con una perspectiva inclusiva y local al entrenar sistemas de IA basados en aprendizaje automático.
- Realizar pruebas rigurosas continuas para identificar y corregir posibles sesgos en los algoritmos mediante evaluaciones y auditorías algorítmicas.
- Apoyar el desarrollo de nuevos métodos para el descubrimiento de sesgos no supervisados en ausencia de información demográfica y anotaciones de expertos, que pueden ayudar a anticipar problemas de equidad al utilizarlos en nuevas poblaciones.
- Promover la participación de personas expertas en ética, equidad y diversidad en el proceso de diseño y evaluación de algoritmos.

⁵ Aunque sus respuestas no resultaron idénticas, compartían ciertas características en términos de las audiencias mostrando una inclinación hacia un sector masculino y politizado y con semejanzas en las respuestas a personas con niveles educativos más altos y de mayor edad (Ortiz de Zárate et al. 2024).

3. Ética y neurotecnologías asociadas a la IA

Los avances en neurotecnología tienen la capacidad de revolucionar el tratamiento de muchas enfermedades, desde lesiones cerebrales y parálisis hasta epilepsia y esquizofrenia. Los desarrollos existentes para el contexto de la medicina clínica y la investigación neurocientífica, incluyen neuroimágenes (técnicas que se utilizan para acceder directa o indirectamente a la estructura y función del sistema nervioso central reflejada por la actividad eléctrica, el consumo de oxígeno, la liberación de neurotransmisores, etc.), neurodispositivos (que pueden reemplazar partes del cuerpo -por ejemplo, prótesis robóticas- o estimular o inhibir funciones del cerebro -por ejemplo, estimuladores cerebrales profundos implantados-), interfaces cerebro-computadoras (que traducen los procesos cerebrales en resultados como, por ejemplo, mover una prótesis o mejorar el estado de ánimo de una persona deprimida mediante la recopilación de datos relacionados con la actividad neuronal a través de sensores) e IA aplicada a las neurociencias (en los que se utilizan algoritmos de IA en la investigación de neurociencia clínica con fines predictivos y de diagnóstico, por ejemplo, para detectar signos tempranos de la enfermedad de Alzheimer partir de patrones de actividad o anomalías estructurales en escáneres cerebrales) (UNESCO 2021, GESDA 2023).

El campo de la aumentación humana (*human augmentation*) se encuentra en veloz desarrollo: mejoras cognitivas y tecnologías para tratar desórdenes neurodegenerativos, aumentación de la conciencia mediante intervenciones electromagnéticas y drogas farmacéuticas moleculares y psicodélicas van a transformar capacidades y concepciones sobre razonamiento, memoria y concentración (Dresler 2019, Schenberg 2018). Asimismo, la cognición artificial, que abarca la simulación de cerebros humanos o hardware que recapitula las funciones biológicas del cerebro, permitirá conocer más sobre las funciones del cerebro y aumentar las capacidades de la IA.

Si bien estos desarrollos prometen grandes avances para el bienestar humano, y sus usos clínicos están guiados por normas de la bioética, **la ágil evolución de las interfaces cerebro-computadora combinadas con IA plantea potenciales desafíos éticos en fines no clínicos comerciales en relación con la privacidad, la potencial manipulación cognitiva del comportamiento, la afectación de la identidad personal y la igualdad de las personas**⁶. El uso de datos neuronales obtenidos por dispositivos neurotecnológicos no invasivos comerciales podría eventualmente permitir a los desarrolladores alterar profundamente algunas características humanas fundamentales como la vida mental privada, la agencia individual y desencadenar reacciones y emociones sin consentimiento (Goering et al.

⁶ Si bien la alteración, modificación o manipulación neurotecnológica es considerada como una violación a la dignidad humana, es fundamental distinguir los contextos y la voluntad, ya que la neurotecnología en el contexto clínico es una herramienta poderosa para restaurar la dignidad humana mediante la rehabilitación y el retorno a la autonomía.

2021, Yuste 2019, UNESCO 2022)⁷. Otra preocupación ética radica en que si el acceso a neurotecnologías se limita a quienes pueden pagar por ella, esto acentuará las desigualdades sociales aumentando aún más las brechas existentes (Veit 2018, López-Silva 2021).

Ante las posibilidades sin precedentes de acceso, recopilación, difusión y manejo de datos del cerebro, **las neurotecnologías plantean importantes nuevos desafíos en materia de derechos humanos**. Trabajos como el de Ienca y Andorno (2020) muestran que el marco actual de derechos humanos no parece ser suficiente para responder a estas problemáticas emergentes y **proponen el reconocimiento del derecho a la libertad cognitiva, el derecho a la privacidad mental, el derecho a la integridad mental y el derecho a la continuidad psicológica** para abordar los nuevos riesgos. Tal como sucedió con la adaptación de las normas de derechos humanos a los desafíos planteados por la genética, a través de la Declaración Universal sobre el Genoma Humano y los Derechos Humanos (1997) y la posterior Declaración Internacional sobre Datos Genéticos Humanos (2003) es posible que se requiera integrar nuevos reconocimientos de derechos en esta materia. Recientemente, UNESCO conformó un equipo internacional para crear el primer marco global sobre la ética de la neurotecnología, que será debatido con los Estados miembros para adoptarlo a fines de 2025.

Hay precedentes de la incorporación de estos desafíos emergentes en marcos ético-legales nacionales como, por ejemplo, **la legislación chilena, pionera en materia de neuroderechos⁸ al consagrar, en 2021, la protección de la actividad cerebral y la información procedente de ella a nivel supralegal** con la modificación de la Constitución Política de la República⁹. Actualmente, el Congreso chileno trabaja en el proyecto de ley "Sobre protección de los neuroderechos y la integridad mental, y el desarrollo de la investigación y las neurotecnologías" que tiene como objetivo proteger la privacidad mental, el libre albedrío y la no discriminación en el acceso de los ciudadanos a la neurotecnología, y establece el rol del Estado en materia de neurociencia y neurotecnología (López Silva y Madrid 2022).

⁷ En la actualidad existen desarrollos en curso como "NeuraLink", de Tesla Company, o la BCI no intrusiva llamada "Brain-to-text", de Facebook, que permitirían a los usuarios escribir en sus teléfonos mediante el pensamiento consciente (Facebook Reality Lab 2020); y el denominado "Kernel-Flow" que utiliza tecnología infrarroja para transmitir información neuronal en tiempo real a cualquier usuario que tenga el dispositivo (López Silva y Madrid 2022).

⁸ Los neuroderechos son un conjunto de nuevos derechos humanos cuyo objetivo es proteger el cerebro y su actividad, a la luz de los nuevos desarrollos en neurotecnología e inteligencia artificial (Ienca y Andorno 2020). El desarrollo del concepto de neuroderechos viene también asociado al concepto de "neurodatos", ya que se sustenta en la idea de que la actual protección normativa del derecho a la privacidad de la información no sería suficiente para la información neuronal (neurodatos).

⁹ Se modificó el número 1° del artículo 19 de la Constitución Política de la República de Chile, agregando el párrafo final: "El desarrollo científico y tecnológico estará al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica. La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella".

En base a lo mencionado sobre la problemática de IA aplicada a **neurotecnologías y neuroderechos**, es posible esbozar las **siguientes recomendaciones**:

- Asegurar la anticipación responsable, esto es, el asesoramiento científico periódico sobre los avances y desarrollos emergentes en materia de neurotecnología con una mirada prospectiva capaz de prever las implicancias de los usos y la protección de derechos.
- Garantizar por medio de políticas públicas inclusivas la equidad en el acceso a las tecnologías de aumentación humana en desarrollo (*human augmentation*) para evitar la profundización de desigualdades existentes.
- Considerar desde una perspectiva intercultural las distintas concepciones sobre la mente, la conciencia, la relación mente-cuerpo y las terapias alternativas para ampliar y nutrir los paradigmas en posible transformación.

4. Sistemas autónomos inteligentes y responsabilidad

En un sentido general, el concepto de autonomía refiere a la capacidad de las personas y las instituciones de tomar decisiones sin intervención ni influencias externas. La filosofía moderna asoció la noción de autonomía al ejercicio de la razón y la conciencia para guiar las propias acciones, concibiendo un sujeto humano ético capaz de darse reglas a sí mismo y, ligado a eso, capaz de asumir responsabilidad por sus decisiones y actos¹⁰. De ahí la ligazón entre autonomía, ética y responsabilidad.

El avance de los sistemas de IA dio lugar, en los últimos años, al desarrollo de máquinas inteligentes autónomas, esto es, sistemas tecnológicos capaces de realizar tareas de manera independiente sin intervención humana directa. Estos sistemas utilizan una combinación de IA, aprendizaje automático, sensores avanzados, sistemas de posicionamiento y navegación por satélite, cámaras de visión artificial y otras tecnologías para tomar decisiones y adaptarse a su entorno, lo que **ofrece ventajas pero a la vez abre problemas éticos ligados a la trazabilidad de la toma de decisiones y a la responsabilidad última de las acciones realizadas.**

¹⁰ En ello radicaría, según I. Kant, autor clásico de referencia para el significado ético de este concepto, la posibilidad de concebir a un acto como moral a partir de una voluntad autónoma. Por el contrario, si un acto fuese determinado por algo ajeno a la voluntad, es atribuido, consiguientemente, a una coacción externa y no es concebido como moral. Posteriores tratamientos filosóficos relacionados con la bioética y la filosofía feminista añaden al concepto de autonomía sentidos ligados a las prácticas subjetivas, los deseos, las expectativas y los comportamientos, asociados a una noción de responsabilidad social (Suarez Tomé 2021).

Los vehículos autónomos (VA) son una clase particular de sistemas autónomos inteligentes que emplean sistemas que llevan a cabo total o parcialmente la tarea de conducción y permiten aumentar significativamente la seguridad vial evitando errores humanos que, según estimaciones, causan más del 90 % de los accidentes de tráfico. Optimizan la organización del flujo de vehículos y extienden el beneficio del uso del automóvil a conductores con necesidades especiales. Tipos especiales de vehículos autónomos permiten, incluso, operar en áreas con condiciones climáticas extremas, como por ejemplo el vehículo SKÚA, desarrollado por científicos argentinos a pedido del Comando Conjunto Antártico¹¹.

Al margen de estas ventajas, este tipo de sistemas plantea una serie de dilemas ligados a la responsabilidad ética y legal en los accidentes provocados por sistemas de conducción autónoma en los que no está aún definido si la responsabilidad por la toma de una decisión con consecuencias lastimosas será del propietario del vehículo, del desarrollador del software o del fabricante. Las decisiones tomadas por vehículos autónomos tienen un carácter moral, tanto en situaciones inevitables donde se deben ponderar vidas humanas como en situaciones de tráfico diario con distintos niveles de riesgo. Estas decisiones pueden diferir de las que tomaría un conductor humano, según mostraron estudios enmarcados en la "ética del riesgo" que expusieron discordancias entre las decisiones autónomas de los vehículos y las preferencias de los participantes del estudio sobre las maniobras de conducción de los VA (Krügel y Uhl 2024)¹². Esto plantea problemas sobre la transparencia y la representatividad moral de los algoritmos utilizados y la certificación de estos vehículos en base a ellos (Bustamante Donas 2022). Por otro lado, el continuo avance de este tipo de tecnologías pone de manifiesto la necesidad de establecer criterios éticos y protocolos de seguridad combinados, por ejemplo, ligados a la vulnerabilidad frente a ciberataques, ampliamente demostrados en varios artículos científicos (Giannaros et al. 2023, Islam et al. 2023).

La interacción con sistemas autónomos avanzados desafía la noción tradicional de atribuir capacidad de acción moral sólo a los humanos; y conlleva a permanecer atentos a qué tipo de capacidades podrán adquirir en el futuro y cuál será la responsabilidad ligada a esa capacidad

¹¹ El Skua, un Vehículo Terrestre No Tripulado (UGV, por sus siglas en inglés) que además puede cumplir funciones de asistencia a las tareas que allí realizan científicos e investigadores, fue desarrollado por la startup argentina American Robotics con el aporte de ingenieros de universidades nacionales como la Universidad Tecnológica Nacional y de la Universidad Nacional de Entre Ríos (UNER).

¹² Utilizando una representación gráfica interactiva de diferentes situaciones de tráfico, esta investigación midió las preferencias de los participantes sobre las maniobras de conducción de los VA en una encuesta representativa en Alemania. Las preferencias de los participantes se desviaron significativamente de la mera evitación de colisiones. Los participantes estaban dispuestos a asumir riesgos ellos mismos en beneficio de otros usuarios de la calle. Las conclusiones a las que arriba el estudio son que, para reflejar las intuiciones morales de la población alemana, el algoritmo de conducción tendría que tener en cuenta el número de víctimas potenciales de colisión y posiblemente el tipo de usuarios de la calle al determinar el comportamiento al volante de los vehículos autónomos. Esto se descuida si la mera evitación de colisiones determina la programación de los algoritmos de conducción (Krügel y Uhl 2024).

de acción. No obstante, la responsabilidad parece seguir siendo una cualidad exclusivamente humana. La **"ironía de la autonomía"** radica, justamente, en que implementamos sistemas autónomos porque pueden hacer mejor el trabajo que el operador/conductor pero, aun así, precisamos que el operador controle que el sistema esté funcionando de manera efectiva, es decir, que la máquina computacionalmente superior en realidad necesita operadores humanos para garantizar su funcionamiento eficiente; y el humano es inevitablemente responsable de los errores, incluso si la máquina es más eficaz o precisa (Ganesh 2020).

Dentro de las **recomendaciones éticas** para para una **transición responsable y segura hacia vehículos autónomos**, se incluyen:

- Monitorear sistemáticamente y comunicar con evidencia y datos científicos sólidos las mejoras en la seguridad logradas por los vehículos autónomos.
- Fomentar una cultura de responsabilidad para los vehículos autónomos, que exceda las consideraciones ligadas a responsabilidad en la indemnización y promueva la responsabilidad de prevenir resultados indeseables en su uso.
- En relación con la ética de datos y algoritmos, la adquisición y el procesamiento de datos estáticos y dinámicos por parte de los CAV deben proteger los derechos básicos de privacidad y deben realizarse mediante procesos que sean accesibles y comprensibles para los sujetos involucrados.

Un caso de aún **mayor preocupación ética aparece con la aplicación de IA en el campo de los sistemas autónomos de armas letales, es decir, en las tecnologías de guerra autónomas.** Se prevé que este tipo de armamento alternativo al tradicional desempeñe un papel cada vez más importante en los conflictos futuros, como se evidenció en la guerra de Ucrania, en la que tanto Rusia como Ucrania buscaron automatizar diversos aspectos de los enfrentamientos con drones, dando lugar a una carrera armamentista para desplegar niveles cada vez mayores de autonomía habilitada por IA tanto en la selección de objetivos como en el ataque¹³ (Rickli y Mantellassi 2024).

Los LAWS (por su sigla en inglés *Lethal Autonomous Weapon Systems*) son sistemas bélicos controlados con algoritmos de aprendizaje automático, con la capacidad de seleccionar blancos y actuar de manera letal sin

¹³ Según los expertos, esta guerra funcionó como un banco de pruebas para muchas de las aplicaciones de la IA cuya potencial militarización se predecía en los últimos años. El dron Saker Scout de Ucrania fue capaz de encontrar, identificar y atacar de forma autónoma 64 tipos diferentes de objetivos militares rusos cuando la interferencia impedía el control del operador. Se informó, además, de una nueva variante de munición merodeadora de fabricación rusa altamente efectiva, Lancet, con capacidad de volar en enjambres de drones, encontrar y atacar objetivos de forma autónoma (D. Hambling, Forbes, 17 October 2023; F. Farrell, Kyiv Independent, 8 November 2023).

intervención humana. Es en esa capacidad de destrucción que radica el mayor grado de problematización de la autonomía adquirida: no es igual de problemática la autonomía para volar que la autonomía para seleccionar un objetivo y proceder. Además de las inquietudes que este tipo de tecnología suscita por tratarse de armas, su uso supone una serie de profundos cuestionamientos éticos que requerirán regulación: ¿cómo atribuir responsabilidad y garantizar la rendición de cuentas si los sistemas de armas autónomas letales se comportan de manera impredecible o cometen un error (por ejemplo, un crimen de guerra)? ¿Cómo conciliar la imposibilidad de atribuir responsabilidad jurídica a una máquina con el uso de sistemas de armas autónomas letales en un conflicto armado? ¿Cómo evitar que el uso de sistemas de armas autónomas letales en conflictos se traduzca en impunidad jurídica?

A la vez, surgen problemas ligados a la ética y la investigación científica, ya que muchas de las tecnologías que impulsan a este tipo de armas se desarrollan en gran medida para aplicaciones civiles, por lo que surge la cuestión de **cómo regular las tecnologías de doble uso.**

Actualmente no hay regulación internacional ni instrumento jurídicamente vinculante para establecer prohibiciones y restricciones claras sobre los sistemas de armas autónomas. El Grupo Gubernamental de Expertos de las Naciones Unidas en el Área de Sistemas de Armas Autónomas Letales presentó, en 2018, una guía de principios para asegurar el control y la responsabilidad humana sobre el uso de la fuerza y las decisiones de vida o muerte, enfatizando **que las máquinas ataquen de forma autónoma a los humanos es una línea moral que la humanidad no debería cruzar.**

Tomando en cuenta aspectos sustantivos de la **guía del Grupo de Expertos de Naciones Unidas sobre Armas Autónomas Letales**, se distinguen **recomendaciones específicas:**

- La responsabilidad humana por las decisiones sobre el uso de los sistemas de armas debe mantenerse, ya que la rendición de cuentas no puede transferirse a las máquinas. Esto debería tenerse en cuenta a lo largo de todo el ciclo de vida del sistema de armas.
- La rendición de cuentas por el desarrollo, despliegue y utilización de cualquier sistema de armas emergentes debe garantizarse de conformidad con el derecho internacional aplicable, incluso mediante la operación de esos sistemas dentro de una cadena responsable de mando y control humano.
- En el estudio, desarrollo, adquisición o adopción de una nueva arma, medio o método de guerra, el Estado debe determinarse si su empleo, en algunas o todas las circunstancias, estaría prohibido por el derecho internacional.

5. Impacto de la IA en el trabajo

En un horizonte futuro, los avances en IA tendrán un impacto significativo y progresivo en la interacción y dinámica social al alterar el mercado laboral, dejando obsoletas profesiones enteras y creando numerosos nuevos empleos y oportunidades comerciales. La IA tiene el potencial de transformar las industrias al automatizar tareas y trabajos actualmente realizados por humanos. **Esta disrupción en la fuerza laboral requerirá una fuerte intervención política y legislativa, ya que se estima que los sistemas de aprendizaje profundo superarán a los trabajadores en una amplia gama de profesiones.** Si la trayectoria actual se mantiene, los sistemas avanzados de IA tendrán también un fuerte impacto ambiental, vinculado a las crecientes necesidades energéticas de modelos con una complejidad cada vez mayor (GESDA 2023).

Tanto a nivel mundial como para el caso de Argentina, se realizaron estudios que aportan una mirada anticipatoria sobre el impacto de la IA en el empleo, con el fin de adaptar el escenario laboral a un nuevo contexto y diseñar políticas y estrategias para gestionar la transformación de una manera inclusiva y justa. **Estos informes calculan la vulnerabilidad de diversas ocupaciones al riesgo de automatización a través de IA, basándose en la proporción de tareas susceptibles de ser automatizadas.** Según un estudio del Ministerio de Trabajo de Argentina, de 2023, sobre el impacto de la IA generativa en el empleo asalariado registrado del sector privado, el 54% del empleo formal del sector (3.000.000 de empleos) tiene al menos la mitad de sus tareas con posibilidad de ser ejecutadas por este tipo de tecnología (Breard et al. 2023). Actualmente, están en curso otros estudios que ampliarán el universo de trabajadores y de tecnologías, pero estos resultados indican la importancia de posicionar en la agenda de políticas públicas el debate sobre la incorporación responsable de la IA en el entramado empresarial, productivo y sindical nacional.

A nivel global, se estima que casi el 9% de la población activa, es decir, 281 millones de trabajadores y trabajadoras, podría ver desaparecer sus puestos de trabajo por la IA generativa. Estos sistemas, a diferencia de desarrollos de IA anteriores que estaban centrados en la automatización de tareas específicas, tienen la capacidad y la viabilidad económica de impactar en empleos de múltiples industrias y diferentes niveles de habilidades, produciendo resultados similares a los humanos en áreas como el lenguaje, la creación de contenido y la interacción con el cliente (GPAI 2023).

El reciente estudio de la OIT sobre el impacto de la IA generativa (específicamente a los transformadores preentrenados generativos-GPT) en la cantidad y calidad de empleos plantea una perspectiva diversa, señalando que probablemente el impacto más importante de la tecnología sea el de “mejorar” el trabajo automatizando en algunas tareas dentro de una ocupación mientras se deja tiempo para otras tareas, en lugar de automatizar completamente las ocupaciones. Señala, además, que los posibles efectos sobre el empleo, ya sea la ampliación o la automatización, varían ampliamente

entre los grupos de ingresos de los países, debido a las diferentes estructuras ocupacionales (en los países de bajos ingresos, solo el 0,4 % del empleo total estaría potencialmente expuesto a los efectos de la automatización, mientras que en los países de altos ingresos la proporción aumentaría al 5,5 %) y subraya los efectos en relación con el género (la proporción de mujeres potencialmente afectadas por la automatización es más del doble) (Gmyrek, Berg y Bescond 2023).

Teniendo en cuenta que la IA es una herramienta que impulsará la productividad del trabajo en el marco de una política de desarrollo productivo nacional, las políticas y legislaciones no deberían restringir el acceso de las empresas y las organizaciones a la IA, sino más bien promoverlas. Pero considerando las implicancias de esta tecnología para los empleos y en consiguiente la calidad de vida de las personas, se enumeran una **serie de recomendaciones para fomentar un desarrollo productivo que sea inclusivo y de alta calidad en la transformación laboral por IA:**

- Contemplar las necesidades de formación e impulsar la capacitación para la reconversión de los perfiles ocupacionales considerando a las personas proclives a ser desplazadas por la adopción de las IA en las empresas y deban, por ello, transitar hacia otros sectores; las que requieran de nuevas habilidades y competencias; y los y las trabajadoras que deban formarse para el desarrollo / adaptación de tecnologías IA en la empresa.
- Con este propósito se requerirá el fortalecimiento de las políticas orientadas a promover la formación profesional impartida en el propio ámbito de las empresas, recomendándose medidas en las que el Estado brinde ciertos beneficios (impositivos o tributarios) como contrapartida de que las empresas lleven a cabo procesos formativos dirigidos a sus trabajadores.
- Fortalecer la negociación colectiva con cambios en los contenidos de los convenios colectivos , la implementación de medidas de preservación del empleo y el desarrollo de nuevos sistemas de información que permitan identificar las demandas de calificaciones y competencias del sector productivo para favorecer la reinserción laboral y fortalecer la asistencia a los trabajadores y las trabajadoras durante el período de formación.
- Impulsar una distribución equitativa para garantizar que las ganancias de productividad de la IA se traduzcan en bienestar para la mayoría, y un seguro ante desplazamiento por IA con colaboración público-privada gestionado por el Estado para ayudar a mitigar los cambios en el mercado laboral.
- Crear un consejo consultivo nacional sobre el futuro de la IA y el trabajo en el marco ético.

6. Gobernanza de la IA y geopolítica

Los sistemas de IA son tecnologías transfronterizas en virtud de sus funciones, estructuras, aplicaciones y capacidad de impactar masivamente en la sociedad global. Teniendo en cuenta las implicancias de su desarrollo para la vida humana, social y ambiental, **se discute actualmente cuál será el tipo de gobernanza adecuado para las tecnologías englobadas en la IA**, teniendo en cuenta tanto la necesidad de democratizar y ampliar su acceso como la de prevenir sus riesgos.

La Organización de las Naciones Unidas (ONU) expresó la necesidad de establecer un organismo regulador global de la IA que se encargue de supervisar su desarrollo, inspirada en lo que fue, para el ámbito nuclear, la creación del Organismo Internacional de Energía Atómica (OIEA). En ese sentido, **se creó en 2023 el Órgano Asesor sobre Inteligencia Artificial de la ONU** orientado a analizar y promover recomendaciones para la gobernanza internacional de la IA. Está conformado por 38 expertos de gobierno, integrantes de la industria y académicos de 33 países, con participación de voceros de empresas como Microsoft, Sony, Google y OpenAI. **En su primer reporte, de diciembre de 2023, el Órgano enfatizó el rol y la legitimidad de las Naciones Unidas para ejercer la gobernanza de la IA que emana de la membresía universal fundada en la Carta de las Naciones Unidas**, acordada universalmente, como base institucional y normativa para la acción colectiva. Asimismo destacó el esfuerzo legítimo de las leyes nacionales y regionales en la materia, aunque considerándolas como partes de un mosaico integrador que debería fundamentarse en acuerdos globales y marcos de gestión de riesgos compartidos (UN AI Advisory Body 2023). El propósito central de este organismo internacional es alcanzar una gobernanza global con participación igualitaria de todos los Estados miembros, para que la competencia geopolítica no impulse una IA irresponsable que garantice la rendición de cuentas de daños y asegure la accesibilidad de los recursos.

Al mismo tiempo, surge otro posicionamiento sobre la gobernanza de la IA que problematiza el enfoque y cuestiona, no solo las capacidades individuales de los Estados en comparación con el poder de la gobernanza multilateral, sino también el propio estatus de los Estados nacionales como los principales actores geopolíticos con autoridad exclusiva en este escenario emergente. En el marco del crecimiento exponencial y veloz de esta meta-tecnología, que permite un relacionamiento directo con los consumidores y tiene un impacto instantáneo sobre las formas de vida e interacción social, **las empresas tecnológicas y los creadores de IA muestran ser, en este escenario, agentes de poder con capacidades de intervención antes reservada a los Estados**. M. Suleyman (investigador y CEO de Microsoft AI) sostiene que para que la gobernanza global de la IA sea posible, el sistema internacional debe superar las concepciones tradicionales de soberanía e incorporar a las empresas de tecnología a la mesa. A la vez, reconoce que la legitimidad de estos actores no proviene de un contrato social pero, sin ellos, la gobernanza eficaz de la IA no tendría ninguna posibilidad (Suleyman 2023).

Parece evidenciarse que la IA no puede gobernarse como cualquier tecnología anterior y tiene un fuerte impacto en el escenario de poder geopolítico. Asimismo el desarrollo veloz de tecnologías cuánticas, neurotecnologías y tecnologías climáticas muestran que los desafíos son cada vez más globales y trascienden los esfuerzos aislados de los gobiernos nacionales. **Aun así, los gobiernos mantienen una posición privilegiada en relación con la IA, ya que determinan las prioridades estratégicas nacionales, las inversiones públicas y la normativa aplicable.**

¿Cuáles son hoy los marcos normativos o legales en relación con la IA? El **primer marco normativo universal sobre ética de la IA** —la **"Recomendación sobre la ética de la inteligencia artificial"**— fue **adoptado por los 193 Estados miembros de la UNESCO en noviembre de 2021, entre los cuales se encuentra Argentina.** Se centra en la aplicación práctica de valores y principios para regir la IA mediante recomendaciones de política concretas, haciendo especial hincapié en la inclusión, las cuestiones de igualdad de género y la protección del medio ambiente y los ecosistemas. **Se dirige a los Estados miembros, tanto en su calidad de actores de la IA autoridades responsables de la elaboración de marcos jurídicos y reguladores a lo largo de todo el ciclo de vida de los sistemas de IA, así como de la promoción de la responsabilidad empresarial.**

A la vez, los países adoptan cada vez más abordajes regionales asumiendo que los temas relativos a la IA trascienden las fronteras. **El Parlamento europeo desarrolló el primer marco legal sobre IA, la "Ley de Inteligencia Artificial de la Unión Europea", adoptada en marzo de 2024, para establecer salvaguardas en torno a la industria de estas tecnologías.** Es una normativa destinada a **regular la IA según un enfoque basado en el riesgo.** Esta ley proporciona a los desarrolladores e implementadores de IA requisitos y obligaciones con respecto a usos específicos de la IA, y **divide a los sistemas de aprendizaje automático en cuatro categorías principales en función del riesgo potencial que supongan para la sociedad.** Los sistemas considerados de alto riesgo estarán sujetos a **normas estrictas que se aplicarán antes de que entren en el mercado de la UE,** correspondiendo a los Estados miembros crear agencias nacionales de supervisión.

Por otra parte, el G7 estableció un conjunto de directrices y principios plasmados en el **"Marco de Política Integral del Proceso de IA de Hiroshima",** orientado al **uso responsable de la IA generativa, el abordaje de la desinformación, la protección de los derechos de propiedad intelectual y la gobernanza de la IA generativa.** Fue discutido y adoptado durante la Cumbre de Ministros de Digitalización del G7 en Hiroshima en abril de 2023, **buscando posicionar a los países del G7 como líderes globales en la gobernanza y regulación de la IA.**

La ciencia juega un rol central en el desarrollo de la IA y es un indicador del lugar estratégico que los países le asignan. La evolución de la producción científica sobre IA registrada a nivel mundial muestra que el país con mayor producción acumulada es China, en segundo lugar —ya con un volumen mucho menor— aparece Estados Unidos, y los

principales cinco países en términos de su producción se completan con India, Reino Unido y Alemania (OEI 2023).

¿Cuál es el **lugar que ocupa América Latina y en particular Argentina en el entramado global de producción, acceso e impacto de la IA?** El desarrollo de esta meta-tecnología se da en un escenario global de desigualdades interdependientes y asimetrías en la circulación, producción e impactos de los conocimientos. A pesar del impacto global, los avances tecnológicos se concentran en unas pocas empresas y países, e incluso, una tercera parte de la humanidad que aún está desconectada del internet.

Varios de los países de la región desarrollaron diversos pasos para una estrategia nacional de IA, pero aun la mayor parte de las regulaciones o iniciativas actuales relacionadas a la IA provienen del llamado norte global. Asimismo, se requieren inversiones y avances en infraestructuras del conocimiento. Según un informe del BID, la falta de infraestructura digital en la región surge como un reto clave en cuanto al desarrollo y la democratización de los beneficios de la IA: independientemente de las diferencias que se registren entre las zonas urbanas y rurales, predomina la falta de conectividad. En materia de género, a nivel regional hay una mujer por cada dos hombres estudiando carreras CTIM (ciencia, tecnología, ingeniería y matemáticas). En cuanto a patentes de IA, ALC se halla en los niveles más bajos en comparación con otras regiones (menos del 1%). No obstante, el ecosistema de emprendimiento está impulsando numerosos casos de uso de IA al servicio del bien social, y la ciencia y la tecnología van ganando espacios destacados (Gomez Mont et al. 2020).

La IA es un nuevo factor de producción que permite originar innovaciones en todos los sectores productivos en el marco de un nuevo paradigma tecnoeconómico y, en este sentido, es mucho más que una simple ola económica. Para competir en el mercado, la producción industrial recurre a la innovación y, de este modo, resultará estratégico impulsar la IA en la industria para procesar grandes cantidades de datos y automatizar procesos, lo que redundará en una mayor eficiencia (Anlló 2023)

El Informe científico-legislativo sobre “IA, desarrollo y bienestar” (2024) elaborado a partir de las contribuciones de destacados referentes científicos de Argentina, aboga por el establecimiento de una estrategia nacional de IA con objetivos claros, plazos y financiamiento, lo que implicará definir una estructura de gobernanza para la función de IA, multisectorial y multidisciplinaria. Propone fomentar la innovación mediante incentivos y apoyos a proyectos de IA que busquen soluciones innovadoras en biomedicina, bioinformática y agricultura entre otros. Asimismo, recomienda establecer políticas que fomenten la sostenibilidad y responsabilidad en el uso de la diversidad biológica, protegiendo tanto los derechos de las comunidades locales como el patrimonio genético del país. El aprovechamiento de la aplicación de IA en el ámbito de la agroindustria nacional dependerá de la promoción y ampliación del número de empresas de base tecnológicas (EBT). En este marco, es preciso aumentar la financiación de la investigación científica en IA, proporcionando la infraestructura y los recursos necesarios para

realizar investigaciones pertinentes. Resulta estratégico apoyar los esfuerzos de la comunidad científica local y su inserción en las dinámicas de colaboración internacional para acrecentar el ritmo de avances, publicaciones y aplicaciones de estas tecnologías, teniendo en cuenta que si bien el crecimiento de la producción científica en IA en la región fue del 209% en los últimos diez años, aún se encuentra por debajo del crecimiento a nivel mundial (+485%), y Argentina se beneficiaría de un mayor nivel de especialización y de colaboración internacional y regional (OEI 2023).

En el escenario geopolítico de discusión sobre la gobernanza de la IA y el estado de desarrollo de estas tecnologías en la Argentina y la región, es posible esbozar una serie de recomendaciones para la mejor inserción nacional:

- **Promover un abordaje regional y colaborativo en materia de IA, siguiendo las estrategias que están tomando otros bloques y potencias globales.**
- **Adoptar estrategias y hojas de ruta nacionales para la IA en el sector público y elaborar una estrategia de datos nacional.**
- **Promover la incorporación de la IA en la industria y establecer mecanismos de apoyo económico/financiero a emprendedores, PyMEs, y startups.**
- **Implementar políticas de gestión de datos que protejan la privacidad y la soberanía nacional.**
- **Desarrollar un marco regulatorio y ético específico sobre el uso ético de la IA que integre estándares globales y perspectiva local.**

6. IA y democracia

El aumento de contenidos falsos creados a partir de los rápidos desarrollos de la IA generativa comporta nuevos riesgos para la calidad democrática. Se entiende por *deepfakes*¹⁴ las imágenes y videos alterados deliberadamente con técnicas de inteligencia artificial para generar información errónea y desinformación. Las herramientas de IA generativa permiten producir imágenes y videos hiperrealistas que son una arista de lo que se conoce como “synthetic media” o “medios sintéticos”. Estas utilizan aprendizaje profundo para copiar con precisión la voz, los patrones de habla y las expresiones faciales de una persona, convirtiéndolos en herramientas poderosas de desinformación pública.

Si bien la investigación sobre el impacto social y humano de los *deepfakes* aún es escasa, ya existen estudios de neurociencia cognitiva e IA que demuestran que las personas tienen una tendencia a confundir los *deepfakes* con videos

¹⁴ Las *deepfakes* se producen mediante la tecnología de redes generativas antagónicas (GAN), un tipo de aprendizaje automático que tiene como objetivo crear contenido nuevo enfrentando entre sí dos redes neuronales de manera competitiva.

auténticos (en lugar de lo contrario), y sobreestiman sus propias habilidades de detección, lo que en combinación hace que sean particularmente susceptibles a ser influenciadas por el contenido falso (Köbis et al. 2021). Un estudio centrado en los procesos cognitivos puso el foco específicamente en el reconocimiento de los *deepfakes* ligadas a contenido científico (cambio climático), mostrando que entre el 27 y el 50% de las personas no pueden distinguir los videos auténticos de los *deepfakes*, y que la vulnerabilidad aumenta con la edad y la confianza en las fuentes de información (Doss et al. 2023).

Los riesgos éticos que implican estos contenidos es múltiple. Entre ellos la erosión de la creencia histórica de que los videos y audios son registros fiables de la realidad. Por otro lado, pueden perjudicar gravemente la imagen de las personas, ya que, justamente, son imágenes o videos manipulados y alterados para retratar falsamente a personas diciendo o haciendo cosas que nunca hicieron. Y, por supuesto, pueden tener como objetivo alterar la opinión pública con fines políticos. Existen muchos ejemplos recientes, como por ejemplo un video del presidente ruso, Vladimir Putin, circulado en junio de 2024, declarando la ley marcial y la movilización militar en las regiones fronterizas con Ucrania transmitido a través de canales de radio y televisión pirateados; o el caso de una llamada automática falsa en la que el presidente de EE. UU. Joe Biden disuadía a los votantes de participar en las elecciones primarias presidenciales de New Hampshire.

En marzo de 2024, la Asamblea General de las Naciones Unidas adoptó por unanimidad su primera resolución global sobre IA centrada, justamente, en la promoción de sistemas de IA “seguros, protegidos y confiables”. Si bien no se incluye el término *deepfake*, se reconocen los riesgos de que el contenido generado por IA sea indistinguible del contenido auténtico y se promueve el desarrollo de herramientas para la “autenticación confiable de contenido”. La resolución también aboga por aumentar la alfabetización mediática e informativa para permitir a los usuarios determinar cuándo el contenido digital fue generado o manipulado por IA (Asamblea General de las Naciones Unidas 2024).

De manera inversa, hay también ejemplos de cómo la IA puede ayudar a fortalecer la democracia. Uno es el algoritmo llamado “Bot Dog”, que funciona como un sabueso que busca discursos ofensivos y violentos y los señala. **Este tipo de desarrollo puede colaborar frente a la proliferación de discursos de odio, es decir, pronunciados en la esfera pública que tengan como objetivo promover, incitar o legitimar la discriminación, la deshumanización y/o la violencia** hacia una persona o un grupo de personas en función de la pertenencia de las mismas a un grupo religioso, étnico, nacional, político, racial, de género o cualquier identidad social (Informe LEDA 2021).

Por otro lado, la cuestión de la **protección de datos suscita una serie de consideraciones ligadas al vínculo entre IA, democracia y ética.** Los datos son fundamentales para el desarrollo y el funcionamiento de los sistemas de IA y, en este sentido, es preciso conciliar la innovación tecnológica con la protección de derechos. El uso responsable de la IA en el sector público

requiere aplicar una ética de los datos garantizando la protección de los datos personales, su uso con integridad, políticas de privacidad, el monitoreo de los datos destinados a fundamentar el desarrollo y entrenamiento de los sistemas de IA y la automatización de las decisiones (OECD-CAF 2022).

La democracia tiene la misión de promover la justicia social y una IA democrática y democratizada debe ser una herramienta para la promoción del desarrollo social y económico, que apoye el mejor desenvolvimiento de la vida humana, el desenvolvimiento intelectual y espiritual, la relación con el planeta y el dialogo y conocimiento entre las culturas. El uso de esta herramienta debe estar guiado hacia los fines de la democracia para mejorarla y hacerla cada vez más igualitaria e inclusiva.

En el marco de un **uso virtuoso de la IA en favor de la democracia y, a la vez, en favor de una democratización de la IA**, se enuncian las **siguientes recomendaciones**:

- Políticas de información ciudadana por parte del gobierno cuando los sistemas de IA perfilan o toman decisiones en el marco de la prestación de servicios públicos.
- Promoción del uso de la IA para la preservación de las lenguas indígenas y minoritarias, del patrimonio cultural y el diálogo entre las culturas.
- Fomento de investigaciones sobre la ética de la IA.
- Normativa de protección de datos que otorgue control a los usuarios, un marco de notificación y consentimiento, requisitos de transparencia con respecto al uso de los datos, reglas específicas sobre información sensible (por ejemplo, datos relativos a la salud).
- Ley o política relativa al impacto de la IA en las redes sociales, incluyendo transparencia, desinformación, información falsa y discursos de odio.

En suma, los avances en la IA abren innumerables puertas para el desarrollo de la Argentina, de la región y el mundo pero, al mismo tiempo, plantean nuevos riesgos. Asegurar una IA ética y responsable implica revisar cuáles son las implicancias éticas que conllevan estas tecnologías emergentes en temas de gobernanza de datos, neurotecnologías, diversidad, privacidad, y cómo estos riesgos pueden abordarse durante los procesos mismos de desarrollo. Además, el progreso de la IA requerirá decisiones políticas urgentes sobre su agencia y su responsabilidad.

Referencias

- Anlló, G. (2023) Inteligencia Artificial. Pistas para entender su revolución en el entramado tecno-productivo. OEI, *El estado de la ciencia. Principales indicadores de ciencia y tecnología iberoamericanos*. OEI, UNESCO, RICYT. Disponible en: <https://www.ricyt.org/wp-content/uploads/2023/12/EL-ESTADO-DE-LA-CIENCIA-2023.pdf>
- Aránguiz Villagrán, M. (2022). *Auditoría algorítmica para sistemas de toma o soporte de decisiones*. Banco Interamericano de Desarrollo. Disponible en: <https://publications.iadb.org/es/auditoria-algoritmica-para-sistemas-de-toma-o-soporte-de-decisiones>
- Asamblea General de las Naciones Unidas (2024). "Aprovechar las oportunidades de sistemas seguros, protegidos y fiables de inteligencia artificial para el desarrollo sostenible". A/78/L.49. Disponible en: https://documents.un.org/doc/undoc/ltd/n24/065/95/pdf/n2406595.pdf?token=maeYujWWr_d8qq3ezhx&fe=true
- Belda, M. J.; Katzalin Olcoz, F.; Tirado, F. (2022). "Optimization of a Line Detection Algorithm for Autonomous Vehicles On a RISC-V With Accelerator". *Journal of Computer Science & Technology*, Vol 22, Number 2. Disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/146924/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- Biblioteca del Congreso de la Nación, (2024). *Dossier Legislativo sobre Inteligencia Artificial. Proyectos de Ley, Instrumentos internacionales, Legislación nacional, Doctrina, Otros*. Biblioteca del Congreso de la Nación, Año XII n° 279, Disponible en: <https://bcn.gob.ar/uploads/adjuntos/Dossier-279-legis-nacional-inteligencia-artificial-feb-2024.pdf>
- Breard, G.; Castillo, V.; Schleser, D.; Robert, V.; Rotondo, S.; Frydman, L.; Wisznia, J. y Marraffini Llorente, G. (2023). *Impacto de la Inteligencia Artificial generativa en el empleo asalariado registrado del sector privado: diagnóstico y respuestas de política*. Subsecretaría de Planificación, Estudios y Estadísticas, del Ministerio de Trabajo, Empleo y Seguridad Social. Disponible en: https://www.argentina.gob.ar/sites/default/files/impacto_de_la_inteligencia_artificial_generativa_en_el_empleo_asalariado_registrado_del_sector_privado_0.pdf
- Buolamwini, J., Gebru, T. (2018). "Gender shades: Intersectional accuracy disparities in commercial gender classification". *Proceedings of Machine Learning Research. Conference on Fairness, Accountability, and Transparency*. Disponible en: <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
- Bustamante Donas, J. (2022). "Dilemas éticos de los vehículos autónomos: Responsabilidad ética, análisis de riesgo y toma de decisiones". *Argumentos de razón técnica*, Núm. 25. Disponible en: <https://revistascientificas.us.es/index.php/argumentos/article/view/22856/20173>
- Costa, F., Mónaco, J. A., Covello, A., Novidelsky, I., Zabala, X. ., & Rodríguez, P. (2023). "Desafíos de la Inteligencia Artificial generativa: Tres escalas y dos enfoques transversales". *Question/Cuestión*, 3(76), e844. <https://doi.org/10.24215/16696581e844>
- Costanza-Chock, S.; Harvey E.; Inioluwa, D., Czernuszenko, M.; Buolamwini, J. (2022). "Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem". *2022 ACM Conference on Fairness, Accountability, and Transparency (FACCT '22)* Disponible en: <https://doi.org/10.1145/3531146.3533213> y <http://arxiv.org/pdf/2310.02521>
- Doss, C., Mondschein, J., Shu, D. et al. (2023). "Deepfakes and scientific knowledge dissemination". *Sci Rep* 13, 13429. Disponible en: <https://www.nature.com/articles/s41598-023-39944-3#citeas>
- Dresler, M.; Sandberg, A; Bublitz, C.; et al. (2019). "Hacking the Brain: Dimensions of Cognitive Enhancement". *ACS Chem Neurosci*. 2019 Mar 20;10(3):1137-1148. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6429408/>
- Estevez Almenzar, M., Fernández Llorca, D.; Gomez Gutierrez, E.;cMartinez Plumed, F.(2022). *Glossary of human-centric artificial intelligence*. Publications Office of the European Union, Luxembourg. Disponible en: <https://op.europa.eu/en/publication-detail/>

[/publication/28702dc2-2f21-11ed-975d-01aa75ed71a1/language-en](#)

Ferrante, E. (2021). "Inteligencia artificial y sesgos algorítmicos. ¿Por qué deberían importarnos?". *Nueva Sociedad*, No 294. Disponible en: https://static.nuso.org/media/articles/downloads/1.TC_Ferrante_294.pdf

Ganesh, M.I.(2020) The ironies of autonomy. *Humanit Soc Sci Commun* 7, 157 (2020). <https://doi.org/10.1057/s41599-020-00646-0>

Geneva Science and Diplomacy Anticipator (2023). *The GESDA 2023 Science Breakthrough Radar*. Geneva Science and Diplomacy Anticipator's Annual Report on Science Trends at 5, 10 and 25 years. Disponible en <https://radar.gesda.global>.

Giannaros, A.; Karras, A.; Theodorakopoulos, L.; Karras, C.; Kranias, P.; Schizas, N.; Kalogeratos, G.; Tsois, D. (2023). "Autonomous Vehicles: sophisticated attacks, safety issues, challenges, open topics, blockchain, and future directions". *J. Cybersecur. Priv.* 2023, 3, 493-543. Disponible en: <https://doi.org/10.3390/jcp3030025>

Goering, S.; Timothy, B. y Klein, E. (2021). "Neurotechnologies and relational agency". *Philosophy Compass* vol. 16, 5.

Gómez Mont et al. (2020). *La inteligencia artificial al servicio del bien social en América Latina*. IADB. Disponible en <https://publications.iadb.org/es/publications/spanish/viewer/La-inteligencia-artificial-al-servicio-del-bien-social-en-America-Latina-y-el-Caribe-Panorámica-regional-e-instantáneas-de-doce-paises.pdf>

GPAI (2024). "Algorithmic Transparency in the Public Sector: A state-of-the-art report of algorithmic transparency instruments". Report, May 2024, Global Partnership on Artificial Intelligence. Disponible en: <https://gpai.ai/projects/responsible-ai/algorithmic-transparency-in-the-public-sector/algorithmic-transparency-in-the-public-sector.pdf>

GPAI (2023). "IA generativa, empleos y respuesta de política pública". *Innovation Workshop*, Montreal 2023. Disponible en: <https://gpai.ai/projects/future-of-work/policy-brief-generative-ai-jobs-and-policy-response-innovation-workshop-montreal-2023.pdf>

Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (2018) *Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*. Geneva, CCW/GGE.1/2018/3 Disponible en: <https://documents.un.org/doc/undoc/gen/g18/323/29/pdf/g1832329.pdf?token=hwAP5ZQibi2moxnaE&fe=true>

Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (2020). *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility*. Publication Office of the European Union, Luxembourg. Disponible en: <https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en>

Ilenca, M.; Andorno, R. (2021). "Hacia nuevos derechos humanos en la era de la neurociencia y la neurotecnología". *Análisis Filosófico*, vol. 41, núm. 1, pp. 141-185. Sociedad Argentina de Análisis Filosófico. Buenos Aires, Argentina. Disponible en: <https://www.redalyc.org/journal/3400/340067606006/340067606006.pdf>

International Science Council (2024). *A guide for policy-makers: Evaluating rapidly developing technologies including AI, large language models and beyond*, Paris, France, International Science Council. Disponible en: <https://council.science/publications/policy-makers-guide-framework-digital-technologies/>

Islam, T.; Sheakh, A.; Jui A.; (2023). "A review of cyber attacks on sensors and perception systems in autonomous vehicle". *Journal of Economy and Technology*, Vol. 1. Disponible en: <https://doi.org/10.1016/j.ject.2024.01.002>

Köbis, N. et al. (2021). "Fooled twice: People cannot detect deepfakes but think they can". *iScience* 24 Vol 24, Issue 11. Disponible en: [https://www.cell.com/iscience/fulltext/S2589-0042\(21\)01335-3#articleInformation](https://www.cell.com/iscience/fulltext/S2589-0042(21)01335-3#articleInformation)

Krügel, S., Uhl, M. (2024). "The risk ethics of autonomous vehicles: an empirical approach". *Sci Rep* 14, 960. Disponible en: <https://doi.org/10.1038/s41598-024-51313-2>

Kulesz, O. (2024). *Artificial Intelligence and International Cultural Relations*. ifa Edition Culture and Foreign Policy. Disponible en: <https://opus.bsz->

[bw.de/ifa/frontdoor/deliver/index/docId/1203/file/ifa-2024_kulesz_ai-intl-cultural-relations.pdf](https://www.bw.de/ifa/frontdoor/deliver/index/docId/1203/file/ifa-2024_kulesz_ai-intl-cultural-relations.pdf)

Larrazabal, A.; Nieto, N.; Peterson, V.; Milone, D.; Ferrante, E. (2020). "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proceedings Of The National Academy Of Sciences Of The United States Of America". *Nat/ Acad Sciences*, vol. 117, no. 23. Disponible en: <https://www.pnas.org/doi/full/10.1073/pnas.1919012117>

López-Silva, P. (2021). "Mejoras cognitivas y el problema de la inequidad: entre bioconservadurismo y bioliberalismo". *Persona & Derecho* vol. 84(1), 183-206.

López Silva, P.; Madrid, R. (2022). "Acerca de la protección constitucional de los neuroderechos: la innovación chilena". *Prudentia Iuris*, N. 94, pp.39-68. Disponible en: <https://repositorio.uca.edu.ar/bitstream/123456789/15662/1/acerca-protección-constitucional.pdf>

Muñoz, J.M., Marinaro, J.Á., Iglesias, J.A. et al. (2024). "Effects of the first successful lawsuit against a consumer neurotechnology company for violating brain data privacy". *Nat Biotechnol*. Disponible en: <https://doi.org/10.1038/s41587-024-02303-2>

OECD (2023). *G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI*. OECD Publishing, Paris, <https://doi.org/10.1787/bf3c0c60-en>.

OECD/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, Paris. Disponible en: <https://doi.org/10.1787/5b189cb4-es>.

OEI (2023) *El estado de la ciencia. Principales indicadores de ciencia y tecnología iberoamericanos*. OEI, UNESCO, RICYT. Disponible en: <https://www.ricyt.org/wp-content/uploads/2023/12/EL-ESTADO-DE-LA-CIENCIA-2023.pdf>

OIT. Gmyrek, P., Berg, J., Bescond, D. (2023) "Generative AI and jobs: A global analysis of potential effects on job quantity and quality". ILO Working Paper 96 (Geneva, ILO). Disponible en: [https://www.ilo.org/publications/generative-](https://www.ilo.org/publications/generative-ai-and-jobs-global-analysis-potential-effects-job-quantity-and-quality)

[ai-and-jobs-global-analysis-potential-effects-job-quantity-and](https://www.ilo.org/publications/generative-ai-and-jobs-global-analysis-potential-effects-job-quantity-and-quality)

Ortiz de Zárate, J. M.; Dias, J. M.; Avenburg, A. y Gonzalez Quiroga, J. I. (2024). *Sesgos algorítmicos y representación social en los modelos de lenguaje generativo (LLM)*. Fundar. Disponible en: https://fund.ar/wp-content/uploads/2024/04/Fundar_Sesgos_algoritmicos_y_representacion_social_en_los_modelos_de_lenguaje_generativo_CC-BY-NC-ND-4.0-1.pdf

Radia-Dixit, E.; Toft Djanegara, N. (2023). *Race and Surveillance Brief, Center for Comparative Studies in Race and Ethnicity (CCSRE)*. Universidad de Stanford. Disponible en: https://ccsre2023-prod.stanford.edu/sites/ccsre2023/files/media/file/race_and_surveillance_brief_final.pdf

Raji, I & Buolamwini, J. (2019). "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products". *Conference on Artificial Intelligence, Ethics, and Society*. Disponible en: <https://www.media.mit.edu/publications/actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/>

Rickli, J.M.; Mantellassi, F. (2024). "The War in Ukraine: Reality Check for Emerging Technologies and the Future of Warfare". Geneva Centre for Security Policy, Geneva Paper 34/24 April 2024. Disponible en: <https://www.gcsp.ch/publications/war-ukraine-reality-check-emerging-technologies-and-future-warfare>

Santurkar, S.; Durmus, E.; Ladhak, F.; Lee, C.; Liang, P.; Hashimoto, T. (2023). *Whose opinions do language models reflect?* Stanford University. Disponible en: <https://arxiv.org/pdf/2303.17548>

Schenberg, E. (2018). "Psychedelic-Assisted Psychotherapy: A Paradigm Shift in Psychiatric Research and Development". *Front Pharmacol*. 2018 Jul 5;9:733. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6041963/>

Secretaría de Innovación Pública (2022) *Manual de Recomendaciones para una Inteligencia Artificial Fiable*. Disponible en: https://www.argentina.gob.ar/sites/default/files/2023/06/recomendaciones_para_una_inteligencia_artificial_fiable.pdf

Simon, J., Wong, P.-H., & Rieder, G. (2022). "El sesgo algorítmico y el enfoque del diseño sensible al valor". *Revista Latinoamericana de Economía y Sociedad Digital*. <https://doi.org/10.53857/tzvn9229>

Stanovsky, G.; Smith, N.; Zettlemoyer, L. (2019) «Evaluating Gender Bias in Machine Translation». Actas de la 57a Reunión Annual de la Asociación de Lingüística Computacional (acl), 7/2019. Disponible en: <https://aclanthology.org/P19-1164.pdf>

Suárez Tomé, D.; Belli, L. (2021). "La autonomía revisitada desde la perspectiva de una bioética feminista". Herrera, M. et al. *Tratado de Géneros, Derecho y Justicia*. Buenos Aires (Argentina): Rubinzal Culzoni. Disponible en: <https://www.aacademica.org/danila.suarez.tome/37.pdf>

Suleyman, M. (2023) *The coming wave. Technology, power, and the twenty-first century's greatest dilemma*. Penguin Random House, New York.

UN AI Advisory Body (2023) Interim Report Governing AI for Humanity. Disponible en: https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf

UNESCO (2022) Recomendación sobre la ética de la inteligencia artificial, Francia. Disponible en:

https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

UNESCO, Comité Internacional de Bioética (2022). *Ethical issues of neurotechnology*. Francia. Disponible en:

<https://unesdoc.unesco.org/ark:/48223/pf0000383559/PDF/383559eng.pdf.multi>

UNESCO (2023). *Metodología de evaluación del estadio de preparación: una herramienta de la Recomendación sobre la Ética de la Inteligencia Artificial*. Francia. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000385198_spa/PDF/385198spa.pdf.multi.page=16&zoom=150,-4,627

Veit, W. (2018). "Cognitive enhancements and the threat of inequality". *Journal of cognitive enhancement* vol. 2, 404-410

Yuste, R. (2019). "Las nuevas neurotecnologías, y su impacto en la ciencia, medicina y sociedad". *Lecciones Cajal*. Universidad de Zaragoza, vol. 1, 7. Disponible en: https://zaguan.unizar.es/record/86978/files/B_OOK-2020-001.pdf