

Uso responsable de tecnología en Latinoamérica y el Caribe

Ejemplos de modelos y herramientas del Norte y Sur global. Oportunidades y desafíos desde una perspectiva ética y de derechos humanos

**Elina Castillo Jiménez y
Cristina Martínez Pinto**
Noviembre 2021

Uso responsable de tecnología en Latinoamérica y el Caribe. Ejemplos de modelos y herramientas del Norte y Sur global. Oportunidades y desafíos desde una perspectiva ética y de derechos humanos

Autoras: Elina Castillo Jiménez y Cristina Martínez Pinto

Contribuidora: Luz Elena González

Coordinadoras: Carolina Aguerre y Maia Levy Daniel

Revisión: Gonzalo Bustos Frati y Matías Jackson

Diseño: Mónica Castellanos

Edición: Paula Álvarez Arbelais

Licencia Internacional Pública de Atribución/ReconocimientoNoComercial-SinDerivados 4.0 de Creative Commons.



Las opiniones expresadas en las publicaciones incumben únicamente a los/as autores/as. No tienen intención de reflejar las opiniones o perspectivas del CETyS, CLD ni de ninguna otra organización involucrada en el proyecto.

Uso responsable de tecnología en Latinoamérica y el Caribe

Ejemplos de modelos y herramientas del Norte y Sur global. Oportunidades y desafíos desde una perspectiva ética y de derechos humanos

Noviembre 2021

Cristina Martínez Pinto

Especialista en tecnología cívica, transformación digital y en nuevas tecnologías con enfoque en innovación e impacto social. Es fundadora y directora general de PIT Policy Lab, el primer *spinout* de C Minds, desde donde explora la intersección entre tecnología de interés público y políticas públicas de tecnología. Cuenta con experiencia como consultora de organismos internacionales y en el servicio público mexicano. Es licenciada en Relaciones Internacionales por el Tec de Monterrey y maestra en Políticas Públicas de la Universidad de Georgetown. Forma parte del Board of Advisors del Beeck Center for Social Impact and Innovation, es Fellow del One Day Technology Policy Accelerator de la Federación de Científicos Americanos y recientemente fue seleccionada como experta en el eje de Gobierno Digital y Acceso a Servicios Públicos en el Grupo Asesor para la conformación del Primer Plan de Gobierno Abierto de la CDMX.

Elina Castillo Jiménez

Abogada feminista experta en derechos humanos y derecho internacional público, comprometida a utilizar su voz y sus habilidades para ayudar a construir un mundo más justo e inclusivo. Combina su expertise temático en derechos humanos y activismo digital para desarrollar políticas públicas, gestionar proyectos, diseñar campañas de incidencia y comunicación y facilitar talleres y otros espacios de diálogo, gracias a sus 7 años de experiencia asesorando a gobierno, sociedad civil y organizaciones internacionales. Forma parte de la Red de Consultorxs de PIT Policy Lab, como especialista en derechos humanos y activismo digital.

Luz Elena González

Especialista en cooperación internacional para el desarrollo, con experiencia en internacionalización de ciudades metropolitanas e incidencia legislativa. Está comprometida con promover el diseño ético de la tecnología en su intersección con la política pública, para crear ciudades más inclusivas, sostenibles y resilientes en América Latina. Ha trabajado como consultora y diseñadora de proyectos en el sector público. Es licenciada en Relaciones Internacionales por el Tec de Monterrey. Actualmente es asistente de proyectos en PIT Policy Lab, integrante de la comunidad MCODER.ai y curadora de contenido en Women in AI México.

Resumen ejecutivo

A nivel global, ha proliferado la publicación de documentos que abordan la importancia y necesidad de anclar principios éticos al desarrollo e implementación de sistemas de inteligencia artificial. Sin embargo, pocos abordan cómo llevar dichos principios a la práctica. El presente trabajo de investigación hace una revisión de literatura sobre el concepto de uso responsable de tecnología, presenta un breve balance de los modelos basados en la ética y en los derechos humanos para su gobernanza y delinea algunas ventajas de su incorporación para organizaciones. En una segunda sección, se realiza una revisión documental de insumos publicados por el Foro Económico Mundial (WEF por sus siglas en inglés) y se presentan ejemplos de herramientas que han sido implementadas por grandes empresas de tecnología en el Norte global. En el plano latinoamericano, se hace una revisión de algunas de las herramientas desarrolladas por el Banco Interamericano de Desarrollo (BID) en el marco de la iniciativa fAIR LAC y se propone una combinación de los modelos basados en la ética y en los DDHH en el contexto de la región. Por último, se exploran buenas prácticas regionales y ejemplos que se han implementado desde el sector privado en México, para concluir con reflexiones acerca de los desafíos que enfrenta la región hacia el futuro.

Contenido

1. Introducción	6
2. Uso responsable de tecnología	9
2.1. Ventajas del URT en organizaciones y empresas	9
2.2. Breve balance de los modelos basados en la ética y en los DDHH para la gobernanza del URT	10
2.2.1. Modelos de gobernanza basados en la ética: beneficios y críticas	11
3. Propuestas para operativizar el uso responsable de tecnología.....	16
3.1. Modelo planteado en el reporte Uso responsable de tecnología	16
3.2. Modelo complementario: acción-habilitación-influencia	18
3.3. Modelo planteado en el reporte Ética por diseño	18
3.4. Ejemplos de herramientas para el URT	19
4. Oportunidades y desafíos para la política pública en América Latina.....	24
4.1. Modelos de gobernanza para América Latina y el Caribe	24
4.2. Buenas prácticas regionales en el URT.....	26
5. Conclusiones / desafíos.....	30
6. Agradecimientos.....	32
7. Referencias	33

1.

Introducción

En los últimos años, en línea con el auge global de la inteligencia artificial (IA) como herramienta transformadora y tecnología de propósito general, proliferó la creación de marcos conceptuales basados en una serie de principios éticos para guiar su desarrollo y adopción. El Centro Berkman Klein de Harvard realizó un análisis comparativo de más de 30 documentos que exponen numerosos principios e interpretaciones de estos, identificando 8 tendencias o ejes temáticos compartidos: privacidad, responsabilidad, seguridad y protección, transparencia y explicabilidad, equidad y no discriminación, control humano de la tecnología, responsabilidad profesional y promoción de los valores humanos (Fjeld & Nagy, 2020).

Si bien hay un consenso en la necesidad de abordar estas temáticas, también se ha cuestionado, principalmente en el sector privado, que la publicación de dichos principios no cuenta con mecanismos de acompañamiento para su operativización, es decir, que son solo buenas intenciones a las cuales aspirar, pero que no se traducen en acciones concretas en el ciclo de vida de productos o servicios de tecnología y, por lo tanto, que no existe una rendición de cuentas sobre potenciales impactos no deseados.

Para hacer frente a este reto, una de las iniciativas más relevantes desde el Norte global se ha dado dentro del el Foro Económico Mundial (WEF, por sus siglas en inglés), donde se creó un grupo de trabajo pionero en 2018 sobre el uso responsable de la tecnología (URT), a través del cual realiza investigación temática respecto a herramientas y metodologías innovadoras a disposición de las empresas para transformar culturas organizacionales e implementar estrategias basadas en aproximaciones éticas y de derechos humanos (DDHH); así como en aprendizajes abstraídos de las ciencias del comportamiento que buscan incidir en la conducta de las personas y de equipos de trabajo. Asimismo, organiza talleres y conversaciones con empresas, eticistas, sociedad civil y la academia para intercambiar conocimiento y ejemplos desde lo privado sobre cómo llevar los principios éticos de la IA a la práctica. Con la participación de empresas globales como Deloitte e IBM, e instituciones como el Centro Markkula para Ética Aplicada de la Universidad de Santa Clara en Estados Unidos, este grupo de trabajo es coliderado por Microsoft y Salesforce, quienes cuentan con equipos dedicados a transversalizar sus respectivos principios éticos de la IA en coordinación con las distintas áreas encargadas del diseño e implementación de dicha tecnología, a través del catálogo de productos y servicios que ofrecen.

En América Latina y el Caribe (LAC), el Banco Interamericano de Desarrollo (BID) lanzó en 2019 la iniciativa fAIR LAC, a través de la cual se plantea la necesidad de abordar nuevas conversaciones y acciones para el aprovechamiento de la IA desde un URT y desde un contexto anclado en la realidad regional. Desde esta plataforma se ha trabajado en el desarrollo de herramientas prácticas y la identificación de casos de uso en LAC donde la IA está al servicio del interés público y en los cuales están involucrados gobiernos, academia, sociedad civil, *start-ups* y empresas (Gómez Mont *et al.*, 2020). Dichos casos demuestran un panorama regional alentador; sin em-

bargo, son escasos los ejemplos en donde se aborda el desarrollo e implementación de IA y su intersección con marcos conceptuales basados en ética y DDHH.

Cabe destacar que en LAC, como en el resto del mundo, el contexto de pandemia por COVID-19 aceleró la transformación digital de empresas hacia la economía digital y de gobiernos hacia la provisión de servicios públicos basados en herramientas digitales y trabajo remoto. En este proceso, se han exacerbado brechas de desigualdad, no solo en acceso a tecnología, sino también en educación en competencias digitales, evidenciando la necesidad de incrementar la inversión en la materia y avanzar hacia una regulación flexible y ágil en torno a nuevas tecnologías, incluida la IA (López Calva & Melguizo, 2021).

Asimismo, desde la academia se ha explorado cómo, a pesar de tener un sistema regional de derechos humanos robusto, todavía hay espacio para fortalecer el vínculo entre el desarrollo de IA para el interés público y la incorporación de marcos éticos y de DDHH (Lara Gálvez, 2020). El contexto pandémico ha puesto en evidencia la necesidad de contar con guías y herramientas adecuadas para facilitar que empresas incorporen elementos de *due diligence*, desde la perspectiva de DDHH, en el diseño e implementación de tecnologías.

En este sentido, existe la posibilidad de trabajar tanto con empresas globales de tecnología para codiseñar soluciones de impacto a escala, como con pequeñas y medianas empresas innovadoras que están teniendo un crecimiento exponencial en la región, mediante el desarrollo de nuevos modelos de negocios basados en sistemas de Inteligencia Artificial, o bien, ofreciendo productos o servicios que utilizan esta tecnología (Everis, 2021). Por tanto, resulta prioritario acompañar la documentación de ejemplos innovadores que surgen desde pequeñas y medianas empresas que están participando en ejercicios exploratorios sobre uso responsable y ético de la IA como parte de sus modelos de negocio, a fin de recopilar aprendizajes, aciertos y retos tanto en las fases de desarrollo como de implementación de esta tecnología emergente.

La primera sección de la presente investigación se enfoca en un análisis conceptual donde se realiza una revisión de literatura sobre el URT, se comparten algunas ventajas del uso de URT en organizaciones y empresas y se presenta un breve balance de los modelos basados en la ética y en los DDHH para su gobernanza, incluyendo beneficios y críticas a ambos.

En la segunda sección se desarrolla un análisis documental de los marcos presentados por la iniciativa del grupo de trabajo de URT del WEF, incluyendo las fases propuestas para su operativización en los reportes *Uso responsable de tecnología* (2019) y *Ética por diseño* (2020). Se incluyen también ejemplos de herramientas que empresas como Microsoft, Salesforce e IBM, líderes en dicho grupo de trabajo, han documentado y puesto a disposición del público para ejemplificar cómo llevan a la práctica el aterrizaje de principios éticos en el ciclo de vida de sus productos de tecnología. Asimismo, desde la mirada latinoamericana, se analizan algunos de los programas y herramientas que el BID ha puesto a disposición de *start-ups* en América Latina y el Caribe en el marco de la iniciativa fAlr LAC.

En la tercera y última sección se plantean desafíos y oportunidades desde una perspectiva de política pública en Latinoamérica para lograr una combinación de los modelos basados en la ética y en los DDHH, y se comparte una selección de experiencias desde el sector emprendedor radicado en México quienes han sido parte de la primera generación de programas impulsados

por el BID antes mencionada. Se concluye con reflexiones para continuar identificando casos de uso y ampliando su documentación en un mayor número de países latinoamericanos.

Así, este trabajo tiene como **objetivo** analizar de forma descriptiva y crítica algunos de los marcos de referencia desarrollados para promover el URT y sus aplicaciones en LAC, desde una perspectiva de gobernanza basada en la ética y en los DDHH. Las autoras pretendemos contribuir con insumos a una conversación regional, para acercar mejores prácticas internacionales y ejemplos de buenas prácticas locales al ecosistema emprendedor y tomadores de decisiones, a la vez de visibilizar herramientas metodológicas para diagnosticar y evaluar los retos que enfrentan sus productos y servicios basados en IA, incluyendo en el diseño de estos de forma inclusiva a las comunidades que impactan y personas usuarias.

También, buscamos aportar algunas ideas respecto de un modelo híbrido de gobernanza basado en la ética y en los derechos humanos, partiendo de algunas de las discusiones que se han dado dentro del grupo de trabajo de uso responsable de tecnología del WEF y en otros espacios internacionales. Por último, buscamos dar a conocer en LAC los marcos conceptuales que ha elaborado dicho grupo de trabajo y dar mayor visibilidad y reconocimiento a las herramientas desarrolladas desde el BID de la mano de organizaciones aliadas en la región, con la finalidad de motivar su impulso y aprovechamiento desde los más altos niveles de liderazgo, tanto en los sectores público como privado.

Para lograr los objetivos planteados, la presente investigación se basa en una combinación de metodologías, incluyendo análisis de literatura, conversaciones con actores clave tanto de instituciones y organismos internacionales como el WEF y el BID para entender de primera mano sus avances y visión con relación al uso responsable de tecnología, así como con los fundadores de las *start-ups* que formaron parte de la selección de experiencias desde el sector emprendedor radicado en México para la identificación de ejemplos y buenas prácticas asociadas al abordaje ético y práctico de la IA, como parte de su participación en la primera generación de programas desarrollados por el BID.

2.

Uso responsable de tecnología

Al hablar de *uso responsable de tecnología* (URT), nos referimos al marco de trabajo propuesto por el Foro Económico Mundial (2019) con la intención de introducir y apropiar conceptos de pensamiento ético y estándares internacionales de derechos humanos en la cadena de valor de las compañías que diseñan, desarrollan, comercializan y usan tecnologías disruptivas. Dentro del universo de marcos de referencia, la mayoría forman parte del grupo de trabajo del *uso responsable de tecnología* del WEF. Este esfuerzo tiene como puntos de partida dos grandes retos: la formulación de nuevas políticas y regulaciones frente al constante desarrollo tecnológico, y las preocupaciones éticas y de DDHH que presuponen la digitalización y adopción de tecnologías disruptivas, combinando las dimensiones de ética y DDHH en inteligencia artificial (IA), aprendizaje de máquina (ML por sus siglas en inglés), tecnologías *blockchain*, políticas de privacidad y manejo de datos e internet de las cosas (IoT por sus siglas en inglés).

Desde la perspectiva del WEF y de quienes participan en el proyecto antes mencionado, el concepto de URT está intrínsecamente relacionado con la construcción de un entorno de confianza (tanto al interior como al exterior de la empresa), la relevancia de articular una visión de impacto hacia la sociedad, la necesidad de realizar análisis de riesgo exhaustivos para identificar y, en la medida de lo posible, mitigar consecuencias negativas, así como con la importancia de formar liderazgos y contar con procesos inclusivos de toma de decisión e impulsar la diversidad en los equipos de trabajo, entendiendo el rol fundamental que juegan el contexto y la cultura organizacional de una compañía como habilitadores de cambio sistémico (WEF, 2020). Estos marcos conceptuales son de sumo valor y utilidad para que las empresas puedan contar con un listado de ejemplos e implementar aquellas acciones que mejor se adapten a sus propias identidades y necesidades. De forma específica, la profesora e investigadora Virginia Dignum, construye el concepto de *inteligencia artificial responsable* como la responsabilidad frente al poder y potencial de la IA en la ejecución de tareas de forma autónoma, asegurando que el propósito codificado en el sistema esté alineado con el propósito que se visualiza para el producto o servicio que utiliza la IA (Dignum, 2012).

2.1. Ventajas del URT en organizaciones y empresas

Desde una perspectiva de valor agregado, se ha demostrado que contar con un espacio para la reflexión ética y la formulación de acciones concretas para su operativización demuestra un mayor conocimiento tecnológico y madurez empresarial (Sniderman, 2020). A partir de una serie de estudios realizados por Deloitte sobre la aceleración de la innovación digital y liderazgos empresariales en el contexto de la transformación digital, así como la carrera internacional para tomar ventaja en el desarrollo de sistemas de IA, la consultora recalcó que la atención al impacto ético y de DDHH de la tecnología refleja un mayor avance en el involucramiento de las

compañías en la transformación digital, en comparación con otras empresas de los mismos rubros que se encuentran en las primeras etapas de asimilación digital. De acuerdo con dicho estudio, el 57% de las personas entrevistadas en compañías en proceso de maduración digital reportaron que los liderazgos internos de sus organizaciones asignaban tiempo adecuado y útil para reflexionar y comunicar los impactos sociales de las iniciativas digitales que emprenden (Kane, *et al.*, 2019).

De forma similar, el mismo estudio señala que el 47% de las organizaciones que cuentan con procesos claros y definidos en la toma de decisiones reconocen la importancia del URT. La cifra desciende al 21% entre aquellas donde se considera que no se cuenta con un proceso tan claro (2019). Por tanto, la proactividad ética en el ecosistema de emprendimiento permite a las empresas contar con un panorama más amplio sobre los retos de desarrollo y uso de tecnología, al convertir estas preocupaciones en una parte intrínseca de sus procesos, con lo cual siguientes iteraciones de productos y servicios contarán con aprendizajes previos y consideraciones específicas a tomar en cuenta de forma transversal.

Además, atender consideraciones éticas en todas las etapas de desarrollo de tecnología, en la cultura organizacional y en el modelo de negocios, representa una oportunidad para reforzar la confianza de las personas clientes y actores involucrados. El Reporte global de tendencias de marketing 2020 de Deloitte apunta que la confianza en las marcas es un aspecto cada vez más importante para el éxito de los negocios y en sus relaciones con clientes, reguladores y medios, a través de la transparencia y apertura sobre los procesos que guían la toma de decisiones (Deloitte, 2020).

Por otro lado, el Índice de compañías éticas 2021 de Ethisphere, que reúne más de 200 puntos de datos sobre la cultura y prácticas sociales, actividades centradas en la ética y acciones de gobernanza y diversidad de 135 empresas, comparó el índice de rendimiento de las compañías rankeadas con el Índice Solactive de alta capitalización de los EE.UU. A través de esta comparativa, se encontró que existe una “prima ética”, beneficiando a las compañías éticas con un crecimiento de 7.1 puntos porcentuales más que sus contrapartes en el periodo de cinco años entre 2016 y 2021. Sobre ello, Ethisphere destaca que este beneficio ha sido consistente desde que comenzó la medición de las compañías y sostiene que el resultado de prácticas organizacionales y de desarrollo éticas se reflejan en un mejor desempeño financiero (Ethisphere, 2021).

En atención a la creciente importancia de la confianza y la ética empresarial, el grupo de trabajo de URT del Foro Económico Mundial plantea modelos de operativización de las perspectivas ética y de derechos humanos, de forma que las ventajas de ambos puedan ser apalancadas en las diversas etapas del ciclo de productos y servicios tecnológicos.

2.2. Breve balance de los modelos basados en la ética y en los DDHH para la gobernanza del URT

Existe consenso en que el diseño y uso de tecnología debe integrar principios mínimos alrededor de distintos ejes temáticos (Fjeld & Nagy, 2020). Ahora bien, para poder llevar estos valores

y principios a la práctica, ha sido necesario construir modelos o directrices que permitan, tanto a actores estatales como no estatales, poder identificar brechas en el URT. Esta sección presenta un breve balance de dos de los modelos principales utilizados a estos fines: los modelos de gobernanza basados en la ética y los modelos basados en los DDHH. Asimismo, abordaremos algunas pautas sobre oportunidades para su operativización en LAC.

2.2.1. Modelos de gobernanza basados en la ética: beneficios y críticas

Los modelos de gobernanza basados en la ética han sido ampliamente utilizados en el URT (WEF, 2019), en parte, por la posibilidad de adaptarlos a distintos contextos geográficos y culturales. Estos modelos permiten incorporar consideraciones desde diferentes sistemas de valores, tradiciones y culturas. Así, modelos basados en la ética permiten tomar decisiones “en situaciones en las que lo correcto y lo incorrecto, lo bueno y lo malo, no están claramente definidos” (WEF, 2019). De esta forma, los modelos basados en la ética resultan particularmente atractivos para empresas multinacionales, ya que permiten facilitar diálogos que implican considerar diferentes perspectivas que no pueden ser descartadas en función de elementos más neutrales (WEF, 2020).

Ahora bien, aunque los modelos basados en la ética priman en la industria, no han estado exentos de críticas. Una de las principales es que, al abrir espacio para consideraciones culturales, se puede habilitar la posibilidad de definir estándares de cumplimiento diferentes, según el sistema de valores de cada contexto (Elsayed-Ali, 2018). Por ejemplo, preguntas como a quién vender, cómo restringir el uso inadecuado de la IA o cómo reducir sesgos, no siempre pueden ser resueltas de forma homogénea usando modelos basados en la ética (WEF, 2019).

Otra de las críticas es la falta de criterios concretos y medibles para monitorear el cumplimiento de estándares de URT. Se ha resaltado cómo, en la práctica, los modelos basados en la ética suelen ser de difícil operativización para equipos técnicos (Burt, 2020). Así, estos modelos corren el riesgo de convertirse en declaraciones aspiracionales y no en criterios específicos de control que deben ser cumplidos. Por ejemplo, de 160 directrices analizadas a nivel mundial en un amplio estudio, la gran mayoría se basan en compromisos voluntarios o recomendaciones con muy pocos “dientes” para garantizar su cumplimiento (Algorithm Watch, 2021).

Como solución ante esta ambigüedad, se ha ofrecido la creación de indicadores o métricas claras, medibles y monitoreadas, para cada principio o criterio basado en la ética que las organizaciones desarrolladoras de tecnología adopten en sus procesos (Burt, 2020). Sin embargo, esta propuesta no permitiría resolver del todo cómo se define el contenido de estos principios basados en la ética. Por ejemplo, ¿A qué nos referimos con “respeto a la igualdad y no discriminación”? ¿Qué marcos conceptuales usaremos, si se depende del contexto local? ¿Cómo se mide su cumplimiento, si no sabemos a qué se refiere en concreto?

Consideramos que aún con el desarrollo de indicadores y métricas, usar modelos basados en la ética para promover el URT de forma exclusiva no resuelve del todo las barreras para el cumplimiento. Tampoco logra resolver la falta de mecanismos de seguimiento: en el estudio antes citado, por ejemplo, se documenta cómo *solo* 10 de las 160 directrices analizadas tienen

mecanismos de aplicación prácticos (Algorithm Watch, 2021). Los modelos basados en la ética, por sí solos, no parecen ser suficientes para incorporar, dentro de los ciclos de diseño y uso de tecnología, recursos disponibles en caso de incumplimiento, así como medidas para proteger y prevenir el riesgo de daño.

Como alternativa, diferentes actores, principalmente desde la sociedad civil, han promovido el diseño de modelos basados en criterios más universales y firmes, que, al igual que la IA, trascienden las fronteras. La Declaración de Toronto, por ejemplo, es un llamado a ir más allá del modelo basado en la ética, para adoptar uno basado en el marco internacional de los DDHH (Amnistía Internacional & Access Now, 2018).

2.2.2. Modelos de gobernanza basados en los derechos humanos: ¿un marco único y común?

En términos generales, la propuesta de modelos de gobernanza basados en DDHH se fundamenta en el argumento de que estos son universales, vinculantes y accionables, es decir, que presentan un marco claro de acción, generalmente aceptado por gobiernos, empresas y sociedad civil, con responsabilidades tanto de los Estados, como de los actores privados. En contraste con la ética, un modelo de URT basado en los DDHH toma, como punto de partida, principios y normas internacionalmente reconocidas (WEF, 2019).

En efecto, los modelos basados en DDHH abarcan criterios legalmente vinculantes, que presentan un mínimo de condiciones a ser respetadas, denominados como derechos y obligaciones, así como a titulares de esos derechos y de las obligaciones. En el contexto del URT, los modelos basados en DDHH permiten identificar roles y estándares a seguir por gobiernos y actores privados, de cara a las personas usuarias de las herramientas de tecnología, así como de personas que podrían ser impactadas por éstas (WEF, 2019). Así, elementos como la privacidad y el libre acceso a la información, por ejemplo, pasan de ser principios aspiracionales a derechos de toda persona, con un contenido específico definido por tratados y otros instrumentos legales, y que deben ser respetados tanto por los estados, como por las empresas.

Si bien los modelos basados en la ética pueden percibirse como endógenos a quienes impulsan el desarrollo de la IA, en el caso de los modelos de DDHH son de reciente debate y se vinculan más a las responsabilidades de los Estados frente al posible impacto de esta tecnología. Desde esta última perspectiva, se ha discutido bastante sobre el rol regulador de los gobiernos. Sin embargo, en las últimas décadas y desde el derecho internacional de Derechos Humanos, se ha reflexionado también sobre el rol de las empresas en su garantía.

Esta construcción de paralelismos para actores no estatales, como lo son las empresas que se involucran en el desarrollo de IA, ha sido un proceso que parte del reconocimiento de que, si bien a nivel normativo, los Estados son los garantes últimos de los DDHH, no *solo* los gobiernos tienen el poder de impactar a las personas a través de la IA y, por lo tanto, no *solo* los Estados deben tener responsabilidad bajo el derecho internacional de Derechos Humanos.

Así, los modelos de URT basados en DDHH incluyen, como una de sus piedras angulares, los Principios de la ONU sobre Empresas y Derechos Humanos, que plantean claramente el deber de las empresas de respetarlos. Bajo este instrumento, las empresas, sin importar el alcance geográfico de sus operaciones, pasan de ser actores pasivos a agentes que tienen un impacto en su cumplimiento y, por tanto, deben seguir ciertas pautas dentro de sus modelos de negocios (ONU, 2011).

En este sentido, los Principios Rectores son un esfuerzo para hacer paralelismos entre la responsabilidad estatal, tal como se define en el derecho internacional, y las responsabilidades que toda empresa debe asumir para participar en el proceso de garantizar Derechos Humanos a nivel doméstico e incluso transnacional.

En este mismo tenor, cabe mencionar que la Declaración de Toronto representa un gran esfuerzo para trasladar conceptos de Derechos Humanos al campo de la IA, partiendo de antecedentes como los Principios Rectores. Esta Declaración delinea responsabilidades tanto de los estados como de actores no estatales respecto al uso de sistemas de ML, tales como la mitigación de los efectos discriminatorios y la provisión de recursos efectivos para individuos o comunidades afectadas negativamente por la tecnología (Amnistía Internacional & Access Now, 2018).

Precisamente, una de las preocupaciones donde el derecho internacional de los Derechos Humanos puede ser particularmente útil, en el campo de la IA, es en temas de igualdad y no discriminación. En efecto, uno de los principales estándares que el modelo de DDHH ofrece al URT se vincula al principio de igualdad y no discriminación que permite identificar, desde la inepción de herramientas de IA, los posibles sesgos y prejuicios que podrían ser replicados, así como entender de qué manera prevenir ciertas discriminaciones.

El derecho internacional de los Derechos Humanos, particularmente desde la tradición latinoamericana y caribeña, ha desarrollado conceptos, así como pruebas o *tests*, que permiten identificar la ocurrencia de discriminación estructural, múltiple o interseccional y hacen espacio al reconocimiento de sesgos como causas estructurales (MacGregor, 2019). En este sentido, al incorporar estos criterios y pruebas en el diseño y uso de IA, se podría decidir el abandono del desarrollo de alguna herramienta en particular, porque el riesgo de que replique situaciones de discriminación es alto y no prevenible (MacGregor, 2019).

Partiendo de este razonamiento, podemos identificar que una de las responsabilidades que las empresas deben cumplir, bajo el modelo basado en DDHH, es conducir análisis de impacto o de riesgos basados en Derechos Humanos durante el inicio del diseño, no *solo* durante el despliegue y uso (WEF, 2019). Esto porque uno de los beneficios que aportan los modelos basados en DDHH, además de estándares mínimos que cumplir, es la necesidad de prevenir posibles violaciones a estos, incluyendo el riesgo de generar o ampliar situaciones de discriminación (Lara, 2020). Para lo anterior, distintos actores dentro del campo de la IA han propuesto el uso de herramientas como las evaluaciones de impacto sobre los Derechos Humanos, o *Human Rights Impact Assessments* (HRIAs por sus siglas en inglés), durante todo el ciclo de vida de los sistemas de IA (Latonero, 2020).

Cabe destacar que los HRIAs están previstos por los Principios Rectores. Además, el uso de estas herramientas es ampliamente considerado como una buena práctica, pero aún no son del todo mandatorios internacionalmente para empresas e incluso para actores estatales que se involucran en el diseño y uso de tecnología (Dawson, 2019). Así, distintas partes interesadas o *stakeholders*, particularmente desde la sociedad civil y la academia, han abogado por que los gobiernos, de manera progresiva, incluyan el análisis de impacto y de riesgo en los Derechos Humanos como una condición obligatoria que debe ser cumplida para poder otorgarse licencias o permisos de uso de tecnología (Element AI, 2019). La Declaración de Toronto es, justamente, un esfuerzo hacia estos fines.

La vinculatoriedad de los Derechos Humanos es la base de otro de los beneficios que ofrecen los modelos de gobernanza basados en estos. Algo que en muchas ocasiones falta en modelos basados en la ética, es la necesidad de definir mecanismos de rendición de cuentas (WEF, 2019). Desde la lógica del derecho internacional de los DDHH, esta protección de derechos iría en doble vía: la prevención de violaciones de DDHH, que ya hemos mencionado, y la garantía de esos derechos. En los casos en donde estos no sean realizados plenamente, corresponde la posibilidad de poder acceder a un recurso efectivo, y en caso de que proceda, recibir reparación por el daño causado.

Aunque los modelos basados en Derechos Humanos podrían suplir muchas de las lagunas no cubiertas por los modelos basados en la ética, no representan una solución perfecta. De manera general, es importante que las empresas desarrolladoras de tecnologías colaboren constantemente con sociedad civil, academia y comunidades impactadas, a fin de operativizar los tecnicismos de los Derechos Humanos y construir procedimientos efectivos que los garanticen. Por ejemplo, esta colaboración podría ayudar a lograr que las empresas desarrolladoras de tecnología puedan establecer recursos o mecanismos de rendición de cuentas, que sean a la vez efectivos y lo suficientemente ágiles para poder responder a la rápida naturaleza del desarrollo y uso de tecnología.

Asimismo, aunque una de las ventajas de los Derechos Humanos es su vinculatoriedad, al ser normas legales, a nivel doméstico, deben ser reforzadas por los gobiernos. En general, las normas de DDHH mayormente referenciadas dentro del campo de la IA son Tratados, como el Pacto Internacional de los Derechos Civiles y Políticos, generados en el seno de organismos multilaterales como la Organización de Estados Americanos (OEA) o la Organización de las Naciones Unidas (ONU). Para que los modelos basados en Derechos Humanos sean más efectivos, es ideal que a nivel nacional o doméstico, se incorpore la obligación de las empresas de proteger los DDHH en las políticas públicas relacionadas con sistemas de IA.

Esta incorporación podría ayudar a hacer frente a otra de las críticas al modelo, vinculada con este último punto. Como mencionamos, una de las fuentes esenciales de las normas de Derechos Humanos son los Tratados, que a su vez, son interpretados por órganos especializados. Algunas voces, como investigadoras del Berkman Klein Center de la Universidad de Harvard, han destacado cómo, a pesar de que los preceptos de DDHH son universales, no proporcionan la precisión y certeza necesarias para facilitar que empresas puedan operarlos, particularmente, debido a que la interpretación de estos estándares, por parte de tribunales y autoridades gubernamentales, no es homogénea a través de los sistemas regionales, como el europeo y el interamericano (Dvoskin, 2019).

Aunque es posible que puedan encontrarse algunas diferencias en la interpretación del contenido de ciertos derechos entre sistemas regionales, cada vez hay un mayor intercambio entre sistemas regionales, como un ejercicio de fortalecer sus trabajos. Asimismo, es importante reiterar que con mayor frecuencia las empresas son llamadas a hacer análisis de riesgo basados en DDHH y tomar decisiones a partir de esos resultados, desde los ordenamientos jurídicos nacionales y desde el Sistema Interamericano. Como ejemplos recientes, se destacan el último informe sobre IA y Derechos Humanos emitido en 2021 por la Alta Comisionada de Naciones Unidas para los Derechos Humanos, así como los estándares de empresas y Derechos Humanos de la Comisión Interamericana de Derechos Humanos (CIDH). Abordamos ambas fuentes más adelante.

Ahora bien, teniendo en cuenta lo expuesto anteriormente, en lugar de optar por uno de los modelos, proponemos que los modelos de gobernanza basados en la ética y los basados en los Derechos Humanos sean complementarios entre sí. En una región como LAC, utilizar ambos modelos de forma complementaria podría ayudar a resolver numerosos desafíos que tanto las empresas como los gobiernos enfrentan al momento de asegurar que la IA y otras tecnologías sean utilizadas de forma responsable. Abordaremos esta propuesta más adelante.

3.

Propuestas para operativizar el uso responsable de tecnología

Desde una perspectiva del Norte global, el Grupo de Trabajo de URT del WEF reconoce que los retos en materia ética de las tecnologías disruptivas son inherentes a éstas, mientras que los procesos de digitalización que hemos emprendido traen consigo la necesidad de cuestionar, repensar y delinear nuestra relación con las nuevas tecnologías para lograr aprovechar sus beneficios (BSR & WEF, 2019). En el documento *Uso responsable de tecnología*, se destacan los roles específicos de las compañías que desarrollan tecnologías disruptivas, así como de aquellas que las adquieren y aplican. Las primeras, cuentan con la oportunidad de delinear las políticas de uso que guíen lo que se puede y no se puede hacer con sus productos, pueden escoger a quien vender y pueden acercar información y prácticas de mitigación de riesgos a sus posibles clientes; las segundas, tienen la posibilidad de comprender el impacto del uso y aplicación de tecnología a través de una evaluación de DDHH, mitigar los riesgos de implementación al decidir cómo se utilizarán las tecnologías y participar en un diálogo constante con las compañías desarrolladoras para evitar que los usos de tecnología afecten a las personas usuarias. Además, se establecen acciones puntuales en tres etapas o fases de la cadena de valor: Diseño y Desarrollo, Despliegue y Aplicación o Uso. A continuación, se describen los modelos de Uso Responsable de Tecnología, Ética por Diseño y el modelo complementario de Acción-Habilitación-Influencia, así como ejemplos de herramientas desarrolladas por Microsoft, Salesforce e IBM para su operativización.

3.1. Modelo planteado en el reporte *Uso responsable de tecnología*

El modelo que plantea el reporte *Uso responsable de tecnología* (BSR & WEF, 2019) tiene por objetivo la implementación combinada de las perspectivas éticas y de derechos humanos en el desarrollo de productos y servicios tecnológicos. Para ello, propone asumir la responsabilidad corporativa con relación al cumplimiento de los derechos humanos y la necesidad de tomar acción frente a posibles impactos desfavorables a estos en la actividad empresarial, como los delinear los Principios de la ONU sobre Empresas y Derechos Humanos, aplicando las nociones de responsabilidad y capacidad de acción en cada etapa de la cadena de valor de los productos y servicios que se ofrecen.

En la fase de diseño y desarrollo propuesta en este modelo, el enfoque es el reconocimiento de implicaciones éticas y de DDHH en los prototipos e ideas de los equipos por medio de cuatro prácticas distintivas. Primero, si las compañías carecen de los conocimientos necesarios para identificar situaciones de impacto negativo, deben contratar personas especializadas y consultoras que estén a cargo de desarrollar políticas de la empresa basadas en el respeto a los DDHH a través de la concienciación de los diferentes equipos involucrados. Segundo, la participación en discusiones activas con actores relevantes a las tecnologías en desarrollo permitirá

conocer los puntos de vista de poblaciones vulnerables que podrían ser mayormente afectadas y tomar decisiones en las fases tempranas del ciclo de creación de productos para mitigar riesgos asociados. En este sentido, ya existen marcos de trabajo que habiliten esta interacción entre empresas desarrolladoras y actores: el diseño centrado en las personas, los DDHH por diseño y la innovación responsable son algunos de ellos (BSR & WEF, 2019). La tercera práctica implica la utilización de herramientas técnicas para responder a posibles escenarios de discriminación y sesgo, como la Caja de Herramientas de Justicia de Accenture o AI Fairness 360 de IBM, o el uso de bases de datos representativas de las poblaciones a quienes afecta un producto. La última práctica son las “pruebas adversariales” en las que un producto o servicio es retirado del contexto en el que se asume que será utilizado y se usa en condiciones diferentes para evaluar los posibles impactos de su despliegue (BSR & WEF, 2019).

Las anteriores propuestas requieren la participación multidisciplinaria de los involucrados en la cadena de valor, empezando por los equipos de políticas y sustentabilidad, en la institucionalización de los marcos de trabajo de diseño centrado en las personas, así como los equipos de ingeniería y diseño de productos para consultar la implementación de nuevos procesos y para dar retroalimentación en las primeras etapas de diseño de productos, servicios y tecnologías (BSR & WEF, 2019).

En la fase de despliegue se propone la implementación de tres medidas en atención a la ética y DDHH en la venta y uso de nuevos productos. Primero, las políticas de uso aceptable (AUP, por sus siglas en inglés) son consideradas un ejercicio pertinente para delimitar los usos éticos y sus límites, pero generan una distancia legal y moral entre las personas y las compañías creadoras de tecnología, de forma que la responsabilidad final sigue recayendo sobre las personas usuarias, además de que hacer cumplir las AUP significa un reto asociado (BSR & WEF, 2019). Segundo, dirigido a actores más comerciales e industriales, crear “listas blancas o negras” delimita el uso permitido de una tecnología en ciertos sectores, a través de la restricción de posibles compradores e industrias de aplicación; esto se logra a través del entendimiento de la naturaleza del cliente y el uso final que tendría el producto o servicio. Las listas blancas registran a consumidores potenciales a quienes se les da luz verde para poder adquirir cierta tecnología, mientras que las listas negras eliminan a ciertas compañías o industrias como posibles clientes. Ambas listas se construyen a través de una evaluación de criterios éticos, de sostenibilidad, equidad, de DDHH, entre otros. Finalmente, proveer entrenamiento ético y en materia de derechos humanos como guías para el mejor uso de los productos durante el proceso de venta incrementa la conciencia de quienes adquieren el producto y abre la puerta para que gobiernos, universidades y sociedad civil se apropien de estos lineamientos (BSR & WEF 2019).

Por último, en la fase de uso, durante el ciclo de creación y producción, la responsabilidad regresa a las personas usuarias y consumidoras, por lo que es importante tomar acciones para distribuir la responsabilidad entre los sectores e industrias involucrados. Las compañías de tecnología crean nuevos productos y servicios, pero su uso ético tiene relevancia en la amplia variedad de las industrias que los adoptan.

Este marco metodológico acepta que existen limitaciones a la aplicación de las prácticas de URT. Primeramente, es necesario incorporar a toda la cadena de valor en estas prácticas, en

vez de realizarlas de forma aislada; en consecuencia, las personas hacedoras de política pública, sociedad civil y otros actores relevantes deben involucrarse en la resolución de retos éticos y de DDHH frente a la tecnología. Este involucramiento requiere de un compromiso de los estados para incorporar y aplicar políticas para el URT que ni las empresas, ni grupos de la sociedad civil o académicos pueden aplicar de forma independiente (BSR & WEF, 2019). De forma paralela, el modelo complementario de acción-habilitación-influencia, se plantea como un camino para generar cambios sistémicos en la cadena de valor y en los actores involucrados, a partir de la iniciativa de las empresas comprometidas con el URT.

3.2. Modelo complementario: acción-habilitación-influencia

El modelo acción-habilitación-influencia, también planteado en la publicación de *Uso responsable de tecnología* del WEF (2019), propone que las empresas tomen acción para crear, de forma interna, estrategias de negocios y gobernanza sobre los modelos basados en DDHH y ética, para posteriormente habilitarlos fuera de los límites de la propia empresa, extendiéndose a las relaciones con actores, clientes y personas usuarias a lo largo de toda la cadena de valor. Por su parte, influencia se refiere a convertir a las empresas en promotoras de la tecnología responsable por medio del apoyo a políticas, regulaciones y leyes que refuercen la relación entre el éxito comercial y la creación de productos con perspectivas de URT.

Al hacer uso de este modelo, las empresas establecen estrategias de negocios que tengan el URT al centro, pues tanto el ciclo de vida de los productos, las relaciones con clientes y la interacción con actores externos de influencia, como gobiernos y sociedad civil, se alinean con el objetivo de respetar los DDHH, actuar de forma ética y prevenir daños por el uso de tecnologías disruptivas.

3.3. Modelo planteado en el reporte Ética por diseño

En seguimiento a las ideas planteadas en el documento *Uso responsable de tecnología*, el reporte Ética por diseño, también del WEF, (2020) retoma el modelo desarrollado por Nicholas Epley y David Tannenbaum (2018), con base en su investigación en el campo de la psicología y las ciencias del comportamiento, para diseñar principios que incentiven la reflexión y acciones éticas. Dicho modelo se divide en tres fases relacionadas: atención, construcción y motivación (WEF, 2020).

El primer pilar asume que un diseño efectivo para el comportamiento ético impulsará a las personas a reflexionar de manera rutinaria; para ello, propone el uso de materiales de concientización, estímulos organizacionales institucionalizados, como recordatorios dentro de los procesos de selección y entrevista, o herramientas técnicas que evalúen posibles riesgos éticos de los algoritmos, así como la creación de lenguajes compartidos que incluyen eslogans, mantras y misiones codefinidas.

El segundo pilar de la *Ética por diseño* del WEF reconoce la influencia de los contextos en la toma de decisión y propone incorporar la diversidad de estos en las experiencias vividas por las personas que integran la organización (2020). Para ello, se propone la creación de equipos socialmente diversos, que desarrollen en conjunto modelos de razonamiento ético y de URT con base en paradigmas éticos y políticas existentes, mientras que dichos modelos son iterados con las personas y comunidades que podrían ser afectadas por las nuevas tecnologías (WEF, 2020).

El último pilar asume que el cambio en el comportamiento de las personas necesita motivaciones psicosociales, para lo cual presenta actividades de introspección histórica de las personas integrantes de la organización, así como su toma de decisiones a partir de los contextos vividos, utilizando la misión previamente definida y los marcos de trabajo ético como rúbrica (WEF, 2020). Con ello, este modelo también busca crear relaciones empáticas entre los individuos de las organizaciones hacia las personas clientes y usuarias, todo dentro de una narrativa general del impacto ético de los productos y servicios que desarrollan.

Como se puede observar, el liderazgo del WEF en el desarrollo de modelos operativos para institucionalizar los modelos basados en ética y en DDHH ofrece resultados aplicables en todo el proceso de creación de productos y servicios tecnológicos. Por ello, en línea con el crecimiento en América Latina y el Caribe del ecosistema de innovación y emprendimiento en tecnologías disruptivas, resulta relevante acercar estos modelos y ejemplos de herramientas a las empresas y *start-ups* de la región. Tomar elementos de dichos modelos y herramientas en el ecosistema emprendedor en América Latina y el Caribe permitirá habilitar un cambio de paradigma en el que su uso se vuelva una práctica generalizada fuera de empresas y organizaciones basadas en los Estados Unidos y Europa, abonando a la construcción de un panorama internacional del URT, sensible a los contextos regionales y a las necesidades específicas de diferentes poblaciones.

3.4. Ejemplos de herramientas para el URT

En línea con principios y guías planteadas en los modelos precitados, Microsoft, Salesforce e IBM, empresas líderes en el grupo de trabajo de URT del WEF, han documentado ejemplos que ofrecen un catálogo de opciones para los equipos involucrados en el ciclo de vida de productos o servicios de tecnología, sumando a un amplio rango de actores para incluir perspectivas diversas sobre los impactos de los sistemas de IA. A continuación, se comparten algunas herramientas que dichas compañías han puesto en práctica.

Juego de juicio (*Judgement Call*)

Retomando los cinco principios de la IA de Microsoft (justicia, seguridad, privacidad, inclusión, transparencia y rendición de cuentas), el juego presenta a los equipos en diferentes fases del ciclo de vida del producto o servicio de tecnología en cuestión con la oportunidad de evaluar el impacto de dicho producto o servicio desde la perspectiva de otro actor involucrado. Las

personas participantes reciben una carta que les asigna el rol de algún actor o grupo al que el producto digital impacta, una carta donde se lee uno de los principios de IA y una carta con un número del 1 al 5, que representa el número de estrellas que el actor hipotético otorgará al producto en una reseña. Con base en esta información, cada miembro del equipo escribe reseñas, explicando el impacto de dicha tecnología (Markkula Center for Applied Ethics & WEF, 2021).

Esta iniciativa ha probado ser de suma eficacia para motivar el pensamiento ético y la colaboración en los equipos dentro de la compañía (Ballard, *et al.*, 2019). Algunas razones por las que esta herramienta ha sido exitosa incluyen que el liderazgo organizacional empujó hacia el desarrollo de soluciones relacionadas a la ética, el juego construye sobre los principios de IA de Microsoft y toma en cuenta prácticas existentes en los equipos de desarrollo de productos, de modo que su incorporación en el proceso no representa una carga de trabajo mayor, sino un cambio de perspectiva sobre los procesos existentes.

Consejo Asesor Externo y Oficina para el Uso Ético y Humano de la IA

Salesforce instauró la Oficina de Uso Ético y Humano de la IA desde 2018 como parte de su Oficina de Igualdad, con la finalidad de asegurar que los servicios y productos tecnológicos no afecten, sino que ayuden a proteger los DDHH y generar valores comunes para la industria de la tecnología a través del liderazgo de la compañía (Salesforce, 2020). A partir de la creación de esta oficina, se asumió el compromiso de generar un espacio donde los distintos actores involucrados en las decisiones de la empresa, sus necesidades y perspectivas éticas sean reconocidas. En paralelo, se creó un consejo asesor externo, con expertos en ética y URT con perspectivas variadas que nutren la toma de decisiones (Salesforce, 2020) (WEF, 2020).

Taller Envision AI

Este ejercicio de Microsoft ayuda a los equipos a reflexionar sobre el desarrollo de un análisis de impacto. En él, las personas que participan revisan escenarios reales del proceso de desarrollo del sistema de IA del Proyecto Tokio, establecido en 2016, cuyo objetivo es aumentar las capacidades humanas a través de la tecnología. El objetivo de la revisión de estos escenarios, a partir de la innovación centrada en las personas y el entendimiento de las necesidades tecnológicas de diferentes grupos, es acercar herramientas y recursos para mitigar posibles riesgos de uso de tecnología e identificar el impacto diferenciado en las personas usuarias (WEF, 2021). En el marco del evento Forum Europe 2021 se contó con una versión corta del taller, en el cual quienes asistieron tuvieron la oportunidad de poner en práctica los principios de inteligencia artificial de Microsoft a través de la discusión interactiva y la reflexión basada en casos (Forum Europe, 2021).

Notificaciones dentro de plataformas de IA

Einstein es la plataforma de IA de Salesforce. Una de las funciones que proporcionan para

alertar a quienes utilizan este servicio con fines comerciales y que podrían estar usando variables sensibles como edad, raza o género es a través de guías llamadas Einstein Discovery. La herramienta indica problemas en los datos que podrían sesgar los resultados, tales como discrepancia entre las clases, alta correlación entre variables sensibles, demasiados *outliers* o la existencia de una variable correlacionada con los resultados (Salesforce, 2020).

Jurado comunitario

La herramienta de Jurado comunitario, propuesta por Microsoft, permite que los equipos de diferentes etapas del ciclo de vida de un producto o servicio de tecnología interactúen con las poblaciones a quienes este podría impactar. El primer paso para llevar a cabo un jurado comunitario es establecer un grupo de actores relevante, representativo y diverso, quienes tendrán el rol de jurado. Los equipos que desarrollan un nuevo servicio o producto presentan el propósito y posibles casos de uso de la tecnología, así como beneficios y peligros potenciales. Las personas participantes revisan esta información y comparten perspectivas sobre el impacto del producto entre ellos, con la ayuda de un moderador neutral que asegure la participación igualitaria de todas. Los temas clave que surjan en la discusión son incorporados como oportunidades y retos para la cocreación de soluciones (Markkula Center for Applied Ethics & WEF, 2021).

InterpretML

Esta herramienta de Microsoft tiene el objetivo de crear modelos de aprendizaje de máquina más transparentes, entendibles e interpretables. A través de InterpretML se pueden generar explicaciones *globales* sobre el comportamiento de un modelo y explicaciones *locales* para predicciones individuales de dicho modelo. Además, incorpora árboles de decisión y mecanismos para generar modelos explicables y transparentes, así como herramientas para descifrar modelos de caja negra (Markkula Center for Applied Ethics & WEF, 2021).

Análisis de error

De manera similar a InterpretML, el análisis de error es una herramienta utilizada en Microsoft que busca áreas de oportunidad para la explicabilidad de modelos de ML, pero esta herramienta se centra en la identificación de sesgos y errores en los datos. Por medio de visualizaciones, los equipos de ciencia de datos pueden identificar clasificaciones dentro de un modelo que puedan contener errores y cuáles son los grupos de datos cuyo margen de error sobrepasa frente a los parámetros establecidos, así como identificar la distribución del porcentaje de error en estos (Markkula Center for Applied Ethics & WEF, 2021).

Herramientas Open Source de IBM

En 2020, IBM donó una serie de herramientas de código abierto a la Fundación Linux. Entre

ellas se encuentran *AI Fairness 360*, que cuenta con más de 70 métricas de justicia de la IA y 10 algoritmos de mitigación de sesgos; *AI Explainability 360*, que utiliza 8 algoritmos para transparentar y explicar el funcionamiento de ML en atención a la variedad de actores involucrados en el ciclo de vida de la IA; la *Caja de Herramientas de Robustez Adversarial 360*, que permite a las personas desarrolladoras de IA defenderse de ataques, a través del aprendizaje basado en la simulación de posibles ataques y evaluación de modelos de respuesta, y la *Caja de Herramientas de Cuantificación de la Incertidumbre* que permite a las personas desarrolladoras conocer los límites de la confianza que se puede depositar en la IA, por medio de la cuantificación de la incertidumbre en la toma de decisiones de un sistema determinado (Markkula Center for Applied Ethics & WEF, 2021).

Herramientas cerradas de IBM

Cloud Pak for Data es un paquete de herramientas de IBM, utilizado en todo el ciclo de análisis de datos al interior de la compañía, que permite a las personas usuarias participar del monitoreo de modelos de IA confiables. Una de las herramientas más importantes es Watson Studio, utilizada de la mano con la solución de IA Watson, para asegurar la calidad de los datos utilizados para el aprendizaje, escanear los modelos para detectar sesgos e identificar poblaciones a las que los algoritmos podrían impactar negativamente, y generar recomendaciones para realizar modificaciones en los modelos de aprendizaje de máquina, todo esto explicado en términos que tanto personas del ámbito de los negocios como la comunidad científica de datos puedan entenderlos (Markkula Center for Applied Ethics & WEF, 2021).

Las herramientas presentadas por Salesforce, IBM y Microsoft en el marco del grupo de trabajo pueden ser de utilidad como punto de partida para entender cómo operativizar el URT, pero como el documento *Caso de estudio de Microsoft (2021)* hace notar, el cambio en la cultura dentro de la organización es crucial para que estos principios, marcos de trabajo y herramientas se vuelvan parte inherente al proceso de desarrollo de nuevas tecnologías. Por ello, al pensar en la posible aplicación de dichas herramientas en América Latina y el Caribe, es necesario adaptar al contexto de la región las expectativas, capacidades y necesidades de las empresas de tecnología multinacionales, nacionales y *start-ups* en línea con sus prioridades.

En el caso de LAC, en concordancia con la necesidad de contar con marcos de referencia adaptados a la realidad política, económica y social de la región, la iniciativa fAIR LAC del Grupo BID fue lanzada como punto de reunión para actores del sector público, privado, academia y emprendedores para generar redes que faciliten la aplicación ética y responsable de la IA en la región para resolver problemas públicos (BID, s.f). La iniciativa surge a finales de 2019, a partir del incremento de proyectos con componentes de uso de datos privados o sensibles, así como un alineamiento con la Visión 2025 del BID, en donde se democratizan herramientas que faciliten la digitalización de las empresas, especialmente las de menor tamaño, y la adopción responsable de nuevas tecnologías por parte de los sectores público y privado.

En mayo de 2021 publicaron la Autoevaluación Ética de la IA para Actores del Ecosistema Emprendedor como el primer producto del proyecto fAIR LAC *Entrepreneurial Journey*. Se trata de una guía orientativa en español para empresas, inversionistas y aceleradoras sobre aspectos

éticos a considerar en el desarrollo e implementación de soluciones tecnológicas basadas en el manejo de datos de IA, de acuerdo con su nivel de madurez. El resultado de este ejercicio se traduce en la identificación de áreas de mejora que permitirán mejorar la calidad de los productos, servicios, procesos, modelos y sistemas en desarrollo, así como identificar las principales áreas de riesgo en materia de ética.

Lo innovador de este documento radica en dos aspectos principales: el primero, que no pone el peso de la autorregulación únicamente en los emprendedores, sino que involucra a dos actores clave para el ecosistema: los financiadores de proyectos y las aceleradoras; y segundo, que las preguntas orientadoras corresponden a tres niveles de desarrollo de las empresas desde etapas tempranas en fases de ideación, hasta pequeñas y medianas empresas (PyMEs) desarrollando productos innovadores. Asimismo, si bien el documento está pensado principalmente para orientar en el desarrollo e implementación de soluciones basadas en IA, también es útil para aquellas soluciones basadas en el manejo de datos (Rosales, Buenadicha & Narita, 2021). Un punto importante de la herramienta es que fue cocreada con emprendedores y está siendo probada dentro de los ecosistemas de emprendimiento en Iberoamérica.

En conclusión, tanto el WEF desde una visión primordialmente eurocentrista, como el Grupo BID con un enfoque en LAC, son organismos internacionales que han sido pioneros en la generación de conocimiento, herramientas, iniciativas y proyectos que buscan operativizar el URT en cercana colaboración con empresas globales de tecnología y con el ecosistema emprendedor de tecnología en la región.

4.

Oportunidades y desafíos para la política pública en América Latina

La generación de marcos éticos y de DDHH para el URT en América Latina es una tarea pendiente. Los marcos propuestos por el WEF responden a un contexto general de las empresas globales y organizaciones que los desarrollan en conjunto, por lo que no toman en cuenta la especificidad de los contextos y necesidades de la región LAC. En ese sentido, si bien nos encontramos ante la oportunidad de utilizar los marcos y herramientas existentes, y documentar sus usos en las empresas globales con presencia en la región, es necesario abrir la conversación hacia el ecosistema emprendedor de América Latina, de forma que las empresas emergentes puedan participar en la construcción de modelos de gobernanza ética y de DDHH en sus iniciativas y a su vez, informen un modelo híbrido para la región, que distinga efectivamente entre los modelos, otorgando claridad y discrecionalidad en su aplicación para el URT.

4.1. Modelos de gobernanza para América Latina y el Caribe

Uno de los principales desafíos para la aplicabilidad de los modelos de gobernanza basados en la ética para el URT en la región es la alta tasa de desigualdad, ya que LAC es la región más desigual del mundo (ONU, 2020). Por lo tanto, no basta con mirar hacia las brechas entre el Norte y Sur global en cuanto al desarrollo de tecnología, sino que también es necesario mirar hacia adentro. El desarrollo de IA en LAC debe hacerse desde el reconocimiento de las desigualdades de nuestra región y desde la posibilidad de contar con estándares normativos que pueden adaptarse, en la forma de manuales y otras herramientas, al diseño y desarrollo de la IA. En este sentido, como ya hemos mencionado anteriormente, los modelos basados en derechos humanos pueden facilitar el camino hacia lograr estos objetivos.

Además, aunque ha habido avances importantes, hacer que la protección de los DDHH sea vinculante para las empresas, de la misma forma en como lo es para los estados, sigue siendo un desafío (De Asís, 2020). Como mencionamos anteriormente, a nivel del sistema universal de derechos humanos, aunque contamos con los principios de la ONU sobre empresas y derechos humanos, este es un instrumento de *soft law*, es decir, no tiene carácter vinculante, como los tratados. De forma similar, en el contexto latinoamericano y caribeño, aunque se reconoce la necesidad de que las empresas asuman los riesgos y el impacto que sus operaciones tienen en los DDHH de las personas, el Sistema Interamericano de Derechos Humanos aún plantea que los estados son los últimos garantes y que deben ejercer su rol regulatorio y fiscalizador para asegurar que las empresas respeten estas obligaciones (CIDH, 2019).

Sin embargo, podemos identificar algunos indicios esperanzadores tanto a nivel regional y universal. En la tradición latinoamericana de DDHH, existe el concepto de “debida diligencia”, para referirse a toda acción que los Estados pueden y deben adoptar a fin de evitar que ocurran

violaciones de Derechos Humanos (CIDH, 2019). En los más recientes análisis de la Comisión Interamericana de Derechos Humanos sobre empresas y Derechos Humanos (2019), este concepto se ha utilizado para establecer qué rol que deben jugar los DDHH en las prácticas empresariales, y en particular, en evaluar y prevenir o mitigar posibles daños a estos causados por sus actividades. Esta obligación es particularmente relevante en el campo de la IA, porque va más allá de evitar el uso indebido o inadecuado de sistemas de IA en el campo público, ya que implica, también, la creación de mecanismos que permitan fiscalizar la actuación de actores no estatales *vis-à-vis* titulares de derechos (CIDH, 2019).

Sin embargo, hasta ahora y a través de toda la región, siguen manteniéndose prescripciones reguladoras prohibitivas, que no responden a dudas sobre qué sí pueden hacer las empresas que quisieran embarcarse en un proceso de desarrollo responsable, de acciones positivas, para incorporar principios de protección en el diseño y uso de sistemas de IA. Ahora bien, si retomamos el concepto de “debida diligencia” de cara al ciclo de diseño y uso de IA, planteamos que además de estas pautas prohibitivas, las empresas involucradas también pueden considerar, de antemano, las implicaciones que el desarrollo de tecnología podría tener con relación a derechos específicos y tomar acciones para reducir ese riesgo.

Todo indica que hacia esta dirección apunta el sistema universal de protección de DDHH. Desde la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), en julio de 2021, se avanzó en el desarrollo de una recomendación para el uso ético de la IA desde la promoción de los DDHH, un importante paso para sus estados miembro (UNESCO, 2021). En septiembre de 2021, la Alta Comisionada de las Naciones Unidas para los Derechos Humanos, en un revolucionario informe, reiteró el argumento de que *tanto* los estados *como* las empresas deben garantizar que se lleve a cabo una “diligencia debida” no *solo* desde el desarrollo, despliegue y operación, pero incluso —y esta es una de sus novedades— reconoce el rol de los estados no solo como reguladores, sino como “clientes”, planteando que deben ejercer la debida diligencia también al evaluar la adquisición de tecnología (ONU, 2021).

Asimismo, propone que se considere una “moratoria” en el desarrollo y uso de IA cuyos riesgos para los derechos humanos representen un serio desafío, por lo menos hasta que haya garantías efectivas para mitigar o prevenir los posibles daños (ONU, 2021). Si bien todavía no habla de una obligación vinculante para empresas a nivel internacional, definitivamente refuerza el rol de los estados a nivel doméstico para asegurar que las empresas desarrolladoras de tecnología adopten modelos de DDHH para el URT (ONU, 2021). Si bien es cierto que no todas las entidades que participan en el proceso de diseño y uso de la IA tienen, hoy en día, las capacidades de incorporar criterios de uso responsable basados en un modelo u otro desde el inicio, no es menos cierto que, visto desde ambos modelos, es imprescindible construir herramientas que apunten hacia ese objetivo.

Priorizar un modelo basado en DDHH permitiría un uso responsable de la IA desde su inicio, reduciendo a largo plazo las posibilidades de incurrir en mayores inversiones para corregir sesgos o, en el peor escenario, pérdidas por falta de uso. Un ejemplo concreto viene desde la intersección entre el uso de tecnología y los derechos a la privacidad y a la libertad de expresión, en particular, en el campo de las redes sociales. Algunos autores han sugerido acercarnos a un modelo de gobernanza basado en los DDHH y sustentado por la responsabilidad de las empre-

sas (Kaye, 2019). En particular, Kaye propone la figura de los “consejos multiactor” como órganos descentralizados, incluyendo a personas usuarias y activistas locales en su composición, y que utilicen como fundamento los estándares de DDHH, de forma que las compañías cuenten con un “lenguaje para articular sus posturas de forma internacional, en formas que respeten las normas democráticas y contrarresten las posturas autoritarias” (Kaye, 2019). Hoy en día, parece ser que empresas como Facebook, con su Consejo de Asesores de Contenidos, compuesto por expertos en Derecho y otras áreas, parecen haber acogido propuestas como esta (Botero, 2020).

En conclusión, es ideal poder combinar ambos modelos, tanto los basados en la ética como en los derechos humanos. Sin embargo, en un contexto de amplia desigualdad como el de LAC, priorizar modelos basados en derechos humanos permitiría seguir criterios no *solo* aplicables a toda la región, sino a otras partes del mundo; permitiría mayor intercambio y construcción de conocimientos a través de fronteras, con las consideraciones de lugar para el contexto de LAC más allá de principios de bienestar o maleficencia. En este sentido, las autoras plantean que los modelos basados en derechos humanos ofrecen mayor homogeneidad y prevención del daño, mientras que los modelos basados en la ética facilitan la toma de decisiones donde varias opciones pudieran ser viables, particularmente en contextos donde se debe priorizar la protección de valores democráticos o de igualdad.

Es aquí donde entra el valor de documentar pilotos y casos de uso que puedan convertirse en buenas prácticas para el uso responsable de la IA. Los modelos basados en la ética y en los derechos humanos proponen criterios que ayudan a discernir las acciones a no tomar o a dejar pasar. El qué sí hacer, el cómo llevar a la práctica esos principios (obligaciones positivas, en lenguaje de DDHH; o mandatos morales, en lenguaje de la ética) es un campo que sigue en desarrollo y, para poder aportar a ello, es necesario mirar a lo que se ha hecho. En la próxima subsección, y desde esta perspectiva, se presentan algunos ejemplos de buenas prácticas en nuestra región.

4.2. Buenas prácticas regionales en el URT

Como se ha mencionado con anterioridad, si bien existen diferencias entre países del Norte y Sur globales en cuanto a desarrollo de tecnología, LAC ha experimentado un crecimiento exponencial en el ecosistema de inteligencia artificial en los últimos años. Es aquí que nos preguntamos, ¿cómo se han incorporado principios de URT en el marco de dicho desarrollo tecnológico? Apostando al desarrollo de conocimiento regional, a continuación se comparten algunos ejemplos que han sido evidenciados en reportes en torno al impacto de la IA en el emprendimiento, e iniciativas implementadas desde un enfoque de colaboración multisectorial (Open Loop, liderada por Facebook) y por *start-ups* en México que han formado parte de programas y herramientas puestas a disposición por el BID en el marco de fAir LAC.

4.2.1. Reportes: el impacto de la inteligencia artificial en emprendimiento

Las ediciones 2018 y 2020 de los reportes realizados por la colaboración entre Endeavor y Everis, compañía de NTT Data, profundizan sobre la evolución de las características de las empresas emergentes que aplican sistemas de IA en sus procesos productivos o que ofrecen la tecnología como propuesta comercial en América Latina. En 2020, el análisis indicó que el *hub* de emprendimiento de IA emplea a más de 38,000 personas en 490 empresas de la región; este número representa un incremento de más del doble de las empresas analizadas en la edición del año 2018 del mismo reporte (Everis, 2018 & Everis, 2020).

De igual forma, la versión más reciente incluye un apartado que describe las principales preocupaciones de las empresas que utilizan IA. El dilema más mencionado fue “encontrar una solución a medio camino entre la automatización e intervención humana” (p. 60), explicado a través de aplicaciones de IA que utilizan *chatbots*, *voicebots* y otras tareas de automatización para interactuar con las personas. En el área de ética de la tecnología y URT, este dilema responde al derecho de las personas, como usuarias, de saber que están interactuando con un sistema automatizado, aunque este pueda parecer humano; y a la dificultad de eliminar interacciones humanas indeseables o abusivas en la comunicación con sistemas de IA.

Se mencionaron otras preocupaciones en menor medida, como sesgos algorítmicos, protección y portabilidad de datos, privacidad, y necesidad de generar confianza en las personas usuarias (Everis, 2020); se destaca que “no se aprecia una visión más profunda de las implicaciones éticas de la IA” (p.66). Todas las cuestiones mencionadas por las personas emprendedoras como dilemas del uso de la IA, se encuentran relacionadas con el *uso responsable de tecnología* y los modelos previamente descritos los retoman de forma interactiva en diferentes etapas del desarrollo de sistemas de IA, para traer conciencia sobre el impacto que tienen en el producto o servicio desplegado. Dado el estado actual del ecosistema de IA en América Latina, en crecimiento y expansión, es de esperar que los modelos basados en ética y DDHH de URT no hayan sido explorados a profundidad. Sin embargo, es una oportunidad para que el joven ecosistema emprendedor genere sus propias herramientas de URT y estas se vuelvan ubicuas al desarrollo de tecnologías de IA en la región.

4.2.2. Open Loop México

La iniciativa Open Loop, liderada por Facebook e implementada de forma global, cuenta con un Prototipo de Política Pública en México (el primero en su tipo en la región), que se enfoca en la transparencia y explicabilidad (T&E) de sistemas de inteligencia artificial. Este esfuerzo es liderado por el Laboratorio de Resiliencia EON de C Minds, el BID por medio de fAir LAC y cuenta con apoyo del Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos (INAI). En ella, once *start-ups* implementaron prototipos de transparencia y explicabilidad, que sentarán las bases para desarrollar recomendaciones de política pública en el contexto mexicano (Facebook, C Minds & BID, 2021).

El tema central del programa es el fortalecimiento de la T&E de los sistemas de IA de las empresas participantes. La T&E son parte de los principios de uso y aplicación ética de la IA, los cuales

permiten en primer lugar, que los usuarios sean conscientes de que están interactuando con un sistema de IA y además, les ayude a entender el proceso, propósito, capacidades y limitaciones de éste. Asimismo, el marco normativo probado dentro del prototipo de política pública tiene en el centro de su creación salvaguardar los derechos de privacidad y de protección de datos personales, así como la importancia de la selección de datos de entrenamiento, los algoritmos y el modelo en sí para garantizar su aplicación confiable y transparente para reducir los impactos negativos en los derechos y libertades de las personas.

Para dicho propósito, desde C Minds se crearon “misiones” de entre una y dos semanas para que las empresas participantes realizaran un estudio interno, análisis o implementación y después compartieran sus aprendizajes en formularios con preguntas definidas por C Minds y Facebook. Esto permitió analizar la efectividad de la política recomendada y generar recomendaciones para el INAI. Es importante mencionar que las empresas que probaron el marco no conocían el tema de T&E; un aprendizaje importante fue que, de forma general, no tenían mucho conocimiento sobre el tema de ética de la IA, por lo que fue necesario rediseñar el programa para cubrir esa base y después hacer el enfoque en T&E. La respuesta de las empresas fue positiva, ya que compartieron numerosos aprendizajes que están implementando en sus sistemas.

Es necesario replicar iniciativas multiactor como esta para crear marcos que encaminan el uso de los sistemas de IA de forma responsable, sobre todo con las personas encargadas del desarrollo de la tecnología. Esto es porque se identificó una falta de comunicación entre desarrolladores y las empresas que utilizan los sistemas de IA. Por otro lado, dentro de los aprendizajes del programa, sus implementadores encontraron que una vez que una empresa participa en este tipo de programa, no hay programas más especializados para continuar apoyándoles.

4.2.3. *Start-ups* en México

A continuación se comparten dos experiencias impulsadas desde el sector emprendedor radicado en México, a partir de conversaciones con sus liderazgos y documentación de ejemplos, respecto a su participación en programas desarrollados por el Grupo BID en el marco de fAIr LAC.

helKi

helKi es una empresa mexicana que nace con el sueño de democratizar la información y mejores prácticas del desarrollo infantil temprano a todas las familias y cuidadores. Para lograr esta visión, han desarrollado una serie de principios inspirados en diferentes organizaciones líderes en impacto social y vanguardia tecnológica a nivel mundial. A fin de operativizar los principios a los que se han comprometido, han formado parte de PASIA, un programa de aceleración de start-ups en inteligencia artificial en el marco de fAIr LAC Jalisco; además, son una de las once empresas que participaron en la iniciativa de Open Loop México y son parte del programa de fAIr LAC Entrepreneurial Journey del Grupo BID.

Para Irene Velasco, CEO y fundadora de la empresa, la Herramienta de Autoevaluación del fAIR LAC Entrepreneurial Journey permite que las empresas en etapa temprana enfoquen los algoritmos hacia el impacto, mitigando riesgos e informando decisiones. Por medio de la autoevaluación, fueron capaces de ampliar la conciencia en la organización sobre las consecuencias negativas de acciones de minimización de impacto sin perspectiva ética; generar un plan de acción para definir temas críticos a resolver a la par del crecimiento de helKI y crear conversaciones abiertas sobre ética de las organizaciones, con aliados, inversionistas y clientes, tomando el rol de influenciadores de la ética.

Entre los principales hitos que han alcanzado a partir de su participación en los distintos programas antes mencionados, han logrado:

- Definir procesos para delimitar el acceso y garantizar la seguridad de la información.
- Cambios en la redacción de información para evitar sesgos injustos.
- Creación de una teoría de cambio donde se deja descrito el riesgo tecnológico para no perderlo de vista en la medición de impacto.
- Comunicar al equipo y vivir como cultura organizacional los valores que guían a la empresa y tener en el centro al beneficiario en la toma de decisiones.

OS City

OS City es una start-up enfocada en la creación de servicios digitales centrados en la ciudadanía. Es parte del portafolio de UNICEF, CAF y Ethereum Foundation. Cuentan alianzas formales con el BID a través de su red de blockchain denominada LACChain y con múltiples comunidades internacionales, como RadicalXChange y Singularity University.

Jesús Cepeda, cofundador de la compañía, ha destacado la utilidad de la herramienta de autoevaluación del fAIR LAC Entrepreneurial Journey del Grupo BID para atender el desconocimiento de sesgos y su mitigación en los modelos que desarrollan, mientras que se ponen en la mesa los riesgos políticos y se aborda la perspectiva de participación y democracia. Si bien no cuentan con un marco de referencia definido formalmente al interior de la empresa, existe apego a los principios y valores de comunidades internacionales como los Young Global Leaders del WEF y el Consejo de Inteligencia Artificial Ética del Grupo BID.

Asimismo, considera que uno de los principales retos para avanzar el URT en la región está relacionado con educar al cliente sobre la importancia y el valor de la explicabilidad y el uso responsable de las herramientas que desarrollan.

5.

Conclusiones / desafíos

Las tendencias en la creación de marcos conceptuales basados en principios éticos para el desarrollo de tecnología apuntan hacia la privacidad, responsabilidad, seguridad y protección, transparencia y explicabilidad, equidad y no discriminación, control humano de la tecnología, responsabilidad profesional y promoción de los valores humanos como ejes temáticos que guían la conversación (Fjeld & Nagy, 2020). Sin embargo, aunque existen buenas intenciones por parte de la comunidad desarrolladora de sistemas de IA, la existencia de pocos insumos y mecanismos de operativización y acompañamiento de los marcos éticos ha representado un freno en su adopción. Por ello, el Foro Económico Mundial creó el grupo de trabajo sobre URT, con el objetivo de realizar investigación temática para poner herramientas y metodologías innovadoras a disposición de organizaciones y empresas, de forma que puedan transformar sus estrategias a través de la combinación de perspectivas éticas y de DDHH en cada fase del ciclo de vida de productos y servicios tecnológicos. En tanto que el ecosistema de IA en América Latina se encuentra en construcción, se genera un espacio para la reflexión sobre las prioridades regionales en el URT, el contexto de su implementación y la necesidad de generar marcos de trabajo que exploren cómo combinar las perspectivas de ética y DDHH para maximizar el beneficio a las personas usuarias.

A partir de la investigación realizada, las autoras identifican algunos indicios de próximos pasos o puntos a profundizar a través de mayor investigación. Uno de esos puntos es que se han detectado oportunidades y desafíos para el ecosistema tecnológico en LAC para incorporar las perspectivas éticas y de DDHH en la cultura organizacional y la creación de nuevos productos y servicios tecnológicos. Con este trabajo, esperamos sentar algunas bases para promover mayor intercambio e investigación sobre cómo diseñar e implementar un modelo híbrido de ética y derechos humanos desde LAC, reconociendo los beneficios que este ejercicio podría plantear, como hemos señalado.

Asimismo, desde el panorama global hacia la región LAC, existe la oportunidad de que las empresas de tecnología que participan en el grupo de trabajo del WEF prueben las herramientas que han desarrollado e implementado en el Norte global en el contexto latinoamericano, por medio de sus operaciones en los países de la región. Por su parte, los emprendimientos tecnológicos en crecimiento en la región de LAC se beneficiarán de la incorporación de las perspectivas ética y de protección a los DDHH como elemento central a la cadena de valor, aplicando las herramientas exploradas en este artículo, generando estrategias propias y documentando los resultados para robustecer el ecosistema. Hay oportunidad de promover mayor intercambio entre estos actores para avanzar hacia la incorporación de dichas herramientas en LAC.

Para lograr estos objetivos, será necesario tomar acción sobre el desconocimiento por parte de las empresas y *start-ups* en la región sobre la diferencia entre modelos basados en ética y en DDHH. La aproximación desde los programas que se ofrecen desde organismos internacio-

nales como el BID no hace distinción entre ambos, por lo que *solo* se hace un planteamiento de perspectiva ética. En respuesta a este reto, la sistematización y documentación de experiencias en la región desde el sector privado (*start-ups*), como se hizo con los ejemplos de México, ofrece una oportunidad para detectar líneas de trabajo existentes, así como prioridades del ecosistema emprendedor, tendencias y áreas de oportunidad. El cuestionario en línea como parte del estudio para conocer la adopción de marcos éticos en el desarrollo de algoritmos en empresas y *start-ups* en América Latina que se está llevando a cabo desde la propia iniciativa GuIA por parte del Centro Latam Digital y CETyS de la Universidad de San Andrés, con el apoyo de fAIR LAC, es un primer paso hacia este propósito.

A partir de un conocimiento en profundidad de los principios rectores de las perspectivas éticas y de DDHH en las iniciativas existentes en materia de URT de la región LAC, será más viable generar un modelo híbrido y a la medida para la región, que contribuya a dar forma a las responsabilidades de los estados, la iniciativa privada, la sociedad civil y academia. Este modelo permitiría relacionar ambas perspectivas con las áreas específicas de incidencia de los actores involucrados, de forma que el contexto regional y las necesidades detectadas guíen su uso. El Observatorio de Inteligencia Artificial, puesto en marcha por el BID junto con la Oficina Regional de Ciencias de la UNESCO para América Latina y el Caribe tiene el potencial de convertirse en un espacio de transversalización de políticas públicas comprometidas con el desarrollo sostenible y el *uso responsable de tecnología* (ONU México, 2021).

Finalmente, el estudio de ambas perspectivas en la iniciativa privada extiende la oportunidad para explorar cómo llevar estas herramientas al sector público como buenas prácticas que podrían implementarse para hacer frente a los retos que conlleva la falta de operativización de ambos modelos (ético y DDHH) y generar una discusión más amplia sobre las responsabilidades de los gobiernos nacionales, así como la necesidad de desarrollar alianzas y acuerdos regionales en la toma de decisiones ética y la protección a los DDHH con relación al desarrollo de productos y servicios basados en tecnologías disruptivas como la inteligencia artificial.

6.

Agradecimientos

Agradecemos a las personas que contribuyeron con insumos, ideas y ejemplos de buenas prácticas internacionales y regionales para la elaboración de este reporte.

Nombre	Institución
César Rosales	BID Lab (organismo internacional)
Tetsuro Narita	BID Lab (organismo internacional)
Irene Velasco	Helki (startup)
Claudia Del Pozo	C Minds (Action Tank)
Jesús Cepeda	OS City (startup)
Daniel Lim	Salesforce (Fellow a cargo del Grupo Responsable de Tecnología del WEF en el periodo 2018-2021)
Emily Ratte	WEF (grupo de trabajo uso responsable de tecnología)

7.

Referencias

Algorithm Watch. (2020). *In the realm of paper tigers – exploring the failings of AI ethics guidelines*. Disponible en: <https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/>

Amnistía Internacional & Access Now (2018). *The Toronto Declaration*. Disponible en: <https://www.torontodeclaration.org/>

Ballard, **et al.** (2019). *Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology*. Association for Computing Machinery. Disponible en <https://dl.acm.org/doi/10.1145/3322276.3323697>

Banco Interamericano de Desarrollo (s.f). *fAIR LAC: Quiénes somos*. Disponible en <https://fairlac.iadb.org/es/quienes-somos>

Botero, C. (2013). *Relatoría Especial para la Libertad de Expresión*. Comisión Interamericana de Derechos Humanos. Disponible en: http://www.oas.org/es/cidh/expresion/docs/informes/2014_04_08_Internet_WEB.pdf

Botero, C., **et al.** (2020). *¿Quién decide cuáles publicaciones se eliminan de Facebook? A partir de ahora, nosotros*. **New York Times**. Disponible en <https://www.nytimes.com/es/2020/05/06/espanol/opinion/facebook-junta-supervision.html>

Burt, A. (2020). *Ethical Frameworks for AI Aren't Enough*. Harvard Business Review. Disponible en <https://hbr.org/2020/11/ethical-frameworks-for-ai-arent-enough>

Business for Social Responsibility & World Economic Forum (2019). *White Paper. Responsible Use of Technology*. Disponible en: http://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology.pdf

Comisión Interamericana de Derechos Humanos, (2019). *Empresas y Derechos Humanos: Estándares Interamericanos*. Disponible: <http://www.oas.org/es/cidh/informes/pdfs/EmpresasDDHH.pdf>

Dawson, P. (2019). *Closing the Human Rights Gap in AI Governance*. Element AI. <https://s3.amazonaws.com/element-ai-website-bucket/whitepaper-closing-the-human-rights-gap-in-ai-governance.pdf>

De Asís, R. (2020). *Inteligencia Artificial y Derechos Humanos*. *Materiales de Filosofía del Derecho*. Universidad Carlos III de Madrid. Disponible en <https://e-archivo.uc3m.es/bitstream/handle/10016/30453/WF-20-04.pdf?sequence=1&isAllowed=y>

Dvoskin, B. (2019). *Why International Human Rights Law Cannot Replace Content Moderation*. Berkman Klein Center for Internet & Society at Harvard University. Disponible en <https://medium.com/berkman-klein-center/why-international-human-rights-law-cannot-replace-content-moderation-d3fc8d-d4344c>

Deloitte (2021). *Tech Trends Report*. Disponible en: https://www2.deloitte.com/content/dam/insights/articles/6730_TT-Landing-page/DI_2021-Tech-Trends.pdf

Dignum, V. (2019). *Responsible Artificial Intelligence*. In *Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-30371-6>

Elsayed-Ali, S. (2018). *New Human Rights principles on artificial intelligence*. *Open Global Rights*. Disponible en <https://www.openglobalrights.org/new-human-rights-principles-on-artificial-intelligence/?lang=English>

Epley & Tannenbaum (2018). *Treating ethics as a design problem*. *Behavioral Science & Policy*. 3(2). Disponible en <https://behavioralpolicy.org/articles/treating-ethics-as-a-design-problem/>

Ethisphere (2021). *Ethisphere Ethics Index*. Disponible en https://ethisphere.com/2021-wme-announcement/?_hstc=146412722.1624468978657.1624468978657.1624468978657.1&_hssc=146412722.3.1624468978658&_hsfp=1162251756

Everis (2021). *El impacto de la IA en el emprendimiento en América Latina*. Disponible en: <https://www.diariosustentable.com/wp-content/uploads/2021/02/ESTUDIO.pdf>

Facebook, C Minds, BID (2021). *Transparencia y explicabilidad de los sistemas de IA: Prototipado de Políticas Públicas en México*. Disponible en: <https://es.cminds.co/prototipo-politica-ia>

Fjeld & Nagy (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center for Internet and Society at Harvard University. Disponible en: <https://cyber.harvard.edu/publication/2020/principled-ai>

ForumEurope (2021). *Envision AI Workshop (short version)*. Disponible en <https://datasciencelawforum.eu/netherlands/>

Gómez Mont *et al.* (2020). *La Inteligencia Artificial al Servicio del Bien Social en América Latina y el Caribe: Panorámica Regional e Instantáneas de Doce Países*. fAir LAC - Banco Interamericano de Desarrollo. Disponible en: <https://publications.iadb.org/es/la-inteligencia-artificial-al-servicio-del-bien-social-en-america-latina-y-el-caribe-panor%C3%A1mica-regional-e-instant%C3%A1neas-de-doce-paises>

Kane, *et al.* (2019). *Accelerating Digital Innovation Inside and Out: Agile Teams, Ecosystems and Ethics*. MIT Sloan Management Review. Disponible en https://sloanreview.mit.edu/projects/accelerating-digital-innovation-inside-and-out/?gclid=CjwKCAjwz_WGBhA1EiwAUAXIcdBVmk5CsScKkeVr6QbP6KmZtIDjx-1Vo2i-Ju7wVtlxTmXdGMPCMWxoCW0UQAvD_BwE

Keller, D. (2019). *Dolphins in the Net: Internet Content Filters and the Advocate General's Glawischnig-Piesczek v. Facebook Ireland Opinion*. Disponible: <https://cyberlaw.stanford.edu/files/Dolphins-in-the-Net-AG-Analysis.pdf>

Kaye, D. (2019). *A New Constitution for Content Moderation*. *One Zero*. Disponible en: <https://onezero.medium.com/a-new-constitution-for-content-moderation-6249af611bd>

Latonero, M. (2018). *Governing Artificial Intelligence: Upholding Human Rights & Dignity*. Data Society. Disponible en <https://datasociety.net/library/governing-artificial-intelligence/>

Lara Gálvez, J. C. (2020). *Perspectivas de Derechos Humanos en la implementación de Marcos Éticos para la Inteligencia Artificial*. En C. Aguerre, (Ed.). *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*. Buenos Aires: CETyS Universidad de San Andrés. Disponible en: <https://guia.ai/wp-content/uploads/2020/05/Lara-Perspectivas-de-derechos-humanos-en-la-implementacion-de-marcos-eticos-para-la-inteligencia-artificial.pdf>

López Calva & Melguizo (2021). *El impulso para digitalizar bien a Latinoamérica*. *El País*. Disponible en: <https://elpais.com/planeta-futuro/2021-02-18/el-impulso-para-digitalizar-bien-a-america-latina.html>

Mac-Gregor, E. (2017). *Acerca de la discriminación cultural histórica en razón de la posición económica de los trabajadores sometidos a trabajo esclavo*. Instituto de Estudios Constitucionales del Estado de Querétaro. <https://archivos.juridicas.unam.mx/www/bjv/libros/10/4745/10.pdf>

Markkula Center for Applied Ethics & WEF (2021). *Responsible Use of Technology: The Microsoft Case Study*. Disponible en: http://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_2021.pdf

Markkula Center for Applied Ethics, Deloitte & WEF (2020). *White Paper. Ethics by Design: an organizational approach to responsible use of technology*. Disponible en: http://www3.weforum.org/docs/WEF_Ethics_by_Design_2020.pdf

Markkula Center for Applied Ethics & WEF (2021). *Responsible Use of Technology: The IBM Case Study. White Paper*. https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_The_IBM_Case_Study_2021.pdf

ONU México (2021). *BID y UNESCO lanzan Observatorio de Inteligencia Artificial*. <https://www.onu.org.mx/bid-y-unesco-lanzan-observatorio-de-inteligencia-artificial/>

Rosales, Buenadicha & Narita. (2021). *Autoevaluación ética de IA para actores del Ecosistema Emprendedor*. Banco Interamericano de Desarrollo. Disponible en <https://publications.iadb.org/publications/spanish/document/Autoevaluacion-etica-de-IA-para-actores-del-ecosistema-emprendedor-Guia-de-aplicacion.pdf>

Salesforce (2020). *How Salesforce Infuses Ethics into its AI*. Disponible en <https://www.salesforce.com/news/stories/how-salesforce-infuses-ethics-into-its-ai/>

Sniderman, Bannister & Buckley (2020). *Ethical tech: Making ethics a priority in today's digital organization*. Disponible en <https://www2.deloitte.com/us/en/insights/topics/digital-transformation/make-ethical-technology-a-priority.html?id=us:2sm:3tw:4di6289:5awa:6di:MMDDYY:HANDLE:dr26:author&pkid=1006571#endnote->

United Nations (2011). *Guiding Principles on Business and Human Rights*. Disponible en: https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

UNESCO (2021). *AI Ethics: Another step closer to the adoption of UNESCO's Recommendation*. Disponible en <https://en.unesco.org/news/ai-ethics-another-step-closer-adoption-unescos-recommendation-0>

United Nations Human Rights Office of the High Commissioner (2021). Artificial intelligence risks to privacy demand urgent action. Disponible en: <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=27469&LangID=E>

Descargo de responsabilidad. Las opiniones expresadas en la publicación incumben únicamente a los/as autores/as. No tienen intención de reflejar las opiniones o perspectivas del CETyS, CLD ni de ninguna otra organización involucrada en el proyecto.