

Ciencia, Conocimiento, Tecnología e Innovación

Nuevos mapas para la diplomacia

Desafíos en un
nuevo ciclo político:
efectos en la Política Exterior

Mensajes iniciales

Los desafíos de la intersección entre diplomacia y ciencia

Desafíos de la Política exterior en el ámbito de la CTCI

Participación y liderazgo de la mujer y de las niñas en la ciencia

Transformación digital y Big Data: ejes de una transformación cultural

Derechos humanos y tecnologías emergentes: Neuroderechos e Inteligencia artificial

Capital humano avanzado: una inversión de futuro

Innovación y emprendimiento en CTCI: un nuevo ciclo de oportunidades

Comunicación y ciencia: una contribución a la apropiación social de la CTCI

Miradas prospectivas para una CTCI de futuro



Academia Diplomática de Chile

La Academia Diplomática de Chile (ACADE), al iniciar un nuevo ciclo de política exterior reafirma la prioridad del conocimiento científico y a la innovación tecnológica como factores que estimulan procesos transformadores- sociales y culturales. En este contexto, ha creado un espacio donde encontrarán una secuencia de separatas temáticas, las que incluyen las colaboraciones de reconocidos especialistas, académicos y diplomáticos.

Cada una de estas secciones abordará temas que requieren especial atención, y pueden ser objeto de iniciativas y acciones concretas. Es posible que ellas conduzcan a nuevas formas de diplomacia.

Este ciclo se iniciará con una presentación de la Directora de la ACADE, Emb. María del Carmen Domínguez, seguido de mensajes introductorios de diversos actores y autoridades nacionales e internacionales vinculados al ecosistema de ciencia, conocimiento, tecnología e innovación (CTCI).

Las separatas abordarán las siguientes áreas temáticas:

- Mensajes iniciales.
- Los desafíos de la intersección entre diplomacia y ciencia.
- Desafíos de la Política exterior en el ámbito de la ciencia, el conocimiento, la tecnología y la innovación.

- Participación y liderazgo de la mujer y las niñas en la ciencia.
- Transformación digital y Big Data: ejes de una transformación cultural.
- Inteligencia artificial, Neuroderechos: derechos humanos y tecnologías emergentes.
- Capital humano avanzado: una inversión de futuro.
- Innovación y emprendimiento en ciencia, conocimiento, tecnología e innovación: un nuevo ciclo de oportunidades
- Comunicación y ciencia: una contribución a la apropiación social.
- Miradas prospectivas para una ciencia, conocimiento, tecnología e innovación de futuro.

Nota:

Los textos y artículos, contenidos en la secuencia de Separatas Temáticas, son producto del generoso y desinteresado aporte intelectual, reflexivo y analítico de sus autoras y autores.

Esos contenidos no reflejan el pensamiento ni comprometen a la Academia Diplomática o al Ministerio de Relaciones Exteriores de Chile.

SEPARATA VI.

INTELIGENCIA ARTIFICIAL Y
NEURODERECHOS:

DERECHOS HUMANOS Y
TECNOLOGÍAS EMERGENTES.

Índice.

Nota Introductoria.

Emb. Pedro Oyarce

Algunas consideraciones acerca de la Inteligencia Artificial y la ética.

Dr. Ramón López de Mantarás.

Ciberdiplomacia, neuroderechos y control social.

Lorena Donoso

UNESCO's Recommendation on the Ethics of Artificial Intelligence as a reflection of global cooperation in diplomacy, science, technology, and policy

Gabriela Ramos

Transformación del Estado e Innovación Pública al servicio de las personas

Román Yosif

Ética e Inteligencia Artificial

Maria Florencia Attademo-Hirt

Nota Introductoria.

Emb. Pedro Oyarce¹.

La Inteligencia Artificial (IA) es uno de los asuntos centrales de nuestra era, marcada por una convergencia de tecnologías con profundas implicancias para las culturas, las sociedades y el medio ambiente. El impacto de esta en la humanidad dependerá de cómo se articulan procesos regulatorios y de la forma que se canalicen avances hacia un mundo más justo y sustentable.

La IA, además de variadas tecnologías emergentes, han abierto realidades y procesos tecnológicos que pueden entregar un valor inédito a la ciencia y el conocimiento. Hay hechos objetivos: máquinas que pueden tomar decisiones cognitivas (vehículos autónomos) o también sugerir cursos de acción (software de diagnóstico médico). Es posible, igualmente, que un programador incorpore en el software, instrucciones que, en la práctica, se traduzcan en decisiones éticas.

El Dr. Ramón López de Mantarás se refiere a las dimensiones éticas de la IA y su vínculo con las tendencias que están impactando la sociedad en la que vivimos. Ello exige una reflexión continua para orientar estos procesos transformadores de las relaciones humanas y la convivencia societal.

El autor menciona dos enfoques básicos: "*top-down*", donde los principios éticos se programan explícitamente en la máquina y esta toma decisiones; y, "*bottom-up*", donde

se espera que la máquina aprenda a decidir, éticamente, mediante la observación del comportamiento humano en situaciones reales, sin que se les enseñe reglas formales ni ninguna filosofía moral concreta. Ya existe una experiencia no relacionada con la ética, como es el aprendizaje automático en vehículos autónomos.

Este artículo es interesante pues evidencia que las dificultades pueden asociarse a temas más fundamentales relacionados a definiciones filosóficas. Concluye con razones para normar o poner límites al desarrollo de la inteligencia artificial y pensar en las máquinas inteligentes como asistentes al ser humano, insistiendo en que estos últimos son los que actúan como agentes morales. Este es un debate abierto, el cual debe considerar la protección de la dignidad y los derechos de las personas.

La profesora Lorena Donoso, en el artículo "*Ciberdiplomacia, neuroderechos y control social*", afirma que el desarrollo científico y tecnológico debe estar al servicio de las personas, respetando la vida, la integridad física y psíquica. Enfatiza el resguardo de la actividad cerebral, así como también la información proveniente de ella. Esto es consistente con los análisis de diversos especialistas que han estudiado la capacidad de procesamiento de los *smartphones* contemporáneos, lo que ha facilitado la denominación de "co-cerebros".

¹ Agradezco la colaboración editorial y de redacción de Kevin Fiegehen (Academia Diplomática de Chile).

En este artículo, la académica hace ver la neuroprotección como una manifestación del desarrollo progresivo de los instrumentos internacionales de Derechos Humanos, a partir de la Declaración Universal de Derechos Humanos y los Pactos, hasta la mencionada Declaración Universal sobre Bioética y Derechos Humanos (2005), y la Declaración Universal sobre Genoma y Derechos Humanos de la UNESCO. Identifica un conjunto de principios que deben atenderse en la neuroética y estudia las implicancias éticas, legales, políticas y sociales de la investigación neurocientífica.

La autora entrega una detallada explicación de la reforma constitucional, materializada a través de la Ley 21.383, la cual establece el desarrollo científico y tecnológico al servicio de las personas. Efectúa una especial mención a la protección del cerebro humano, la cual debe tener el carácter de *Jus Fundamental*. Recuerda la atención que se le debe conceder a los imperativos éticos en el uso de las neurotecnologías: protección de los neurodatos, derechos a la identidad personal y a la autodeterminación, equidad en el acceso a las mejoras de las capacidades cerebrales, resguardo de las personas frente a los riesgos de discriminaciones en las aplicaciones de Inteligencia Artificial (conocido como “sesgos algorítmicos”).

Concluye señalando que el nuevo texto constitucional era necesario porque responde a los avances tecnológicos de la neurociencia y, además, es coherente con los instrumentos de Derechos Humanos,

suscritos y ratificados por Chile. Este es un tema inevitable de abordar en la sociedad contemporánea, donde debe haber un cuerpo jurídico que atienda los aspectos éticos de estas tecnologías, en términos de derechos fundamentales, de responsabilidad, de transparencia y de rendición de cuentas (IA – robótica y tecnologías asociadas, consideradas de alto riesgo). Este es un esfuerzo que tiene una proyección universal, donde es fundamental salvaguardar los derechos de las personas.

Gabriela Ramos, Directora General Adjunta para Humanidades y Ciencias Sociales de la UNESCO, reflexiona en torno a cómo estructurar un marco regulatorio global sobre la inteligencia artificial. Recuerda la Recomendación sobre la Ética de la Inteligencia Artificial, adoptada en el 41º Período de Sesiones de la Conferencia General².

Este texto, no vinculante, constituye una guía u orientación para asegurar el Estado de Derecho en el dominio digital, incorporando una perspectiva de Derechos Humanos y un conjunto de estándares tendientes a considerar los efectos de la disrupción tecnológica. La idea era abordar, fundamentalmente, las implicancias para la privacidad y la protección de datos. El artículo se refiere a la precaución respecto del uso potencial de información sensible, más allá de sus objetivos fundamentales.

Es interesante recordar que UNESCO es la única organización internacional con un mandato en asuntos éticos como parte del

² El texto de la Recomendación se encuentra disponible en <https://unesdoc.unesco.org/ark:/48223/pf0000380455>

proceso de cooperación científica. En esta dirección, puede ser apropiado señalar que esta institución busca maximizar las oportunidades que ofrece la Inteligencia Artificial para fortalecer la democracia, el desarrollo y los derechos humanos.

Es claro que debemos mirar esta agenda desde el norte y el sur global, procurando converger en recomendaciones universales e inclusivas, que incorporan diferentes perspectivas culturales, asumiendo siempre la perspectiva ética de la ciencia. Esta conlleva referencias a la bioética y al manejo del genoma humano, de acuerdo a la Declaración Universal sobre Genoma y Derechos Humanos³: dado el avance de tecnologías de edición genética como CRISPR, estos son campos que adquieren una relevancia creciente.

La autora hace referencia a la política chilena en materia de IA y al Plan de Acción adoptado en 2021, indicando que nuestro país, junto a un grupo de Estados ha dado respuesta a una de las áreas más necesarias de abordar con acciones concretas en los procesos de transformación que las tecnologías generan en nuestras sociedades. Ello requiere, imperativamente, de certidumbres regulatorias. El artículo concluye convocando a una cooperación global en ética e inteligencia artificial, proceso que necesariamente debe entenderse en el contexto de un diálogo intercultural, indispensable para encontrar acuerdos frente a realidades complejas.

Una dimensión de alto interés que contiene esta separata es la reflexión de Román Yosif, quién, desde el Laboratorio de Gobierno, unidad especializada del Ministerio de Hacienda, plantea la necesidad de profundizar la innovación pública al servicio de las personas, invitando a ser parte activa de la discusión global sobre la transformación del Estado.

Describe un proceso tecnológico, altamente disruptivo, que genera cambios en la convivencia, acentuados por la pandemia. Yosif califica la emergencia de bienes digitales como uno de los principales productos de la cuarta revolución industrial. Redes sociales como nuevas formas de construir identidades, el almacenamiento de grandes volúmenes de datos en la nube (*cloud*) y la tecnología para el respaldo del intercambio de activos digitales (*blockchain*) son prácticas que modifican las formas de vincularnos con el entorno, donde la versión digital es claramente tangible.

El autor plantea que estas determinantes nos convocan como Estado para producir valor público y desarrollar soluciones para enfrentar estos procesos transformadores. En este contexto, sugiere el enfoque de innovación del sector público y la contribución del Laboratorio de Gobierno, como una institución capaz de aportar al diseño, desarrollo e implementación de políticas y gestión pública: una institución *usuario-céntrica* que procura cruzar todo el proyecto de carácter público, instalando a la innovación como una dimensión de gobernabilidad, crecientemente relevante

³Disponible en: <https://en.unesco.org/themes/ethics-science-and-technology/human-genome-and-human-rights>

para la calidad de la democracia. Se exponen ejemplos que permiten ver la transversalidad de temas de políticas públicas que han sido abordados a través de la metodología de trabajo del Laboratorio de Gobierno.

El artículo concluye con una reflexión, en torno a los principales desafíos del actual ciclo político en Chile. Hace ver que la materialización de transformaciones estructurales, fundamentalmente en el ámbito de los Derechos Económicos, Sociales y Culturales (DESC), exige un sector público con nuevas capacidades de innovación. Sin un Estado amigable, ágil, eficiente, digitalizado y más profesional, muy probablemente, sería difícil canalizar las aspiraciones ciudadanas.

Esta separata concluye con algunos comentarios del trabajo multilateral regional en inteligencia artificial (IA). La representante en Chile del Banco Interamericano de Desarrollo, María Florencia Attademo-Hirt, analiza el potencial de la IA para hacer más eficiente tanto el uso de los recursos como la gestión de las organizaciones. En definitiva, la calidad de vida de la gente, simplificando procesos para los usuarios y adelantándose a necesidades futuras.

Este es un tema en el cual el BID ha asumido los lineamientos y estándares de la OCDE, en contextos y realidades regionales. La idea es contribuir a una aplicación efectiva de un área en constante evolución: la civilización de la inteligencia artificial. Esta incluye diversas tecnologías y complejidades en torno a ejes habilitantes como la infraestructura, las definiciones de acciones para adoptar la IA y la dimensión ética.

Se trata de abordar procesos de una revolución industrial en acelerada evolución transformadora, con las perspectivas del sur global para evitar reproducción de sesgos y otros efectos poco favorables, para sectores amplios de la región. Es imperativo entender que estos escenarios conllevan profundos cambios de la geopolítica y del poder, donde es determinante el control de los datos, el desarrollo de los algoritmos y la hegemonía de la IA.

Conviene también tener presente que esto puede ir acompañado de un proceso de obsolescencia de las instituciones que carecen de capacidad de seguimiento y acción en esta esfera.

El artículo efectúa igualmente una referencia a los algoritmos como parte esencial de los sistemas de IA, ya que están encargados de procesar los datos para realizar diversas funciones (predicciones, clusterizaciones y caracterizaciones). Se plantea, además, una dimensión esencial al abordar la IA: la necesidad de determinar si un proyecto cumple con estándares éticos en la aplicación de esta. Este es un tema que la UNESCO ha analizado, aprobando una Recomendación sobre la ética de la IA (abordada, de igual manera, en esta separata).

Como una forma de cautelar el diseño, desarrollo e implementación ética de los algoritmos, el BID viene trabajando la iniciativa “fAIr LAC”, cuyo objetivo es el uso responsable de la IA basado en los ya mencionados principios y orientaciones de la OCDE. Se advierte una preocupación por

operacionalizar tres dimensiones: la articulación de una red público-privada, generar capacidades para una adopción responsable de la tecnología y la mitigación de los riesgos.

El BID, a través de “fAIr LAC” ha generado una red de expertos y hubs en diferentes países de la región, desarrollando iniciativas y proveyendo soluciones a través de programas que han generado capacidades orientadas al sector público y privado.

Se aspira a fortalecer el ecosistema emprendedor, una de las principales fuentes de soluciones algorítmicas, trabajando desde la oferta de startups tecnológicas. La representante del BID menciona también iniciativas en países de la región, patrocinadas o impulsadas por el Banco, donde la IA está mejorando la calidad de vida de las personas, en el ámbito de las enfermedades crónicas, la mitigación del contagio del COVID-19 y del empleo.

Este artículo permite observar cómo un banco regional es capaz de entregar capacitación e intercambio de buenas prácticas a través de proyectos, para profundizar un posicionamiento con pertenencia a los países de la región, siempre asumiendo que la IA agrega valor a las estructuras económicas, organizacionales y sociales, generando condiciones para la gobernabilidad regional de estos procesos con alcance global.

La diplomacia científica debe estar en condiciones de actuar en los procesos regulatorios, en la construcción de proyectos regionales y globales, y en cautelar diversos principios éticos que han sido objeto de un trabajo multilateral.

Algunas consideraciones acerca de IA y Ética

Ramon López de Mantarás Badia⁴

Algunos investigadores de la Inteligencia Artificial (IA) piensan que los computadores deberían ser capaces de razonar y tomar decisiones en base a criterios morales explícitos. Por ejemplo, Susan Anderson y Michael Anderson (en *Machine Ethics*, 2011), dicen: "idealmente, nos gustaría poder confiar en máquinas para tomar decisiones éticas correctas por sí mismas, y esto requiere que creemos una ética para las máquinas".

Su argumento es que si estas máquinas pueden tomar miles de decisiones cognitivas por sí mismas (por ejemplo, en el caso de los vehículos autónomos, cuando frenar, cuando detenerse, cuando cedió el paso, etc.), también deberían poder tomar decisiones éticas. Este punto de vista es debido a que estos investigadores no distinguen entre razonar sobre asuntos fácticos y cuestiones morales, porque creen, erróneamente que las decisiones morales se pueden siempre racionalizar. De hecho, en 1853, John Stuart Mill, en su obra *"Sobre el Utilitarismo"*, ya afirmaba que la facultad de tomar decisiones morales es una rama de nuestra razón.

Se han sugerido dos enfoques generales para que las máquinas inteligentes tomen decisiones éticas por sí mismas: el enfoque "top-down" (de arriba a abajo) y el enfoque "bottom-up" (de abajo a arriba). En el

enfoque de *top-down*, los principios éticos se programarían explícitamente en la máquina, por ejemplo, en el sistema de conducción automática del vehículo. Así, las máquinas podrían obedecer "leyes" al estilo de las de la Robótica de Asimov (aunque no creo que ninguna máquina pueda realmente entender las leyes de Asimov) o "leyes" más ad-hoc, o incluso una filosofía moral general, como el imperativo categórico de Kant, el utilitarismo de Stuart Mill o alguna otra forma de consecuencialismo.

El punto principal es que un programador incorpore instrucciones, por ejemplo, en el software de conducción automática de vehículos, de modo que la máquina proceda en condiciones específicas de la manera más ética posible. Es decir, la máquina tome decisiones éticas basadas en la filosofía moral que se ha implantado en su software basado en inteligencia artificial.

Los críticos del enfoque *top-down* reconocen las dificultades inherentes de adherirse a cualquier filosofía moral en particular, dado que cualquiera de ellas, en algún momento u otro, conducirá a acciones y resultados que muchos encontrarán moralmente inaceptables, por el hecho de que no hay principios éticos universales. Para tomar un ejemplo familiar: surgirían preocupaciones obvias si, de acuerdo con el

⁴ Miembro del Instituto de Estudios Catalanes, del Instituto de Investigación en Inteligencia Artificial y del Consejo Superior de Investigaciones Científicas (CSIC).

consecuencialismo, un automóvil concluyera que sería preferible chocar contra el menos lujoso de dos automóviles aparcados, con el objetivo de minimizar el coste del daño causado cuando el daño es inevitable.

Las filosofías morales han desarrollado variantes que intentan abordar estos problemas. Sin embargo, entre estas escuelas éticas, también dentro de ellas, hay debates importantes que resaltan las dificultades que enfrentan recurriendo a filosofías particulares para que sirvan como sistemas de guía moral por las máquinas. Por ejemplo, hay un desacuerdo bien conocido y significativo sobre si es posible cuantificar la utilidad e incluso sobre si hay que tener en cuenta o no distintos grados de utilidad. Los consecuencialistas continúan enfrentándose a desafíos tales como estimar los efectos a largo plazo o determinar para quién se deben tener en cuenta estas mismas. Por ejemplo, la mayoría de los experimentos mentales del dilema del tranvía asumen que arrollar a cinco personas es obviamente peor que arrollar a una. Sin embargo, la gente no concede el mismo valor moral al hecho de que las posibles víctimas sean niños o un asesino psicópata convicto.

Es pues prácticamente imposible programar una máquina que sea capaz de tomar decisiones morales por sí misma, ya sea utilizando una filosofía moral concreta o una combinación de filosofías morales. Pero uno podría preguntarse: "Si los humanos pueden hacerlo, ¿por qué no las máquinas inteligentes?". Además, si bien se puede argumentar que los individuos tomamos decisiones morales sobre la base de esta o aquella filosofía, los humanos, primero,

adquirimos valores morales de quienes nos educan y luego modificamos estos valores al estar expuestos a diversas aportaciones de nuevos grupos y culturas, desarrollando, gradualmente, nuestra propia moral personal. Además, estos valores están influenciados por principios sociales particulares que no se limitan necesariamente a ninguna filosofía moral concreta. En resumen, el enfoque de arriba hacia abajo es muy poco verosímil.

En el segundo enfoque, el *bottom-up*, se espera que las máquinas aprendan a tomar decisiones éticas mediante la observación del comportamiento humano en situaciones reales, sin que se les enseñe ninguna regla formal ni estén equipadas con ninguna filosofía moral en particular. Este enfoque se ha aplicado ya a aspectos no relacionados con la ética como el aprendizaje automático en vehículos autónomos. Por ejemplo, varios investigadores han utilizado el aprendizaje automático para mejorar la capacidad de detección de peatones. En particular, se ha logrado que un automóvil sin conductor aprenda aspectos básicos de conducción automática observando sólo 72 horas de datos de conducción humana. Sin embargo, considerar estos experimentos como precedentes para aprender la conducta ética es suponer que no hay una diferencia significativa entre aprender a responder de manera diferente a los semáforos verdes, rojos y amarillos, y aprender a comprender y apreciar el imperativo moral de reducir la velocidad, o incluso se detenga el automóvil, para evitar salpicar a un peatón que camina por la acera cuando hay grandes charcos de agua por donde circula el automóvil.

El problema es que situaciones como ésta son extraordinariamente numerosas y cada una puede pasar en contextos diferentes. A veces es un perro el que cruza y provoca un accidente, a veces un autobús escolar, etc. Un automóvil sin conductor necesitaría demasiados años de aprendizaje observando comportamiento humano como para aprender comportamiento ético de esta manera.

Para acelerar este proceso, algunos han sugerido que los automóviles sin conductor podrían aprender de las decisiones éticas de millones de conductores humanos, mediante algún tipo de sistema de agregación, como una especie de razonamiento grupal o basándose en lo que se conoce como "*wisdom of the crowd*" (sabiduría de las multitudes). Cabe señalar, sin embargo, que esto puede dar como resultado que los automóviles adquieran preferencias muy poco éticas, ya que no está nada claro que la mayoría de los conductores sigan un estándar de comportamiento digno de ser emulado por los vehículos autónomos. Si aprenden lo que hacen muchas personas, los vehículos inteligentes aprenderán por no respetar los límites de velocidad, llevar a cabo adelantamientos peligrosos, etc. También se debe tener en cuenta que, cuando se enfrentan al tipo de elecciones planteadas por el dilema del tranvía, las personas ejecutan acciones basadas en reflejos automáticos, en lugar de basarse en deliberaciones y la toma de decisiones éticas. Es decir, observar a las personas no les enseñará a estas máquinas lo que es ético, sino lo que es común.

En resumen, tanto el enfoque *top-down* como el *bottom-up* se enfrentan a serias dificultades. Estas no son tecnológicas, sino que se refieren a la esencia de las filosofías éticas utilizadas por los humanos. Aun así, estas dificultades no son nada en comparación con las planteadas por la cuestión de si las máquinas inteligentes pueden convertirse o no a agentes morales, ya que para ello deberían ser inteligencias artificiales "fuertes" en el sentido del filósofo John Searle. Ello, en mi opinión, es altamente improbable por no decir imposible.

Otra fuente importante de dificultad en el problema de incorporar ética en las máquinas es el concepto de autonomía. La autonomía es, de hecho, un concepto gradual. Algunas herramientas no tienen autonomía y uno debe explicar plenamente sus acciones mediante fuerzas externas a ellas. Por ejemplo, un martillo que golpea un clavo no tiene autonomía incluso cuando falla, porque se puede demostrar que el error fue debido a la torpeza de la persona que lo usó. Pero se puede afirmar que un sistema GPS rudimentario tiene un pequeño grado de autonomía para que, cuando se utiliza para saber cuál es la mejor manera de ir de un punto A un punto B, compara varias opciones y recomienda una, pero su recomendación se basa en un algoritmo creado por humanos, que calcula la ruta más corta, o la que menos tiempo tardará en recorrer, o algún otro criterio preprogramado.

Una cantidad significativa de autonomía se da cuando la máquina recibe pautas, incluso algunas contradictorias, y se le ordena que se base en la información que adquiere a

medida que avanza, extrayendo conclusiones por sí misma: por ejemplo, una versión más avanzada de un sistema GPS, que identifica la situación del tráfico y recomienda una ruta alternativa. Se considera que las máquinas equipadas con IA pueden actuar de forma mucho más autónoma que las que no están equipadas con ella.

Efectivamente, sabemos que los algoritmos de aprendizaje automático toman cada vez más decisiones sobre, por ejemplo, la concesión o no de créditos, diagnósticos médicos, recomendaciones personalizadas, publicidad y oportunidades laborales. El problema es que no pueden dar explicaciones sobre los motivos de sus decisiones.

Algunos creen que las máquinas pueden llegar a la autonomía total. Un (mal) ejemplo de ello son las armas letales autónomas, que eligen sus propios objetivos sin intervención humana y por consiguiente sus decisiones no se pueden abortar. De hecho, incluso estas máquinas están limitadas a las misiones que les asigna un humano y sólo son "libres" para elegir sus objetivos, desde la base de su programación. Su autonomía es pues también limitada: el especialista en ética militar George Lucas Jr., en *"Killing by remote control: the ethics of an unmanned military"*, señala que en los debates sobre la ética de las máquinas a menudo se confunde la autonomía de las máquinas con la autonomía moral.

Las aspiradoras autónomas y el misil Patriot son autónomos en el sentido de que realizan sus misiones, adaptándose y respondiendo a

circunstancias imprevistas con una supervisión humana mínima, pero no en el sentido de que puedan autónomamente decidir cambiar o abortar su misión en base a objeciones morales. Brad Templeton en *"Machines of loving grace: the quest for common ground between humans and robots"*, lo expresó bien cuando afirmó que un robot sería verdaderamente autónomo el día que se le ordene llevar a cabo una tarea y, en lugar de hacerla, decida ir a pasear. Las máquinas son pues, en última instancia, herramientas de los seres humanos que las diseñan y fabrican.

Otra fuente importante del concepto erróneo que parece subyacer en gran parte de las discusiones sobre ética de las máquinas, se encuentra en las discusiones públicas sobre la IA, incluso en entornos académicos. Me refiero a la suposición de que existe esencialmente un solo tipo de IA. En realidad, hay dos tipos diferentes de IA. El primer tipo involucra software que busca razonar y tomar decisiones cognitivas de la misma manera en que lo hacemos las personas y, por lo tanto, es una IA que aspira a poder reemplazar a los humanos. Esta IA busca reproducir en el ámbito digital los procesos en los que se involucre a los cerebros humanos cuando deliberan y toman decisiones. En otras palabras, serían inteligencias artificiales "fuertes", dicho de otro modo, no simulan ser mentes, sino que serían mentes.

Sin embargo, hay otra: la IA específica. Esta, simplemente, busca proporcionar asistencia inteligente a los humanos. Este tipo de IA solo requiere que las máquinas sean mejores que los humanos en algunas tareas muy

concretas, y que lo hagan de manera efectiva dentro de los parámetros establecidos por los humanos o bajo su estrecha supervisión. Por ejemplo, la IA que colabora en el cuidado de los niños junto con los padres durante periodos cortos de tiempo dentro del hogar, o las diseñadas para el cuidado de las personas mayores, trabajando en equipo con el personal humano, y haciendo algunas tareas por su cuenta (por ejemplo, recordando a los pacientes el momento de tomar la medicación) y alertando al cuidador humano en determinadas condiciones, como, por ejemplo, cuando un paciente que abandona la habitación cuando no debiera.

Los términos, todos ellos sinónimos, utilizados en referencia a esta IA como herramienta sofisticada de ayuda a los humanos son "Inteligencia Aumentada", "Inteligencia Amplificada" y "Cognición Aumentada".

En el contexto de los vehículos autónomos, parece que Tesla simplemente busca, al menos por ahora, proporcionar a los conductores humanos ayuda basada en inteligencia artificial que haga la conducción más segura. Pero se advierte al conductor humano que, incluso cuando el automóvil está en modo de piloto automático, debe mantener las manos en el volante en todo momento y estar siempre atento a la carretera y al entorno del vehículo. Hay una remota posibilidad de que la diferencia entre los dos tipos de IA pueda reducirse, o incluso desaparecer, a muy largo plazo.

Desafortunadamente, todavía prevalece la opinión de que las IA actuales, con capacidad

enormemente limitadas, ya son capaces de actuar como si fueran IA generales y fuertes. Todo ello es pertinente porque si (a pesar de la muy alta improbabilidad de que ocurra) las máquinas inteligentes llegan a tener mentes, entonces podrían efectivamente reemplazar a los humanos y actuar realmente por sí mismas. Esto significa que podrían dejar a los humanos fuera de los procesos de toma de decisiones (por ejemplo, en el caso de armas letales autónomas, máquinas que, una vez tomada la decisión de disparar no se pudiera revertir dicha decisión). En este escenario, más cercano a la ciencia ficción que a la ciencia real, las máquinas inteligentes deberán tomar decisiones morales por sí mismas y, por lo tanto, deberán ser tratadas como si fueran agentes morales que asuman la responsabilidad de sus decisiones. Ya no se podría considerar solo a los programadores, como fabricantes y usuarios sino como agentes morales. Bajo este escenario, la cuestión, aparentemente sencilla, de quién o qué multar cuando un automóvil sin conductor infrinja las leyes de tráfico se vuelve una cuestión compleja.

Hay razones muy poderosas para poner límites al desarrollo de la IA y pensar en las máquinas inteligentes en tanto que asistentes, en lugar de agentes con una mente que les permita funcionar por sí mismos. Una de las principales razones es que, aunque las máquinas inteligentes parecen muy buenas para llevar a cabo algunas funciones que los humanos realizamos (por ejemplo, memorizar), son muy malas en otras funciones (por ejemplo, preocuparse por aquellos a quien sirven). Por lo tanto, los robots para el cuidado de

ancianos son buenos para recordarles a los pacientes que tomen sus medicamentos, pero no para consolar cuando se afligen o tienen miedo. Esto lo deben hacer los cuidadores humanos.

En definitiva, en una sociedad formada por máquinas inteligentes y seres humanos debería exigirse que sean estos últimos los que actúen como agentes morales. Los seres humanos somos y debemos seguir siendo los únicos agentes morales. Aunque a muy largo plazo la IA fuerte fuera posible, convendría plantearse si deberíamos desarrollarla. Esa es, en mi opinión, la primera y más importante decisión ética en relación con la IA que debemos tomar.

Aprecio que la Academia Diplomática de Chile (ACADE) reflexione sobre temas que involucran a la tecnología y a la ética, lo que ciertamente exigirá definiciones en donde debería prevalecer la filosofía moral y la toma de decisiones por parte de seres humanos.

Es interesante advertir cómo en el sistema multilateral, tanto a nivel global como regional, se están planteando estas materias: de hecho, la propia UNESCO, a través de la Recomendación para la Ética de la Inteligencia Artificial entrega orientaciones que deberán estimular procesos reguladores en diversas sociedades y contextos. Estos son ámbitos donde la acción diplomática es también necesaria, para lo cual se requiere

una formación a fin de comprender los alcances del desarrollo del conocimiento científico y la innovación tecnológica.

Ciber diplomacia, neuro derechos y control social

Lorena Donoso Abarca⁵

Este trabajo busca realizar una reseña explicativa del sentido y alcance de la reforma constitucional coloquialmente denominada “consagración de los neuroderechos” o simplemente “neuroderechos”. Para ello, esbozamos algunos conceptos de base, y analizamos los instrumentos internacionales de derechos fundamentales que cimentan las bases de la reforma tramitada en Chile. Sobre esta base, revisamos las ideas matrices del nuevo texto del inciso final del artículo 19 n° 1, y las analizamos en concordancia con normas vigentes en materia de derecho sanitario. No realizamos un análisis acabado de la normativa de protección de datos personales ni a la ley de neuroderechos, porque, al estar en tramitación, no contamos aún con textos afinados.

Antecedentes

“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.”

Esta nueva disposición constitucional chilena no es una construcción ajena a nuestro sistema jurídico, sino que es una manifestación del desarrollo progresivo de derechos fundamentales internacionalmente reconocidos y reflejados en nuestra legislación interna.

Para entender el sentido y alcance de esta norma, en primer lugar, daremos noticia de la preocupación del legislador, para la memoria histórica de la reforma, luego acudiremos al elemento gramatical y

científico para entender el objeto regulado, para luego referirnos al elemento sistemático, a través del análisis de cuerpos legales relacionados.

El elemento histórico

Las capacidades de las tecnologías para leer, interpretar y condicionar la actividad cerebral o más ampliamente del sistema nervioso central de la persona⁶, a través de interfaces cerebro-cerebro, cerebro – máquina y otros dispositivos dotados de inteligencia artificial ha llamado la atención a los juristas, desde la óptica de los derechos fundamentales. En especial se ha llamado la atención respecto de que el empleo de estos dispositivos podría afectar indebidamente la libertad, igualdad y dignidad de las personas.

En materia de libertad se advierte que el libre albedrío y la autodeterminación pueden verse afectados por manipulaciones

⁵ Abogada, Magister en Informática y Derecho, Profesora Asociada de la Universidad de Chile.

⁶ El Sistema nervioso central comprende el cerebro, las meninges y la médula espinal.

indebidas de la persona a través de estos mecanismos.

En relación con la igualdad se ha llamado la atención en la necesidad de garantizar el acceso equitativo a las mejoras cognitivas o de calidad de vida que proveen estos equipos, con independencia a la situación económica y lugar en que la persona habite. Adicionalmente, se teme la necesidad de proteger a la persona frente a intrusiones ilegítimas en su información personal y de los sesgos algorítmicos que pueden conllevar los procesos automatizados de adopción de decisiones.

Desde la ciencia se ha solicitado que se establezca una regulación estricta respecto del empleo de neuro tecnologías e inteligencia artificial en ser humanos. Se sugiere adoptar medidas tales como limitar la posibilidad de renunciar a los datos neuronales o autorizar la implantación de nuevos datos en el cerebro o sistema nervioso central. Proscribir los pagos asociados a este tipo de pruebas y experimentaciones. Establecer requisitos estrictos para la cesión a terceros de los datos neuronales, y establecer obligaciones de minimización de datos, esto es, limitar la información que los dispositivos podrán transmitir a la máquina a aquellos estrictamente necesarios para la satisfacción de la finalidad legítima e informada de que se trate.

Rafael Yuste ha señalado que es necesario discutir y acordar un tratado internacional que defina las acciones que se considerarán prohibidas en relación con las neuro tecnologías e inteligencia artificial, en la que

se visibilice los derechos neuronales y que proteja a las personas para que la información de sus cerebros goce del estatus de “dato de salud”, y proscribiendo su explotación con fines de lucro o manipulación de las personas.

El elemento gramatical

Para comprender los alcances de la necesidad, es importante tener a la vista algunos conceptos. En primer lugar “ciencias neuronales”, entendidas como aquellas que se dedican al estudio del funcionamiento del cerebro; luego, “neuro ética”, que estudia las implicancias éticas de la investigación sobre el cerebro. En tercer lugar, “neuro tecnologías”, concepto que engloba a las ciencias y técnicas que se ocupan del desarrollo de dispositivos capaces de leer, analizar, o escribir en el cerebro.

Cortina (2010) nos llama la atención respecto de que “las neurociencias son ciencias experimentales que intentan explicar cómo funciona el cerebro, sobre todo el humano, y dieron un paso prodigioso al descubrir que las distintas áreas del cerebro se han especializado en diversas funciones y que a la vez existe entre ellas un vínculo”.

En el plano de las aplicaciones o usos de este tipo de desarrollos, se han mencionado los siguientes:

- **Neuromarketing:** Análisis de las decisiones de compra y diseño de estrategias de inducción.
- **Neuro política:** Análisis de los procesos de decisión política de las personas y diseño e implementación de estrategias de inducción.

- **Ciencias neuronales:** Bajo esta denominación se incluyen las disciplinas que se ocupan del estudio del funcionamiento del sistema nervioso, el cual está compuesto por el cerebro, la médula espinal y las redes de células sensitivas o motoras, en todo el cuerpo, a fin de comprender su funcionamiento, en sus distintas dimensiones, tanto fisiológicas como cognitivas y emocionales. Serán tales la neuro farmacología, neuro genética, neuro inmunología, neuro epidemiología, neuro intensivismo, etc.
- **Neuro tecnología:** "el conjunto de métodos e instrumentos que permiten una conexión directa de dispositivos técnicos con el sistema nervioso"⁷.
- **Neuro tecnología clínica:** con ello se hace referencia a exámenes tales como el electroencefalograma, La estimulación magnética transcraneal, las neuro imágenes funcionales.

La consagración de la neuro protección como manifestación del desarrollo progresivo de los Instrumentos Internacionales de Derechos Humanos.

Frente a esta problemática, desde el derecho se ha recomendado analizar este problema desde los tratados y declaraciones de derechos humanos que se han adoptado a lo largo de la historia.

A vía ejemplar, la Declaración Universal de Derechos Humanos de 1947 consagra un andamiaje en el cual cimentar la construcción de la obligación de que el desarrollo

tecnológico no atente contra la persona humana.

Es así como en su artículo 3, el derecho a la vida, la libertad y a la seguridad de su persona; en el artículo 5, protege a la persona respecto de la tortura o tratos crueles, inhumanos o degradantes; en el 6, consagra el derecho al reconocimiento de la personalidad; en el 7, el derecho a la no discriminación; en el 12, la protección de la persona respecto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, y de ataques a la honra o a la reputación de la persona; en el 18, el derecho a la libertad de pensamiento, de conciencia y de relación, tanto adoptar un credo como a cambiarlo y manifestarlo tanto en público como en privado; y el 19, el derecho a la libertad de opinión y de expresión, que incluye el de no ser molestado a causa de sus opiniones, el de investigar y recibir informaciones y opiniones, y el de difundirlas, sin limitación de fronteras, por cualquier medio de expresión. Finalmente, como norma de clausura, el artículo 30 establece que "nada en la presente declaración podrá interpretarse en el sentido de que confiere de derecho alguno al Estado, a un grupo o a una persona, para emprender y desarrollar actividades o realizar actos tendientes a la supresión de cualquiera de los derechos y libertades proclamados en esta Declaración".

Como podemos apreciar, este instrumento fundamental, que dota de un marco común a los países miembros de las Naciones

⁷ Müller, O., & Rotter, S. "Neurotechnology: Current Developments and Ethical Issues". *Frontiers in systems neuroscience*, 11, 93. 2017. Disponible en:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5733340/> (Abril, 2019).

Unidas en materia de derechos humanos, mantiene su plena vigencia hoy en día, pues sus normas son de carácter general y su redacción es neutra y atemporal. Sus imperativos se ven complementados con los instrumentos sectoriales, de los cuales sólo nos referiremos a aquellos que nos permiten visualizar los principios atinentes al objeto de nuestro análisis, cual es la modificación constitucional en la que se establece el derecho a la neuro protección.

Otro instrumento relevante en lo que importa al acceso equitativo de las mejoras cognitivas o de otra índole, asociada a las neuro tecnologías, es el Pacto Internacional de Derechos Económicos, Sociales y Culturales (PIDESC), que en su artículo 12 que reconoce el derecho a toda persona al disfrute del más alto nivel posible de salud física y mental, el artículo 13, relacionado al derecho a una educación que se oriente hacia el pleno desarrollo de la personalidad humana y del sentido de su dignidad, y debe fortalecer el respeto por los derechos humanos y libertades fundamentales.

Igualmente, relevante en este ámbito es que este artículo, a continuación, dispone que los Estados Parte convienen en que “la educación debe capacitar a todas las personas para participar efectivamente en una sociedad libre” (art.12) y reconocer el derecho de toda persona a “gozar de los beneficios del progreso científico y de sus aplicaciones (art. 15), para lo cual se deberán adoptar las medidas necesarias para la conservación, el desarrollo y la difusión de la ciencia y la cultura, respetar la indispensable libertad para la investigación científica y para la actividad creadora y reconocer los

beneficios que derivan del fomento y desarrollo de la cooperación y de las relaciones internacional en cuestiones científicas y culturales.

En primer lugar, la Declaración de Helsinki de la Asociación Médica Mundial, AMM, las Pautas Éticas Internacionales para la Investigación Biomédica en Seres Humanos (2002) del Consejo de Organizaciones Internacionales de las Ciencias Médicas (CIOMS), la Declaración Universal sobre Bioética y Derechos Humanos (2005); y la Declaración Universal sobre el Genoma Humano y Derechos Humanos, de la UNESCO, de 1998, aportan importantes elementos a la discusión.

Respecto de este último instrumento destacamos los principios que se reconocen respecto de la investigación en materia de genoma humano, a saber:

- **Dignidad**, en base al cual se propugna la preminencia de lo humano y la no discriminación de las personas en el marco de este tipo de investigaciones.
- **Libertad**, conforme al cual se debe exigir el consentimiento libre e informado de la persona de manera previa a su participación en una investigación de esta naturaleza. Adicionalmente, habrá de procurarse los mecanismos que permitan asegurar la protección a las personas en situación de vulnerabilidad, tales como los niños, niñas y adolescentes, personas aquejadas por una enfermedad o condición de vida que les impida decidir libremente. Asimismo, en base a este principio, la

persona podría decidir libremente si quiere saber o no saber.

- **Información**, que exige que se informe a la persona sobre las finalidades de la investigación, los efectos que pueda tener sobre su persona, los lineamientos de los protocolos de investigación, los usos que se dará a la información que emane de la investigación, a quien se comunicarán dichos resultados.
- **Adecuación o razonabilidad**, conforme al cual se debe contar con protocolos de investigación aprobados por comités ético-científicos, la elección de medios y procedimientos que minimicen los riesgos para la salud de las personas.
- **Confidencialidad**: la reserva de la información de la persona, de forma tal que solo tenga acceso a ella las personas que, dentro de sus facultades, la requiere, estableciendo mecanismos de seguridad eficientes.
- **Justicia**: Se trata de garantizar el acceso universal y equitativo al nuevo conocimiento, establecer mecanismos de cooperación científica y cultural a nivel internacional. Reconocimiento a las identidades culturales y hacerse cargo de la responsabilidad por los efectos nocivos que se pueda producir en las personas a raíz de las investigaciones.

Tratándose de las Pautas CIOMS, adoptadas en el seno de la Organización Mundial de la Salud (2002) rescatamos los siguientes principios complementarios:

- **Respeto a las personas**: “Toda investigación con seres humanos debe

llevarse a cabo con el debido respeto y preocupación por los derechos y el bienestar de las personas participantes y las comunidades donde se realiza la investigación” (p.2). Ello implica el reconocimiento de la autodeterminación del ser humano, poniendo en la base de la legitimación el consentimiento informado de la persona, el cual debe obtenerse de acuerdo con los requerimientos legales y cumpliendo los estándares culturales de la comunidad en que se realizará la intervención. De este principio, además, hacen emanar el requisito de protección a las personas que se encuentren en situación de vulnerabilidad, por cualquier causa.

- **Beneficencia / no maleficencia**: Que exige que se busque el máximo beneficio con el mínimo riesgo de perjuicio a las personas, lo cual debiera aplicarse tanto al diseño de las investigaciones como a la selección de personal que participa de ellas, además del control ético de los resultados. En aquellos casos en que no sea factible obtener el consentimiento informado de los participantes, las intervenciones o los procedimientos de investigación que no ofrecen beneficios individuales, los riesgos para los participantes no deben sobrepasar el mínimo, salvo que un comité de ética de la investigación determine que no existe otro método que permita recolectar los datos necesarios en otra población o de una manera menos riesgosa.
- **Justicia distributiva**: Se exige la diseminación de los resultados de la investigación, incluidos los resultados

negativos. Asimismo, se exige que los beneficios de la investigación alcance a las comunidades en las cuales se llevan a cabo las investigaciones, ayudando a crear capacidad local de investigación, tanto personal como en infraestructura local. Asimismo, de este principio se deriva el requerimiento de que los productos que emanen de las investigaciones queden a disposición de la comunidad a precios y condiciones accesibles a la población.

- **Pertinencia, razonabilidad, adecuación:** En la selección de investigaciones, se trata de realizar aquellas investigaciones que proporcionan una información que no podría obtenerse de otra forma. En el diseño de los protocolos de investigación, la elección de métodos adecuados a los objetivos de la investigación y al campo de estudio y la autorización previa de los protocolos de investigación por comités de evaluación científica y ética. Adicionalmente, la idoneidad profesional será un factor crítico a la hora de que asegurar de que se concrete el valor científico y social proyectado. En este ámbito se exige además que las investigaciones tengan una capacidad de respuesta a las necesidades o prioridades de salud presente en la comunidad o en la región en la que dicha comunidad se inserta. Finalmente, este requisito tendrá aplicación a la selección de las personas que participen de una investigación, y en el acceso de ellas a los beneficios de la investigación. Asimismo, los riesgos a que se someta a las personas deben ser

proporcionales a los beneficios individuales y valor social esperados.

- **Valor social y científico:** se trata de que las investigaciones que tengan relación con la salud tengan un diseño científicamente sólido, de forma de asegurar “la calidad de la información que se producirá, su pertinencia para abordar problemas de salud importantes y su contribución a la formulación o evaluación de intervenciones, políticas o prácticas que promuevan la salud pública o la salud de las personas” (p.1)

En base a estos instrumentos, desde la neuro ética se busca analizar y estudiar las implicancias éticas, legales, políticas y sociales de la investigación neurocientífica y proponer lineamientos o pautas de conducta que permitan delimitar y guiar el trabajo no sólo de los neuro científicos y neuro técnicos, sino que, además, las implicancias éticas y sociales que atañen a otras disciplinas tales como el derecho, la ingeniería, la medicina y las disciplinas a que dan lugar.

Desde esta perspectiva se señalan dos aristas relevantes: la ética de la neurociencia, que busca desarrollar un marco ético de las actividades de investigación y desarrollo en el ámbito de las neurociencias y neuro técnicas; en segundo lugar pero que además la neurociencia de la ética, que busca explicar la concepción moral de una persona o incluso de una comunidad a partir de la base neuronal, esto es, como impacta la actividad cerebral tanto en la concepción y comprensión del bien y el mal, además de la capacidad de discernimiento a estos respectos por el ser humano.

En nuestro caso, nos centraremos en las cuestiones legales que surgen a raíz de la evidencia del desarrollo de la neurociencia, y cómo ha sido abordado, principalmente a partir de la experiencia de Chile. Asimismo, nos centraremos en la reforma constitucional propuesta y aprobada, que busca fortalecer los derechos fundamentales de la persona, por lo que no nos referiremos a cuestiones tales como la legitimación del uso de la neurociencia y neuro tecnología con finalidades forenses (detectores de mentiras e imágenes cerebrales como pruebas en juicio), ni la necesidad de analizar la normativa que se refiere a la determinación de la responsabilidad penal, aun cuando la norma constitucional aprobada podría tener repercusiones en estas materias.

La modificación constitucional

Las ideas matrices.

La ley N° 21.383, que modifica la carta fundamental para establecer el desarrollo científico y tecnológico al servicio de las personas, del Ministerio de Ciencia y Tecnología, publicada recientemente el 25 de octubre de 2021, se inició por una moción parlamentaria presentada con fecha 07 de octubre de 2020. En su elaboración participaron científicos, académicos y representantes de la sociedad civil. Dentro de las ideas matrices del proyecto destacan la "necesidad de proteger la dignidad humana frente al uso de nuevas técnicas, en especial en lo tocante a la protección del "cerebro humano", concepto que no se agota sólo en una dimensión física, sino que más bien se expande hacia su dimensión de

potencialidad mental que envuelve los misterios de la existencia humana y es por esa razón que debe tener la máxima protección jus fundamental"⁸. Se declara luego, que los imperativos éticos del correcto uso de las neurotecnologías incluyen cuatro elementos, a saber:

- **Protección de los neurodatos:** Derecho a la Privacidad de la información producida por la actividad cerebral a la cual es posible acceder a través de la neurotecnología. Entendemos por datos neuronales, Aquella información obtenida, directa o indirectamente a través de los patrones de actividades de las neuronas cuyo acceso está dado por neurotecnología avanzada, incluyendo sistemas de registro cerebrales tanto invasivos como no invasivos. Estos datos contienen una representación de la actividad psíquica, tanto consciente como subconsciente, y que corresponden al más íntimo aspecto de la privacidad humana. Será tal, la información electrofisiológica, la actividad mental consciente y la actividad mental inconsciente.

En todo caso, si bien no toda esta información es dato de salud, no cabe duda de que es información sensible desde el punto de vista del análisis de riesgo desde la protección de datos personales. Adicionalmente, es importante considerar que esta información difícilmente puede ser anonimizada, pues a través de procesamiento de datos complejos podría Re identificarse a su titular.

⁸ Historia de la ley N° 21.383, <https://www.bcn.cl/historiadelaley/nc/historia-de-la-ley/7926/>.

- **El derecho a la identidad personal y la autodeterminación:** la protección de la persona frente al uso de las neurotecnologías como herramientas que inhiba, modifique o altere indebidamente la identidad de la persona o que disminuya o anule sus posibilidades de autodeterminarse. Dicho, en otros términos, se trata de que “nadie puede alterar o manipular sin consentimiento los estados mentales de una persona que la constituye como tal, que la hacen ser quien es”⁹

En este ámbito, quedarían proscritas las neurotecnologías que afecten la libertad de pensamiento, la libre deliberación política, la libertad ideológica de la persona y su capacidad de adoptar decisiones en base a sus propias convicciones, tales como aquellas que tienden a manipular a las personas con diversas finalidades.

En el ámbito de autodeterminación, además, se trata de que los dispositivos que conectan el cerebro a una máquina, por ejemplo, para articular un implante o una prótesis, sean consistentes con la voluntad de la persona y no fuercen acciones indeseadas, puesto que lo contrario, además de vulnerar el libre albedrío, tendría graves consecuencias a la hora de imputar responsabilidad a la persona por actos realizados a través de dichos implantes o prótesis. En un ejemplo, pensemos en una persona que cuenta con una interfaz cerebro

máquina asociada a una prótesis de brazo; si esa prótesis es susceptible de ser hackeada o por un tercero, u opera en base a un algoritmo malicioso podría infligir daño a un tercero, ya sea en la persona o sus bienes. En este caso, ¿a quién debiéramos atribuir la responsabilidad por los daños causados?, para que la respuesta sea a la persona que tiene implantada o conectada el dispositivo sería necesario poder asegurar que esas acciones son indubitadamente imputables a esa persona.

- **Equidad en el acceso a las mejoras de las capacidades cerebrales:** establecimiento de condiciones de acceso equitativo a todos los segmentos de la población a los beneficios de las neurotecnologías. A vía ejemplar, el acceso a neurotecnologías que restablecen o reparan la conexión entre un órgano y el cerebro, para permitir o potenciar su función, no podrá ser accesible solo a las personas con más recursos, sino que debiera propenderse a su acceso universal.
- **Protección de la persona frente al sesgo algorítmico:** establecimiento de resguardos que minimicen los riesgos de discriminaciones arbitrarias en las aplicaciones de inteligencia artificial a las neurotecnologías. A este respecto, estimamos que los sesgos que se incluyen en la programación de un dispositivo o en soluciones de inteligencia artificial pueden constituir

⁹ Wajnerman P. Abel. “Neuroderechos: Un Código Ético”, en línea <https://www.uahurtado.cl/wp-images/uploads/2021/06/neuroderechos-codigo-etico.pdf>

una discriminación, que será arbitraria si no responde a criterios o intereses legítimos. A este respecto, una de las pretensiones es la transparencia de los algoritmos o al menos su auditabilidad. Asimismo, el proyecto de ley de protección de datos en tramitación, siguiendo a la normativa europea, pretende establecer el derecho de oposición a las decisiones automatizadas, como una herramienta al servicio de la protección de la persona frente a las decisiones arbitrarias que puedan emanar de estas operaciones de tratamiento de datos personales.

El texto inicial y el aprobado

El texto de la moción parlamentaria era del siguiente tenor: "Intercálese el siguiente inciso segundo en el artículo 19 Numeral 1º de la Constitución Política del Estado, pasando el actual inciso segundo a ser tercero y así sucesivamente «La integridad física y psíquica permite a las personas gozar plenamente de su identidad individual y de su libertad. Ninguna autoridad o individuo podrá, por medio de cualquier mecanismo tecnológico, aumentar, disminuir o perturbar dicha integridad individual sin el debido consentimiento. Sólo la ley podrá establecer los requisitos para limitar este derecho, y los requisitos que debe cumplir el consentimiento en estos casos», el texto finalmente aprobado es el siguiente:

"Artículo único.- Modifícase el número 1º del artículo 19 de la Constitución Política de la República, de la siguiente forma:

Agregase el siguiente párrafo final, nuevo:

«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;»".

Como podemos apreciar, el texto aprobado es consistente con los instrumentos de derechos humanos a que nos referimos al inicio de nuestro análisis, además de guardar coherencia con las ideas matrices declaradas y con el actual texto constitucional.

En efecto, la primera parte del nuevo inciso final del artículo 19 N°1, que declara que el desarrollo "científico y tecnológico estará al servicio de la persona humana", tiene una directa relación con el principio de "dignidad" contenido en el tratado de la Declaración Universal sobre el Genoma Humano y Derechos Humanos, de la UNESCO (1998); además, de constituir una manifestación de los principios de "respeto a las personas" y de "beneficencia / no maleficencia", contemplados en las Pautas CIOMS, adoptadas en el seno de la Organización Mundial de la Salud (2002).

La segunda parte, en cambio, emana del imperativo de coherencia interna del texto constitucional, especialmente en lo previsto en el artículo 19 N° 26 que dispone que "la seguridad de que los preceptos legales que por mandato de la Constitución regulen o complementen las garantías que ésta establece o que las limiten en los casos en que ella lo autoriza, no podrán afectar los derechos en su esencia, ni imponer

condiciones, tributos o requisitos que impidan su libre ejercicio”.

Es así como mientras la constitución sienta las bases esenciales del derecho, deja luego al legislador la dictación de la normativa que desarrolle el derecho, pero siempre respetando su contenido esencial. Es por ello por lo que, además del proyecto ley aprobado, se tramita otro proyecto de ley (Boletín 13.828), que busca cumplir con el mandato del nuevo inciso a que nos referimos en estas líneas. Ello sin perjuicio de las necesarias referencias a la modificación a la ley 19.628, mal llamado de protección de la vida privada, que se tramita en el congreso nacional (BOLETINES Nos 11.092-07 y 11.144-07, refundidos).

Otro aspecto relevante dice relación con la labor que deberán cumplir los tribunales de justicia en la interpretación y aplicación práctica de la nueva norma, especialmente a través del recurso de protección, consagrado en el artículo 20 del texto constitucional, pues la nueva disposición queda dentro del ámbito de aplicación del recurso de protección, frente a actos u omisiones arbitrarias o ilegales de los que la persona sufra privación, perturbación o amenaza en el legítimo ejercicio del derecho de las personas a que el desarrollo científico y tecnológico cumpla con los imperativos de la nueva norma constitucional.

El análisis sistemático

Sin perjuicio de lo anterior, no podemos olvidar que la nueva norma se inserta en un marco jurídico de larga data, en el cual se contemplan algunas normas de utilidad a los efectos del problema que nos ocupa.

Es así como la Constitución Política ya había sido reformada el año 2018 a través de la ley 21.096, para consagrar el derecho a la protección de datos personales, incorporado en el 19 N° 4, en los siguientes términos: “4º.- El respeto y protección a la vida privada y a la honra de la persona y su familia, y, asimismo, la protección de sus datos personales. El tratamiento y protección de estos datos se efectuará en la forma y condiciones que determine la ley”. No cabe duda que los datos neuronales debieran ser considerados datos sensibles en los términos de la ley 19.628, sin embargo, se debe tener la precaución de incluirlos en esta categoría, máxime que en el actual texto del proyecto de ley que modifica a la ley se establece una enumeración cerrada de los datos que sensibles.

Adicionalmente, cuando nos referimos al empleo de neuro tecnologías en el ámbito de las acciones de promoción, protección y recuperación de la salud, cobra relevancia lo previsto en la normativa sanitaria. En primer lugar, el Código Sanitario, en su artículo 111 se dispone que “los instrumentos, aparatos, dispositivos y otros artículos o elementos destinados al diagnóstico, prevención y tratamiento de enfermedades de seres humanos, así como el reemplazo y modificación de sus anatomías” y que no correspondan a sustancias farmacéuticas, deberán cumplir con las normas y exigencias de calidad que les sean aplicables según su naturaleza.

En particular, “las personas naturales o jurídicas que, a cualquier título, fabriquen, importen, comercialicen o distribuyan tales elementos deberán realizar el respectivo

control y certificación de su calidad en servicios, instituciones laboratorios o establecimientos con autorización sanitaria expresa, otorgada por el Instituto de Salud Pública de Chile". Estas certificaciones, previstas en el artículo 7 de este mismo código, son condición necesaria para iniciar el funcionamiento de la empresa. Se trata de certificaciones temporales, por el plazo que establezca el reglamento, con un mínimo de tres años.

Mirada desde los sujetos.

A continuación, trataremos de sistematizar las normas, desde una perspectiva subjetiva, de forma que permita al lector comprender el rol que cabe a cada uno de los sujetos o instituciones que participan de los procesos asociados a estos dispositivos;

- **Instituto de Normalización Nacional (INN):** a quien corresponde establecer las especificaciones técnicas a través de normas oficiales chilenas, aprobadas por el Ministerio de Salud. A falta de una norma chilena, el Ministerio de Salud podrá fijar dichas especificaciones, a proposición del INN y sobre la base de la información generada por organismos internacionales o entidades extranjeras especializadas (art. 111, letra c))
- **Instituto de Salud Pública (ISP):** a quien corresponde autorizar y fiscalizar a las entidades que realicen el referido control y certificación. Adicionalmente, a falta de organismos privados que desarrollen dichas tareas, deberá ejecutarlas por sí mismo. (art. 111 letra b). Adicionalmente, al ISP corresponde llevar un registro público de todas las investigaciones científicas en seres

humanos con elementos de uso médico autorizados para realizarse en el país, el cual quedará sujeto al deber de transparencia, conforme a las disposiciones de la ley 20.285.

En el caso que se pretenda usar un dispositivo en investigaciones científicas en seres humanos, al ISP le corresponderá evaluar y, en caso de que proceda, otorgar una autorización especial para su uso provisional, la cual es aplicable tanto a dispositivos que no cuenten con el respectivo registro o, respecto de aquellos que cuentan con dicho registro, pero se pretende utilizarlos de una manera distinta a la registrada. A la solicitud correspondiente se deberá cumplir con lo previsto en el artículo 10 de la ley 20.120, esto es, se deberá adjuntar el protocolo de investigación, el formato de consentimiento informado, la póliza de seguros y todo otro antecedente que establezca el reglamento. La autorización será temporal, limitada al máximo de un año y específica, por lo que no podrá aplicarse el dispositivo a una investigación distinta. (art. 111 A.-).

- **Ministerio de Salud (Minsal):** además del rol que señalamos antes, a Minsal le corresponde disponer, a proposición del ISP, por decreto fundado, las diferentes clases o tipos de instrumentos, aparatos, dispositivos, artículos y elementos que quedarán sometidos a estas exigencias. El decreto deberá indicar las especificaciones técnicas a que se sujetará el control de calidad y las entidades que cuentan con autorización oficial para ejecutarlo o la inexistencia de interesados en obtener

esta autorización. Asimismo, el Ministerio de Salud podrá establecer, mediante decreto supremo, la exención de la exigencia, respecto de aquellos elementos de uso médico que no conlleve un riesgo relevante para las personas.

Adicionalmente, al Ministerio le corresponderá disponer, por decreto supremo, "la clase de los elementos de uso médico para cuya investigación será obligatoria la presentación de pólizas de seguro"¹⁰.

- **Secretaría Regional Ministerial (SEREMI):** Las infracciones a estas disposiciones podrán ser objeto de un Sumario sanitario, de la Secretaría Regional Ministerial del territorio en que se cometan y los elementos y dispositivos "que se comercialicen o distribuyan a cualquier título sin contar con el certificado de calidad establecido en esta disposición serán decomisados, sin perjuicio de las demás medidas que pueda adoptar la autoridad sanitaria" (art. 111 letra f)).

Las infracciones que se comentan con ocasión del empleo de elementos de uso médico en el marco de investigaciones en seres humanos, serán sancionadas conforme a las disposiciones del libro décimo del Código Sanitario y las que establece la ley N° 20.120, sobre investigación científica en el ser humano, su genoma y prohíbe la clonación humana.

Utilización de dispositivos en investigaciones en seres humanos.

La ley 20.120, de 2006, regula la investigación científica en seres humanos. Si bien esta ley se dictó con ocasión de los avances de la investigación científica biomédica, estimamos que debe ser tenida a la vista en el análisis sistemático de la nueva normativa. En especial lo que dispone el artículo 10, en tanto que se ajusta a las ideas matrices del nuevo texto constitucional, como podemos ver:

"Artículo 10.- Toda investigación científica en seres humanos que implique algún tipo de intervención física o psíquica deberá ser realizada siempre por profesionales idóneos en la materia, justificarse en su objetivo y metodología y ajustarse en todo a lo dispuesto en esta ley.

No podrá desarrollarse una investigación científica si hay antecedentes que permitan suponer que existe un riesgo de destrucción, muerte o lesión corporal grave y duradera para un ser humano.

Toda investigación científica biomédica deberá contar con la autorización expresa del director del establecimiento dentro del cual se efectúe, previo informe favorable del Comité Ético Científico que corresponda, según el reglamento."

En esta ley además se consagra el principio de consentimiento, conforme al cual toda investigación científica en un ser humano debe contar con el consentimiento de la persona o, en su defecto, el de aquel que deba suplir su voluntad. El consentimiento será revocable, sin que la revocación pueda

¹⁰ Art. 111 F.- Código Sanitario.

acarrear perjuicios o pérdida de beneficios a la persona.

Para que sea válido, el consentimiento deberá ser previo, expreso, libre e informado y específico, en relación a la investigación y su protocolo, de forma tal que si se modifica alguno de estos elementos deberá volver a recabarse de la persona o de quien le represente.

Para que el consentimiento sea informado, la persona debe conocer y entender los aspectos esenciales de la investigación, en especial su finalidad, beneficios, riesgos y los procedimientos o tratamientos alternativos. (art.11).

En la materia que nos ocupa, con el nuevo texto constitucional debiéramos sostener que habrá casos en los cuales el consentimiento de la persona no será suficiente, puesto que la investigación deberá garantizar que se ajusta al mandato conforme al cual “la ciencia y la tecnología están al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica.”, además de resguardar la actividad cerebral, así como la información proveniente de ella.

Mención especial merece, en las investigaciones científicas en personas humanas realizadas con dispositivos que cuenten con autorización provisional, el rol del titular de la autorización, la entidad patrocinante, el investigador principal y el respectivo centro donde se realice la investigación, a quienes corresponde la

obligación de utilizar los dispositivos de acuerdo a la autorización y al protocolo de investigación aprobado, además del deber de notificar al ISP, y al Comité Ético Científico que corresponda “en el plazo y según la forma que establezca el reglamento respectivo, de las reacciones adversas y los eventos adversos producidos con ocasión del estudio. Asimismo, serán responsables del cumplimiento de las normas sobre farmacovigilancia y tecnovigilancia conforme al reglamento”.¹¹

Asimismo, en este caso, es preciso hacer presente que el paciente sujeto del ensayo clínico tendrá derecho a que, una vez terminado el mismo, “el titular de la autorización especial para uso provisional con fines de investigación y, con posterioridad en su caso, el titular del registro sanitario del producto sanitario de que se trate, le otorgue sin costo para el paciente la continuidad del tratamiento por todo el tiempo que persista su utilidad terapéutica, conforme al protocolo de investigación respectivo” Obligación que pesará sobre el titular del registro sanitario, aun cuando no haya sido titular de la autorización provisional o haya adquirido con posterioridad el registro sanitario¹².

Además, los titulares de las autorizaciones para uso provisional con fines de investigación serán responsables por los daños que causen con ocasión de la investigación, aunque estos se deriven de hechos o circunstancias que no se hubieren podido prever o evitar según el estado de los conocimientos de la ciencia o de la técnica,

¹¹ Art. 111 B.- Código Sanitario.

¹² Art. 111 C.- Código Sanitario.

existentes en el momento de producirse los daños; incluso, acreditado el daño, se presumirá que se ha producido con ocasión de la investigación. La acción para perseguir esta responsabilidad prescribe en 10 años contados desde la manifestación del daño, salvo que se haya decretado una alerta sanitaria con ocasión de una epidemia o pandemia y durante la vigencia de esta, caso en el cual el plazo se contará desde el término del respectivo ensayo, cuando se trate de dispositivos médicos destinados a enfrentar las circunstancias que sirvieron de fundamento al decreto de alerta sanitaria de que se trate¹³.

Adicionalmente, el Código establece la obligación de los titulares de las autorizaciones especiales para uso provisional para fines de investigación de elementos de uso médico, contar con una póliza de seguro por responsabilidad civil, conforme al reglamento que se dicte a través del Ministerio de Salud.

Clasificación de dispositivos médicos según riesgo.

A las normas antes señaladas, estimamos que es de suma relevancia añadir el análisis de la guía para la clasificación de dispositivos médicos según riesgo, elaborada por el ISP¹⁴, a efectos de orientar tanto a fabricantes, importadores y distribuidores como a los profesionales del área clínica.

Según este organismo, un modelo de clasificación de dispositivos basado en riesgos "considera la vulnerabilidad del

cuerpo humano teniendo en cuenta los riesgos potenciales asociados con este tipo de productos" (p.4).

En primer lugar, el documento define un dispositivo médico, en los siguientes términos:

"Se entenderá por tal cualquier instrumento, aparato, implemento, máquina, equipo, artefacto, implante, reactivo, para uso in vitro, software, material u otro artículo similar o relacionado, que cumpla con las siguientes condiciones copulativas:

- 1.- Que no se trate de las sustancias descritas en los artículos 95 inciso primero, 102 y 106 del Código Sanitario;*
- 2.- Que no logre su acción principal en o sobre el cuerpo humano por mecanismos farmacológicos, inmunológicos o metabólicos, aunque pueda ser ayudado en su función por tales mecanismos; y,*
- 3.- Que su uso previsto en los seres humanos, solo o en combinación, se refiera a uno o más de los siguientes fines:*
 - Diagnóstico, prevención, monitoreo, tratamiento, alivio o cura de una enfermedad;*
 - Diagnóstico, monitoreo, tratamiento alivio, cura o compensación de un daño o lesión;*
 - Investigación, reemplazo, modificación o soporte anatómico o de un proceso fisiológico;*
 - Soporte o mantenimiento de la vida;*
 - Control de la concepción;*
 - Desinfección de dispositivos médicos;*

¹³ Art. 111 E.- Código Sanitario.

¹⁴ Instituto de Salud Pública (2018). "Guía para la clasificación de dispositivos médicos según riesgo"

https://www.ispch.cl/sites/default/files/Guia_de_Clasificacion_de_Dispositivos_Medicos_Segun_riesgo_Formato_Institucional.pdf

- *Suministro de información para propósitos médicos o diagnósticos a través de un examen in vitro de especímenes derivados del cuerpo humano."*

Como podemos ver, los elementos incluidos dentro de la concepción de neuro tecnología, que son los que nos interesan, caben dentro del concepto de dispositivos médicos, tanto en los términos del Código Sanitario como en la guía en análisis.

Luego, clasifica y conceptualiza los distintos tipos de dispositivos, en las siguientes categorías, en lo que nos interesa:

- **Dispositivos médicos activos:** que son aquellos cuyo funcionamiento depende de la energía eléctrica -o de cualquier otra fuente de energía distinta de la generada directamente por el cuerpo humano o por la gravedad- y que actúa mediante la conversión de dicha energía;
- **Dispositivos médicos combinados:** que incorporan como parte integrante una sustancia que puede ejercer sobre el cuerpo humano una acción auxiliar o accesoria a la del dispositivo médico, , la cual según la definición legal que le compete en sus respectivas normativas cuando se utilice independientemente, puede ser un producto farmacéutico, biológico o cosmético.
- **Dispositivos médicos implantables:** que son aquellos indicados para ser totalmente introducidos "dentro del cuerpo humano o para sustituir una superficie epitelial o la superficie ocular, mediante intervención quirúrgica, destinado a permanecer allí después de

la intervención. También se considera dentro de esta categoría a cualquier dispositivo médico indicado para ser parcialmente introducido dentro del cuerpo humano, mediante intervención quirúrgica, destinado a permanecer al menos treinta días después de la intervención".

- **Dispositivo médico invasivo:** Dispositivo médico que penetra parcial o completamente en el interior del cuerpo a través de un orificio corporal o de la superficie corporal.
- **Dispositivo médico invasivo de tipo quirúrgico:** Dispositivo médico que penetra en el interior del cuerpo por medio de una intervención quirúrgica o en el contexto de una intervención quirúrgica.
- **Dispositivo médico para diagnóstico in vitro:** Dispositivo consistente en un reactivo, producto reactivo, calibrador, material de control, kit, instrumento, aparato, pieza de equipo, programa informático o sistema u otro artículo, utilizado solo o en combinación, destinado por el fabricante a ser utilizado in vitro para el estudio de muestras procedentes del cuerpo humano, incluidas las donaciones de sangre y tejidos, única o principalmente con el fin de proporcionar información sobre uno o varios de los elementos siguientes: - relativa a un proceso o estado fisiológico o patológico; - relativa a deficiencias físicas o mentales congénitas; - relativa a la predisposición a una dolencia o enfermedad; - para determinar la seguridad y compatibilidad con posibles receptores; - para predecir la respuesta o reacción

al tratamiento; - para establecer o supervisar las medidas terapéuticas.

- **Dispositivo médico quirúrgico reutilizable:** Dispositivo médico destinado a fines quirúrgicos para cortar, perforar, aserrar, escarificar, raspar, pinzar, retraer, recortar, inmovilizar u otros procedimientos similares, sin estar conectado a ningún dispositivo médico activo, y que puede ser utilizado nuevamente una vez efectuados todos los procedimientos pertinentes.
- **Dispositivo médico activo terapéutico:** Dispositivo médico utilizado, solo o en combinación con otros dispositivos médicos, para sostener, modificar, sustituir o restaurar funciones o estructuras biológicas en el contexto del tratamiento o alivio de una enfermedad, lesión o discapacidad.
- **Nanomaterial:** Material natural, fabricado u obtenido accidentalmente que contiene partículas separadas o formando un agregado o aglomerado, y que, en la distribución de tamaños granulométricos, el 50 % o más de las partículas presenta una o más dimensiones externas en el intervalo de tamaños comprendido entre 1 y 100 nm. Los fullerenos, los copos de grafeno y los nanotubos de carbono de pared simple con una o más dimensiones externas inferiores a 1 nm también se considerarán nanomateriales.
- **Software como dispositivo médico:** creación intelectual que comprende los programas, procedimientos, reglas y cualquier documentación asociada con la operación de un sistema de

procesamiento de datos con propósitos médicos, de forma tal que ejecute funciones clínicas o ejerza un efecto o acción directa sobre el diagnóstico, la terapia o el proceso preventivo. Se extiende tanto al que forma parte de un equipo médico computarizado, como al suministrado de forma independiente como producto en sí mismo.

Luego, en las reglas aplicables a cada uno de los dispositivos, la regla 9.3 señala, declara que pertenecen a la clase III, todos los dispositivos médicos activos destinados a emitir radiaciones ionizantes con fines terapéuticos, incluidos los que se utilizan para controlar o supervisar dichos dispositivos médicos o que influyan directamente en el funcionamiento de los mismos, ejemplificando con los equipos de rayos X fijos, móviles o portátiles, equipos de braquiterapia manuales y automáticos de baja, media y alta tasa de dosis; y en la regla 9.4, señala que pertenecen a la clase iv.- Todos los dispositivos médicos activos destinados a controlar, supervisar o influir directamente en el funcionamiento de dispositivos médicos implantables activos, tales como "el firmware de las partes no implantables de un dispositivo médico implantable activo (tal como en un implante coclear); software utilizado en la programación de un marcapasos cardíaco".

Adicionalmente, nos interesa la regla 10.1, que incluye los dispositivos médicos activos con fines de observación y diagnóstico. Que corresponden a equipos para diagnóstico u obtención de imágenes ultrasónicas, captura de señales fisiológicas, radiología

intervencionista y radiología de diagnóstico, de los cuales, en la clase IV se ubican aquellos que presentan un potencial medio o elevado de exposición interna, tales como los “nanobiosensores para detección de glucosa; chips con nanosensores para el torrente sanguíneo para detección de síntomas previos a un ataque al corazón; micro-robots dirigidos electromagnéticamente para cirugías de precisión (tal como los utilizados en cirugías oculares).”

Luego, también cobrará relevancia la regla 22, que asigna a la clase IV, los dispositivos médicos terapéuticos activos con una función diagnóstica integrada o incorporada, que determine de manera importante el manejo del paciente por el dispositivo médico, tales como los “sistemas con retroalimentación (*closed loop systems*); desfibriladores externos automáticos; bomba de insulina con retroalimentación”.

De estas reglas entendemos que mientras más alta sea la clase a la que se asigna el dispositivo, se le asigna un factor de riesgo más alto en relación con la salud de la persona en la cual se utilizan.

Conclusión.

Contrariamente a lo que algunos sostienen, el nuevo texto constitucional no sólo es necesario, dado el avance tecnológico alcanzado en el ámbito de las neurociencias, sino que además es coherente con los instrumentos de derechos humanos suscritos y ratificados por nuestro país.

Entendemos que, además, la preocupación sobre la necesidad de regular en estas

materias es compartida en otros entornos. Es así como en el contexto europeo se ha presentado una propuesta de resolución del Parlamento Europeo, denominada “inteligencia artificial: cuestiones de interpretación y de aplicación del derecho internacional en la medida en la que la UE se ve afectada en los usos civil y militar, como de la autoridad del Estado fuera del ámbito de la Justicia penal.

Dentro de los lineamientos base, el documento reconoce que el uso de la IA, la robótica y las tecnologías relacionadas, debe garantizar el respeto de la dignidad humana y los derechos humanos, consagrados en la Carta de los Derechos Fundamentales de la Unión Europea (consid. D), con lo cual se reconoce el principio de dignidad humana como línea guía y delimitadora del avance de estos desarrollos.

En segundo lugar, se considera la necesidad de que la IA, “la robótica y las tecnologías relacionadas –pues pueden tener un uso transfronterizo– estén centradas en el ser humano, es decir, que se conciban básicamente para ser utilizadas al servicio de la humanidad y el bien común, a fin de contribuir al bienestar y el interés general de sus ciudadanos” (consid. E), con lo que se reconocen los principios de beneficencia o no maleficencia, y primacía de lo humano a que nos referimos antes.

Asimismo, la norma aprobada en nuestro país es consistente con el imperativo que se reconoce en la propuesta de resolución, que prevé que la normativa debe establecer un “marco jurídico adecuado y completo que atienda a los aspectos éticos de estas

tecnologías, así como a la responsabilidad, la transparencia y la rendición de cuentas (en particular para la IA, la robótica y las tecnologías relacionadas consideradas de alto riesgo); que tal marco debe reflejar que los valores humanistas intrínsecamente europeos y universales son aplicables a toda la cadena de valor en el desarrollo, la implementación y los usos de la IA; que tal marco ético debe aplicarse al desarrollo (incluida la investigación e innovación), el despliegue y la utilización de la IA, respetando plenamente el Derecho de la Unión Europea y los valores establecidos en la Carta de los Derechos Fundamentales de la Unión Europea”.

Estimamos, en todo caso, que el camino normativo recién comienza, pues hay un largo camino que recorrer, si bien el ámbito sanitario cuenta con más andamiaje normativo, las neurotecnologías tienen aplicaciones en distintos ámbitos, tales como los bienes de consumo, el marketing en sus distintas aplicaciones, la educación, la justicia, especialmente en el ámbito de la investigación forense, en todos los cuales habrá de realizarse los análisis que permitan analizar la suficiencia de nuestros cuerpos normativos, y en el caso que sea necesario, realizar las adecuaciones que permitan salvaguardar los derechos de las personas, sin perjuicio de permitir el avance científico y tecnológico.

Este esfuerzo debe ser de alcance global, puesto que el desarrollo tecnológico deja en evidencia que la compatibilidad internacional de los marcos normativos resulta esencial para salvaguardar adecuadamente los derechos de las personas.

Es aquí donde un trabajo coordinado entre ciencia y diplomacia puede contribuir a generar procesos regulatorios y normativos que cuenten con los necesarios consensos, fundamentalmente multilaterales, para favorecer una implementación más armónica en los órdenes jurídicos internos. De allí, el valor que desde la Academia Diplomática de Chile (ACADE) se esté reflexionando en torno a estos temas emergentes de la política exterior y de las relaciones internacionales contemporáneas.

Bibliografía

- Chnadler, Jennifer et. Al. “enfoques legales internacionales de la neuocirugía para trastornos psiquiátricos”, disponible en línea en <https://www.frontiersin.org/articles/10.3389/fnhum.2020.588458/full>
- Cortina, Adela (2010). Neuroética: ¿Las bases cerebrales de una ética universal con relevancia política? ISEGORÍA. Revista de Filosofía Moral y Política N° 42, enero-junio, 2010, 129-148. ISSN: 1130-2097. En línea en <https://doi.org/10.3989/isegoria.2010.i42.687>.
- Declaración de derechos humanos, 1947
- Declaración de Helsinki de la Asociación Médica Mundial, AMM, las Pautas Éticas Internacionales para la Investigación Biomédica en Seres Humanos (v.2016) del Consejo de Organizaciones Internacionales de las Ciencias Médicas (CIOMS). https://cioms.ch/wp-content/uploads/2017/12/CIOMS-EthicalGuideline_SP_INTERIOR-FINAL.pdf

- Declaración Universal sobre el Genoma Humano y Derechos Humanos, de la UNESCO, de 1998.
- Declaración Universal sobre Bioética y Derechos Humanos (2005)
- Instituto de Salud Pública (2018). "Guía para la clasificación de dispositivos médicos según riesgo" https://www.ispch.cl/sites/default/files/Guia_de_Clasificacion_de_Dispositivos_Medicos_Segun_riesgo_Formato_Institucional.pdf
- Müller, O., & Rotter, S. "Neurotechnology: Current Developments and Ethical Issues". *Frontiers in systems neuroscience*, 11, 93. 2017. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5733340/> (Abril, 2019).
- Pacto Internacional de Derechos Económicos, Sociales y Culturales (PIDESC)
- Yuste, Rafael "Four ethical priorities for neurotechnologies and AI" (2017). Disponible en línea en <https://www.nature.com/articles/551159a>
- Wajnerman P. Abel. "Neuroderechos: Un Código Ético", en línea en <https://www.uahurtado.cl/wp-images/uploads/2021/06/neuroderechos-codigo-etico.pdf>

Legislación

- Código Sanitario.
- Ley 19.628, sobre protección de la vida privada.
- Ley. 20.120, sobre la investigación científica en el ser humano, su genoma, y prohíbe la clonación humana.
- Ley 20.584, que regula los derechos y deberes que tienen las personas en relación con acciones vinculadas a su atención en salud.

UNESCO's Recommendation on the Ethics of Artificial Intelligence as a reflection of global cooperation in diplomacy, science, technology, and policy.

Gabriela Ramos¹⁵¹⁶

Introduction

Artificial intelligence (AI) is one of the central issues of the era of converging technologies, with profound implications for humanity, cultures, societies and the environment. The impact of AI will depend on how humanity frames it and masters it, and the extent to which it prioritizes its contributions to a fairer world.

The COVID-19 pandemic has exemplified how AI technologies can help, but it has also shown the risks, the unintended consequences and the malicious uses of AI. Current AI technologies are systems with the capacity to process large amounts of data and information in a way that resembles intelligent behaviour. Their decisions cannot be fully predicted or understood due to the opacity of algorithms and how they process the data. Moreover, these technologies are developed by humans, and people are making choices in every step of AI's development. Therefore, AI technologies are not value-neutral,¹⁷ and the research, development and deployment of AI needs to be accompanied by ethical reflection.

As AI is a distributed technology, whose current practical governance is spread across numerous institutions, organizations and companies, its use provokes questions about what type of future we want for all. This reflection needs to address the main challenges in the development and use of AI technologies related to the biases embedded in data and algorithms, the protection of people's privacy and personal data, the lack of diversity and inclusion in the digital business model, the issues of fair distribution of benefits and risks, accountability, responsibility, impacts on employment and the future of work, human rights, security and risks of dual use.

These issues are recognized on multiple levels, as various frameworks are being adopted throughout the world. This includes Chile's first National Policy on Artificial Intelligence and the Action Plan adopted in October 2021.¹⁸ The Chilean National Policy positions ethics in AI as one of three substantial pillars for policy action and enshrines the principle of multinational

¹⁵ Assistant Director-General for Social and Human Sciences at UNESCO. The author would like to thank Maksim Karliuk, Programme Specialist of the Bioethics and Ethics of Science and Technology Section of UNESCO, for helping me prepare this paper.

¹⁶ The author would like to thank Maksim Karliuk, Programme Specialist of the Bioethics and Ethics of Science and Technology Section of UNESCO, for the assistance in preparing this paper.

¹⁷ Contemporary academic philosophers and theorists of technology from different schools widely agree with the claim that technology in general is not morally and politically neutral. See Miller, B. (2021). Is Technology Value-Neutral? *Science, Technology, & Human Values*, 46(1), 53–80.

¹⁸ See *Chile presents the first National Policy on Artificial Intelligence*. (2021, October 28). Government of Chile. <https://www.gob.cl/en/news/chile-presents-first-national-policy-artificial-intelligence/>

collaboration to guide its evolution. Furthermore, this instrument prioritizes the country's policies for a ten-year horizon, addressing long-standing challenges in human capital and data protection regulation, and declares an ambition to lead the region in digital infrastructure, innovation and scientific development.

However, there are limits to what national policies can achieve without an international framework. Until recently there was no global ethical framework for AI technologies. On 23 November 2021, the General Conference of UNESCO – the main governing body of the Organization – at its 41st session adopted the Recommendation on the Ethics of Artificial Intelligence (hereinafter ‘the Recommendation’).¹⁹ It has become the first-ever global normative instrument in this field, adopted by 193 Member States. The fact that it was approved by acclamation, with its solid provisions, shows how much countries felt the need for it. The Recommendation was developed in a pluralistic, multidisciplinary, multicultural and multi-stakeholder manner, and constitutes an ambitious, yet policy-friendly, blueprint to guide the development of AI technologies, providing an overarching ethical framework to protect people and to ensure the rule of law in the digital domain. It includes existing human rights and extends ethical reflection to incorporate new threats posed by these technologies that are not addressed in the current legal architecture.

We hope this international normative instrument will help the world reach an ethical consensus on the most pressing

challenges raised by AI, which represents not just a technological turning point but also an anthropological disruption. This is a necessary step towards more proactive thinking about the interrelations of science, technology and policy. Importantly, the Recommendation goes beyond a declaration of ethical values and principles to include concrete policy actions for their operationalisation, considering all stages of the AI system life cycle. The Recommendation provides for global consensus on the ‘what,’ as well as the ‘how’ of ethical regulation of this game-changing technology, and serves as a shared reference point for leaders around the world on how to control the risks, and harness these new technologies as a force for good. It is also proof of the ability of countries to come together and ensure productive dialogue and outcomes.

This text raises key issues in the relation between science and society. UNESCO celebrates the initiative of the Diplomatic Academy of Chile to convene these reflections, allowing us to realize the value of scientific diplomacy and its imperative relation with ethics, in the context of this multilateral organization. Certainly, this is a subject which demands standard-setting instruments, convinced that good governance of AI technologies must consider the ethical dimension.

What can go wrong with AI that is developed without ethical guardrails? Unfortunately, AI has the potential to reproduce biases and discriminatory practices present in the

¹⁹ The text of the Recommendation is available here: <https://unesdoc.unesco.org/ark:/48223/pf0000380455>

analogue world. It is up to us, therefore, to ensure that AI contributes to addressing the challenges, lest it exacerbates them. In the case of women, it seems that it is doing exactly this. The reasons are multiple, including the fact that only 22% of AI professionals globally are women, and AI technologies developed by male-dominated or male-only teams can contribute to perpetuating harmful gender stereotypes.²⁰ For instance, some AI-powered recruitment software has been found to discriminate against women due to the use of male-dominated datasets. Almost all voice assistants are given female names and voices, and express a 'personality' that is engineered to be uniformly subservient, as highlighted in UNESCO's ground-breaking report, "I'd Blush if I Could".²¹

Many institutions are relying more and more on AI systems to take decisions. When deciding how to allocate loans, the financial sector is no exception. Without ethical guardrails, the results are frequently inequitable, particularly for women – who are often determined by algorithms as less creditworthy than men.²² Moreover, there are instances, where a woman with a higher credit score than her husband would be given a lower credit limit by an algorithm.²³ These biases are rooted in stark gender imbalances in digital skills education and are exacerbated by the gender imbalances of the

technical teams developing AI technologies – these teams being within companies with significant gender disparities in their corporate boards.

Further, in the real world, if you are affected by a decision, there are clear ways to seek justice. In the digital world, this is more difficult because you may not know that you are affected; you do not know how your data is being used, and additionally, you cannot retrieve it. There is too much opacity and too many conflicting objectives (for example, protecting intellectual property rights versus transparency; privacy versus creating larger datasets to eliminate bias).

One of the major areas of impact of AI technologies is labour and skills, as new technologies create winners and losers in the labour market due to changing relative demands for occupations. By 2025, 85 million jobs may have been displaced by a shift in the division of labour between humans and machines, while 97 million new roles may emerge that are more adapted to the new division of labour between humans, machines and algorithms.²⁴ Companies estimate that around 40% of workers will require reskilling of six months or less and 94% of business leaders report that they expect employees to pick up new skills on the job – a sharp uptake from 65% in 2018. For those workers set to remain in their roles, the

²⁰ See <https://www.weforum.org/agenda/2018/12/artificial-intelligence-ai-gender-gap-workplace/>

²¹ West, M., Kraut, R., & Chew, H. E. (2019). *I'd Blush If I Could: Closing gender divides in digital skills through education*. UNESCO. <https://en.unesco.org/ld-blush-if-i-could>

²² USAID. (2021, March 24). *How Can We Address Gender Inequity in Artificial Intelligence?* Medium. <https://medium.com/usaaid-2030/how-can-we-address-gender-inequity-in-artificial-intelligence-41367b0e1f87>

²³ Telford, T. (2019, November 11). Apple Card algorithm sparks gender bias allegations against Goldman Sachs. The Washington Post. <https://www.washingtonpost.com/business/2019/11/11/apple-card-algorithm-sparks-gender-bias-allegations-against-goldman-sachs/>

²⁴ World Economic Forum. (2020). *The Future of Jobs Report 2020*. <https://www.weforum.org/reports/the-future-of-jobs-report-2020>

share of core skills that will change in the next five years is 40%, and 50% of all employees will need reskilling (up 4%).

In view of these and other identified challenges, the Recommendation contains strong provisions to promote gender parity, eradicate gender stereotyping, protect agency and ensure privacy. It also calls for the reform of the data governance model and for support for Member States in their work to safeguard individuals' safety and privacy throughout the AI lifecycle. The Recommendation makes a case for stronger consent rules and telling people how their data is being used. Similarly, strong provisions are provided in various other fields.

1 Towards UNESCO standard-setting on the ethics of AI

1.1 *The work of UNESCO leading to the mandate on ethics of AI*

UNESCO has a longstanding role at the United Nations (hereinafter 'UN') level and globally in promoting ethical science, which harnesses technological and scientific advancements for the benefit of all, protects the planet and constitutes a solid basis for peaceful cooperation. The Social and Human Sciences sector at UNESCO, which I am proud to lead, has been the source of many standards that have changed the course of several technological disruptions. It is supported by the work of its consultative

organs – the International Bioethics Committee (IBC, created in 1993)²⁵ and the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST, 1998)²⁶, in coordination with the Intergovernmental Bioethics Committee (IGBC, 1998)²⁷. UNESCO is the only international organization that has ethics as a core mandate, as part of the promotion of scientific cooperation.

The decision to proceed with the mandate on the ethics of AI followed an explicit request of the Director-General of UNESCO, Audrey Azoulay, as soon as she took office. This mandate built on the expertise and support of COMEST, a multidisciplinary scientific advisory body of UNESCO, made up of independent experts. The work of COMEST built on and complemented the work on AI being done elsewhere in the United Nations system,²⁸ other international organizations, nongovernmental organizations, academia and others.

Starting from its very first Ordinary Session (Oslo, Norway, 1999), COMEST devoted significant efforts to the ethics of new technologies. This work has led to the adoption of a number of major COMEST documents directly linked to AI: "Report of COMEST on Robotics Ethics" (2017)²⁹ and "Preliminary Study on the Ethics of Artificial Intelligence" (2019).³⁰ In the latter, COMEST identified that the highest ethical issues from

²⁵ See <https://en.unesco.org/themes/ethics-science-and-technology/ibc>

²⁶ See <https://en.unesco.org/themes/ethics-science-and-technology/comest>

²⁷ See <https://en.unesco.org/themes/ethics-science-and-technology/igbc>

²⁸ For an overview of the work of UN institutions in the field of AI see *United Nations Activities on Artificial Intelligence (AI)*. (2021). [ITU

Publication]. International Telecommunication Union. <https://www.itu.int/en/publications/gs/Pages/publications.aspx?parent=S-GEN-UNACT-2021&media=electronic>

²⁹ World Commission on the Ethics of Scientific Knowledge and Technology. (2017). *Report of COMEST on Robotics Ethics*. (UNESCO). <https://unesdoc.unesco.org/ark:/48223/pf0000253952>

³⁰ World Commission on the Ethics of Scientific Knowledge and Technology. (2019). *Preliminary Study on the Ethics of Artificial*

UNESCO's fields of competence concern "its implications for culture and cultural diversity, education, scientific knowledge, and communication and information." However, it put the emphasis on an aspect that is core to the social and human agenda: the need to achieve fair, inclusive and sustainable outcomes, and to ensure that AI technologies support the advancement of peaceful societies, in line with UNESCO's global mandate. Indeed, UNESCO's constitutional aim is to advance international peace and the common welfare of humanity through strengthening its "intellectual and moral solidarity".³¹ The 2019 study also recognized the need to support developing countries, particularly Africa, given the global priority of the Organization.

The combination of these factors led the General Conference of UNESCO to mandate the Director-General "to prepare an international standard-setting instrument on the ethics of artificial intelligence (AI) in the form of a recommendation".³² The process to deliver the Recommendation engaged the full membership of UNESCO in the drafting and negotiating of the text, as well as the active engagement and support of the Secretariat, led by the Assistant Director-General for Social and Human Sciences.

1.2 *UNESCO's added value*
UNESCO's universality in membership and multidisciplinary expertise were strong bases

for the development of this global standard. The mandate was also a recognition that for a long-time technological progress has been centre-stage, with negative consequences for people. The call is therefore to align the technologies with human goals, and not the other way around. Therefore, UNESCO's humanistic approach matters.

By the time UNESCO started developing the Recommendation, there were many national and international standards underscoring the need to develop AI for good. UNESCO's Recommendation benefited from them, but went beyond, with practical recommendations for implementation and for policy and regulatory work. The Organization's perspective, by calling diverse and inclusive set of actors, matters, as the building of the Recommendation brought together both developed and developing countries, different cultural perspectives, as well as various stakeholders within the public and private spheres. UNESCO also builds on its long track record of developing ethical frameworks on science, including bioethics, such as the work on the human genome that resulted in the widely acclaimed Universal Declaration on the Human Genome and Human Rights.³³

UNESCO's voice is strong in ensuring that the good governance of AI technologies is based on ethical considerations and on a strong respect of human rights and human dignity.

Intelligence. UNESCO.
<https://unesdoc.unesco.org/ark:/48223/pf0000367823>

³¹ Recital 5 of the Constitution of UNESCO. See Basic texts, <https://unesdoc.unesco.org/ark:/48223/pf0000380873?25=null&q=urid=b6193443-5e7b-49a9-acb8-c16f91780009>

³² 40 C/Resolution 37, [https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000372579&file=/in/rest/annotationSVC/Download](https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000372579&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_f95aa79b-711f-472f-9bc6-36b4349a12b0%3F_%3D372579eng.pdf&locale=en&multi=true&rk=/ark:/48223/pf0000372579/PDF/372579eng.pdf#%5B%7B%22num%22%3A110%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C54%2C182%2C0%5D)

[adWatermarkedAttachment/attach_import_f95aa79b-711f-472f-9bc6-36b4349a12b0%3F_%3D372579eng.pdf&locale=en&multi=true&rk=/ark:/48223/pf0000372579/PDF/372579eng.pdf#%5B%7B%22num%22%3A110%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C54%2C182%2C0%5D](https://unesdoc.unesco.org/ark:/48223/pf0000372579/PDF/372579eng.pdf#%5B%7B%22num%22%3A110%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C54%2C182%2C0%5D)

³³ See <https://en.unesco.org/themes/ethics-science-and-technology/human-genome-and-human-rights>

For example, the current COVID-19 pandemic has seen the explosion of AI-powered applications for monitoring and contact tracing. However, this has implications for privacy and data protection. In some cases, there are no precautions against the potential use of the personal and sensitive data collected for purposes beyond pandemic control — for instance, for mass surveillance of the population. As explored further, the Recommendation comes with strong principles and concrete policy actions to operationalize them, particularly stipulating that AI technologies should not be used for the purposes of mass surveillance, social scoring or abuses of cognitive biases. It also makes a broader call to enhance the capacities of governments to deal with these technologies, and to develop effective legislative and regulatory frameworks to deal with the downsides.

Furthermore, while it seeks to counter the risk of a growing digital and knowledge divides that could leave behind those who are relatively disadvantaged, or excluded - such as people in least developed countries, women and girls, youth, people with disabilities and the marginalized - UNESCO aims to foster policies that maximize the opportunities offered by AI for enhancing democracy, development and human rights. It helps strengthen the capacity of Member States to harness AI for the benefit of humanity and for achieving the Sustainable Development Goals (hereinafter 'SDGs') in line with ethical principles. This is all the more

important considering that half of the world's population is still not connected to internet, and the majority of these technological developments are made in a few advanced economies.

Finally, the ethical framework allows for innovation and experimentation, as it does not aim to codify every aspect of the development of the technologies, but to encourage the ethical reflection and awareness to happen at every step of the way.

2 Developing the Recommendation on the Ethics of AI

2.1 *Ad Hoc Expert Group and multi-stakeholder consultations*

In line with UNESCO's expertise, the development of the Recommendation was facilitated by a multicultural and multi-disciplinary Ad Hoc Expert Group (hereinafter 'AHEG') appointed by the UNESCO Director-General in March 2020. The AHEG was composed of 24 independent experts from all UNESCO regions³⁴; it produced the first draft, with the support of the Social and Human Sciences sector of UNESCO's Secretariat.

This text was based on UNESCO's previous research³⁵ and proposals. The AHEG's discussions concentrated primarily around the risks and challenges inherent to the rapid development of AI technologies, with specific

³⁴ See the Ad Hoc Expert Group's composition here, <https://unesdoc.unesco.org/ark:/48223/pf0000372991>

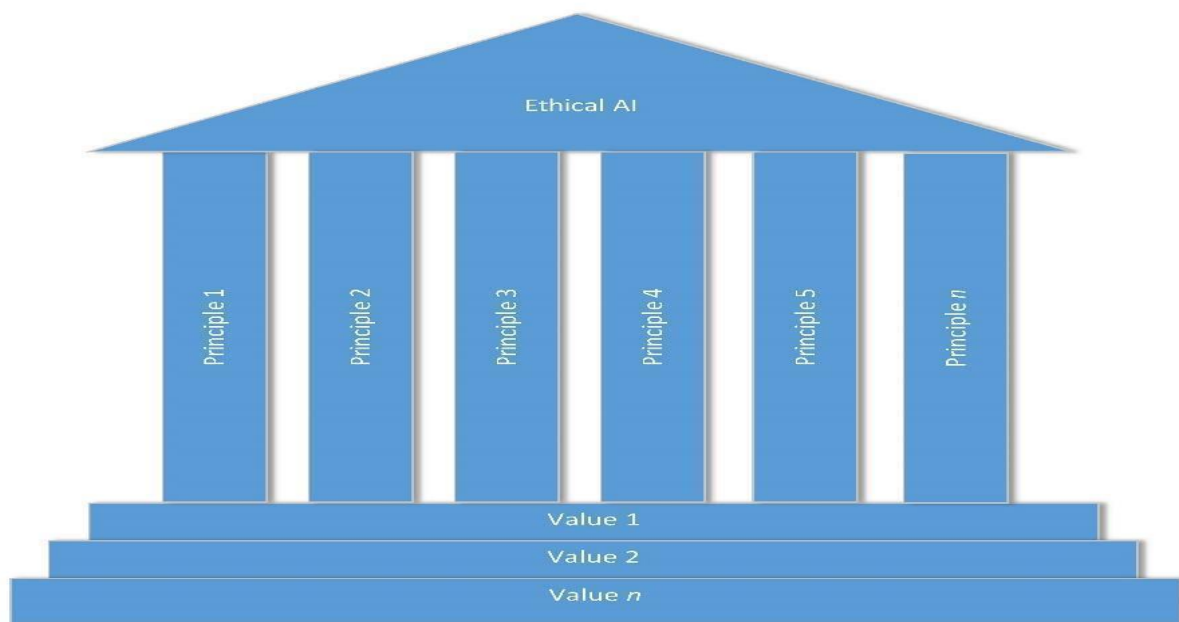
³⁵ In addition to the COMEST's reports indicated above, see the working document "Toward a Draft Text of a Recommendation on

the Ethics of Artificial Intelligence". <https://unesdoc.unesco.org/ark:/48223/pf0000373199?posInSet=1&queryId=3428aff7-f290-425b-a8cd-57715396ccca>

attention to how AI technologies are already impacting our world. Thus, the AHEG identified and conceptualized the issues that need to be corrected and the outcomes to achieve with the proposed ethical framework. Their deliberation was one of the most fascinating discussions I have witnessed, as it encompassed different perspectives, cultural, and geographical diversity. It included experts on AI, but it also gathered together philosophers, psychologists and historians. It was led by an effective and inclusive Chair from South Africa, Emma Ruttkamp-Bloem, Head of the Department of Philosophy at the University of Pretoria, and AI Ethics lead at the South African Center for AI Research (CAIR). The team also included, and benefited from the engagement of the Chair of COMEST, Peter-Paul Verbeek, Professor at the University of Twente. On the Secretariat side, which I was proud to lead, the team of the Bioethics and Ethics of Science and Technology section, led by Dafna Feinholz, was composed of a group of highly talented staff, that already had a strong track record of supporting the elaboration of ethical standards in UNESCO.

The AHEG stressed two crucial dimensions: 1) The AI technologies' role in solving complex challenges and bringing immense economic value; and 2) the concerns that AI creates regarding gender equality, diversity and inclusion, biases, privacy and data protection, increase of disinformation, digital exposure by populations at risk, to name a few. This showed the need for a general universal ethical guidance in terms of core values that must underpin the development of AI systems.

Having agreed that AI technologies are not value-neutral, the AHEG took the approach of identifying foundational values and closely interlinked principles, which inform further outlined policy actions. Foundational values have a role of necessary preconditions or prerequisites for principles and policy actions to work. Taken together they form a coherent framework to ensure ethical AI. The basics of the overall approach can be visualized as follows:



The values identified include the full respect and promotion of human dignity, human rights and fundamental freedoms, leaving no one behind, living in harmony, trustworthiness. The draft Recommendation included principles such as human oversight, privacy, fairness, transparency and explainability, safety and security, responsibility and accountability, and others. Apart from that, the AHEG proposed new principles, which have not been previously seen elsewhere, such as proportionality, which helps ensuring that AI methods correspond to the aims, do not bear excessive risks and are appropriate to the context. The difference between this work and many other AI standards, is that the group of experts developed very clear definitions of each of the principles, clarifying their meaning and proposing policy actions that would ensure their implementation in practice. The AHEG then introduced concrete

policy actions for implementation, and attempted to address the identified problems also by introducing innovative approaches, e.g., in the areas of governance, impact assessment and capacity-building. I enjoyed participating in the discussions of the group, and also supporting the actionable nature of the Recommendation, particularly by calling on the need for strong regulatory frameworks at the national level, in line with UNESCO's Recommendation.

The document prepared by the experts, was then submitted to a comprehensive and inclusive global and regional multi-stakeholder consultations, conducted from June to August 2020, led by the Social and Human Sciences sector. In light of the challenges posed by the pandemic, worldwide partners³⁶ came together with UNESCO to support this consultation process based on three components: (i) public online consultation; (ii) regional and sub-regional virtual consultations co-organized with host

³⁶ See <https://en.unesco.org/artificial-intelligence/ethics/partners>

countries/institutions and UNESCO's Field Offices in all of UNESCO's regions; and (iii) open, multi-stakeholder, and citizen deliberation workshops organized by partners. The public online consultation received completed responses from more than 600 participants, generating more than 50,000 comments. Furthermore, eleven regional and sub-regional virtual consultations, including two with youth, were held in all of UNESCO's regions, involving more than 500 participants. The majority of the consultations were led by the Assistant Director-General for Social and Human Sciences.

Finally, the open, multi-stakeholder, and citizen deliberation workshops drew approximately 500 participants. There was also a targeted effort to consult with the private sector, and particularly the major digital platforms, which also provided extensive suggestions. A dedicated session was organized with the International Chamber of Commerce, with the participation of a large group of companies. Consultation on the first version of the draft Recommendation was also held with the UN system through the intersessional meeting of the High-Level Committee on Programmes (hereinafter 'HLCP') in July 2020.³⁷

The consultations beefed up the content of the Recommendation and asked for it to be bolder, particularly in terms of addressing gender issues, and broader inclusion, justice and integrity. Greater emphasis was needed to reflect the differential access to AI and its

impact across subgroups and individuals - especially women, girls, older persons, youth and children - which should run through the entire document. It was suggested for the Recommendation to address more the issues of diversity of country groups, as well as differences within countries in terms of level and diffusion of AI technologies. It was recognized that the draft Recommendation highlighted the trade-offs, which are not always taken into account in other documents: it was suggested to elaborate on those furthermore.

Suggestions were also made to further elaborate on strategies to be followed in helping the low- and middle-income countries - including countries in Africa, Latin America and the Caribbean, and Central Asia, as well as Small Island Developing States - to access AI systems in various fields. There was a need identified for principles, actions and other recommendations aimed specifically at bridging gaps, capacities, resources, and power imbalances between countries and regions with regard to the proliferation of AI technologies and capacity to regulate between AI-producing and AI-consuming countries. There was a strong call not to have a "people-centred" standard, but a broader perspective including the impact of AI systems on the environment. The consultations were fascinating, and I want to thank, particularly, all the gender-equality champions that attended our call.

Consultations also underscored that the additional value added of the UNESCO

³⁷ The HLCP is the principal mechanism for UN system-wide coordination and policy coherence in the programme area. It is responsible for fostering coherence, cooperation and coordination

on the programmatic dimensions of strategic issues facing the United Nations system. See <https://unsceb.org/high-level-committee-programmes-hlcp>

instrument is linked to the more action-oriented sections, advising on “how” to foster ethical AI developments, besides the “what” related to the definitions of what is ethical. It was also suggested that the draft Recommendation be more ambitious.

The AHEG revised the text accordingly and the first draft of the Recommendation was sent to Member States in September 2020. Forty-eight Member States and two observers submitted their comments and observations, which were integrated by the Social and Human Sciences sector working with other sectors of UNESCO.

2.2 *Member States' Cooperation*

The next stage of the elaboration was to constitute a negotiating body to review and approve the Recommendation, in the form of a Special Committee of representatives of Member States that met from April to June 2021. The Chair of the Committee, Ambassador Adam Al Mulla from Kuwait, was key in ensuring that the text was approved on time. Working hand-in-hand with the Secretariat's team, he ensured that the level of ambition of the text was not watered down, and brought 193 Member States to achieve consensus. He was supported by the Vice-Chairs from Spain, Slovenia, Mexico and Bangladesh, and the Rapporteur Immolatrix Linda Geingos from Namibia.³⁸ The negotiations benefited from the commitment of many Member States, and champions from different regions. It also helped having some members of the AHEG as representatives of their countries (Egypt, France, Latvia, Poland) and with the

presence, as advisor, of the Chair of the AHEG.

The Chair of the negotiations also suggested to have some criteria that were helpful in guiding the discussions of amendments, namely, those of high importance for the countries, focusing on the substance of the Recommendation, preserving the effectiveness of this non-binding instrument, and avoiding issues that were already debated or issues that might be addressed elsewhere in the text, as well as issues regarding specific language, style, or other details that are not related to the substance of the instrument. He called on us to intervene as many times as were deemed appropriate so as to advance the debate. Having had many years of experience in international negotiations, for me this was one of the most interesting ones because of its length and scope, and the clear commitment of Member States to achieve a strong final outcome.

It is worth mentioning that the whole negotiation of the 150+ paragraphs was done in the context of the COVID-19 pandemic confinement. This was perceived as an insurmountable challenge at the beginning but proved to be a blessing in disguise, as Member States used the technology (the chat in the digital platform), to agree or disagree on the presented proposals, and this helped significantly to advance in the negotiations. The decisions that we took with the Chair not to stop the negotiations until the document was done were also key, as we were ready to extend

³⁸ See <https://en.unesco.org/artificial-intelligence/ethics#intgvtprocess>

the end point beyond June and organized intersessional discussions. Going beyond the allocated time was not necessary, as the Chair suggested not having days without negotiations.

However, the debates were not straightforward. The number of comments by Member States came to about 500, and positions were highly diverging – from those who wanted to keep the status quo, and the light regulatory stance towards AI technologies, to those who wanted to change the business model for more sustainable outcomes. There were also strong disagreements between a large group of countries that wanted a strong follow-up and monitoring chapters, and those that would have preferred to have a good narrative but not concrete follow up. In the course of the discussions, some countries from the north, who considered that the Recommendation would add little to the international list of instruments, ended up recognizing its value added.

A major debate focused on to what extent ethics should replace or reinforce human rights. Some Member States feared that the ethical framing may be used as an excuse to ignore the internationally agreed framework for human rights, and therefore their original position was to strengthen the language in this domain. However, the draft Recommendation was already strong on this aspect, and the group of supporters focused on retaining the original draft, and they succeeded. It helped that the Recommendation considers AI ethics as a dynamic basis of reflection that can guide societies in dealing responsibly with the

known and unknown impact of AI, aiming to ensure the protection of human dignity, human rights, well-being and the prevention of harm.

Technologies in general, and AI technologies in particular, evolve very quickly and it is not always easy to anticipate their impact. That is why having a fixed set of principles or ethical codes is not enough. An inclusive ethical reflection to ensure that these technologies continue to be used without infringing on human dignity, human rights and fundamental freedoms is constantly needed. The description of ethics as a dynamic process does not refer to a change in values. On the contrary, it aims at ensuring that values are always respected and existing frameworks, including legal frameworks, are adapted if needed, to ensure the protection of human dignity, human rights and fundamental freedoms, while technologies advance. Ethics is about values, and it is the source of every decision and legal framework; ethics, explicitly or implicitly, is what informs policy and law. Human rights are a way of codifying these principles through law. However, as we know, not all ethical questions can be addressed only by law, or could be codified or regulated, and decisions might need to be made in specific cases to ensure full adherence to the human rights framework, but where no regulation is provided. In such instances, ethics remains the only available tool.

The obligatory character of human rights law is underlined throughout the Recommendation. This includes ensuring that human rights are not subject to trade-offs. Generally, it follows the international

human rights framework and does not provide new interpretations thereof, including when it concerns possible adverse human rights impacts of AI technologies. Examples of such impacts are multiple ranging from possible breaches of privacy as in cases of AI-powered applications that collect and use data for purposes of monitoring and contact tracing during the COVID-19 pandemic to discrimination on various grounds, including gender, as in cases of AI recruiting tools that discriminate against women due to use of male-dominated datasets or biased programming. In addition, it refers to obligations under international law throughout the text, including a general clause at the end of the Recommendation providing that the instrument is without prejudice to international law and human rights obligations.

Nevertheless, disagreements among members appeared on the human rights content, with one group wanting to increase reference to these rights, while others wanting to reduce them. A spirit of compromise was found, in this, and in many other issues.

The chapter that brought more intense negotiations was the one related to the ethical governance and the data policy. The Recommendation is clear that the data should be owned and managed by data subjects who should always remain in control of their information, including in deleting their digital footprint if needed. It also establishes the need to be informed, and to know, when a decision that concerns an individual is taken with the support of AI

systems. There is an obligation to compensate those who may be harmed by certain developments in AI, making sure that the rule of law is enforced online – just as it is offline. The majority of the provisions remained.

Further, several Member States supported more effective regulation as articulated in the policy areas of the Recommendation, while others expressed their concern that regulatory measures were proposed. It must be noted that this instrument outlines aims and shared objectives to be achieved, and proposes ways to achieve them, while the means through which to do so are decided at the national level. At the same time the Recommendation includes strong follow-up mechanisms to ensure implementation, including Ethics Committees, and the Ethical Impact Assessment and Readiness Assessment Methodology tools. These important features of the instrument are further elaborated later in this paper. I, for one, am convinced that it is about time to ensure that the rule of law prevails online as it does offline, and that effective regulations are needed. One of the major contributions is the call for redress mechanisms for those that have been affected.

Although the debates were often heated and special rules of interaction had to be devised, state representatives managed to achieve consensus on truly strong and ground-breaking provisions. For example, members took a strong stance on the unresolved issues of ethical and legal responsibility: it can only

be attributed to natural or legal persons,³⁹ while AI systems cannot be provided with legal personality.⁴⁰ Further, the Recommendation states that control should always be in the hands of human, and that AI systems can never replace ultimate human responsibility and accountability.⁴¹ This requires that the development of AI systems should be done in parallel to strengthening human capacity in order to ensure human oversight. Nevertheless, it is recognized that there can be situations where human decision is not possible, but the decision to cede control in limited contexts remains that of human. It was agreed that in scenarios where decisions are understood to have an impact that is irreversible or difficult to reverse or may involve life and death decisions, final human determination should apply,⁴² and as a rule, life and death decisions should not be ceded to AI systems.⁴³ Importantly, the Recommendation also takes a very strong stance that AI systems should not be used for social scoring or mass surveillance purposes.⁴⁴

These are just some examples of the issues addressed (an overview of some other important provisions is provided in the next section) and of the outcomes of inter-state cooperation. It shows the ability to bring humanity to be conscious of what is at stake, to reflect and understand the things we care about and value.

3 Inside the Recommendation

The Recommendation aims to provide a basis to make AI systems work for the good

of humanity, individuals, societies, and the environment and ecosystems, and to prevent harm. For these purposes it approaches AI ethics as a systematic normative reflection, based on a holistic and evolving framework of interdependent values, principles and actions, that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies and offers them a basis to accept or reject AI technologies.⁴⁵

³⁹ Para. 35 of the Recommendation.

⁴⁰ Para. 68 of the Recommendation.

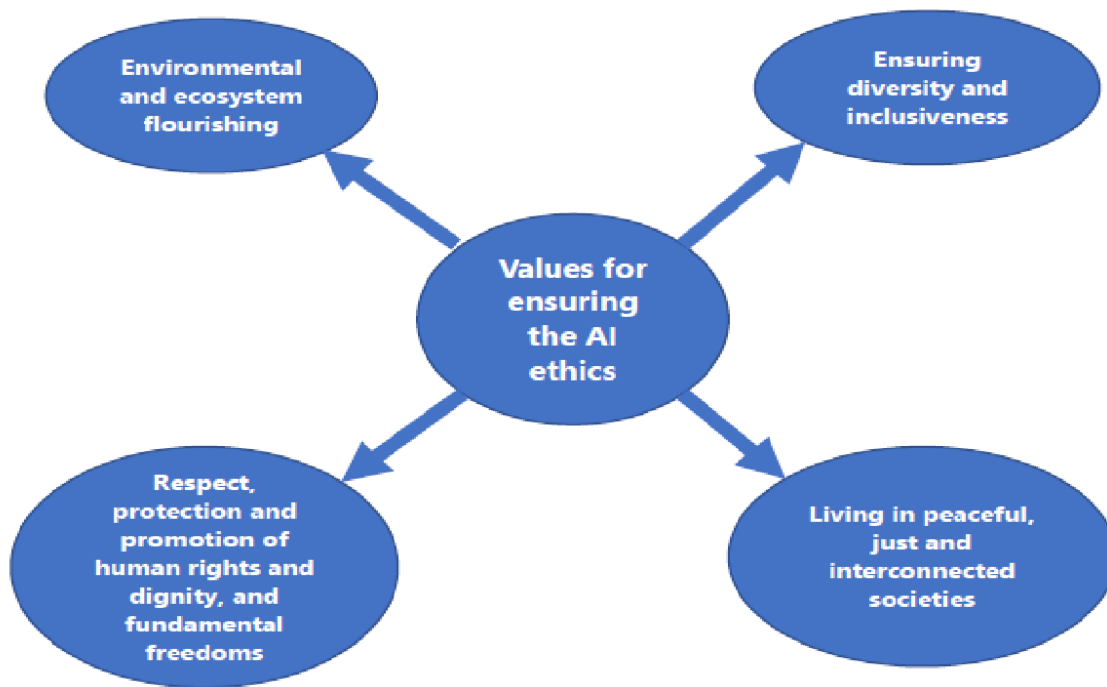
⁴¹ Para. 36 of the Recommendation.

⁴² Para. 26 of the Recommendation.

⁴³ Para. 36 of the Recommendation.

⁴⁴ Para. 26 of the Recommendation.

⁴⁵ Para. 1 of the Recommendation.



The document is future-looking in many respects. It provides an anticipatory framework to help mitigate future risks and offers concrete policy actions designed to resolve and redress existing and emerging harms. It also does not limit its scope by providing one single definition of AI, in recognition that technologies evolve and trying to cover such future developments, making the Recommendation a living document adaptable to changing circumstances.⁴⁶

As explored above, the Recommendation is composed of a set of human rights-based values, principles and policy areas, all synchronously operating together to put people and their individual rights at the centre. Values and principles are inter-related in the following manner: while the set

of foundational values inspires desirable behaviour and represents the foundations of principles, the principles unpack the values underlying them more concretely so that the values can be more easily operationalised in policy statements and actions.

Overall, there are *four values*, shown in the following scheme:

It is important to note that environmental concerns have been deliberately elevated to values to underline those goals for sustainable and inclusive development should prevail. The analysis of various documents proposing principles for AI showed that the protection of the environment receives little attention or is

⁴⁶ Para. 2 of the Recommendation.

overlooked,⁴⁷ despite the enormous effect that AI technologies have on it. Indeed, data extraction alone consumes about 10% of energy globally,⁴⁸ and it is estimated that training one large deep-learning model produces about 300,000 kg of carbon dioxide,⁴⁹ (i.e., 125 round-trip flights between New York and Beijing).⁵⁰ This brings specific added value on the side of UNESCO to the global agenda to ensure that these technologies help us solve our problems rather than create more.

Furthermore, the Recommendation identifies *ten principles* in accordance with which all actions must be undertaken. It must be noted that although many other documents provide for a principle of trustworthiness, in this Recommendation it has been framed as an outcome of the operationalisation of all of the principles, and the proposed policy actions are all directed at promoting trustworthiness in all stages of the AI lifecycle. The principles range from safety and security – addressing unwanted harms and vulnerabilities to attack – to awareness and literacy in recognition of the necessity to ensure public awareness and understanding of AI technologies and the value of data to ensure effective public participation so that all members of society can take informed decisions about their use of AI systems and be protected from undue influence.

Critically, the Recommendation looks at very specific ways to operationalise such principles like transparency, accountability and responsibility. The underlying aim is that, to ensure the beneficial character of new technologies, it is important for the rule of law to be respected (and built when there are gaps) in the digital world as it is in the physical world. It is well-known that regulations are often behind market developments, but in the case of AI, it is even more so, and the focus has been on maximizing profit or geopolitical competition. With its principles, the Recommendation creates a re-equilibrium in favour of people and moves to ensure that no major decisions are left in the hands of developers and platforms.

To do so practically, the Recommendation goes beyond a declaration of ethical values and principles to include concrete policy actions for their operationalisation grouped into *eleven policy areas*, while maintaining focus on all stages of the AI system life cycle. This is crucial because often, bias and discrimination manifested downstream as experienced by the users is the product of structural inequalities and exclusion in the upstream, such as the lack of diversity in the management or technical teams developing these AI technologies.

The Recommendation provides for strong policy actions on the management of data,

⁴⁷ See e.g. A. Jobin, M. Ienca and E. Vayena, “The global landscape of AI ethics guidelines”, *Nature Machine Intelligence* 1, 389-399 (2019), p. 396.

⁴⁸ Corcoran, P., & Anders, A. (2013). Emerging Trends in Electricity Consumption for Consumer ICT (Global Forecasting of ICT Footprints and Handprints Project). https://www.researchgate.net/publication/255923829_Emerging_Trends_in_Electricity_Consumption_for_Consumer_ICT

⁴⁹ Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. 57th Annual Meeting of the Association for Computational Linguistics. <https://arxiv.org/abs/1906.02243>

⁵⁰ Dhar, P. (2020). The Carbon Impact of Artificial Intelligence. *Nature Machine Intelligence*, 2, 423–425. <https://www.nature.com/articles/s42256-020-0219-9>

privacy and access to information. It highlights the importance of data governance, promotes a better understanding of the role of data in developing secure and equitable algorithms. It sets the rules for keeping control over the data in the hands of users, allowing them to access and delete information as needed. It also calls on Member States to ensure that appropriate safeguards for the processing of sensitive data and effective accountability schemes are devised, and to provide redress mechanisms in the event of harm.

The Recommendation further proposes protections against bias and discrimination – it ensures that as we develop these technologies, we have inclusive and representative datasets, and algorithms that are not biased or discriminatory. Thus, it has a strong chapter to advance gender equality in the digital and AI worlds.

Indeed, women, and particularly those from developing countries, are once again being left behind. For example, in terms of connectivity, the global internet penetration rate is 45% for women as compared to 51% for men, and this gap further widens in rural areas.⁵¹ 165 million fewer women than men own a mobile phone and women are 20% less likely than men to own a smartphone. Worldwide, women are on average 20% less likely than men to access mobile internet – that is roughly 300 million fewer women. This digital gender gap is even more prominent in the Global South: at 51%, South Asia has the world's widest digital gender gap, followed by Sub-Saharan Africa at 37%. In education,

female students are underrepresented in STEM- and ICT-related fields. Less than a third of female students choose to study STEM courses in higher education and this figure dips to just 3% of ICT courses, as stereotypes go against women's engagement with technologies and sciences. These divergences are then translated into the working environment.

Gender gap still exists today and is predicted to stay for decades to come. The most recent estimates say that it will take 135.6 years to close the gender gap worldwide.⁵² The gender gap specifically in economic participation and opportunity will take 267.6 years to close. The Middle East and North Africa region remains the area with the largest gap (60.9%). The largest gender gap is in AI and data – women make up 32%. This is followed by other sectors that require disruptive technical skills: in cloud computing women make up 14% of the workforce, in engineering 20%. This reinforces the gender pay gap since the AI sector offers lucrative jobs, which is not the case of female-dominated sector (services, care work, etc.). Therefore, the Recommendation highlights the need to tackle gender biases which are currently replicated by AI systems due to biased programming; it also recommends adopting very concrete measures to promote gender parity in this field, and ensure that women are not only users, but also shapers of these technologies. In particular, the Recommendation calls on Member States to promote policies, which include dedicating funds from the public budgets linked to financing gender-related

⁵¹ See <https://www.weforum.org/agenda/2020/02/stem-gender-inequality-researchers-bias>

⁵²

schemes, which unfortunately is often lacking or insufficient; ensuring that national digital policies include a gender action plan; encouraging female entrepreneurship, participation, engagement and leadership in all stages of the AI life cycle; investing in targeted programmes to increase the opportunities of girls' and women's participation in STEM and ICT disciplines; and eradicating gender stereotyping and ensuring that discriminatory biases are not translated into the AI systems.⁵³

We also know that AI systems can have a terrible impact on children, on democracies, on misinformation, and can exponentially divide our societies. Therefore, particular attention is paid to the psychological and cognitive impact that these systems can have on children and young people. It is provided that Member States should invest in and promote digital, media and information literacy to strengthen critical thinking and competencies. AI systems must be understood in order to mitigate and counter disinformation, misinformation and hate speech. The Recommendation pays particular attention to the possibility of using AI to manipulate and abuse human cognitive biases.⁵⁴

Given the impact of AI technologies on labour and skills, the Recommendation calls for fair transition for at-risk employees, including upskilling and reskilling, and "safety net" programmes. It calls for relevant regulations, such as tax regimes, to counteract the consequences of unemployment caused by AI-based

automation. States should assess and address the impact of AI systems on labour markets and its implications for education requirements, in all countries and with special emphasis on countries where the economy is labour-intensive. Such assessment can include the introduction of a wider range of "core" and interdisciplinary skills at all education levels to provide current workers and new generations a fair chance of finding jobs in a rapidly changing market and to ensure their awareness of the ethical aspects of AI systems. Skills such as "learning how to learn", communication, critical thinking, teamwork, empathy, and the ability to transfer one's knowledge across domains, should be taught alongside specialist, technical skills, as well as low-skilled tasks. Being transparent about what skills are in demand, and updating curricula around these are key.

Throughout the Recommendation specific attention is paid to low- and middle-income countries (LMICs), including Least-Developed Countries (LDCs), Landlocked Developing Countries (LLDCs) and Small Island Developing States (SIDS), as they have their own capacity, but have been underrepresented in the AI ethics debate, which raises concerns about neglecting local knowledge, cultural and ethical pluralism, value systems and the demands of global fairness to deal with the positive and negative impacts of AI technologies. The high concentration of such technologies leads not only to the issue of access to them, but also who is shaping the digital revolution and how, which is currently done by a handful of

⁵³ Paras. 87-93 of the Recommendation.

⁵⁴ Para. 125 of the Recommendation.

firms. The Global South, in particular, is largely absent from such discussions.

Indeed, we know that much of AI technology is currently concentrated in the hands of a few companies in a limited number of countries. For example, companies, in particular those from Japan, the US and China, dominate patenting activity. Companies represent 26 of the top 30 AI patent applicants, while only 4 are universities or public research organisations. Of the top 20 companies filing AI-related patents, 12 are based in Japan, 3 are from the U.S. and 2 are from China. Chinese organisations make up 17 of the top 20 academic players in AI patenting as well as 10 of the top 20 in AI-related scientific publications. Together, the US, China and Japan offices account for 78% of total patent filings. Big tech companies continue to dominate the field of AI. IBM has the largest portfolio of AI patent applications with 8,290 inventions, followed by Microsoft with 5,930.⁵⁵ Over 60 AI start-ups have been acquired by the largest US tech companies, with Apple being the top acquirer since 2010, followed by Google, Microsoft and Facebook.⁵⁶

The Recommendation is generally addressed to all AI actors, but the responsibility of implementing it is in the hands of governments, which necessitates them to increase capacities. Therefore, the Recommendation considers ethical implications for all AI actors at all stages of

the AI system life cycle within its scope, and emphasises the principle of multi-stakeholder and adaptive governance and collaboration. Member States are encouraged to promote inclusive AI governance mechanisms and to set out appropriate frameworks to ensure that other stakeholders, in particular private sector companies, adhere to them. Many companies continue to think that self-regulation of AI is enough. Given AI's far-reaching impact on all aspects of society, self-regulation is insufficient for keeping AI in check and such voluntary standards could easily be disregarded in the pursuit of profits. Therefore, governments need to adopt a regulatory framework to mitigate risks and avoid the harmful consequences of AI, in particular by ensuring that there are compensatory mechanisms in place. There is a need to rebalance the power, and the Recommendation calls on governments to ensure the rule of law so that AI is developed for good.

We worked closely with the private sector and will continue to do so for the implementation. Many analysts have a negative assessment of the role of big tech, and the monopolies that may emerge from having few big digital platforms. We are of the view that the existence of monopolies or non-competitive practices is allowed by the lack of regulation. So, it is there where the emphasis should be put. The private sector, though, is a very important partner, and there are now many multinational companies

⁵⁵ World Intellectual Property Organization. (2019). *WIPO Technology Trends 2019: Artificial Intelligence*. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055_exec_summary.pdf

⁵⁶ Bass, D., & Brustein, J. (2020, March 16). *Big Tech Swallows Most of the Hot AI Startups*. Bloomberg. <https://www.bloomberg.com/news/articles/2020-03-16/big-tech-swallows-most-of-the-hot-ai-startups>

that are partnering with UNESCO to advance implementation. They are also sharing their experiences on how to ensure transparency of algorithms among others. We also worked closely with civil society and will continue to do so.

Therefore, it is important to underline that - unlike the ethical frameworks regarding AI that have already been developed by various organizations all over the world - this Recommendation brings about the first global-level normative instrument also focused on the practical realization of the outlined values and principles via concrete policy recommendations, making it especially strong and innovative. It is also unique as it puts a strong emphasis on remedial action or redressing the damage caused by the use of the technologies. It is one of the very few instruments that is so comprehensive and detailed, that offers policy guidance, and that request the development of concrete tools for its implementation.

4 The impact of the Recommendation

4.1 *The 'non-binding character' issue*

In the context of UNESCO, recommendations are instruments "in which the General Conference formulates principles and norms for the international regulation of any particular question and invites Member States to take whatever legislative or other

steps may be required – in conformity with the constitutional practice of each State and the nature of the question under consideration – to apply the principles and norms aforesaid within their respective territories".⁵⁷ Adopted by the General Conference, the norms contained in a recommendation are not subject to ratification, but Member States are invited to apply them. Recommendations are intended to influence the development of national laws and practices.

Therefore, the Recommendation is a soft law instrument. In the field of international law, soft law and hard law have their own benefits. Beyond recommendations, treaties with soft obligations, voluntary resolutions, codes of conduct formulated and accepted by international and regional organizations and statements prepared by individuals in a non-governmental capacity can be soft law norms.⁵⁸ Nevertheless, soft law should not be disregarded as ineffective. In fact, soft law can be described as "rules of conduct which in principle have no legally binding force but which nevertheless may have practical effects."⁵⁹ It has been also argued that soft law is more suitable in the construction of various international regimes.⁶⁰ Particularly, soft law is more suitable to making some experimental movements when compared with hard law.

⁵⁷ Article 1(b) of UNESCO's Rules of Procedure concerning recommendations to Member States and international conventions covered by the terms of Article IV, paragraph 4, of the Constitution.

⁵⁸ See generally Boyle, A. E. (1999). Some reflections on the relationship of treaties and soft law. *International and Comparative Law Quarterly*, 48(4), 901–913.

⁵⁹ Snyder, F. (1993). The effectiveness of European community law: Institutions, processes, tools and techniques. *The Modern Law Review*, 56(1), 19–54.; Snyder, F. (Ed.). (1994). *Soft Law and*

Institutional Practice in the European Community. In *The Construction of Europe: Essays in Honour of Emile Noël* (pp. 197–198). Kluwer.

⁶⁰ E.g., in the case of the international economic regime see Chaisse, J., & Ji, X. (2018). 'Soft Law' in International Law-Making—How Soft International Taxation Law is Reshaping International Economic Governance. *Asian Journal of WTO & International Health Law and Policy*, 13(2), 463–509.

There are multiple benefits in having soft law documents.⁶¹ One of the biggest, especially relevant for such a dynamic field as AI, is greater flexibility. Soft law provides greater flexibility in renegotiating agreements when things change.⁶² At the same time, it can accommodate diverse legal systems and cope better with uncertainty.⁶³ Thus, actors cannot predict all the possible results if the problems met, cannot be understood well.⁶⁴ Flexibility is particularly beneficial in the current world with fast changing conditions.⁶⁵ Under certain conditions, soft law's flexibility may be a primary reason for its use.⁶⁶ It facilitates innovation and experiment, and it can provide the essential preparatory phase before countries are ready for hard law commitments.⁶⁷

Further, soft law provides for greater simplicity and speed, as it does not require domestic ratification processes and attracts as many countries, no matter what their domestic legal systems are, as possible; soft law can be functional here.⁶⁸ If there is

something unforeseen, soft law is useful in dealing with this condition to reach agreements.⁶⁹ In sum, although soft law is non-binding, it still has legal relevance,⁷⁰ and especially translates into domestic legal changes if combined with follow up mechanisms and peer pressure. Indeed, peer pressure and peer learning, with strong monitoring infrastructure can deliver change on the ground. Also, it should not be overlooked that soft law also helps to build shared international narratives; that is not a minor contribution that the UNESCO Recommendation is making.

Moreover, there are historical examples of soft law non-binding documents that became authoritative and eventually binding through various means. One of the prominent examples is the 1948 Universal Declaration of Human Rights (hereinafter 'UDHR').⁷¹ Although it is a non-binding declaration, its provisions became binding through customary law⁷² and codification in various international treaties.⁷³ There are also

⁶¹ For an overview see Chaisse, J., & Ji, X. (2018). 'Soft Law' in International Law-Making—How Soft International Taxation Law is Reshaping International Economic Governance. *Asian Journal of WTO & International Health Law and Policy*, 13(2), 477–478.

⁶² Chinkin, C. M. (1989). The Challenge of Soft Law: Development and Change in International Law. *The International and Comparative Law Quarterly*, 38(4), 850, 852–53.

⁶³ Shelton, D. (Ed.). (2000). Interdependence, Globalization, and Sovereignty: The Role of Non-Binding Legal Accords. In *Commitment and Compliance: The Role of Non-binding Norms in the International Legal System* (pp. 94–95). Oxford University Press.

⁶⁴ Chinkin, C. M. (1989). The Challenge of Soft Law: Development and Change in International Law. *The International and Comparative Law Quarterly*, 38(4), 850, 855, 861.

⁶⁵ Shelton, D. (Ed.). (2000). Interdependence, Globalization, and Sovereignty: The Role of Non-Binding Legal Accords. In *Commitment and Compliance: The Role of Non-binding Norms in the International Legal System* (pp. 94–95). Oxford University Press.

⁶⁶ Guzman, A., & Meyer, T. (2010). International Soft Law. *The Journal of Legal Analysis*, 2(1), 171–184.

⁶⁷ Kirton, J. J., & Trebilcock, M. J. (Eds.). (2004). Hard and Soft Law in International Institutions: Complements, Not Alternatives. In *Hard*

Choices, Soft law: Voluntary Standards in Global Trade, Environment, and Social Governance. Ashgate, 347, 351.

⁶⁸ Lipson, C. (1991). Why are some international agreements informal? *International Organization*, 45(4), 495–501.

⁶⁹ Lipson, C. (1991). 495–501.

⁷⁰ Guzman, A., & Meyer, T. (2010). International Soft Law. *The Journal of Legal Analysis*, 2(1), 171–184.

⁷¹ <https://www.un.org/en/about-us/universal-declaration-of-human-rights>

⁷² See e.g., *Filartiga v. Pena-Irala*, 630 F.2d 876 (2nd Cir. 1980). The US Court of Appeals regarded prohibition of torture as customary international law as per UDHR. On formation of customary law see e.g.

<https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199552153.001.0001/oxfordhb-9780199552153-e-19> and

<https://www.law.umich.edu/facultyhome/drwcasebook/Documents/Document/ILA%20Report%20on%20Formation%20of%20Customary%20International%20Law.pdf>

⁷³ E.g., International Covenant on Civil and Political Rights (1967) https://treaties.un.org/doc/Treaties/1976/03/19760323%2006-17%20AM/Ch_IV_04.pdf and

specific instances of UDHR being applied on national level⁷⁴ and even overruling conflicting domestic legislation.⁷⁵

There are also UNESCO examples of the application of soft law instruments, and specifically recommendations. This is the case of the Recommendation on the Means of Prohibiting and Preventing the Illicit Export, Import and Transfer of Ownership of Cultural Property.⁷⁶ Thus, a German court, in its *Nigerian Cultural Property* case of 1972, applied the terms of the recommendation when deciding whether the export of cultural property could be legitimately insured.⁷⁷

Therefore, the choice of addressing the issues of the ethics of AI via the means of a recommendation can be explained in multiple ways, as explored above, not least because of the complexity of the subject matter and virtual impossibility to achieve a binding document of such depth in such a short time, and with so many countries subscribing to it. Nevertheless, this document has all the advantages outlined above and sets solid groundwork for its provisions, attaining the binding status in whatever form in the coming future.

4.2 Implementation mechanisms

The implementation depends to a large degree on the substance of the document. As has been shown, the Recommendation on the Ethics of AI has very profound provisions, with concrete follow-up actions. Importantly, it is not just a set of goals, but very concrete

steps on how to ensure them in practice, and UNESCO has a follow-up system for these purposes.

First of all, Member States are called on to translate the content of the Recommendation in regulatory changes in their countries. This is important, as it puts the emphasis on the role of governments and their duty of care. Strong regulatory frameworks are core to this exercise, and jointly with the request to UNESCO to develop specific tools, analysis, and assessment of policy effectiveness, the Recommendation has a strong emphasis on implementation and enforcement.

The Recommendation also calls for an increase in the capacities of governments to deal with these technologies and on the establishment of Ethics of AI Committees. This is in line with the practice of UNESCO working with Member States, that had led to strong institutional innovation in the form of the bioethics committees.

Further, the periodic reports mechanism serves as a monitoring mechanism to identify best practices, gaps, challenges for implementation, emerging risks and new principles that are needed as AI develops. UNESCO offers to be a partner and support Member States in the development, as well as monitoring and evaluation, of policy mechanisms and changes into national legislation and institutions.

⁷⁴ Klabbers, J. (1998). *The concept of treaty in international law*. Kluwer Law International: p. 161.

⁷⁵ See the *Klahr Ehlert*-decision of the Guatemalan Supreme Court, 9 May 1952. As reported in Skubiszewski, K. "Recommendations of

the United Nations and municipal courts", 46 *BYIL* (1972-73), 353-364.

⁷⁶ See http://portal.unesco.org/en/ev.php-URL_ID=13083&URL_DO=DO_TOPIC&URL_SECTION=201.html

⁷⁷ Klabbers, J. (1998): p. 162.

Secondly, in order to help Member States implement the Recommendation and unpack the specific steps that need to be taken, the Recommendation provides for UNESCO to develop specific tools: Readiness Assessment Methodology and Ethical Impact Assessment.

It is envisaged that the readiness assessment methodology will identify the status of countries in terms of their preparedness to implement the Recommendation in its scientific, technological, economic, educational, regulatory, infrastructural, societal, cultural and other dimensions. Thus, it will help in assessing whether Member States have the institutional capacity to develop regulatory frameworks and policies, carry out administrative processes, to benefit and contribute to ethical development and deployment of AI. The ethical angle will be included in national AI strategies, data governance standards, and monitoring and evaluation mechanisms to ensure that they are in line with the Recommendation. In particular, the readiness assessment methodology can help enhance the institutional capacity of Member States to implement the Recommendation and identify if they have effective *ex ante* and *ex post* assessments when regulating AI-relevant fields. In addition, the readiness methodology will help determine whether the existing legislation in the country, such as data protection regulation, is equipped to address the changes brought by AI systems, and if there are appropriate audits and controls in place. The technological dimension would include internet access, e-skills, and protocols for data collection and

analysis, which is critical given the reliance of AI technologies on data.

The ethical impact assessment (hereinafter 'EIA'), in its turn, will assess, monitor and evaluate the ethical impacts of AI systems on societies and assist Member States to take action. It will cover the entire AI life cycle and assess both the direct and indirect impact on humans, societies, the environment and the human rights. The EIA will be a tool that can be deployed by developers and deployers of AI systems, particularly States and business enterprises. A wide range of AI actors will be engaged in the development and implementation of EIAs. In particular, after developers and deployers utilize the EIA, it will be made public, to enable AI actors, including the end-users, to provide input and feedback for the further refinement of the tool.

While UNESCO acknowledges that the process of developing mechanisms for such a broad array of circumstances is not easy, it is an important follow up step that will help in supporting the Recommendation by giving it a more practical angle. UNESCO will support Member States in carrying out such assessments on a demand-driven basis, as it has a long track record of methodologies to help in the implementation of international instruments, and it is committed to an inclusive process that will take into account the diverse perspectives of different Member States, as well as different stakeholders. Moreover, a High-Level Expert Group on the Implementation of the Recommendation has been convened that will help to deliver and implement the tools.

Third, the implementation strategy is also complemented by a number of initiatives that have been developed to support Member States. UNESCO has established the group of early adopters for the implementation of the Recommendation, which consists of Member States that have expressed their interest to pioneer the implementation of the Recommendation, with various activities planned, including the roundtables for implementation. The first meeting had a record participation of 60 Member States.

Moreover, regional initiatives are being launched for the implementation of the Recommendation, with the support of regional institutions. This is the case of Latin America and the Caribbean, led by Colombia, and Southern African. More advanced economies and institutions are supporting implementation in less-developed countries; this is particularly the case of Japan and the European Commission.

UNESCO will also rely on its intersectoral work to deliver on the different chapters related to education, culture and communications. While the Social and Human Sciences sector is overseeing the whole implementation and following up on the different policy chapters, it is relying on the expertise and leadership of other sectors to advance these important areas of work.

Besides, a number of expert networks are being established, which include:

- The Global Forum on the Ethics of Artificial Intelligence as a central point to debate the challenges of these technologies, taking as a basis the

Recommendation and the analytical contributions of many institutions and governments that are leading in this domain.

- The Global Observatory of the Ethics of Artificial Intelligence will serve as a digital platform to inform about policies in this domain, the implementation of the Recommendation, but also advances on the ground.
- The AI Ethics Experts Without Borders will ensure the availability of dedicated interdisciplinary expert group for deployment in Member States on needs basis to assist in the implementation, particularly in developing countries.
- The envisaged Global Network of Knowledge Centres on Ethics of AI will mobilize some of the world's leading universities and academic centres to support the capacity-building programme on ethics of AI, and especially to contribute towards the elaboration of innovative tools and methodologies, and to conduct research in order to explain and understand the technologies better. The Instituto de Empresa in Spain, and the London School of Economics in the UK are cases in point.
- The Women for Ethical AI Network will mobilize the prominent women champions of gender equality for the implementation of the ground-breaking gender policy area of the Recommendation. The members of this network will be influential women who are today's industry, government and civil society leaders driving transformations towards gender equality in and through AI.

Finally, UNESCO is actively pursuing implementation on the global level. In particular, UNESCO is co-leading, with the International Telecommunication Union, the HLCP Inter-Agency Working Group on Artificial Intelligence. The working group's main goal is to deliver concrete and relevant outcomes for the UN system, and it combines the ethical and technological parts of the UN to provide a solid foundation for current and future system-wide efforts on AI with a view to ensuring respect for human rights and accelerating progress on the SDGs. UNESCO further actively contributes to the roadmap towards the Development of a Global Digital Compact as proposed in the Secretary-General's "Our Common Agenda".⁷⁸

In all this, UNESCO will be bringing its multidisciplinary competencies to assist Member States in fulfilling the Recommendation, and partner with regional institutions for these purposes. The Recommendation will also provide a solid foundation for the entire UN system in its collective response to the ethical challenges posed by AI technologies in various fields, with UNESCO playing a leading role in this area.

UNESCO is already analysing the ethics of neurotechnology, another emerging area where the ethical framework will contribute to its positive development. As all of these technologies are converging, the work on this area will also support the effective delivery of the Recommendation.

Conclusion

AI is a very dynamic field with considerable impact throughout various dimensions and across the world. Innovations in AI are occurring at speeds faster than ever witnessed before. The number of AI patents filed in 2021 is more than 30 times higher than in 2015, showing a compound annual growth rate of 76.9%. Private investment in AI in 2021 totalled around \$93.5 billion, which is more than double the total private investment in the previous year.⁷⁹ Therefore, it is not only pervasive, but fast, and the Recommendation on the Ethics of AI can have a great impact.

UNESCO plays an important role as a space for cooperation and action to protect people and to ensure the rule of law in the digital domain. Inclusion is at the core of what it does, and it is immensely important in the far-reaching technological developments of AI for us to enact policies and regulatory frameworks that matter. The complexity of the ethical issues surrounding AI requires equally complex responses that necessitate the cooperation of multiple stakeholders across the various levels and sectors of the international, regional and national communities.

With its National Artificial Intelligence Policy and Action Plan, Chile has joined the group of countries responding to one of the most pressing challenges for the transformation of our societies. Having established ten-year actions and goals in ethics of AI, the country could continue to lead the region by aligning

⁷⁸ United Nations. (2021). *Our Common Agenda: Report of the Secretary-General*. https://www.un.org/en/content/common-agenda-report/assets/pdf/Common_Agenda_Report_English.pdf

⁷⁹ Stanford's AI Index for 2022. https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf

its policies with a swift implementation of the Recommendation. The Recommendation's mechanisms could purposely be oriented to serve the goals of the Policy of bringing regulatory certainty to AI systems, boosting algorithmic transparency, strengthening data protection, eliminating discrimination from algorithm-driven services, strengthening cybersecurity and empowering women's role in AI.

Global cooperation on the ethics of AI and global inter-cultural dialogue are indispensable for arriving at complex solutions. The global nature of the Recommendation on the Ethics of AI is essential in ensuring inclusion. At first glance, the complexity of issues at stake can seem to make this an insurmountable task. Nevertheless, history is replete with examples of humanity coming together to address a critically important issue, and this can be another shining example of what we can accomplish with a shared vision and coordinated action.

In this context, the interaction between science, diplomacy and society is evident. With this Recommendation, there is a great room for strengthening how national legislation in emerging subjects like AI is being increasingly tightened in line with global state-of-the-art frameworks. As shown, this is a growing soft law topic where diplomacy can be an effective player.

Innovación Pública al servicio de las personas: Una invitación a Chile a liderar la discusión global sobre la transformación del Estado.

Roman Yosif Capdeville⁸⁰

¿Estamos viviendo una revolución?

Esta es una de las principales preguntas que interpela a quienes trabajamos y vivimos dentro del mundo de la innovación.

Durante las primeras dos décadas del siglo XXI, se han presentado variadas señales de necesidad de cambios profundos en los paradigmas y puntos de referencia de nuestra sociedad. Sin embargo, la pandemia del COVID-19 tuvo un impacto multidimensional que la constituyó en un punto de inflexión para la convivencia de la humanidad.

Este proceso transformador se explica al menos a partir de tres fenómenos muy marcados, profundos, por sí solos, pero que, en conjunto, están cambiando nuestra manera de entender la vida en sociedad, la economía, la política y nuestros constructos culturales más hondos.

Vivimos, en primer lugar, un cambio tecnológico radical: la disrupción digital está afectando exponencialmente todas las áreas de la vida. La generación de *bienes digitales* es uno de los principales y más extraordinarios productos de la cuarta

revolución industrial, planteando nuevas formas de comprender el mundo que nos rodea y de interactuar.

Las redes sociales son un claro ejemplo de las nuevas maneras de construir identidades individuales y colectivas donde, por una parte, podemos compartir nuestras ideas, adherir a causas, denunciar hechos que nos parecen incorrectos y potenciar los que valoramos. Por otra parte, nos enfrentamos al profundo desafío de mantener relaciones sociales con quienes tienen perspectivas diferentes a las propias⁸¹.

También, con la creación de las nubes de almacenamiento de datos (*cloud*) o con la conformación de redes distribuidas que soportarán el intercambio de activos digitales (como el *blockchain*), se modificarán las formas en que nos vinculamos con el entorno. Será otro el mundo que nos rodeará: su versión digital, contra todo sentido común, es claramente tangible, además de que lo experimentaremos transversalmente, aunque de maneras muy diversas según nuestro género, edad, zona de residencia, y otros factores que inciden en el acceso que tenemos a este nuevo mundo.

⁸⁰ Ingeniero Comercial. Exdirector Ejecutivo del Laboratorio de Gobierno, Ministerio de Hacienda.

⁸¹ La expresión máxima de este punto podrá verse en un horizonte temporal no muy lejano, de la mano de los "metaversos" o realidades inmersivas que sean diseñadas desde los soportes computacionales de las gigantes tecnológicas (Google, Amazon, Microsoft, Facebook, Apple).

Sumado a lo anterior, somos parte de una generación que demanda mayor igualdad de género desde una perspectiva multidimensional. Ello involucra a diversos componentes sociales, económicos, jurídicos y, sobre todo, culturales, junto con la inclusión de la diversidad en su amplio espectro. Existe una presión por relacionarse y construir sociedad de forma diferente, más horizontal, superando los roles de género históricamente arraigados, y que cada vez hacen menos sentido al pensar en el futuro.

Finalmente, la crisis climática es una de las mayores amenazas a nivel global. Si bien esta está propiciando transiciones dentro de nuestros hábitos de consumo y de la matriz productiva, aún es insuficiente y lenta si queremos hacernos cargo de las graves consecuencias que se anuncian desde la evidencia científica⁸². Cabe recalcar que sus efectos no sólo serán observables como desastres “puramente” naturales, sino también desde una perspectiva social, puesto que su impacto está estrechamente relacionado con variables sociodemográficas.

Esta conjunción de factores nos convoca como humanidad, demandando acciones concretas. Pero sobre todo nos convoca como Estados desde la necesidad de generar valor público para desarrollar soluciones que permitan afrontar estas transformaciones. Probablemente, muchas personas consideren que en el Sector Público cuesta anteponernos - o incluso reaccionar

tempranamente - a estas señales. Algo de cierto hay en esa apreciación porque, a pesar de que las alertas se vienen haciendo patentes a lo largo de todo el siglo XX, tuvimos que necesitar de una pandemia para abrir los ojos de manera definitiva y comprender la necesidad urgente de acelerar los cambios. La discusión en torno a la pertinencia y legitimidad del proceso de diseño, implementación y operación de las políticas públicas se torna cada vez más necesaria: esta es la manera a través de la cual los poderes públicos pueden incidir positivamente, ante los múltiples desafíos que enfrentamos.

Vivimos en un contexto en el cual los problemas públicos son complejos por su alcance y multidimensionalidad. Frente a ello, el diagnóstico internacional es compartido: la forma del Estado moderno resulta insuficiente para abordarlos debido a su rigidez y hermetismo estructural. Más aun considerando que las tensiones se presentan con mayor agudeza en lugares como Chile, inmerso de cierta manera en la *trampa de los países con ingresos medios*. Además, el proceso constitucional que estamos experimentando puede potencialmente redibujar las relaciones de poder y redefinir las estructuras políticas sobre las que se ha planteado la provisión de bienes y servicios públicos de las últimas décadas.

En este sentido, un enfoque de innovación para el Sector Público implica un cambio radical y profundamente urgente. Contra

⁸² IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [MassonDelmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M.

Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press.

Disponible en: <https://www.ipcc.ch/report/ar6/wg1/>

toda resistencia a las transformaciones globales y nacionales -observables desde los aspectos más macro hasta los más micro de nuestra existencia - desde la innovación pública, recibimos esta vorágine como una oportunidad inmensa para la generación de conocimiento nuevo; reforzando el vínculo histórico entre las situaciones complejas y la creatividad humana para abordar los desafíos contemporáneos.

Estos son temas que tienen que ver con la gobernabilidad y donde la política exterior puede colaborar para proyectar nuestras experiencias y buenas prácticas, promover esquemas de cooperación e incluir la innovación en el sector público en la agenda regional y global. Ello plantea desafíos técnicos, pero, fundamentalmente, una comprensión innovadora del servicio público y una voluntad política permanente.

El Laboratorio de Gobierno es la principal agencia mandatada para liderar la transformación del Estado chileno, desde el punto de vista de la innovación pública, de cara a este contexto profundamente desafiante, por medio de la instalación de la innovación como una nueva perspectiva para el diseño, desarrollo, implementación de políticas y gestión pública.

El Laboratorio nació en 2015, al alero de la Corporación de Fomento de la Producción (Corfo), lo que permitió dar el “puntapié inicial” para impulsar este paradigma. El atrevimiento propio de los nuevos enfoques fue percibido en un inicio como algo alejado de lo tradicional y de compleja aplicación a las más de 850 instituciones públicas del país. Sin embargo, se instauraba un hito relevante

sobre cómo concebir el Estado y su necesaria transformación.

En 2018 con el cambio de administración y producto de una profunda reflexión sobre el potencial alcance, incidencia y relevancia que debiera tener la innovación pública en el país, el Laboratorio fue reubicado institucionalmente en el Ministerio Secretaría General de Presidencia, conectando su quehacer con la agenda de gobierno y permitiendo el involucramiento en proyectos de alto impacto público. En coherencia con este nuevo rol, existió una reformulación estructural de su modelo de servicios, incorporando otro ángulo: esta mezcla la agilidad para implementar soluciones, una descentralización de prácticas de innovación y la medición e instalación de capacidades para innovar en los Servicios Públicos.

Los aprendizajes derivados de esta experiencia institucional son múltiples, principalmente relacionados con el hecho de estar situados en el centro del gobierno. A casi cuatro años de esa reformulación estructural de la gobernanza y modelo de servicios, para impulsar un Estado más innovador de cara a las personas, observamos múltiples lecciones desde los aciertos y fracasos. También podemos proyectar lo que viene respecto al rol de la innovación en la construcción de un nuevo sector público que aspire transversalmente a ser amigable, ágil, resolutivo y eficiente en cada una de sus interacciones cotidianas con las personas.

En primer lugar, es fundamental considerar permanentemente las necesidades de las personas, de los usuarios y usuarias finales, y

de las innovaciones por desarrollar. Imprimir esa visión usuario-céntrica en nuestro quehacer no sólo nos permite avanzar en la legitimidad y pertinencia de las soluciones, sino también en una estrategia válida para alinear distintos intereses que se cruzan en todo proyecto de carácter público y en toda arena política. Este es también un argumento que - abordado de manera astuta -resulta contundente y reúne a autoridades, tecnócratas, funcionarios y funcionarias públicas. Constituye una reconexión con el sentido más profundo de la vocación pública: si bien pone al usuario al centro de todas las decisiones, no es un componente exclusivo de la innovación, es claramente observable desde sus métodos, situándose como una práctica, no sólo como una teoría.

Poner a las personas en el centro de todo proceso de innovación implica involucrarlas en las diversas fases, desde el entendimiento de sus problemas y subjetividades, hasta asignarles espacio para co-crear soluciones. En muchos casos, sus puntos de vista ponen en tela de juicio las ideas preconcebidas, donde nuestra tarea es abordar todas las mutaciones y ajustes posibles, hasta lograr iniciativas que efectivamente logren un nivel de legitimidad y pertinencia que permita su adopción por parte de las personas.

De esta manera, avanzamos en la instalación de una nueva forma de hacer política pública, desde y para las personas. En ningún caso, este enfoque desconoce las metodologías tradicionales y sus diagnósticos o diseños. Más bien propone un ingrediente adicional que permite conectarnos con los desafíos públicos más complejos y cambiantes y, al mismo tiempo,

nos permite conectarnos con experiencias individuales y colectivas que nos guían con respecto a cómo transformar el Estado de manera adecuada y sostenible.

En segundo lugar, para que las iniciativas de innovación mantengan su vigencia y conexión con las necesidades de las personas es fundamental que su implementación sea ágil. Este es un desafío no menor considerando por una parte la velocidad tradicional del Sector Público y también la histórica dificultad de los Laboratorios de innovación de transformar ideas o prototipos en soluciones concretas e implementadas a gran escala.

Cuando se trata de resolver problemáticas públicas, se debe considerar la imposibilidad de predecir el resultado de los proyectos. Esto, debido a la potencial complejidad que pueden poseer en su gestión, no solo por un eventual nivel técnico alto, sino por la diversidad de las interacciones que se dan entre las personas que participan en él. Se debe destacar que en la esfera pública confluyen constantemente los intereses de distintos actores, que hacen que un proyecto que puede parecer sencillo en su ejecución, pueda verse complejizado por aquellas interacciones e intereses, implicando tal vez que soluciones que funcionan desde el diseño, no vean nunca la luz.

En este ámbito, un desafío permanente es hacernos cargo de la incertidumbre producida por dicha interacción, flexibilizando las fases de un proceso de innovación para intentar mitigar al mínimo esos riesgos. Tuvimos que revisar de manera acuciosa nuestras prácticas de innovación

para que nos permitieran responder, con sentido de urgencia, a las situaciones cambiantes y a las problemáticas más relevantes, sin perder la capacidad de mantener las necesidades de las personas en el centro y promover el desarrollo de capacidades de experimentación dentro de los servicios públicos.

Fue así como uno de los aprendizajes más relevantes es la incorporación de un enfoque de trabajo ágil, estructurado en ciclos de trabajo cortos y que aseguren avances progresivos en cada proyecto y entregables en corto tiempo para materializar la consideración de su valor, frente a los distintos actores involucrados. Luego de muchas iteraciones, sabemos que para alcanzar resultados concretos (denominado "*delivery*") debemos lograr gobernanzas y métodos de trabajos que combinen diversidad, flexibilidad, con agilidad y lógica experimental.

En definitiva, el Estado chileno, por medio del Laboratorio de Gobierno, ha aplicado un nuevo paradigma de transformación de las instituciones públicas. Este busca generar soluciones sostenibles y de alto impacto, mediante un proceso que combina el enfoque de innovación desde el diseño, poniendo como centro a las usuarias y usuarios, asegurando que las soluciones desarrolladas respondan efectivamente a sus necesidades, a fin de entregar soluciones de forma oportuna bajo un ritmo sostenido de trabajo colaborativo y que por sobre todo

comprometa a todos los actores relevantes con su implementación.

Lo interesante de estos aprendizajes es que provienen de la práctica misma y no de ejercicios teóricos. Así, las experiencias más relevantes y recientes de innovación pública del Estado chileno dan cuenta de la consideración de otro ángulo.

Por cierto, el desafío para el Estado chileno en cuanto a mejorar la calidad de sus servicios públicos, su gestión y conexión con las necesidades de las personas continúa. Tal como recomienda la OCDE⁸³ a nuestro país, es fundamental transversalizar un nuevo enfoque para la entrega de servicios públicos, desde una institucionalidad que asuma el rol mandado de *Head of Design*, en la generación de servicios compartidos, entre otros desafíos.

⁸³ OCDE (2020), Gobierno Digital en Chile – Mejorar el diseño y entrega de servicios públicos, OCDE Estudios de Gobierno Digital, Paris. <https://www.oecd.org/chile/digital-government-in->

[chileimproving-public-servicesdesign-and-delivery-b94582e8-en.html](https://www.oecd.org/chile/digital-government-in-chileimproving-public-servicesdesign-and-delivery-b94582e8-en.html)

Experiencias más relevantes de innovación pública

Nombre: Red de Protección Social: Un proceso inédito de co-creación en el Estado
Año de ejecución: 2018 - 2019
Actores participantes: Ministerio de Desarrollo Social y Familia, Fondo Nacional de Salud, Ministerio de Salud, Ministerio de Educación, Ministerio del Interior y Seguridad Pública, Servicio Nacional del Adulto Mayor, Servicio Nacional de Capacitación y Empleo, Superintendencia de Salud, Municipalidades de Independencia y Peñalolén, Junta Nacional de Auxilio Escolar y Becas y Ministerio Secretaría General de la Presidencia.
Descripción: El Ministerio de Desarrollo Social y Familia se planteó el desafío de idear un sistema que reuniera los beneficios sociales ya existentes y diera espacio a los que se crearían más adelante, con el objetivo de que la ciudadanía pudiera obtener información clara y precisa. Junto al Laboratorio de Gobierno, se conformó un equipo para coordinar un proceso de co-creación interministerial e institucional que tomó 9 meses. Más de 500 personas participaron en el proyecto que involucró a 55 instituciones públicas y civiles. En este proceso de co-creación se descubrieron los desafíos de una política pública basada en la interoperabilidad del Estado, para acceder a información personalizada sobre beneficios entregados, disponibles y aplicables según las características socioeconómicas de cada familia. El portal de Red de Protección Social ha recibido más de 11 millones de visitas, teniendo un acceso directo de ciudadanos que utilizan su clave única. Asimismo, el portal ha sido utilizado para brindar más de 18.000 asesorías por parte de los más de 1.000 asesores integrados de la red de Chileatiende, siendo clave también en su utilización de datos para soportar el corazón del servicio de la plataforma Mi Chileatiende. En la actualidad, la plataforma proteccionsocial.gob.cl cuenta con información, en lenguaje claro, sobre más de 160 beneficios y programas para enfrentar diferentes eventos adversos que podrían llevar a una familia a una situación de vulnerabilidad.

Nombre: Whatsapp Mujer: Canal silencioso para orientar a mujeres que sufren violencia doméstica en confinamiento
Año de ejecución: 2020
Actores participantes: Ministerio de la Mujer y Equidad de Género, Servicio Nacional de la Mujer y la Equidad de Género (SernamEG), Facebook Latinoamérica, Whatsapp Latinoamérica.
Descripción: Al comienzo de la pandemia (2020) se observó un aumento del 70% en la cantidad de llamadas recibidas al número 1455 del Servicio Nacional de la Mujer y la Equidad de Género (SernamEG) con respecto al mismo período del año 2019. Además, el confinamiento planteó la dificultad que las víctimas de violencia compartían la cuarentena con sus agresores, limitando la posibilidad de pedir ayuda. La urgencia dio paso a la oportunidad y el Ministerio de la Mujer y EG, SernamEG, en conjunto con el Laboratorio de Gobierno iniciaron un proyecto de innovación pública con la necesidad de crear un medio silencioso para ayudar a mujeres que sufren violencia. El 28 de abril de 2020 se lanzó el servicio de atención silencioso, seguro y confidencial el cual funciona de lunes a viernes las 24 horas del día. El canal de silencioso combina la atención por parte de ejecutivas (casos de orientación, contención emocional y situaciones de emergencia inminente); y un robot de atención (casos de búsqueda de información general). De este modo, es posible concentrar los esfuerzos de las ejecutivas en atender los casos más complejos, ya que el robot sólo se activa si la ejecutiva define que el caso es atendible de esa manera. Se destaca en este proyecto el apoyo de Facebook, empresa propietaria de WhatsApp, el cual se manifestó en la flexibilidad de sus políticas de funcionamiento y privacidad, y en financiar el desarrollo y los tres primeros meses de soporte del servicio. A partir de estos resultados, el canal de WhatsApp que se pensaba inicialmente como un dispositivo de emergencia, se transformó en un canal de atención permanente, vigente a la fecha. En su primer año de implementación se realizaron más de 32.000 atenciones por medio de este canal, correspondientes al 13% de las atenciones realizadas por SernamEG.

Nombre: Subsidio al empleo: Un apoyo en pandemia con foco en las y los usuarios
Año de ejecución: 2020
Actores participantes: Servicio Nacional de Capacitación y Empleo, Ministerio del Trabajo y Previsión Social, Ministerio de Hacienda
Descripción: La pandemia del COVID-19 generó importantes efectos económicos negativos en nuestro país, que derivaron en la mayor crisis de empleo en Chile de los últimos 30 años, implicando 1,5 millones menos de puestos de trabajo en comparación al año anterior. A raíz de esta compleja situación, el Gobierno anunció la generación de un masivo e histórico Subsidio al Empleo con un fondo de USD 2.000 millones en ayuda directa a la contratación de nuevas personas y al retorno de trabajadores que se encontrasen suspendidos bajo la Ley de Protección del Empleo. Habiendo investigado y teniendo claras las necesidades de las personas, se trabajó en el diseño e implementación del Subsidio desde cuatro ámbitos de intervención definidos por el Laboratorio de Gobierno: Estrategia, Servicio, Operación y Organización, desarrollando un sistema completo de información, postulación, seguimiento y pago del subsidio, 100% diseñado pensando en los usuarios directos, las MIPYMES y en los beneficiarios finales, que finalmente quedó disponible en www.subsidioalempleo.cl El Subsidio fue lanzado en septiembre de 2020, con los siguientes componentes: - Servicio digital elástico, preparado para una alta demanda con acceso y atención personalizada y con información precargada para una mejor orientación en la postulación. - Coordinación e interoperabilidad entre varias instituciones del Estado y privados para evitar trámites anexos adicionales a la postulación misma del subsidio. - Múltiples canales de atención digitales, presenciales y telefónicos con un foco en resolver problemáticas concretas, no solamente informar. - Pago rápido dentro del mismo mes de postulación pensando en la liquidez, lo cual se valida y genera de manera automática. - Comunicación clara dentro de todo el proceso para fácil entendimiento y acceso para cualquier persona. Durante el primer semestre, se han entregado más de 800.000 subsidios, de los cuales un 83% corresponde a PYMES.

Nombre: Fonasa: Un nuevo modelo de atención centrado en las personas
Año de ejecución: 2018-2019
Actores participantes: Fondo Nacional de Salud, Fonasa
Descripción: El proyecto nace de la oportunidad de transformar Fonasa en un servicio de atención integral y proactivo. Para ello, era necesario disminuir sustancialmente el tiempo de espera para la resolución de problemas, reducir la cantidad de reclamos e impactar positivamente en la satisfacción de las y los usuarios. Se identificó que los atributos más relevantes para una experiencia satisfactoria son: resolutivez, omnicanalidad y proactividad. Sobre la base de estos atributos se inició el proceso de co-creación y desarrollo de soluciones. La primera es Fonasa Resuelve, co-creada por más de 40 funcionarios de la institución. Su rol es intermediar entre los asesores de atención de los cuatro canales y las áreas de negocio, para asegurar una respuesta ágil, idealmente en el primer contacto. Esto, considerando que con la lógica anterior la resolución demoraba en promedio 30 días. Con este nuevo servicio, se proyecta abordar 1.758 consultas al año y de esta manera disminuir en 38% los reclamos y 28% las solicitudes escalables. La segunda es el Departamento de Experiencia de Usuarios, encargado de sostener el nuevo modelo de atención, dándole gobernanza a la innovación a través de un comité permanente. La tercera solución, Hoy nací y soy Fonasa, entrega cobertura de salud a los recién nacidos de forma automática e inmediata una vez que son inscritos en el Registro Civil, a partir de la interoperabilidad de ambas instituciones. Así, las familias ahorran tiempo y dinero ya que no requieren realizar el trámite presencialmente en Fonasa. Desde su implementación, más de 301 mil familias han sido beneficiadas. Por último, Elige Tú es un buscador web de prestadores de salud que permite a usuarios tomar decisiones informadas según los costos y la ubicación donde se entregan los servicios que requieren. Además, permite unificar la entrega de información de los canales de atención.

Hemos también asumido, desde el Laboratorio, el desafío de documentar las diversas metodologías de trabajo que constituyen nuestro acervo al colaborar en cualquier institución pública que tenga un requerimiento de implementación de una solución innovadora. Para ello, el Laboratorio también dispone de una completa biblioteca abierta, donde todas las usuarias y usuarios podrán consultar y conocer diversas publicaciones, con la esperanza de que constituyan una guía en cualquier proceso innovador. La generación de este conocimiento en innovación pública puede constituir un ejemplo de “exportación no tradicional” desarrollada desde el sector público y un activo que nuestra diplomacia debe promover

Estos referentes, con los cuales se apunta a alcanzar un alto impacto ciudadano, han consolidado la posición de liderazgo de Chile en materia de innovación pública a nivel regional y también global. Una Red de Innovadores Públicos de casi 20.000 miembros, encargada de descentralizar las capacidades para innovar entre funcionarios y funcionarias públicas, disponer de un ecosistema de aprendizaje y colaboración, además de la implementación del Índice de Innovación Pública que establece un estándar de medición para los servicios públicos y guías de acción concretas, son todos elementos que constituyen el modelo de innovación pública chileno.

Con estos antecedentes, podemos permitirnos soñar en grande y pensar a nuestro país como un gran exportador de servicios no tradicionales para otros

gobiernos. En este sentido, la diplomacia juega un rol fundamental.

De partida, como bien fue anunciado en esta contribución, la humanidad se enfrenta a un número importante de megatendencias, las cuales están siendo abordadas por diversas instancias de gobernabilidad a nivel regional y global. Allí, la diplomacia chilena ha asumido posturas propositivas y activas para alcanzar nuevos acuerdos, reglamentaciones e instrumentos. Tan solo pensemos en la presidencia de nuestro país durante la COP 25 o en las diversas iniciativas de Chile para el establecimiento de un Tratado internacional que permita enfrentar, coordinadamente, futuras pandemias.

En este sentido, como Director Ejecutivo del Laboratorio de Gobierno, quisiera agradecer las diversas invitaciones que la Academia Diplomática ha extendido al Laboratorio y que nos han permitido dialogar cada año entorno a experiencias y aprendizajes de innovación pública, con sus estudiantes.

Esta es una manera muy relevante de que nuestras futuras funcionarias y funcionarios del Servicio Exterior tengan presente, desde temprano en sus carreras, el valor de las diversas metodologías del Laboratorio y que, además, puedan constituirse en voceros internacionales de nuestro trabajo. Estamos ciertos que “nuevos problemas demandan nuevas soluciones” por lo cual, esperamos, sinceramente, que de estas conversaciones los estudiantes se apropien de todos los productos, acciones y contenidos del Laboratorio.

La formación de alianzas con instituciones e instancias homologas⁸⁴ constituye también un ámbito donde nuestra diplomacia puede apoyar significativamente el trabajo del Laboratorio. Al igual como la Política Exterior ha asumido la tarea de insertar a Chile en las redes internacionales del conocimiento, estas son instituciones con las cuales el Laboratorio puede compartir buenas prácticas y crear sinergias para una acción colaborativa.

Mensajes conclusivos

Antes de la pandemia ya veíamos cómo la colaboración entre países se hacía más necesaria para temas tan diversos como la generación de energía, el cambio climático, la migración, entre otros. Hoy se hace absolutamente necesario y evidente que las fronteras que nos hemos dibujado como seres humanos son cada vez más permeables a la globalización, sobre todo desde el punto de vista del conocimiento. Es en ese contexto que Chile puede ser un gran referente en materia de replicar instituciones, modelos, procesos y metodologías de innovación pública que sirvan a otros países de la región y del mundo en vías de desarrollo.

Uno de los principales desafíos del actual cambio de ciclo en Chile es entender y materializar que la única manera para implementar transformaciones estructurales, de alto impacto ciudadanos, sean estas en salud, educación, seguridad, transporte, pensiones o vivienda, es habilitando al sector

público de nuevas capacidades para poder implementar esas innovaciones y gestionarlas en el día a día.

En definitiva, la innovación está en el corazón del presente y futuro del Sector Público. Para su fortalecimiento, necesitamos generar un movimiento global de colaboración pública-pública entre países, gobiernos e instituciones afines, que nos permita aprender rápido, replicar bien y, sobre todas las cosas, darle prioridad política a estos temas que son de la mayor relevancia para que las demandas ciudadanas sean abordadas por políticas públicas efectivas, eficientes y centradas en la persona humana.

Sin un Estado amigable, ágil, eficiente, digital, más profesionalizado y usuario-céntrico, muy probablemente, las legítimas aspiraciones de la ciudadanía queden frustradas.

⁸⁴ Algunas de las instituciones más relevantes a nivel global en innovación pública, con las cuales el Laboratorio ha abierto espacios de colaboración: "27é Région" en Francia; NESTA en Reino Unido; el "Observatory for the Public Sector Innovation",

perteneciente a la OCDE; GovLab de New York University; Subsecretaría de Innovación Pública de Argentina; Equipo de Innovación Pública del Depto. Nacional de Planeación de Colombia; "011lab" del Gobierno de la ciudad de Sao Paulo, Brasil, entre otros.

Ética e Inteligencia Artificial

Maria Florencia Attademo-Hirt⁸⁵

La inteligencia artificial (IA) y los sistemas automatizados de decisión han experimentado un acelerado desarrollo en los últimos años. Hoy podemos ver acciones sencillas, desde que el celular nos recuerde la hora en la que solemos poner la alarma cada día para despertar, a otras más complejas donde se manejan grandes cantidades de datos y los algoritmos recomiendan decisiones tales como el reclutamiento de personal, el desplazamiento de vehículos autónomos o la asignación de beneficios de carácter público o privado.

El potencial es enorme y apunta a hacer más eficiente en el uso de recursos y la gestión de las organizaciones, mejorando la vida de las personas, simplificando procesos para los usuarios o adelantándose a las necesidades en distintas esferas.

Pero ¿qué es la IA? No existe una definición única, ya que se trata de un área en constante evolución que incluye diversas tecnologías y diversas complejidades. En el grupo BID hemos adoptado los lineamientos de la OECD⁸⁶, que describe la IA como “un sistema computacional que puede, para un determinado conjunto de objetivos definidos

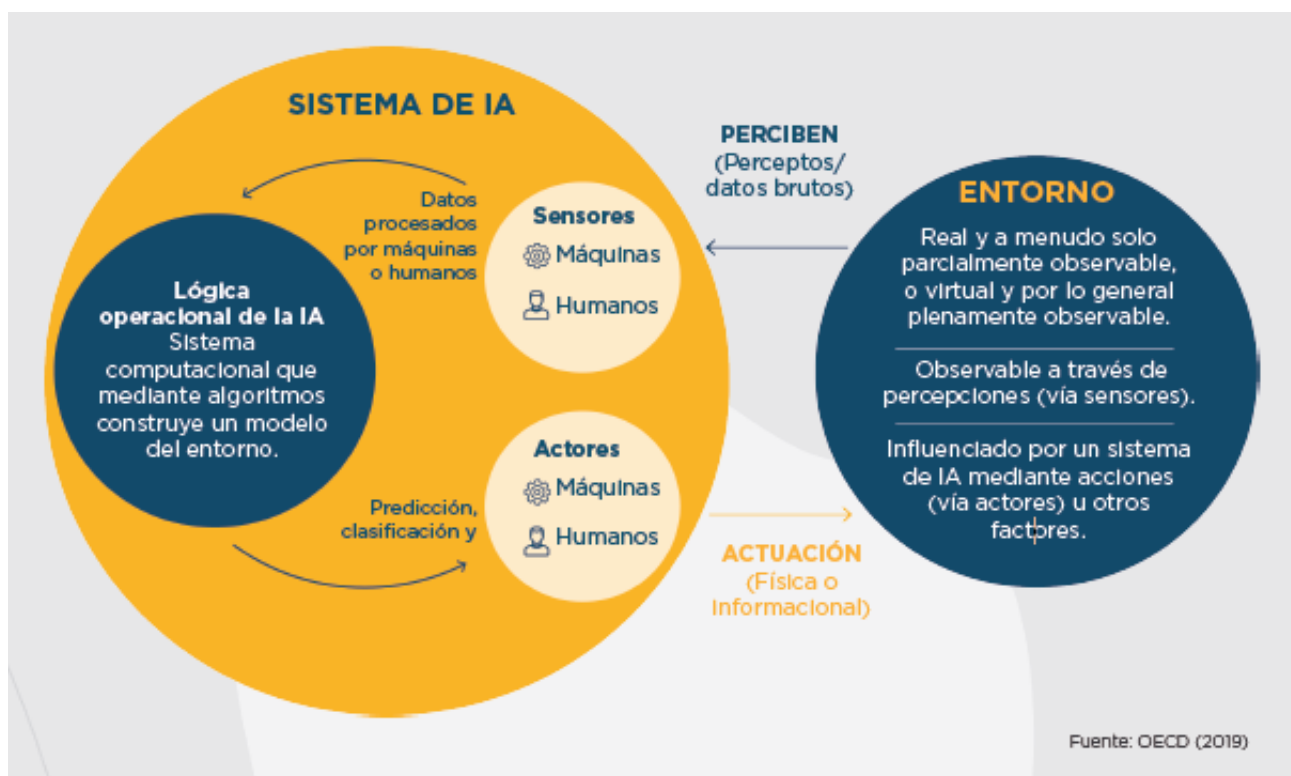
por humanos, hacer predicciones y recomendaciones o tomar decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con distintos niveles de autonomía”.

De esta manera aparece otro concepto clave cuando de IA se trata: los algoritmos, es decir, secuencias de comandos que permiten obtener un resultado a partir de la introducción de unos datos determinados (OECD, 2018) reglas o instrucciones y que, aplicados sistemáticamente, facilitan la toma de decisiones (sea esta o no automatizada). Los algoritmos son la estructura de los sistemas de IA, pues son los encargados de procesar los datos para realizar diversas funciones, entre las que se incluyen predicciones, clusterizaciones y categorizaciones, así como toma de decisiones automatizadas (MGI 2017; Google 2018; IMDA & IDPC 2020). Entrenar a una máquina requiere de responsabilidad tanto en el uso de datos personales como en cautelar que las decisiones resultantes sean beneficiosas y no dañinas para la sociedad.

⁸⁵ Gerente General del Departamento de Países del Cono Sur del Banco Interamericano de Desarrollo y Representante en Chile del Grupo BID.

⁸⁶ Organización Económica de Cooperación para el Desarrollo.

La Inteligencia Artificial: Visión de la OECD.



A la hora de implementar soluciones a través de algoritmos, es necesario que éstos sean entrenados sin sesgos que repliquen o profundicen desigualdades presentes en nuestra sociedad, que velen por la privacidad y protección de los datos, sean incluyentes, se centren en las personas, respetando los derechos humanos, la libertad, la autonomía, la igualdad, la diversidad y la justicia social. Además, deben ser transparentes en su actuar, para que quienes estimen que se han vulnerado sus derechos, puedan demandar al sistema.

Las decisiones relativas al desarrollo y uso de las tecnologías de la IA deben adoptarse dentro de un marco de derechos humanos y con una visión holística e inclusiva.

Para evaluar si un proyecto cumple con las normas éticas en la aplicación de la IA, hay

distintas preguntas que se deben resolver. Primero, definir con precisión cuál es el rol y el aporte de la IA, de manera que esta sea aplicada cuando es pertinente hacerlo y cómo se compara respecto de otras metodologías o tecnologías posibles en cuanto a costos, beneficios, huella de carbono, impacto social. El segundo paso es valorar la robustez que tendrá el sistema, tanto en infraestructura tecnológica, como de pertinencia legal y riesgos de operación. (Galdon 2020).

La transparencia es otro de los factores que ayudan a minimizar riesgos en el desarrollo de sistemas de IA. Para velar por una adecuada transparencia se pueden incorporar marcos de gobernanza, mecanismos de control interno, reportes periódicos y mecanismos para la resolución

de controversias. La selección del modelo a utilizar dependerá de la estructura de la organización, tamaño y recursos disponibles, así como del sector particular en el que opera (Omidyar 2019; WEF & IMDA 2020). Cualquiera sea el modelo de gestión de transparencia que se adopten, es importante mantener actualizado un modelo de gestión de riesgos, identificando de manera temprana posibles factores que afecten el desempeño de los sistemas, así como un plan de mitigación de riesgos y de acción si éstos llegan a concretarse, generando algún potencial daño.

Otra consideración tiene relación con el uso y privacidad de datos. Los datos son el insumo base de todo sistema automatizado, por lo que sistemas más robustos serán aquellos que cuenten con grandes cantidades de datos que les permita entrenarse y realizar predicciones más precisas, minimizando el riesgo de error. Un desarrollo responsable de IA debe contener una data de calidad comprobable y medidas adecuadas que mitiguen riesgos de sesgos y discriminación, por el hecho de contar con datos no representativos o mal gestionados.

Tan importante como lo anterior es el adecuado respeto a la privacidad de la información, contar con procedimientos claros y rigurosos respecto al tratamiento de datos personales es también un principio clave. No podemos esperar a que las leyes sean las encargadas de regular estos ámbitos, la tecnología se desarrolla más rápido que los marcos legales, por lo que

consensuar normas y estándares éticos autogestionados es fundamental.

Considerar los riesgos que conlleva la tecnología es importante para no dañar ni afectar el logro de los objetivos para lo cual fue creada. Sin embargo, aunque parezca una obviedad, vale la pena mencionar que el uso y el desarrollo de tecnología no es algo que podamos limitar. La innovación tecnológica genera enormes beneficios, mejora el acceso a productos y servicios tanto en tiempo, forma y costo, ayuda a tomar mejores decisiones sobre asignación de recursos, mejora la productividad y la competitividad de los países y las instituciones. En definitiva, contribuye a catalizar las metas de desarrollo sostenible. En definitiva, diariamente la tecnología está cambiando la forma como nos relacionamos, trabajamos, consumimos y está en poder de cada uno de nosotros el cautelar que sus impactos sean positivos para las personas y el medioambiente, generando bienestar social.

América Latina y el Caribe necesitan de soluciones innovadoras, éticas y responsables para contribuir a superar las problemáticas sociales y ambientales y mejorar la provisión de servicios a la ciudadanía. Si bien el grueso del desarrollo tecnológico es liderado por otras regiones del mundo, en nuestros países también se está generando innovación tecnológica de primer nivel.

Conviene mencionar que resulta oportuno propiciar nuevos ángulos en la creación de normativas regulatorias en torno a la IA y a

los sistemas autónomos de decisión, por ejemplo, la experimentación regulatoria a través de sandboxes, que permiten fomentar la innovación, atraer la inversión y potenciar el co-aprendizaje público-privado.

Cautelar que este desarrollo, desde el diseño, contemple principios éticos, será una forma de mejorar oportunidades de acceso

a mercados y financiamiento que permitan el crecimiento más allá de sus países de origen.

Otro ámbito fundamental para que la región pueda robustecer sus desarrollos tecnológicos, es el de capital humano en manejo de datos, ciencia de datos e IA. Orientar políticas públicas en este ámbito es parte central del quehacer del Grupo BID.

Principios éticos de la OCDE para el desarrollo de IA ⁸⁷	
	Crecimiento inclusivo, desarrollo sostenible y bienestar
	Valores centrados en el ser humano y la equidad
	Transparencia y explicabilidad
	Robustez, seguridad y protección
	Rendición de cuentas

Como forma de cautelar el diseño, desarrollo e implementación ética de los algoritmos de IA, desde 2019 el Grupo BID viene trabajando en la iniciativa fAIr LAC, cuyo objetivo es promover un uso responsable de la IA, basados en los principios elaborados por la OECD, buscando facilitar su operacionalización, considerando tres dimensiones estratégicas: (i) desarrollo de

una red diversa de actores públicos y privados, (ii) formación de capacidades para una adopción responsable de la tecnología y (iii) promoción de calidad y mitigación de riesgos.

Para la operacionalización de estos principios en la región de América Latina y el Caribe, fAIr LAC ha identificado los desafíos que enfrenta el desarrollo de la IA en nuestra

⁸⁷ <https://oecd.ai/en/ai-principles>

región y que pueden resumirse de la siguiente forma⁸⁸:



El objetivo en la identificación de estos retos es el de orientar el desarrollo de

herramientas de mitigación de fácil implementación para todos los actores

⁸⁸ fAIr LAC, Adopción ética y responsable de la IA en América Latina y el Caribe, BID, 2020.

involucrados. Para ello se ha avanzado en: (i) la creación de espacios de dialogo y de implementación aplicada de la IA que permitan relevar buenas prácticas, aprendizajes y, sobre todo, sensibilización en cuanto a los beneficios y riesgos de su adopción; (ii) generación de capacidades para el diseño, desarrollo, implementación y adopción de mecanismos de IA al tiempo de que se sistematizan productos de conocimiento (guías, marcos, herramientas) que faciliten la adopción responsable de la IA; y (iii) apoyo en la formulación de políticas públicas y mecanismos de autorregulación que velen por la calidad y la mitigación de riesgos en la IA.

En sus tres años de implementación, fAIr LAC ha avanzado en todos los frentes.

En el ámbito de **desarrollo del ecosistema** y diálogo fAIr LAC cuenta con importantes aliados, expertos y Hubs en diferentes países de la región, que guían tanto el desarrollo de la iniciativa como el despliegue de soluciones.

Para la generación de capacidades, se han desplegado diferentes programas de formación en temas como ciencia de datos, auditoria algorítmica, autoevaluación y evaluación ética de soluciones de IA, formulación de proyectos basados en IA, protección de datos, ciberseguridad, interoperabilidad y otros recursos sistematizados en productos de conocimiento, que están siendo aprovechados por el sector público, el sector

privado, la sociedad civil y el ecosistema de emprendimiento tecnológico. El desarrollo de las herramientas orientadas al sector público ha sido liderado por el Sector Social del Banco. Las herramientas orientadas al sector privado han sido lideradas por BID Lab, el laboratorio de innovación del Grupo BID. La idea es fortalecer el ecosistema emprendedor, una de las principales fuentes de soluciones algorítmicas, trabajando no solo desde la oferta (las propias startups tecnológicas), sino también la demanda y entidades de apoyo (inversionistas, aceleradoras de negocios, expertos en datos, academia, sociedad civil, gobiernos, grandes empresas), de manera de alinear la visión respecto a la ética de datos e IA.

Finalmente, el piloteo de la implementación de todas estas prácticas, a través de proyectos específicos, Hubs de países, y pilotaje de las herramientas.

Un ejemplo concreto de este despliegue es el proyecto algoritmos éticos⁸⁹ en Chile, que es cofinanciado por BID Lab. El proyecto, ejecutado por el laboratorio de gobierno de la Universidad Adolfo Ibañez, Goblab, forma parte de la iniciativa fAIr LAC y trabaja, en alianza con el sector público y privado en dos frentes:

- I. Desarrollo de capacidades para incorporar estándares éticos entre las empresas tecnológicas que proveen de algoritmos de decisión automatizada e IA al sector público, con un interés

⁸⁹ <https://goblab.uai.cl/algoritmos-eticos/>

particular en las startups tecnológicas y el ecosistema de apoyo a su crecimiento (incubadoras, aceleradoras, fondos públicos, venture capital).

- II. Fomento a que el sistema de compras públicas considere la gestión ética de datos y de la IA dentro de sus estándares y requisitos de compra.

A la fecha se han desarrollado dos ciclos de programas de formación a emprendedores tecnológicos y funcionarios públicos, se está realizando acompañamiento técnico a 4 pilotos de desarrollo o compra pública de algoritmos de decisión automatizada, incorporando principios éticos y se está apoyando el desarrollo de guías y estándares de compras públicas de este tipo de tecnología.

Actualmente los aliados del proyecto, que reciben acompañamiento técnico, aportan recursos y participan de su gobernanza son ChileCompra, la División de Gobierno Digital

del Ministerio Secretaría General de la Presidencia, el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, la aceleradora de negocios Magical, el Consejo para la Transparencia, Fonasa, la Defensoría Penal Pública, el Instituto de Previsión Social. Las últimas cuatro entidades aliadas, se encuentran implementando pilotos de desarrollo o compra pública de algoritmos de decisión automatizada, incorporando principios éticos, con los cuales se busca validar el potencial de impacto social del uso de estas tecnologías.

De manera adicional, el BID ha apoyado la creación de la División de Estudios e Inteligencia de Negocio en la Dirección de Compras y Contratación Pública, y prestó apoyo al alineamiento de la estrategia de inteligencia artificial de ChileCompra con la estrategia nacional.

Otras iniciativas que están siendo impulsadas por el Grupo BID en el ámbito del desarrollo ético de la IA son las siguientes:

Prevención de Retinopatía diabética, Jalisco, México. Liderado por el Gobierno Jalisco y el Tecnológico de Monterrey.

- La diabetes es una de las enfermedades crónicas más comunes en el mundo. Una de sus consecuencias es la retinopatía diabética, un padecimiento crónico y progresivo tiene una prevalencia de 31.5 por ciento en México, y en los países industrializados se ha convertido en la causa principal de pérdida visual parcial y de ceguera en los adultos.
- Esta una enfermedad no presenta síntomas en sus inicios. La detección oportuna facilita el tratamiento para reducir la pérdida visual y por ende la ceguera, por lo tanto, se recomiendan exámenes cada seis meses para su detección precoz.
- La solución propuesta basada en IA sugiere un programa de tamizaje en la atención primaria donde los pacientes podrán llegar a una revisión visual y un sistema informático transferirá sus fotografías de fondo de ojo al Departamento de Graduación de Retinopatía Diabética. Este departamento es asistido por graduadores certificados y un sistema de IA, que ayudan a determinar el nivel de RD del paciente con la finalidad de transferir a un hospital de segundo nivel aquellos casos que tengan un nivel avanzado para un diagnóstico temprano y tratamiento oportuno, considerando la gobernanza de datos, sensibilización del personal y de los pacientes.

Distancia2, Quito-Ecuador y Buenos Aires, Argentina.

- Distancia2 es una plataforma que el Grupo BID pone a disposición de los países y que aprovecha la infraestructura de vídeo disponible en las ciudades. Gracias al uso de IA se busca mitigar el riesgo de contagio de COVID-19 en áreas con grandes aglomeraciones. En Ecuador opera a nivel nacional con 5.500 cámaras monitoreadas y en Argentina con 11.
- La plataforma puede generar alarmas, según los parámetros que la ciudad determine (por ejemplo % de personas a menos de 2 metros de distancia en cada unidad de tiempo), o entregar conjuntos de datos que registran la respuesta cotidiana de los ciudadanos ante las medidas dispuestas.
- Este desarrollo es escalable y se enfoca en adicionar inteligencia a la infraestructura disponible. Funciona en infraestructura informática que debe ser proveída por parte de las ciudades o entidades locales, ya sea en sitio o en nube; el BID licencia su uso sin ningún costo. La plataforma se puede usar para realizar análisis en tiempo real o para

realizar análisis históricos de comportamiento previamente registrado por las cámaras y tomar medidas concretas como rediseño de los espacios o señalética.

- No se utiliza información privada de los individuos ya que no se analizan rasgos particulares ni características únicas, solamente se detectan personas de manera anónima y se estima la distancia entre ellas.

ParaEmpleo, Asunción, Paraguay. Liderado por el Ministerio del Trabajo, Empleo y Seguridad Social.

- Para buscar empleo, los usuarios de la Dirección General de Empleo de Paraguay debían entregar sus antecedentes en papel en la única oficina de la entidad ubicada en la capital del país. El proceso era lento y costoso. Por ejemplo, en 2010 sólo se brindó orientación laboral individual a 562 personas.
- El piloto busca apoyarse en nueva tecnología con el fin de facilitar el emparejamiento entre oferta y demanda laboral en forma eficiente, precisa, sin papeles, generando valor tanto a los solicitantes como a los potenciales empleadores.
- Gracias al uso de IA se está desarrollando un sistema de intermediación laboral que incluye un motor de emparejamiento que aprende y tiene una base de relaciones utilizando ontología y semántica para proponer emparejamientos basados en múltiples dimensiones que no sean solo la coincidencia de palabras. El sistema realiza un emparejamiento en varias dimensiones: oferta y demanda laboral basada en habilidades duras, blandas, geografía y cercanía, horarios, experiencia, certificaciones, idiomas y otras dimensiones generando emparejamientos múltiples que permiten a las partes gestionar la oferta y demanda de empleo. En base a esta información los consejeros laborales realizan preguntas a los candidatos para ajustar la información si no se ha introducido experiencia relevante y proponen alternativas a cada buscador de empleo.
- El sistema de IA este certificado con la regulación europea GDPR (General Data Protection Regulation por sus siglas en inglés) y separa en el portal los datos sensibles del perfil de los usuarios, de aquellos no sensibles. Los usuarios, buscadores de empleo o empleadores cargan datos personales tales como tales como género, nombre, edad, nacionalidad, etc., los cuales no son incluidos en ningún emparejamiento para evitar sesgos. Por otro lado, los datos de empleabilidad tales como sus habilidades, capacidades, experiencia, localidad, horario y otros, sí forman parte de los emparejamientos en forma innominada. El sistema cumple los requisitos establecidos

las regulaciones locales para el desarrollo y la puesta en marcha de sistemas de información en temas de seguridad de la información, así como de infraestructura.

- La IA es una herramienta útil en este caso, pero requiere de la participación de humanos en el proceso que ayuden a los buscadores de empleo a entender sus resultados y además que ayuden a las empresas a registrar vacantes y cargar información. Adicionalmente, es necesario tener opciones de apoyo de carga por personal especializado (facilitadores) que guíe a los buscadores de empleo con menos habilidades digitales a fin de garantizar el acceso a servicios de toda la población.

Como se puede observar, la IA puede ayudar en diferentes ámbitos a mejorar la vida de las personas y hacer más eficientes la gestión de las organizaciones. Está en nuestras manos hacer posible que ello ocurra, si velamos por un desarrollo ético y responsable de la tecnología.

¡En el Grupo BID estamos comprometidos con esta gran tarea!