



ESTRATÉGIA BRASILEÑA DE INTELIGENCIA ARTIFICIAL (EBIA) Y POLÍTICAS PÚBLICAS: PROPUESTAS PARA EFECTIVIZACIÓN DE LOS EJES LEGISLACIÓN, REGLAMENTACIÓN Y USO ÉTICO Y GOBIERNO DE IA

Sthéfano Bruno Santos Divino¹

Resumen: Este artículo tiene como problema de investigación la siguiente pregunta: ¿cómo y qué políticas públicas pueden ser apropiadas para la efectividad e implementación de la Estrategia Brasileña de Inteligencia Artificial (EBIA) en relación a los ejes Legislación, Regulación y uso ético y Gobernanza de la Inteligencia Artificial (IA)? El objetivo es presentar la referida estrategia y sus ejes de implementación y, posteriormente, contextualizarla según la clasificación de Brasil en materia de Inteligencia Artificial en el escenario internacional. Para ello, se utiliza el Índice Global de Innovación elaborado por la Organización Mundial de la Propiedad Intelectual y el Informe del Índice de Inteligencia Artificial elaborado por la Universidad de Stanford. La metodología utilizada es la de la investigación monográfica. Al final, se concluye que el EBIA es un mecanismo capaz de insertar a Brasil en la carrera regulatoria de la IA, pero necesita de políticas públicas fuertes y efectivas dirigidas a su implementación.

Palabras clave: Estrategia Brasileña de Inteligencia Artificial; gobernanza; inteligencia artificial; políticas públicas; regulación.

1 Introducción: La Estrategia Brasileña de Inteligencia Artificial (EBIA) y sus ejes de implementación

La Inteligencia Artificial² (IA) ha sido rápidamente evolucionada y desarrollada con el objetivo de ofrecer un conjunto de beneficios del orden económico y social. En los últimos años, surgió la necesidad de adoptar de medidas reglamentarias y políticas públicas³ por el ejecutivo para implementación de la IA tanto en el segmento doméstico como en el segmento industrial. Canadá (2017), China (2017), Dinamarca (2021), Comisión Europea (2021), Finlandia (2017), Francia (2018), India (2018), Italia (2020), Japón (2018), México (2018), Países Nórdicos (2018), Singapur (2020), Corea del Sur (2019), Suecia (2018), Taiwán (2018) y Reino Unido (2020) lanzaron estrategias para promover el uso y el desarrollo de la IA (DUTTON, 2018). Aunque ninguna de las estrategias sea semejante, se enfoca en diferentes aspectos de la IA: investigación científica; desarrollo y captura de talentos; habilidades y educación; adopción en los sectores públicos y privados; ética e inclusión; patrones de

¹ Doctorando (2020 - Becario del Programa de Excelencia Académica - Proex - Capes/Taxa) y Máster (2019) en Derecho Privado por la Pontificia Universidad Católica de Minas Gerais. Bachiller en Derecho por el Centro Universitario de Lavras (2017). Profesor Adjunto del Curso de Derecho del Centro Universitario de Lavras (2020 - actual). Profesor sustituto de Derecho Privado de la Universidad Federal de Lavras (03/2019 - 03/2021). Abogado. Orcid iD: <https://orcid.org/0000-0002-9037-0405>. Lattes: <http://lattes.cnpq.br/5133514180104561>. E-mail: sthefanoadv@hotmail.com.

² Una de las definiciones más aceptadas en el ramo científico – aunque también criticada – es la de Russell y Norvig, que definen Inteligencia Artificial como “[...] *the study of agents that receive percepts from the environment and perform actions*”. (RUSSELL; NORVIG, 2010, p. VIII).

³ Sobre definiciones e historia de las políticas públicas, ver más en (SOUZA, 2002)

reglamentación; e infraestructura digital (DUTTON, 2018).

Hasta a nivel internacional, el 22 de noviembre de 2021, la cátedra de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2021) lanza el primer acuerdo Internacional sobre Ética en las aplicaciones de IA. Las recomendaciones también tienen como objetivo la protección de datos, la prohibición de la práctica social de *score* (CITRON; PASQUALE, 2014; SERASA, 2021) y vigilancia en masa, el monitoreo y evaluación de los sistemas durante su implementación y ejecución y a la protección ambiental, teniendo en cuenta la posibilidad de que una IA vuelva más eficiente el uso de la energía y otros servicios primarios.

Reconociendo que el gran aumento en el poder computacional por medio de avances prácticos en el aprendizaje de la máquina (*machine learning*⁴) permite éxitos en una gama de dominios aplicados, llamando la atención al desarrollo de políticas públicas y de empresas para unirse a la carrera por el liderazgo mundial en IA, Brasil (2021) lanza su Estrategia Brasileña de Inteligencia Artificial (EBIA)⁵ objetivando su actuación en campos como trabajo, educación, tributación, investigación, desarrollo e innovación y ética.

EBIA (BRASIL, 2021) pretende guiar el papel de las acciones estatales para el desarrollo de acciones, cualesquiera que sean sus vertientes, para estímulo a la investigación, innovación y desarrollo de soluciones en IA, así como garantizar que su uso sea consciente, ético, legal y en beneficio de un futuro mejor. EBIA fue construida en tres etapas:

- 1) Contratación de consultoría especializada en IA;
- 2) *Benchmarking* nacional e internacional;
- 3) Proceso de consulta pública.

La contratación de consultoría especializada fue realizada por el Ministerio de Ciencia, Tecnología, Innovaciones y Comunicaciones (MCTI), por medio del Proyecto de Cooperación Técnica Internacional (PRODOC) en conjunto a la UNESCO. La consulta pública fue realizada por medio de plataforma electrónica del Gobierno Federal, entre el 12 de diciembre de 2019 y el 3 de marzo de 2020, período en el cual fueron recibidas alrededor de mil contribuciones que fueron utilizadas como base a la construcción de EBIA (BRASIL, 2021).

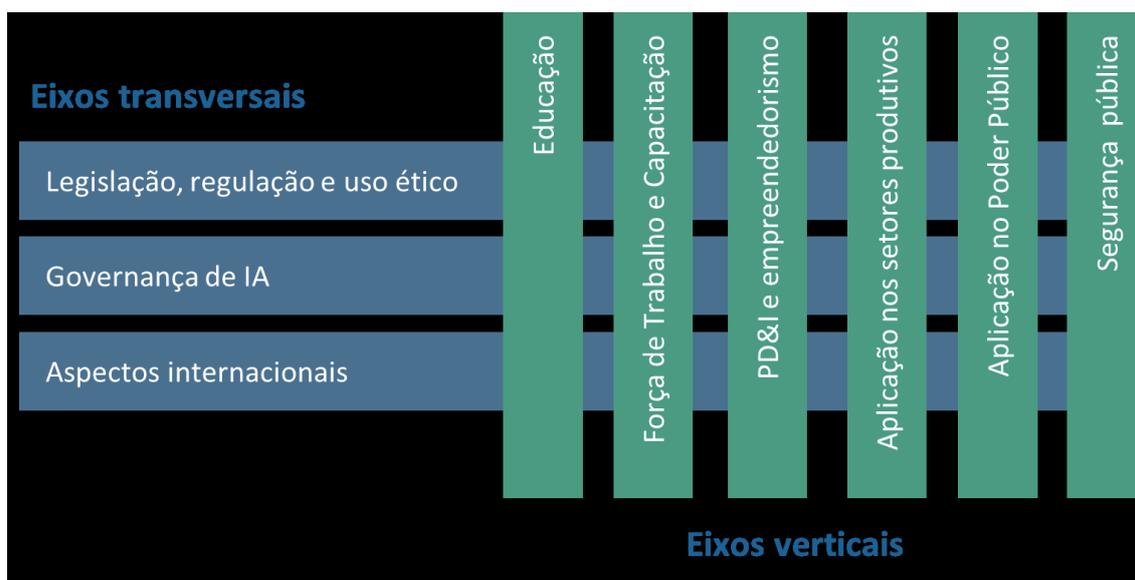
EBIA sigue las recomendaciones de la Organización para Cooperación y Desarrollo Económico (OCDE, 2019) sobre IA, destacándose, pero no solo: a) su beneficio a las personas y al planeta; b) el respeto al Estado de Derecho, a los derechos humanos, a los valores democráticos y a la diversidad; c) transparencia destinada a la comprensión general sobre los sistemas de IA; y d) funcionamiento robusto a lo largo de su ciclo de vida.

⁴ “*Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data*” (EL NAQA; MURPHY, 2015. p. 3-11).

⁵ Instituida por la Ordenanza MCTI nº 4.617, del 6 de abril de 2021, modificada por la Ordenanza MCTI nº 4.979, del 13 de julio de 2021 (Anexo).

EBIA, por lo tanto, trabaja con nueve ejes temáticos, con tres ejes transversales – 1) Legislación, reglamentación y uso ético; 2) Gobierno de IA⁶; 3) Aspectos internacionales –, y seis ejes verticales – 1) Educación; 2) Mano de obra y capacitación; 3) PDI y emprendedorismo; 4) Ampliación en el Poder Público; 5) Aplicación en los sectores productivos; y 6) Seguridad pública. Los ejes transversales funcionan como base para la construcción de los demás sectores verticales. Todos estos deben tener presentes los presupuestos de aquellos (ver Figura 1). Aunque aparenten ser independientes porque los temas son diversificados y abarcan particularmente cada sector, en el objetivo de la creación de los ejes transversales todos deben ser lo suficientemente abstractos para abarcar todos los sectores de IA.

Figura 1 – Ejes de EBIA



Fuente: Brasil (2021).

En ese sentido, se puede afirmar que los algoritmos⁷ deben ser neutros (*agnostics*) y no enfocarse específicamente en una modalidad/especie de IA o de metodología de análisis de datos. La estrategia debe ser aplicada en la constitución, aplicación y en el uso de toda y

⁶ Para Floridi, “el gobierno digital es la práctica de establecer e implementar políticas, procedimientos y patrones para el desarrollo, uso y administración adecuados de la *infosphere*. A través de la supervisión humana, se pretende garantizar que un sistema de IA no coloque en causa la autonomía humana ni produzca efectos negativos. El gobierno digital puede incluir directivas y recomendaciones que se superpongan a la reglamentación digital, pero no son idénticas a ella. Esa es solo otra manera de hablar sobre la legislación relevante, un sistema de leyes elaborado y aplicado por medio de instituciones sociales o gubernamentales para regular el comportamiento de los agentes relevantes en la *infosphere*” (FLORIDI L. 2018).

⁷ “Algoritmos son la base del proceso de desarrollo de *software* y forman parte de las herramientas por las cuales los programadores crean estrategias para fraccionar problemas en etapas y procesos que pueden ser traducidos computacionalmente. En la tecnología, hay ejemplos de todos los niveles de complejidad. Una aplicación más simple de algoritmo es el proceso de inicialización de computadora: hay un *software* – básicamente la traducción computacional de un algoritmo – encargado de testear todos los componentes de su computadora para saber si está todo en orden y, después de eso, buscar el sistema operativo en el disco para cargarlo” (GARRET, 2020).

cualquier IA. En otros términos, la tecnología es neutra y no debe enfocarse en sistemas, *softwares* o técnicas específicas, debiendo ser aplicada independientemente del desarrollo del lenguaje computacional y de las técnicas de almacenamiento de datos. La adopción de esas prácticas sirve como líneas de base para una serie de consideraciones y medidas para las organizaciones que operan en cualquier sector y adoptan concretamente la estrategia. Sectores específicos u organizaciones específicas pueden elegir adoptar consideraciones adicionales y medidas técnicas para adaptar su línea de producción de acuerdo con sus necesidades. De esa forma, la estrategia no debe enfocarse en organizaciones públicas o privadas según su tamaño o modalidad constitutiva.⁸

La propuesta para efectivización de esas pretensiones es delineada por medio de objetivos. EBIA pretende contribuir a la elaboración de principios éticos destinados al desarrollo y uso de IA más responsables. Al mismo tiempo, se presupone el equilibrio entre inversiones sostenidas en investigación y desarrollo en IA para eliminar barreras a la innovación en IA. Se pretende capacitar y formar a profesionales para el ecosistema de la IA para estimular la innovación y el desarrollo de la IA brasileña en ambiente internacional. Así, se puede promover cooperación entre entes nacionales e internacionales, públicos y privados, industria y centros de investigación para el desarrollo de la IA.

Ocurre que tales objetivos, en la coyuntura actual, no son fáciles de ser alcanzados. Este artículo tiene como **problema de investigación** la siguiente pregunta: ¿cómo y cuáles políticas públicas pueden ser adecuadas para efectivización e implementación de EBIA en cuanto a los ejes Legislación, regulación y uso ético y Gobierno de IA?

Después de contextualizarla, la primera sección es responsable por el análisis de la clasificación de Brasil en materia de IA en el escenario internacional. Se utiliza, para ello, el Índice Global de Innovación elaborado por la Organización Mundial de Propiedad Intelectual y el *Artificial Intelligence Index Report* elaborado por la Universidad de Stanford. En este punto, se realiza investigación cuantitativa en los sitios Conecta Startup y Startup Brasil, objetivando verificar la adecuación de los índices internacionales al grado de innovación en territorio nacional. Como resultado, se verifica que, de un total de 268 *startups*, predominan los sectores: agronegocio (14), industrial (10), salud y bienestar (15), educación (23), salud (17), TI y telecomunicaciones (19), comercio minorista (13), finanzas (11) y medios y comunicación (12).

En contribuciones originales, se constata que las políticas públicas destinadas al avance científico en tecnología de la información y, específicamente, en IA no están según las pretensiones de EBIA. Vale resaltar una limitación de este estudio. Según lo señalado, la investigación cualitativa demuestra los sectores de actuación como TI y telecom y *hardware*. En este sentido, se vuelve inviable verificar si las *startups* inscritas en esos ramos actúan

⁸ Críticas extraídas de SINGAPORE (2020).

específicamente en el sector de IA. O sea, el resultado obtenido puede ser más negativo si es verificado concretamente y de forma pormenorizada.

Así mismo, la segunda sección es responsable por las contribuciones al eje Legislación, regulación y uso ético, especialmente en la temática responsabilidad civil de los actos prácticos por IA. El principal resultado obtenido postula que la responsabilidad adecuada para los actos ilícitos practicados por los entes inteligentes artificialmente es la responsabilidad subjetiva, pues permite la reducción del deber indemnizatorio con base en los costos marginales de precaución. Para llegar a ese resultado, se utiliza el Análisis Económico del Derecho (AED) y una breve revisión en las teorías de la responsabilidad civil (subjetiva y objetiva) prescritas por el Código de Defensa del Consumidor (CDC)⁹ y por el Código Civil (CC).

Finalmente, en cuanto al eje Gobierno de la IA, se presentan propuestas de estructuras y medidas de gobierno interno, determinación del nivel de participación humana en la toma de decisiones con base en la IA, gestión de operaciones, interacción y comunicación de las partes interesadas, tanto en instituciones públicas como privadas, con fundamento en el *Artificial Intelligence Governance Framework* de Singapur. La metodología utilizada para ello es la de investigación monográfica. Al final, se concluye que la EBIA es un mecanismo capaz de insertar a Brasil en la carrera reglamentaria de la IA, pero necesita de fuertes y eficaces políticas públicas destinadas a su implementación.

2 Clasificación de Brasil en materia de IA en el escenario internacional

La clasificación de Brasil según índices en nivel internacional puede traer consideraciones sobre los impactos positivos y negativos que la nueva economía y la sociedad de la información traen consigo en materia de IA. En otros términos, ese análisis permite ampliar la capacidad de postular políticas públicas y delimitar cuales sectores, personas, instituciones o empresas poseen la mayor capacidad productiva o producen productos y servicios ligados a IA. Esas experiencias internacionales demuestran bajo otro punto de vista que tales procesos pueden o no estar elevando la productividad a nivel local con impacto global.

Como la innovación es un punto básico y fundamental para el desarrollo económico del país, además de estar ligada nítidamente al desarrollo de la IA por ser técnica dependiente de investigación, infraestructura y desarrollo, el primer paso es recurrir al Índice Global de Innovación referente al año 2021¹⁰ (WIPO, 2021). En la clasificación general, Brasil ocupa la posición 57 del ranking global, la mejor desde el año 2012, lo que representa un avance superior al esperado si se compara al año 2019, cuando ocupaba la 66° posición, y al año 2020, en la 62° posición. Relativo a la región de América Latina y el Caribe, solo Chile (53°), México (55°) y

⁹ Para más, ver en: (DIVINO, 2021).

¹⁰ Este índice posee 132 países evaluados, entre los cuales: Suiza; Suecia; Japón; Grecia; Rumania; Egipto; Bangladesh; Mali; Togo; etc.

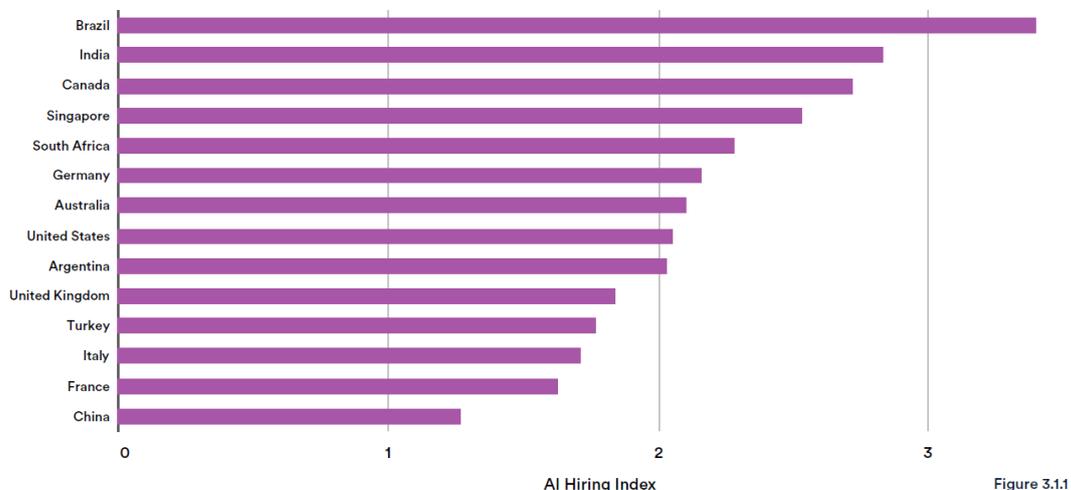
Costa Rica (56°) están al frente de Brasil (WIPO, 2021).

Con relación al aparato institucional (78°), infraestructura (69°), productos creativos (61°) y sofisticación del mercado (75°), son esos los principales puntos a ser desarrollados para los próximos años. Los recursos de capital humano destinados a la investigación (48°), sofisticación empresarial (34°) y productos de conocimiento y tecnología (51°) son puntos que deben ser también mejorados, pero que presentan buenos posicionamientos en el ranking (WIPO, 2021).

Aún en términos optimistas, según el *Artificial Intelligence Index Report* de 2021 elaborado por la Universidad de Stanford (UNITED STATES, 2021), en 2020, Brasil, en conjunto con India, Canadá, Singapur y Sudáfrica, estuvo entre los países con mayor tasa de empleabilidad en el ramo de la IA, según puede ser visto en la Figura 2.

Figura 2 – Índice de Empleabilidad en el sector de IA

AI HIRING INDEX by COUNTRY, 2020
Source: LinkedIn, 2020 | Chart: 2021 AI Index Report



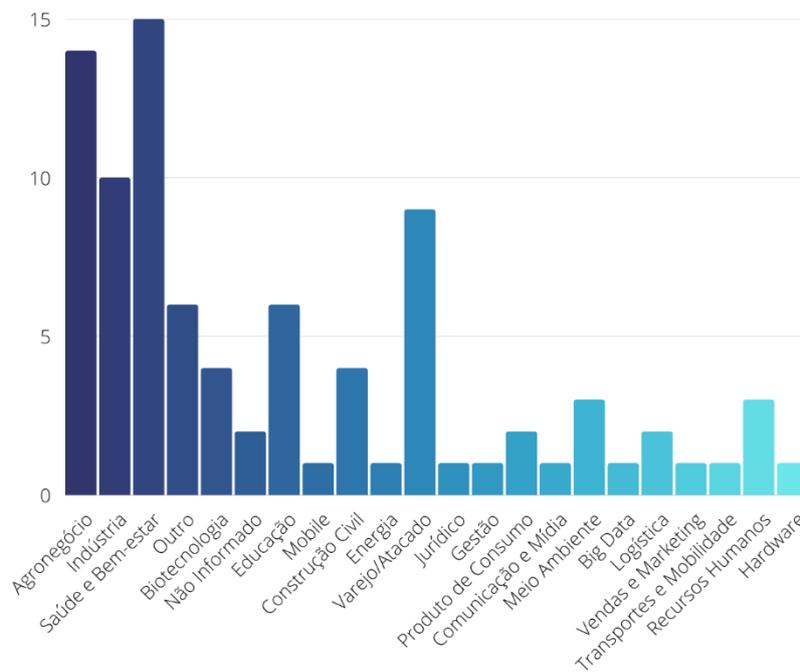
Fuente: United States (2021).

Pero el camino aún es largo. Cuando son analizados los datos aportados por la EBIA, se constata que, en 2020, Brasil (2021) poseía alrededor de doce mil *startups* activas, pero solo 26 de ellas eran relativas al sector de IA. Objetivando la actualización de los datos, utilizándose de los mismos programas gubernamentales de fomento a la innovación y emprendedorismo en nivel de *startups* presentados por la EBIA – Conecta Startup (2021) y Programa Startup Brasil (BRASIL, 2021) –, se presentan datos actualizados con respecto al año 2021.

Según el Conecta Startup, fueron identificadas 94 *startups* registradas. Como puede ser visto en las Figuras 3 y 4, predominantemente ellas ocupan el sector de agronegocio (14), industrial (10) y salud y bienestar (15), con la mayor parte del total (y no de esos sectores) localizada en el Distrito Federal (9), Rio de Janeiro (11), Santa Catarina (9), Amazonas (7) y

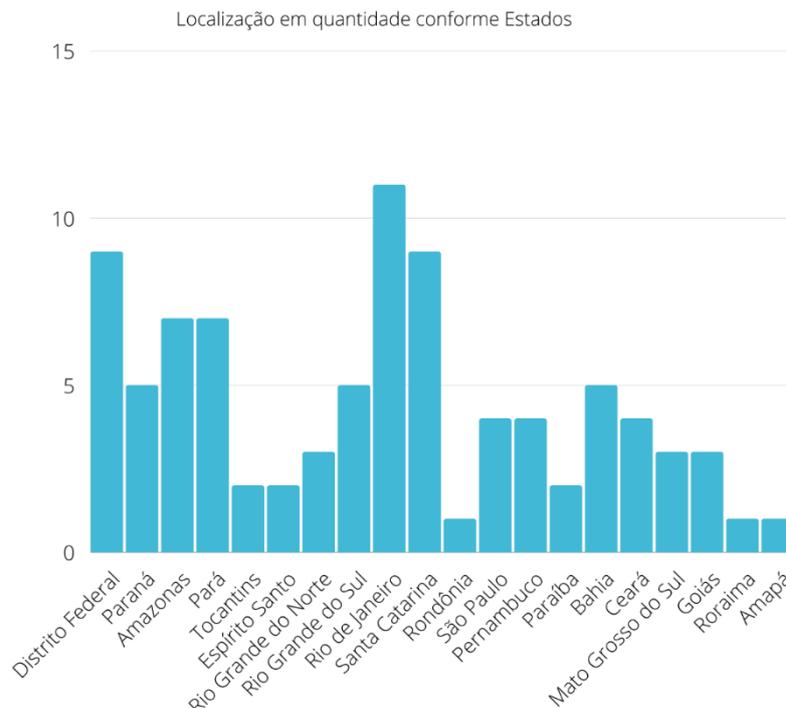
Paraná (7).

Figura 3 – Áreas de actuación (en cantidad) de las *startups* según el programa Conecta Startup



Fuente: Elaborado por el autor con base en los datos suministrados en Conecta Startup (2021).

Figura 4 – Estados de actuación (en cantidad) de las *startups* según programa Conecta Startup



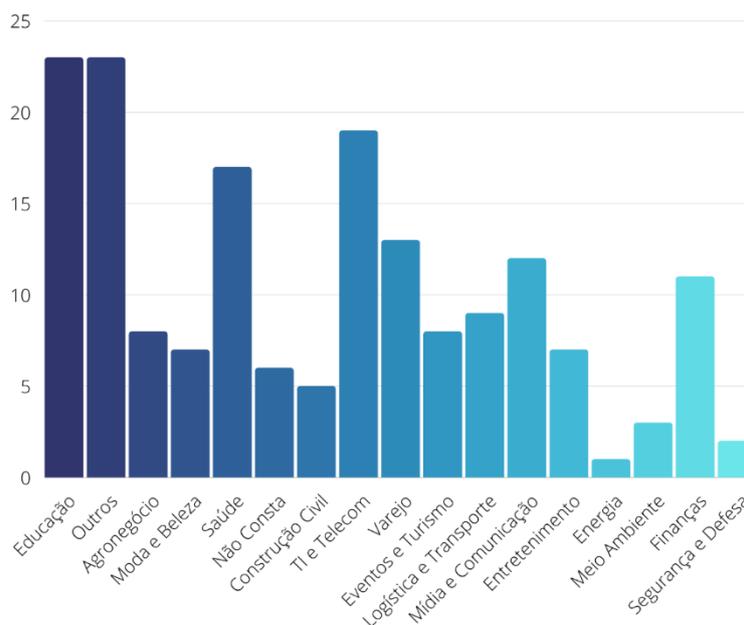
Fuente: Elaborado por el autor con base en los datos suministrados en Conecta Startup (2021).

Las figuras demuestran una serie de limitaciones y puntos que necesitan mejoras, entre ellos el mayor esfuerzo en el desarrollo de políticas públicas dirigidas al fomento de actividades ligadas a la IA teniendo en cuenta la presencia de una única *startup* ligada a la tecnología (*hardware*).¹¹

El resultado obtenido en la plataforma anterior curiosamente se mezcla cuando se analiza el Programa Startup Brasil. Los datos obtenidos demuestran la existencia de 174 *startups* registradas en su plataforma. Predominantemente ellas están en el sector de educación (23), salud (17), TI y telecomunicaciones (19), comercio minorista (13), finanzas (11) y medios y comunicación (12), estando la mayor parte de ellas concentrada en São Paulo (54), Rio de Janeiro (16), Pernambuco (12) y Minas Gerais (10). Las informaciones están disponibles en las Figuras 5 y 6.

Si bien el área de TI y telecomunicaciones es considerablemente mayor con relación al programa anterior (19 para 1), el estudio también es limitado porque no consigue verificar si las referidas *startups* actúan efectivamente en el sector de IA. Sin embargo, los resultados parciales que pueden ser enviados toman en cuenta que, por la gama del territorio y de la población brasileña, los números son ínfimos comparados con las pretensiones y resultados a largo plazo.

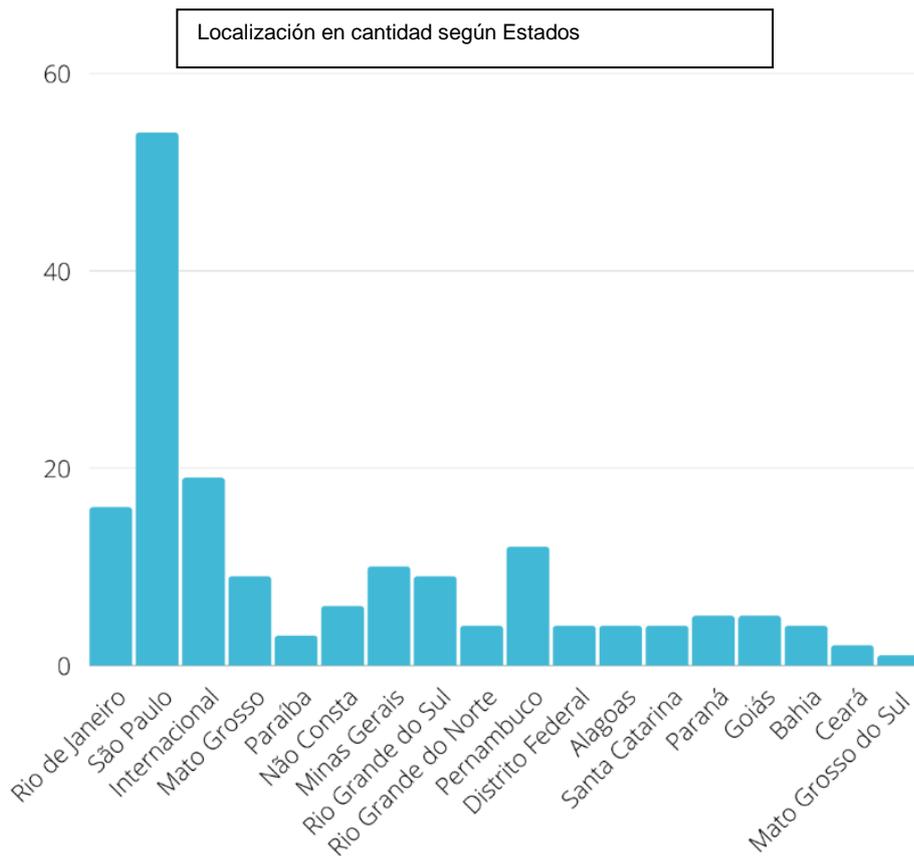
Figura 5 – Áreas de actuación (en cantidad) de las *startups* según el Programa Startup Brasil



Fuente: Elaborado por el autor con base en los datos suministrados en Startup Brasil (BRASIL, 2021).

¹¹ En este punto, el estudio es limitado por no definir expresamente si actúa en el ramo de la IA.

Figura 6 – Estados de actuación (en cantidad) de las *startups* según Programa Startup Brasil



Fuente: Elaborado por el autor con base en los datos suministrados en Startup Brasil (BRASIL, 2021).

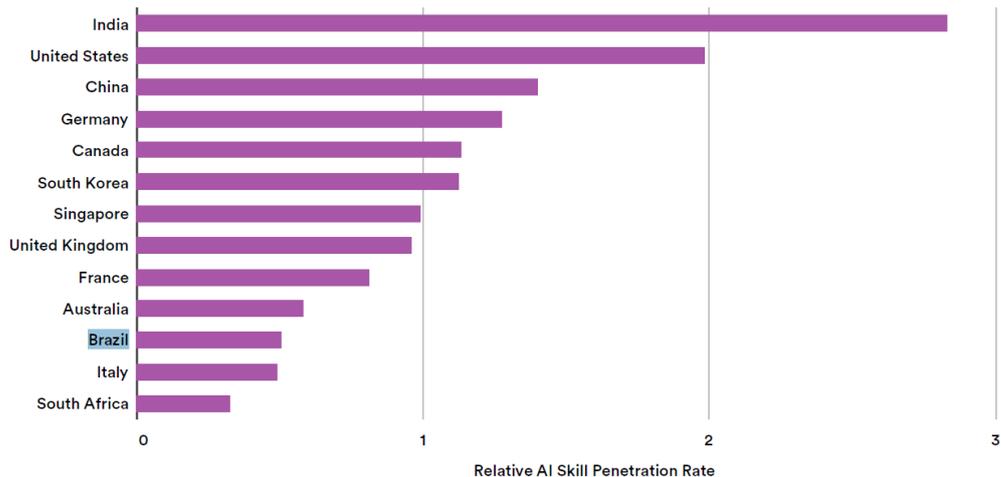
Uno de los importantes resultados obtenidos por medio de la obtención de datos es la constatación de difusión de *startups* por todo Brasil, lo que demuestra considerable potencial emprendedor en la cultura brasileña. Sin embargo, los principales problemas enfrentados por las *startups* brasileñas son la escasez de mano de obra calificada, la alta carga tributaria y la burocracia (BRASIL, 2021). El principio básico para su funcionamiento es el desarrollo de soluciones innovadoras y disruptivas para clásicos o nuevos problemas con el fin de romper estándares en relación a empresas del mismo segmento. Ocurre que esa característica necesita expresamente de una relación inventiva y creativa en que Brasil, a pesar de su continuo desarrollo, aun es precario.

Según el *Artificial Intelligence Index Report* de 2021 elaborado por la Universidad de Stanford (UNITED STATES, 2021), las habilidades en IA de los brasileños presentan baja adhesión, desarrollo, penetración y destaque, aunque aún estén en el ranking 50 en nivel global (ver Figura 7), siendo la Universidad Federal de Rio Grande do Sul referencia en producción relativa a la ética de IA, *machine learning* y conferencias sobre robótica.

Figura 7 – Índice de Habilidad en IA

RELATIVE AI SKILL PENETRATION RATE by COUNTRY, 2015-20

Source: LinkedIn, 2020 | Chart: 2021 AI Index Report

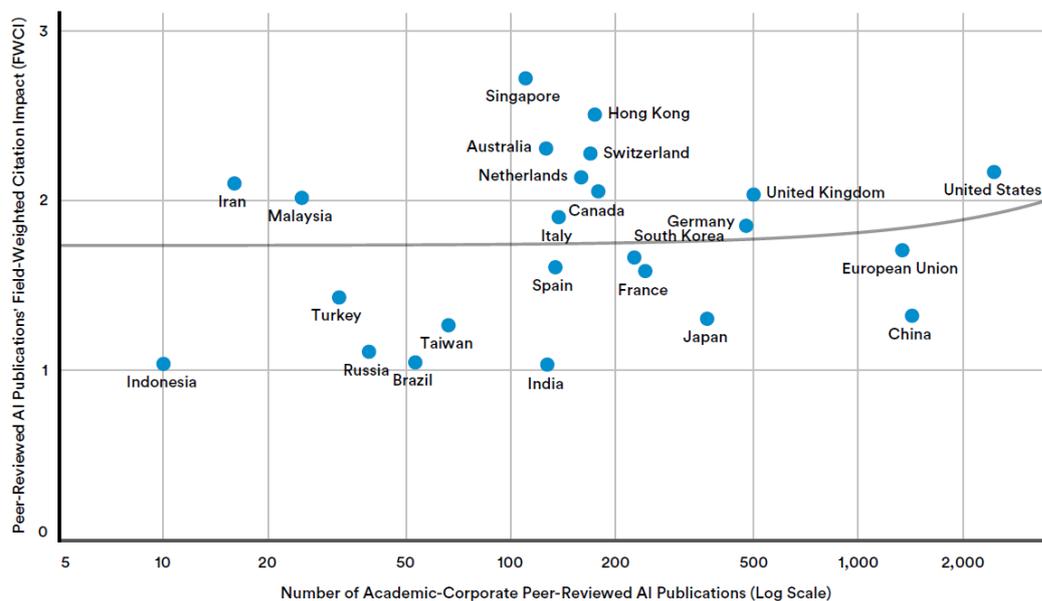


Fuente: United States (2021).

Figura 8 – Publicaciones realizadas por el sistema *double-blind review* en la temática de IA

PEER-REVIEWED AI PUBLICATIONS' FIELD-WEIGHTED CITATION IMPACT and NUMBER of ACADEMIC-CORPORATE PEER-REVIEWED AI PUBLICATIONS, 2019

Source: Elsevier/Scopus, 2020 | Chart: 2021 AI Index Report



Fuente: United States (2021).

Como la EBIA se fundamenta en los cinco principios definidos por la OCDE para una gestión responsable de los sistemas de IA, que son: 1) crecimiento inclusivo, el desarrollo sustentable y el bienestar; 2) valores centrados en el ser humano y en la equidad; 3) transparencia y explicabilidad; 4) robustez, seguridad y protección; y 5) la responsabilización o la rendición de cuentas (*accountability*), cada uno de sus ejes pueden y deben ser desarrollados por medio de políticas públicas adecuadas a sus pretensiones. Al final, los resultados anteriores

demuestran que el sector brasileño, por más prometedor y por más expectativas que existan, necesita de mejoras para posibilitar el alcance de rankings superiores y, primordialmente, mejor satisfacción de la población.

3 Eje temático 1: Legislación, regulación y uso ético

En esa área temática, las mayores preocupaciones de la EBIA están relacionadas a: 1) la protección de datos personales; 2) prevención de discriminación y sesgo algorítmico; 3) equilibrio destinado a la preservación de estructuras adecuadas de incentivo al desarrollo de la IA; 4) creación de parámetros legales destinados a la seguridad jurídica como a la responsabilidad de los diferentes actores integrantes de su cadena de producción.

Las preocupaciones 2 y 3 pueden ser tratadas inicialmente por la Ley General de Protección de Datos (LGPD) brasileña. Cuando el tratamiento de datos es realizado en su forma automatizada, podrá el titular solicitar al controlador o al operador la explicación de cómo él fue realizado, así como sus finalidades.¹² La defensa por la explicabilidad en sistemas de IA parte del supuesto que sus decisiones pueden no ser consideradas necesariamente objetivas, justas o imparciales (BECKER; FERRARI, 2018; BURELLI, 2016; DIAKOPOULOS, 2013; PASQUALE, 2015). De esa forma, la creación de parámetros para efectivizar la explicabilidad y la transparencia son mecanismos básicos que deben estar fundamentalmente vinculados para evitar sesgos en decisiones automatizadas y la discriminación en la recolección y tratamiento de datos (FJELD, 2020).

Su implementación parte del supuesto de exposición de la lógica fundamental determinante, de forma clara y objetiva, responsable por la elaboración de la decisión. Se trata de la presentación de una descripción que sea comprensible para el titular para delimitar como el operador o el controlador utilizó las técnicas de IA para llegar al resultado obtenido (DOSHI-VELEZ; KORTS, 2017). En otros términos, lo que se pretende es la verificación de los parámetros y criterios que fueron utilizados en el tratamiento automatizado y si fueron correctamente utilizados para evitar errores o discriminaciones de la decisión. En la constitución (*by design*) algorítmica se debe verificar: 1) las principales razones que llevaron a aquella decisión; 2) si el cambio, sustitución, alteración o modificación de los factores alteraría la decisión; y 3) si existen casos semejantes con diferentes decisiones etc. (DOSHI-VELEZ; KORTS, 2017)

Una de las formas de concretarse la explicabilidad en la toma de decisión automatizada es por el establecimiento de la obligatoriedad de que los responsables por el desarrollo dejen el código de su software abierto (DIAKOPOULOS, 2013). En caso esa práctica no sea posible, se

¹² Art. 20. El titular de los datos tiene derecho a solicitar la revisión de decisiones tomadas únicamente con base en el tratamiento automatizado de datos personales que afecten sus intereses, incluidas las decisiones destinadas a definir su perfil personal, profesional, de consumo y de crédito o los aspectos de su personalidad. (LGPD)

debe auditar particularmente el algoritmo por medio de una cooperación entre empresas y gobierno, objetivando la comparación del resultado con el comportamiento esperado en la constitución de la IA (SANDVIG, 2014).

La relación entre transparencia y *accountability*, por lo tanto, implica la necesidad de adopción de medidas destinadas a la comprensión de los procesos asociados a la toma de decisión automatizada para posibilitar la identificación de sesgos involucrados en el proceso decisorio. Ocurre que esa práctica puede llevarse a cabo e implementada partiendo de la idea de que los sistemas de IA deben estar centrados en el ser humano (*human-centric AI*), lo que podría facilitar su auditoria, así como la creación de sistemas más confiables (*trustworthy AI*).

Uno de los mayores desafíos destinados a la implementación de este objetivo es la falta de conocimiento técnico por parte de la población en general para comprender, interpretar y leer el código responsable por la ejecución de la IA. Una solución paliativa, que no debe ser tomada como definitiva, es el establecimiento del deber de informar por los proveedores y por los responsables acerca de las decisiones que están siendo tomadas, si son de forma automatizada, como, donde y cuando se toman (OSOBA; WELSER, 2017).

Los esfuerzos de la EBIA para alcanzar este objetivo alinean su marco regulatorio de gobierno con la creación de políticas públicas relacionadas al tema y presenta iniciativas consideradas relevantes para su implementación. La primera es la Estrategia Brasileña para Transformación Digital (e-Digital) (BRASIL, 2018a), la cual busca coordinar las diversas iniciativas gubernamentales ligadas al tema en torno de una singular visión, que sea sinérgica y coherente, para apoyar la digitalización de los procesos productivos y la capacitación para el ambiente digital, promoviendo, en conjunto, la generación de valor y crecimiento económico (BRASIL, 2018a). En total, están previstas 100 acciones, que van desde: 1) conectar 22.000 escuelas públicas, urbanas y rurales, con accesos a banda ancha de alta velocidad, en red terrestre o vía satélite, en el ámbito del Programa Educación Conectada; y 2) ampliar la participación de los centros de investigación y desarrollo en las instancias multilaterales de definición de los estándares internacionales y de bandas de radiofrecuencia a ser establecidos para la quinta generación de la telefonía móvil (5G) la optimización de políticas dirigidas al sector para expandir (más que proporcionalmente) la inversión privada en PD&I en TICs, mejorar la competitividad de la economía brasileña, generar más empleos con valor agregado y promover mayor desarrollo social (BRASIL, 2018a).

La EBIA también indica el Programa Startup Brasil, destinado al apoyo de *startups* brasileñas e internacionales para desarrollo de *software*, *hardware*, servicios de TI o que utilicen esas tecnologías para innovar. Sin embargo, según lo visualizado anteriormente, son pocas las empresas ligadas al programa, y las destinadas al desarrollo del área pretendida aún son pequeñas.

En ese contexto, se cree que la creación de un Centro Nacional de Inteligencia Artificial

(CNAI) sea adecuada para ayudar a la conexión y el contacto entre la sociedad y las instituciones gubernamentales por medio de expertos en IA. Además, el CNAI podrá actuar específicamente en proyectos destinados a la creación, al uso y a la mejora de la IA en múltiples sectores de la economía. Esa práctica facilitaría la colaboración entre los sectores industriales e los investigadores, con el objetivo de atraer mayores inversiones debido a la existencia de mano de obra calificada.

Otro desafío en el sector regulatorio relacionado a la IA es el aspecto ético y de responsabilidad civil. La EBIA no postula como debemos resolver esa cuestión, pero se presentarán consideraciones puntuales a continuación.

3.1 Contribuciones y propuestas en responsabilidad civil de la IA

Para tratar la temática, destaco que no se aborda sobre la posibilidad de que una IA sea o no sujeto de derecho, sino sobre la responsabilidad ética y jurídica que se deriva de sus actos. La atención, en este objetivo, será dirigida a un panorama disruptivo en que la IA incide en la praxis y puede causar daños a la sociedad civil. Incluso cuando intentamos demostrar los requisitos 1) agente; 2) daño; 3) culpa; y 4) nexo de causalidad, el primero es básicamente inverificable. De esta manera, entiendo que una IA no puede ser personalmente responsable por los daños causados a terceros, al menos en este momento. La intención es verificar como las teorías de la responsabilidad civil pueden o no ser suficientes para hacer frente a este problema planteado actualmente

La empresa Microsoft había insertado un *chat-bot* denominado Tay para interactuar autónomamente con los usuarios de esa red social. Su desarrollo fue pretencioso: por medio de ML, Tay debería ser capaz de compilar informaciones en su base de datos para crear discursos comprensibles (PEREZ, 2016). Sin embargo, la empresa desactivó la IA después de denuncias de que Tay estaría haciendo *tweets* racistas y misóginos (VINCENT, 2016). En ese sentido, poco tiempo después el proyecto de Facebook semejante al de Microsoft fue desactivado, pues percibió que sus dos IAs (Alice y Bob) estaban interactuando entre sí de una forma ininteligible.¹³

Aunque fuesen situaciones controlables y fáciles de ser verificadas y remediadas, las IAs tenían un potencial ofensivo que podría desarrollarse en cadena en dos grandes redes sociales. Ese desarrollo podría llegar a grupos extremistas y servir de fundamento para discriminar a otros grupos en la red.

Cuando analizamos los casos de aplicabilidad de Responsabilidad por el Hecho del Producto en el CDC, normalmente lo vinculamos a daños personales y patrimoniales de orden

¹³ Parte del diálogo puede ser expresado: "A few days later, some coverage picked up on the fact that in a few cases the exchanges had become - at first glance - nonsensical: Bob: "I can can I I everything else" Alice: "Balls have zero to me to" (BARANIUK, 2017).

físico. En algunos casos, hasta la pérdida o deterioro del producto por mal funcionamiento. Ese producto normalmente es físico, aunque existan posibilidades de productos digitales. Pero, cuando insertamos la IA en ese contexto, vamos más allá de daños del orden físico. Una IA, por ejemplo, puede causar un daño a la privacidad del titular de aquel producto. Amazon, por ejemplo, recopila detalles íntimos sobre sus usuarios por medio de Alexa, su IA.¹⁴

Otro ejemplo que se puede traer es el de cerraduras electrónicas e inteligentes que pueden cortar la autonomía de sus titulares, en caso de que sean utilizadas por terceros. En 2017, un Hotel en Australia sufrió un ataque de *hackers* que lograron invadir el sistema de seguridad de la empresa, en especial el sistema de cerradura electrónica. Para entrar nuevamente en su cuarto y recuperar sus pertenencias, se exigió de los huéspedes la cantidad de \$ 1.800 (BILEFSKY, 2017).

Sucede que la responsabilidad en el CDC posee algunas peculiaridades, entre ellas la objetividad y el fundamento en el riesgo. Comienza por la segunda. La relación entre riesgo y actividad es siempre una relación probabilística. En otros términos, el pilar central de la teoría del riesgo en el CDC, según entendimiento de este autor, es la posibilidad de prever o no la existencia de los daños. En el ámbito de la IA, esa previsión no es fácil de realizar. Vimos hace poco la relación entre Bob y Alice como algo totalmente fuera de los estándares pre-establecidos. El *Deep Learning* (DL) y el *Machine Learning* (ML) son realizados para, necesariamente, no prever los comportamientos de una IA. Al final, si la intención es replicar el comportamiento del ser humano, nada más adecuado e inspirador que replicar su aleatoriedad.

De cierta forma, no se niega la previsibilidad por medio de los sistemas de ML. Existen reglas que pueden ser insertadas en el sistema de IA para que actúen según lo pre-determinado. Pero esas reglas poseen límites semánticos y sintácticos. Además, cuanto mayor el número de reglas, mayor la complejidad y, consecuentemente, mayor los costos de operación de esa IA. Completar el contenido de una IA con todas las reglas posibles en contenidos lingüísticos y probabilísticos computacionales parece ir más allá de las pretensiones del ML.

Cabe señalar que no todos los actos de la IA son necesariamente impredecibles, tal como de un vehículo autónomo que se detiene en una senda peatonal para que pase un transeúnte. La cuestión en sí es que el riesgo de causar daños no es un riesgo meramente

¹⁴ “Amazon recolecta datos sobre los consumidores por medio de su asistente de voz Alexa, las compras en su marketplace, los lectores electrónicos Kindle, además de la plataforma de música. Así, la empresa reúne una gran gama de informaciones sobre sus clientes en Estados Unidos y comenzó a poner a disposición esos datos para todos mediante solicitud a inicio del año pasado, después de intentar y no lograr ganar una medida de 2018 de California que exigía tales divulgaciones. [...] El expediente de un reportero reveló que Amazon recolectó más de 90 mil grabaciones Alexa de familiares entre diciembre de 2017 y junio de 2021 - un promedio de alrededor de 70 por día. Las grabaciones incluían detalles como los nombres de los hijos y sus canciones favoritas. Amazon atrapó a los niños preguntando cómo podrían convencer a sus padres para dejarlos jugar y recibiendo instrucciones detalladas de Alexa sobre cómo convencer a sus padres a comprar videojuegos. Algunas grabaciones incluían conversaciones entre miembros de la familia usando dispositivos Alexa para comunicarse en diferentes partes de la casa. Varias grabaciones capturaron a niños de siete a 12 años haciendo preguntas a Alexa sobre términos como "pansexual". El reportero no sabía que Amazon estaba almacenando las grabaciones”. (O GLOBO, 2021).

técnico, sino también un riesgo de orden normativa, fuera del objetivo del propio CDC en cuanto a su análisis.

Se cree, con eso, que el riesgo de la actividad en desarrollo de IA trasciende en mucho los riesgos amparados por el CDC. Pero ¿de qué forma eso ocurre? Primero, necesitamos rever algunos conceptos. La responsabilidad por el riesgo del producto intenta resolver algunos problemas, de los cuales: 1) defectos provenientes del proyecto, fabricación, construcción, montaje, fórmulas, manipulación, presentación o acondicionamiento de sus productos; y 2) informaciones insuficientes o inadecuadas sobre su utilización y riesgos. Se trata de responsabilidad objetiva por los daños, aunque no intencionales, al consumidor, sean ellos oriundos del producto en sí, estando defectuoso, o por la falta de información en lo que respecta a su uso y riesgo.

Si una IA no funcionase como inicialmente ella debería, bastaría verificar en su constitución, proyecto o fabricación, por ejemplo, que se trata de un defecto del producto y, en consecuencia atraería responsabilidad al fabricante, al productor, al constructor, nacional o extranjero, y al importador.¹⁵ El defecto es la clave para concretar la responsabilidad de los sujetos antes identificados ante el consumidor.

Existen algunos abordajes establecidos por el CDC para verificar la ocurrencia de defectos en el producto: presentación, uso y riesgos y cuando fue puesto en circulación. Eso significa que el fabricante deberá comprobar la inexistencia de ese defecto en al menos tres modalidades establecidas legalmente. Bajo el aspecto económico, el responsable debe adoptar medidas que reduzcan el riesgo impuesto al desarrollo del producto por medio de técnicas de precaución y prevención de daños que posiblemente serían la clave de un acto ilícito. Se trata de los costos de precaución defendido por la AED, cuya inobservancia trae la responsabilidad objetiva al agente.

Resulta que la IA, puede no presentar ningún defecto y, aún así, causar daños al consumidor. Es aquí donde el desafío comienza. Una de las exclusiones de la responsabilidad objetiva del art. 12, §3º del CDC es la comprobación de que el producto fue colocado en el mercado y el defecto no existe.

Además, existe una **expectativa de seguridad sobre el producto** (art. 12, §1º del CDC), que se vuelve casi imposible de verificar en la práctica debido a la novedad tecnológica. Y, dentro de ese tema, ¿cómo trazar informaciones suficientes y adecuadas sobre los posibles riesgos del producto (art. 12, caput, parte final, CDC), siendo que el principal problema de la IA es su imprevisibilidad y, consecuentemente, la incertidumbre de los riesgos en términos probabilísticos (LEMLEY; CASEY, 2019)? En otras palabras, si la imprevisibilidad es parte del

¹⁵ En los términos del art. 12, §1º del CDC, “El producto es defectuoso cuando no ofrece la seguridad que legítimamente se espera de él, tomándose en cuenta las circunstancias relevantes, entre ellas: I - su presentación II - el uso y los riesgos que razonablemente se esperan de él; III – el momento en que fue puesto en circulación.”

producto, ¿cómo insertar y aceptar esa imprevisibilidad como parte de los riesgos de la actividad, siendo que es desconocida? Sería difícil y económicamente inviable incluir en la base de base de datos todos los riesgos operativos de esa IA. En caso contrario, si se sancionara al fabricante por los riesgos imprevisibles estaríamos transformándolos en previsibles y reconociendo su responsabilidad en todos los casos posibles. En otras palabras, él nunca lograría concretar la exclusión de ilicitud contenida en el art. 12, §3º, II del CDC, pues la tecnología sería en si defectuosa. El raciocinio al contrario puede ser verificable y también inteligible. Rachum-Twaig prescribe que

ya que los riesgos relacionados a la IA son imprevisibles por naturaleza y, por lo tanto, no pueden ser cubiertos por el defecto de proyecto o por el deber de advertencia y doctrina de instrucción, puede haber casos de daños fuera del alcance de la doctrina de responsabilidad por productos; pero esos casos pueden ser compensados por otras formas de responsabilidad en delitos (RACHUM-TWAIG, 2020, p. 1141).

Otro problema oriundo del análisis de riesgo que visualizo no solo en el CDC es la responsabilidad objetiva. A diferencia de la responsabilidad subjetiva basada en la culpa, la objetividad es esencialmente verificable en el riesgo.¹⁶ El problema general de la ausencia de previsibilidad en el comportamiento de la IA es crucial para una interpretación normativa adecuada, ya que su incidencia es indispensable para adecuar y reasignar los riesgos bajo el prisma de la responsabilidad objetiva. Además, existe una asimetría informacional latente entre fabricantes, desarrolladores, programadores y todos los involucrados en la cadena. En gran parte, los desarrolladores conocen y poseen mayores informaciones sobre los riesgos de la IA que los fabricantes. Esos también poseen un nivel considerablemente superior de informaciones sobre la IA que los destinatarios finales. Rachum-Twaig (2020, p. 1163) trae ejemplos de un robot trabajando en medicina que posee más informaciones sobre los riesgos del procedimiento a ser realizado que el propio paciente a ser operado. Resalta el autor que, como la principal característica de los robots basados en IA es que pueden actuar de una forma imprevisible o inexplicable por los humanos, en muchos casos ninguno de los interesados estará mejor situado para evaluar los riesgos que implica su operación, y el problema de la información imperfecta se aplicará igualmente a todos los interesados. En esos casos, los principios orientadores de aplicabilidad de la responsabilidad estricta no pueden ser aplicados.

En la cadena productiva, la reducción de los costos será dirigida exclusivamente a los *designers* y desarrolladores de tecnología, que tendrán que obtener todo y cualquier tipo de información relacionada al riesgo para mitigarlo y erradicarlo, en la medida de lo posible. Bajo el prisma de relación entre costos y beneficios, se vuelve inviable la programación de esa forma. La imprevisibilidad está intrínseca en la programación electrónica, y sus riesgos deben ser

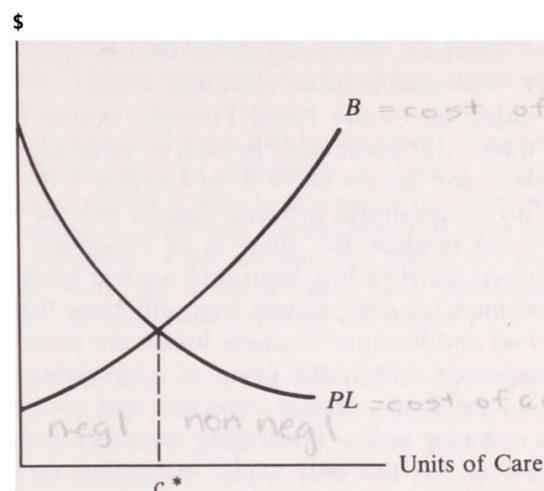
¹⁶ “*fault liability makes wrongful agency the fundamental basis of responsibility for harm accidentally done; strict liability makes agency itself the fundamental basis of responsibility*” (KEATING, 2001, p. 1285).

comprendidos más allá de aquellos establecidos en el CDC y en la responsabilidad objetiva por no ser previsible.

Entiendo, por lo tanto, que la adopción del CDC en esos casos solamente dificultaría y transformaría el ambiente tecnológico en un ambiente normativo dotado de reglas y preceptos jurídicos sin aplicabilidad práctica y solamente con reflejos negativos. Teóricamente, podríamos concluir que el régimen de la responsabilidad subjetiva sería ideal, ya que la objetividad debe ser dejada de lado. En el prisma de la AED, la fórmula de Hand puede ser suscitada como mecanismo dirigido a solucionar esa problemática. Veremos, sin embargo, que no es tan fácil.

La fórmula de Hand, según Posner, impone a las partes de la relación jurídica costos marginales de precaución (B) para evitar posibles daños esperados (PL) derivados de una multiplicación entre la probabilidad de que ese daño realmente se produzca (c) y los daños propiamente dichos (d) (POSNER, 2012).¹⁷ Se representa de la siguiente forma:

Figura 9 – Costos de Prevención



Fuente: Posner (2012).

Se puede extraer del gráfico anterior que, para Hand, la conducta será culpable cuando la inversión en precaución (B) sea menor que los posibles daños esperados (PL). Así, $B < PL \times c$ (POSNER, 2012, p. 148). La curva decreciente representada por PL representa un cambio marginal en los costos de los daños esperados en función de la precaución adoptada. Dicho eso, se verifica que la adopción de prácticas de prevención y precaución tiene un efecto decreciente en la prevención de accidentes. Por otro lado, la curva señalada por B ejemplifica el costo marginal de los cuidados y asciende en la medida en que las prácticas de precaución son escasas y, por lo tanto, se eleva a medida que más productos y servicios son ofrecidos en el mercado de consumo. En la intersección entre las dos curvas, se verifica el deber de cuidado ideal donde

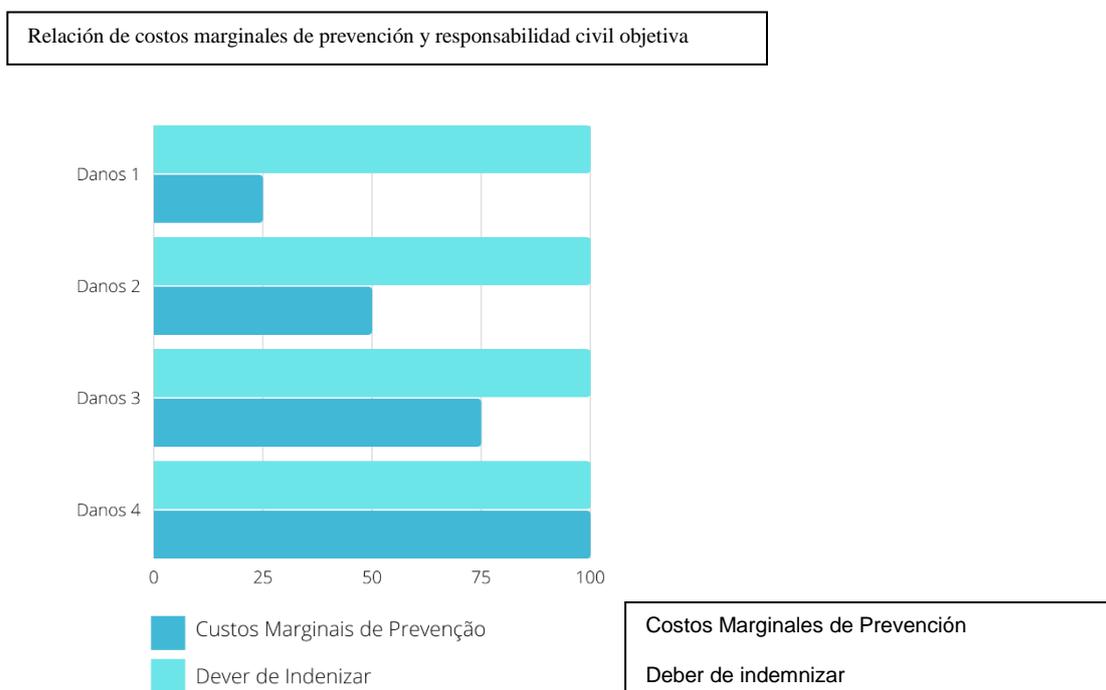
¹⁷ Para más sobre AED ver en (SHAVELL, 2009)

existe separación entre la negligencia y la prudencia/precaución (POSNER, 2012).

Cuando se aplica ese análisis bajo la óptica de la responsabilidad objetiva, si se determina que la reparación de los daños debe realizarse de forma incondicional e invariablemente por su desarrollador, productor, fabricante o responsable por la tecnología, sin análisis de culpa, se negaría y se ausentaría los intereses para la implementación de nuevas prácticas destinadas a la precaución contra la ocurrencia de nuevos daños.

Por lo tanto, se trata de un desestimulo y de desincentivo para adopción de costos marginales de prevención. En caso de que la sociedad empresaria siguiera todos los preceptos recomendados por la legislación y adoptara las prácticas de gobierno y precaución y, aún así, se responsabilizara por todo y cualquier daño ocurrido en su ámbito contractual, los costos en esas acciones serían prescindibles, teniendo en cuenta que la responsabilidad ocurriría con o sin ellos. De esa forma, adoptándose la responsabilidad objetiva, se verifica que los costos marginales de precaución (B) para evitar posibles daños esperados (PL) son indiferentes a los daños propiamente dichos (d).¹⁸ Los expreso de la siguiente forma:

Figura 10 – Costos Marginales x Responsabilidad Civil Objetiva



Fuente: Elaborada por el autor.

Cabe señalar que en la responsabilidad civil objetiva, independientemente de las acciones tomadas, el criterio indemnizatorio siempre permea la totalidad de los daños causados.

¹⁸ Para mejor comprensión de la AED, especialmente en Posner, se recomienda (GAROUPA; PORTO; FRANCO, 2019).

De esa forma, en caso de que esa modalidad (objetiva) sea adoptada, no tendría sentido alguno para el desarrollador o para una sociedad empresaria invertir y adoptar costos de transacción destinados a la prevención, pues, de cualquier forma, serían obligados a indemnizar a los titulares lesionados en el futuro. Así, se trata de una relación de pérdida efectiva y un gasto innecesario en la política empresarial.

Bajo el contexto económico y legal, si la fórmula de Hand es correcta, se verifica que la posibilidad de adopción de la responsabilidad civil en su modalidad subjetiva parece ser la más adecuada, pues permitiría que los costos destinados a la precaución no fuesen reducidos a cero en razón de una responsabilidad incondicional.

Glaubitz y Raymond (2021) reconocen la posibilidad de la fórmula de Hand en la solución del problema de atribución de responsabilidad por los actos de la IA e insertan una tercera variable en el cálculo probabilístico: el deber de cuidado.

Fórmula de Glaubitz y Raymond (2021) para verificación de responsabilidad de IA:

$$B - \frac{k}{\min\{A\}} < P \cdot L$$

En la fórmula anterior, B es el deber y el cuidado en tener la tarea automatizada completada manualmente por un humano; k es una constante que determina la influencia de A ; A es el costo mínimo de una solución algorítmica alternativa (recalificación del algoritmo existente, creación de un nuevo algoritmo o utilización de un algoritmo existente que logre el mismo propósito legítimo con un menor impacto en las clases protegidas); P es la probabilidad de un deber de cuidado preexistente, determinado por una evaluación de impacto dispar que evalúa la extensión de cualquier impacto adverso desproporcional en una clase protegida (identificado a través de la auditoría del algoritmo); y L es la pérdida o daño proveniente del sesgo algorítmico experimentado por una clase protegida (GLAUBITZ; RAYMONG, 2021, p. 31).

En términos prácticos, A sería como una sanción de naturaleza civil. Si el costo marginal de prevención es alto, entonces el agente responsable por el ilícito sería menos penalizado. Si el costo gastado en el deber de cuidado fuera bajo, la indemnización deberá ser valorada según su culpa. Y, según Glaubitz y Raymond (2021), la responsabilidad variará según los grados de automatización del producto y su supervisión. Cuanto más automatizados, mayor deber de cuidado es necesario para mitigar el potencial ofensivo a la sociedad. Ese deber de cuidado sería complementario al costo de prevención P previamente existente en Hand.

Sin embargo, tanto en Hand como en Glaubitz y Raymond (2021) existe un factor indispensable a ser verificado en sus fórmulas: el riesgo. El riesgo es, en suma, un elemento que

debe ser objetivamente medible para que se produzca su reducción. Cuando hablamos de la programación de IA, ese riesgo es difícil – y en algunos casos imposible – de ser delimitado en su totalidad. No se logran verificar acciones *ex ante* para preverlas y evitarlas *ex post*. Las evaluaciones de impacto quedan comprometidas como instrumento adecuado para la aplicación de la fórmula de Hand y, consecuentemente, decaímos en una actividad completa de riesgos e incertidumbres.

Por lo tanto, en materia de responsabilidad civil, la modalidad subjetiva parece ser más adecuada para la implementación de los preceptos de la EBIA en cuanto al mantenimiento de la innovación y al estímulo a su incentivo. Por ese motivo que políticas públicas destinadas al financiamiento de proyectos de investigación que busquen aplicar soluciones no discriminatorias, basadas en la equidad/no discriminación (*fairness*), responsabilidad/rendición de cuentas (*accountability*) y transparencia son las matrices para el estímulo de asociaciones con empresas que estén investigando soluciones comerciales y sociales de esas tecnologías.

En otras palabras, el Estado es responsable por establecer estándares y requisitos técnicos que sean compatibles con la promoción de una IA responsable, pero al mismo tiempo que cumpla con las formas de responsabilidad previamente definidas en estándares legales. En caso de que sea necesario, el mapeo de las barreras legales y regulatorias será fundamental para la identificación de la actualización destinada a la promoción de la seguridad jurídica. Esa práctica solamente será posible cuando sean realizados estudios más profundos sobre la temática.

Por lo tanto, según lo establecido por la EBIA, la creación de políticas de control de calidad de datos para efectivización del ML y del DL bajo parámetros intervencionistas humanos puede ayudar al encuentro de resultados efectivos con bajo grado de riesgos para la sociedad y para el individuo. Y, nuevamente, la creación de un centro específico para la IA, con expertos en la temática, es esencial para esa concreción.

En la medida en que analizamos como la estructura de responsabilidad puede ser implementada, nos encontramos en el sector de gobierno. En este punto, el aspecto fundamental es la creación de prácticas de gestión para monitoreo y supervisión de los sistemas de IA.

4 Eje temático 2: Gobierno de IA

La EBIA, en este punto, postula dos principios básicos y lineamientos de la estructura de gobierno: transparencia y *accountability* (aquí traída en términos de administración pública y comprendida como responsabilidad y rendición de cuentas). Esos dos parámetros imponen al responsable por la aplicación de IA establecer estructuras destinadas a asegurar que su implementación pueda ser analizada por el principio de la precaución, identificándose las aplicaciones de alto riesgo que pueden impactar significativamente la sociedad en un contexto de aplicación determinado, tal como salud o monitoreo del espacio público.

Según la EBIA, la intervención regulatoria en cuanto al gobierno debe ser equilibrada entre el grado de riesgo relacionado a la determinada aplicación de IA y las posibles limitaciones que puedan restringir sus usos. Para ello, la EBIA propone la elaboración de informes periódicos de impacto de protección de datos (RIPDs), que pueden ser identificados en cada sector de actuación de IA, tales como: Informe de Impacto de Seguridad (RIS); Informe de Impacto Ambiental (RIA); o Informe de Impacto de Derechos Humanos (RIDH).

En ámbito legal, el Decreto N 8.777/2016, responsable por el Portal Brasileño de Datos Abiertos, así como el Portal del Software Público Brasileño, regido por la Ordenanza N 46/2016, son medidas y políticas destinadas a la transparencia de los proyectos. Sin embargo, el sistema de gobierno debe partir de un supuesto de colaboración estructural del sector público y del sector privado para el desarrollo de estándares de gestión de riesgos asociados al uso de la IA.

En otras palabras, utilizando el *Artificial Intelligence Governance* de Singapur, existen cuatro medidas claves a ser adoptadas para efectivizar un modelo de gobierno de IA:

- a. Estructuras y medidas de gobierno interno destinadas a adaptar las estructuras y medidas ya existentes o establecer nuevas para incorporar valores, riesgos y responsabilidades relacionados con la toma de decisiones algorítmicas.
- b. Determinar el nivel de participación humana en la toma de decisiones con base en la IA: Una metodología para ayudar a las organizaciones a establecer su apetito por el riesgo para el uso de IA, o sea, determinar riesgos aceptables e identificar un nivel apropiado de participación humana en la toma de decisión con base en IA.
- c. Gestión de operaciones: Cuestiones a ser consideradas al desarrollar, seleccionar y mantener modelos de IA incluyendo gestión de datos.
- d. Interacción y comunicación de las partes interesadas: Estrategias para la comunicación con las partes interesadas de una organización, y la gestión de las relaciones con ellas (SINGAPORE, 2020).

Se trata de un modelo flexible en que las organizaciones públicas o privadas pueden modificarlo según sus necesidades y aspectos culturales. Lo importante, en este caso, es que las directivas pueden asistir y ayudar a las organizaciones a comprender como implementar cada una de las prácticas antes descritas.

4.1 Estructuras y medidas de gobierno interno

Las estructuras y medidas de gobierno interna tienen como objetivo garantizar una supervisión robusta sobre el uso de la IA por una organización, sea ella pública o privada. Las estructuras de gobierno interno existentes pueden ser adaptadas, y/o nuevas estructuras pueden ser implementadas, si es necesario. Por ejemplo, los riesgos asociados al uso de IA pueden ser gestionados dentro de la estructura de gestión de riesgo empresarial, mientras que las consideraciones éticas pueden ser introducidas como valores corporativos y gestionadas a través de consejos de revisión ética o estructuras similares (SINGAPORE, 2020).

Algunos factores son relevantes para su creación y efectivización:

1) Funciones y responsabilidades claras para el uso ético de la IA: como su desarrollo es realizado por periodos, etapas y actividades, en cada una de ellas puede existir la asignación individual de responsabilidad a su desarrollador. De esa forma, los departamentos que poseen una estructura interna de gobierno pueden conocer sus funciones y estar totalmente alertas sobre sus funciones y sobre sus responsabilidades. Entre las prácticas adoptadas están: a) el mantenimiento, monitoreo, documentación y revisión de los modelos de IA desarrollados; b) revisión de los canales de comunicación e interacción con los usuarios para proveer un efectivo *feedback*; y c) garantizar que los empleados/desarrolladores sean técnicamente aptos y propiamente capacitados para lidiar con los sistemas de IA (SINGAPORE, 2020).

2) Gestión de riesgos y controles internos: se trata de una opción para identificar, corregir y lidiar con riesgos internos en cuanto al desarrollo de la IA. Se pueden utilizar esfuerzos razonables para asegurar que los conjuntos de datos utilizados para el entrenamiento del modelo de IA sean adecuados para el propósito pretendido, y para evaluar y gestionar los riesgos de imprecisión o parcialidad, así como rever las excepciones identificadas durante el entrenamiento del modelo. Prácticamente ningún conjunto de datos es completamente imparcial. Las organizaciones deben esforzarse para comprender las formas por las cuales los conjuntos de datos pueden ser tendenciosos y abordar eso en sus medidas de seguridad y estrategias de implementación (SINGAPORE, 2020).

Uno de los ejemplos capaces de ilustrar la implementación de medidas y estructuras de gobierno interno es la adoptada por la empresa Mastercard. Para garantizar una supervisión robusta del uso de IA por Mastercard, la empresa estableció un Consejo de Gobierno para revisar y aprobar la implementación de aplicaciones de IA que son determinadas como de alto riesgo. El Consejo de Gobierno es presidido por su Vice-Presidente Ejecutivo del Centro de Excelencia en Inteligencia Artificial, y cuyos miembros incluyen el *Chief Data Officer*, el *Chief Privacy Officer*, el *Chief Information Security Officer*, científicos de datos y representantes de equipos comerciales (SINGAPORE, 2020).

4.2 Determinación del nivel de participación humana en la toma de decisiones con base en la IA

La EBIA propone un abordaje *Human-Centric AI*. En términos simples, ese abordaje coloca la IA como una herramienta en que los agentes humanos deciden cuando y como usarla. Según Muller, “necesitamos un abordaje HIC de la IA, donde las máquinas sigan siendo máquinas y las personas mantienen el control sobre esas máquinas todo el tiempo” (UNIÃO EUROPEIA, 2018). En su concepción, “los agentes humanos pueden y deben tener el control sobre sí, cuando y como la IA es usada en lo cotidiano, así como cuales tareas le transferimos, cuán transparente es, y el respeto a los aspectos éticos” (UNIÃO EUROPEIA, 2018). Se

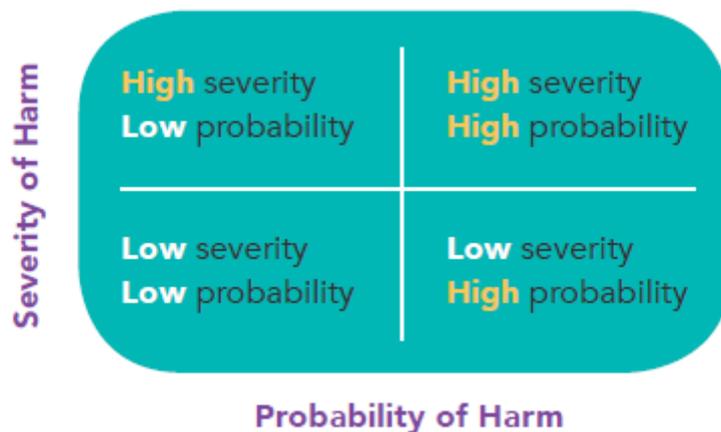
presupone, por lo tanto, el monitoreo y la supervisión del sistema de IA desde la concepción del sistema (*privacy by design, security by design, human rights by design, ethics by design*).

Sucede que, para lograr esos objetivos, antes de lanzar a IA al mercado, las empresas y el gobierno deben decidir sobre sus objetivos comerciales y verificar cuales son los riesgos en su utilización y en la toma de decisiones. En cuanto a las multinacionales, se debe considerar diferentes factores culturales y diferentes sistemas normativos o de valores. Además, algunos riesgos para determinados sujetos pueden manifestarse solo cuando son aplicados en un determinado grupo (tal como la oferta automatizada de productos que puede causar un cambio de comportamiento y, consecuentemente, compras compulsivas).

Además, como la identificación de los objetivos y riesgos comerciales, así como la determinación de actuación humana en la toma de decisiones automatizadas, son procesos interactivos y continuos, es esperado que las organizaciones continúen identificándolos y reviéndonlos constantemente para incrementar mejores soluciones en sus tecnologías, mitigando, por lo tanto, riesgos y manteniendo una efectiva respuesta a acciones previamente fallidas.

Según las consideraciones extraídas de la EBIA, se vuelve interesante pensar en la automatización no como ausencia de la participación humana en una determinada tarea, sino incluirla selectivamente en la expectativa de que el resultado sea un proceso efectivo de las características de la automatización inteligente (WANG, 2019).¹⁹ Ese modelo estructural parte de dos grandes ejes: 1) probabilidad y 2) capacidad de causar daños a un individuo (u organización) como resultado de la decisión obtenida en el proceso.

Figura 11 – Relación daños x probabilidad



Fuente: Singapur (2020).

La propuesta de los autores extranjeros es estandarizada y semejante a la ya existente en

¹⁹ Para más (DIVINO; MAGALHAES, 2020).

la teoría general de la administración pública. En el sector público, de la misma forma que en el sector privado, el gestor debe utilizar herramientas que puedan ayudar en la toma de decisiones. Las herramientas son aún más importantes cuando se trata de riesgos ambientales, de salud, económicos, ergonómicos y otros tantos que necesitan ser gestionados y controlados en las instituciones de todas las áreas de la administración pública, ya que el bienestar de la sociedad es la esencia del servicio público. Se vuelve indispensable tomar decisiones y usar medidas correctas con relación a las políticas y programas públicos, siendo fundamental adoptar estrategias efectivas de gestión de riesgos.

Una de las principales reglamentaciones para ayudar al gestor público en el ejercicio de su cargo y de su función como persona del pueblo es la Instrucción Normativa Conjunta CGU/MP N° 1 de 2016. Esa normativa intenta asegurar que los responsables por la toma de decisiones, en todos los niveles del organismo o entidad, tengan acceso tempestivo a informaciones suficientes en cuanto a los riesgos a los cuales está expuesta la organización. Para ello, se debe comprender que la postura del jefe del ejecutivo diseminada en sus redes sociales o en su ambiente “privado” no se refleja solo en su esfera subjetiva, sino también en el de aquellos que adoptan la misma posición expresada por él. Por ese motivo, las conductas practicadas, aunque bajo el espectro de la libertad de expresión, deben tomar en cuenta el impacto en la sociedad, principalmente en el ámbito crítico de la salud.

La gestión de riesgos parte del supuesto impacto x probabilidad. Cuando se expresa una posición negativa de la vacunación, se debe verificar cual es el impacto que surtirá en la sociedad, así como en la probabilidad de que eso suceda. De acuerdo con el Tribunal de Cuentas de la Unión, se puede ejemplificar de la siguiente forma:

Figura 12 – Método de evaluación de riesgos

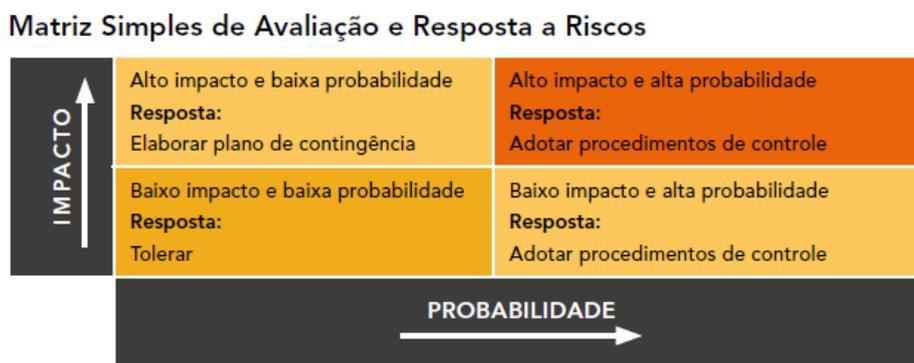


Figura 3: Matriz 2 x 2 de resposta a risco (INTOSAI GOV 9130, traduzido e adaptado)

Fuente: Brasil (2018b).

IMPACTO	Alto impacto y baja probabilidad Respuesta: Elaborar plan de contingencia	Alto impacto y alta probabilidad Respuesta: Adoptar procedimientos de control
	Bajo impacto y baja probabilidad Respuesta: Tolerar	Bajo impacto y alta probabilidad Respuesta: Adoptar procedimientos de control
PROBABILIDAD		

Cuanto mayor el impacto y mayor su probabilidad, mayor será el esfuerzo necesario para evitarlo en la medida de lo posible. Se trata, por lo tanto, de una relación directamente proporcional destinada a la mitigación de riesgos.

Figura 13 – Modelo de gestión de Riesgos

Modelo de Gerenciamento de Risco

		AÇÕES DE GERENCIAMENTO DE RISCO		
IMPACTO	Alto	6 Considerável esforço de gerenciamento é necessário	8 Indispensável gerenciar e monitorar riscos	9 Indispensável extensivo gerenciamento de risco
	Médio	3 Riscos podem ser aceitos, com monitoramento	5 Esforço de gerenciamento é necessário	7 Esforço de gerenciamento exigido
	Baixo	1 Aceitar Riscos	2 Aceitar, mas monitorar riscos	4 Gerenciar e monitorar riscos
		Baixa	Média	Alta
		PROBABILIDADE		

Figura 4: Matriz 3 x 3 de gerenciamento de risco (Secretaria do Tesouro do Canadá)

Fuente: Brasil (2018b).

		ACCIONES DE GESTIÓN DE RIESGO		
IMPACTO	Alto	Es necesario un considerable esfuerzo de gestión.	Indispensable gestionar y monitorear riesgos	Indispensable extensiva gestión de riesgo
	Medio	Los riesgos pueden ser aceptados con monitoreo	El esfuerzo de gestión es necesario	Esfuerzo de gestión exigido
	Bajo	Aceptar riesgos	Aceptar, pero monitorear riesgos	Gestionar y monitorear riesgos
		Baja	Media	Alta
		PROBABILIDAD		

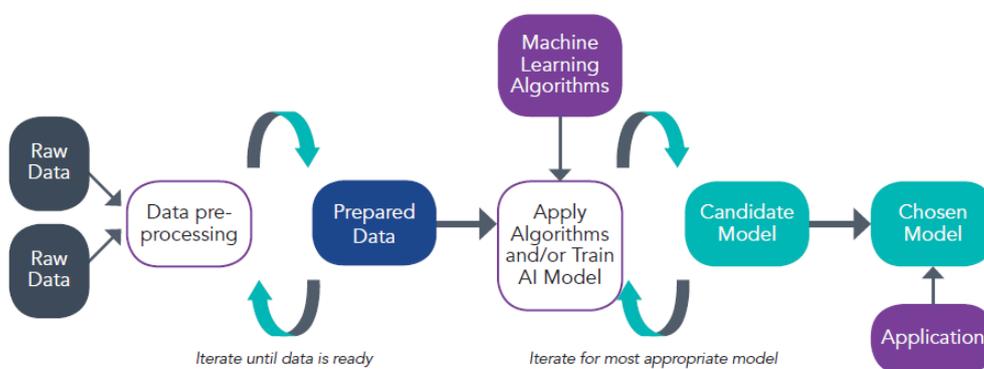
Con esas prácticas en acción, hay una considerable posibilidad de mejora en la calidad de los sistemas de IA a niveles gubernamentales y privados. En lo que respecta a las políticas públicas del sector estatal, los informes de impacto de protección de datos pueden adquirir nuevas características e incrementar resultados concretos destinados a la promoción de la IA en conjunto con su responsabilidad y ética. Se trata de un diálogo social y multisectorial indispensable para impulsar las prácticas de *accountability* relacionadas a la IA en las organizaciones.

4.3 Gestión de operaciones

Para que las prácticas del tópico anterior sean implementadas, una serie de procesos y operaciones deben estar intrínsecamente ligadas a la organización de datos, algorítmico y de la propia constitución de la IA. La EBIA, sin embargo, no establece tales criterios. Pero, el *Artificial Intelligence Governance Framework* de Singapur propone la organización en tres pasos: 1) preparación de los datos; 2) análisis del algoritmo; y 3) elección del modelo adecuado al propuesto.

La primera etapa, la recolección de datos es realizada, formateada y tratada para obtener conclusiones más asertivas. En este caso, la precisión y los *insights* aumentan según la cantidad de datos en la base de entrenamiento de la IA. Posteriormente, esos modelos son entrenados en base de datos y será realizado un análisis algorítmico. Aquí se puede incluir análisis estadísticos o modelos de aprendizaje de máquina extraído de redes neurales. Los resultados son examinados e insertados en los modelos más apropiados. Para ello, el análisis será probabilístico e incorporado en aplicaciones para ofrecer predicciones, tomar decisiones y resolver problemas.

Figura 14 – Modelo de Gestión Operativa en IA



Fuente: Singapur (2020).

Todo el proceso pasa por la comprensión lineal de todos los datos, desde cómo, cuándo y de donde ellos llegaron, así como fue recolectado, tratado y transferido, y hacia adonde, cuándo y cómo ellos irán. La clave para un procedimiento transparente ante los demás organismos será la existencia de registros que permiten a la organización, sea ella pública o privada, garantizar la calidad de los datos según su origen. En este medio término, serán plenamente adoptadas las prácticas destinadas a minimizar sesgos discriminatorios en la toma de decisiones. Uno de los sesgos más comunes a ser tratados se refiere a la cuestión racial y estereotipo y a la omisión de determinadas personas según su ascendencia en un grupo económico o social. Por esa razón, la relación entre diferentes bases de datos para entrenamiento, prueba y validez de la IA en instituciones públicas y privadas es esencial para garantizar una revisión periódica y actualizar constantemente los resultados obtenidos.

La propuesta reglamentaria de Singapur dirige a las organizaciones una aplicación modular algorítmica basada en riesgos representada en dos etapas. La primera debe identificar el subconjunto de características o funcionalidades que tienen el mayor impacto sobre las partes interesadas para las cuales tales medidas son relevantes. En segundo lugar, identificar cuáles de esas medidas serán más eficaces para crear confianza con sus partes interesadas. Algunas de esas medidas, como explicabilidad (o repetibilidad, cuando se usa modelos que no son fácilmente explicados), robustez y sintonía regular, son suficientemente esenciales para que puedan, en diferentes grados, ser incorporadas como parte del proceso de implementación de IA de la organización. Otras medidas, tales como reproductibilidad, rastreabilidad y auditoria, son más intensivas en recursos y pueden ser relevantes para características específicas o en escenarios específicos (SINGAPORE, 2020).

En este mismo sentido, la explicabilidad puede ser alcanzada en cuanto a la delimitación de como los algoritmos funcionan o como el proceso de toma de decisiones incorpora modelos de predicción. Se trata de una correlación entre comprensión y confianza en los sistemas que están siendo desarrollados e implementados. Pero aún existen limitaciones tecnológicas y sociales en cuanto a la explicabilidad, según lo ya mencionado anteriormente. Donde la explicabilidad no puede ser alcanzada, las organizaciones pueden considerar documentar la repetibilidad de los resultados producidos por el modelo de IA. La repetibilidad se refiere a la capacidad de ejecutar una acción o tomar una decisión de forma consistente, dado el mismo escenario. Aunque la repetibilidad (de los resultados) no sea equivalente a la explicabilidad (del algoritmo), algún grado de garantía de consistencia en el desempeño podría proporcionar a los usuarios de IA un grado mayor de confianza (SINGAPORE, 2020).

Otros factores, como robustez²⁰, mejora continua²¹, rastreabilidad²², estandarización y

²⁰ *“Robustness refers to the ability of a computer system to cope with errors during execution and erroneous input and is assessed by the degree to which a system or component can function correctly in the presence of invalid input*

posibilidad de reproducción²³ y auditoria²⁴ de la IA, son indispensables para ser implementados y destinados a la comprensión procesual lineal y de seguridad.

Finalmente, las políticas públicas destinadas al desarrollo de la IA solamente son posibles mediante interacciones y comunicaciones de las partes interesadas que, en este caso, es la sociedad civil y las instituciones públicas y privadas.

4.4 Interacción y comunicación de las partes interesadas

Como la incorporación de la IA en el poder público tiene como fundamento la referencia de la potencialización de las actividades humanas, el poder público, en este caso, no tiene interés en la satisfacción de intereses particulares, lo que relegaría el interés público. En otras palabras, la implementación de la IA en ese sector debe ser realizada por las vías de la equidad e inclusión social. Para ello, las organizaciones gubernamentales deben poner a disposición informaciones generales sobre cuales sectores, productos y servicios aplicaciones de IA serán efectivamente utilizadas. En este punto, políticas y términos de uso son fuertemente recomendados para ser desarrollados. Ellos deben contener explicaciones para ayudar a sus usuarios a comunicarse con la IA, así como solicitar informaciones al ente estatal responsable por su desarrollo y uso. Además, la política también puede contener sus diferentes funciones y la responsabilidad individualizada de cada miembro responsable por su elaboración.

Cabe señalar que la explicación detallada puede ayudar a la comunicación entre el sector público y aumentar considerablemente la transparencia en la toma de decisiones. Así, tanto la explicabilidad como la transparencia son mecanismos apropiados e indispensables para interacción y comunicación entre los sectores y, consecuentemente, aumento de la confianza en

or stressful environmental conditions. Ensuring that deployed models are sufficiently robust will contribute towards building trust in the AI system” (SINGAPORE, 2020).

²¹ *“Establishing an internal policy and process to perform regular model tuning is effective for ensuring that deployed models cater for changes to customer behaviour over time. This allows organisations to refresh models based on updated training datasets that incorporate new input data. Model tuning may also be necessary when commercial objectives, risks, or corporate values change” (SINGAPORE, 2020).*

²² *“An AI model is considered to be traceable if (a) its decisions, and (b) the datasets and processes that yield the AI model’s decision (including those of data gathering, data labelling and the algorithms used), are documented in an easily understandable way. The former refers to traceability of AI-augmented decisions, while the latter refers to traceability in model training. Traceability facilitates transparency and explainability and is also helpful for other reasons. First, the information might also be useful for troubleshooting, or for an investigation into how the model was functioning or why a particular prediction was made. Second, the traceability record (in the form of an audit log) can be a source of input data that can be used as a training dataset in the future” (SINGAPORE, 2020).*

²³ *“While repeatability refers to the internal repetition of results within one’s organisation, reproducibility refers to the ability of an independent verification team to produce the same results using the same AI method based on the documentation made by the organisation. Reproducibility can influence the trustworthiness of the AI product and the organisation deploying the AI model. As implementing reproducibility entails the involvement of external parties, organisations can take a risk-based approach towards identifying the subset of AI-powered features in their products or services that requires external reproducibility testing” (SINGAPORE, 2020).*

²⁴ *“Auditability refers to the readiness of an AI system to undergo an assessment of its algorithms, data and design processes. The evaluation of the AI system by internal or external auditors (and the availability of evaluation reports) can contribute to the trustworthiness of the AI system as it demonstrates the responsibility of design and practices and the justifiability of outcomes. It should, however, be noted that auditability does not necessarily entail making information about business models or intellectual property related to the AI system publicly available” (SINGAPORE, 2020).*

esos tipos de aplicaciones. Como el sector gubernamental es amplio, el primer paso para delimitar esa política y los términos de servicio es identificar al público objetivo de la prestación de servicios y, después, insertarlos en un contexto en que sus pretensiones y propuestas sean compatibles con las aplicaciones de IA.

Las informaciones más generales pueden ser dirigidas a potenciales usuarios para decidir si adhieren o no al servicio ejecutado por la IA. Sucede que, en caso de que sea aceptado, las informaciones más específicas deben ser dirigidas a demostrar cómo funciona la aplicación. Finalmente, en caso de que sea necesaria la intervención humana, un paso a paso debe ser indicado a sus usuarios, así como, si ellos solicitan informaciones sobre el funcionamiento de la toma de decisiones automatizadas, deben estar incluidas en un parámetro de *feedback*.

En otras palabras, perciban que la creación de canales de retorno (*feedback*) destinados a la evaluación del usuario es indispensable para la organización, principalmente para que el *Data Protection Officer* tome sus decisiones y verifique la adecuación de las prácticas la legislación. Así debería ser, pues, según la Comisión Europea, todos los gobiernos en democracias constitucionales deben ser limitados por la ley, inclusive aquellos que utilizan sistemas de IA para mantener o expandir sus procesos democráticos.

Se advierte que los parámetros establecidos para implementación de políticas públicas, aunque individualizados, constituyen un conjunto procesal indisociable para garantizar la inclusión apropiada del sistema en la sociedad. Seguridad y transparencia, según se ha mostrado, son instrumentos básicos para maximizar las innovaciones entre regulación y economía.

Ante lo expuesto, se verifica que la EBIA posee grandes expectativas a nivel internacional de implementación, pero necesita de un elevado y continuo trabajo a nivel de políticas públicas para su efectivización.

5 Conclusión

El problema de investigación inicialmente propuesto fue como y cuales políticas públicas pueden ser adecuadas para efectivización e implementación de la EBIA en cuanto a los ejes Legislación, regulación y uso ético y Gobierno de IA. Se constató en la primera sección que la EBIA, en nivel nacional, posee grandes desafíos a ser cumplidos, entre ellos el mayor incentivo a la investigación, tecnología, innovación y desarrollo a nivel nacional y para *startups*. En ese caso, los programas dirigidos al fomento están residualmente aplicados en sectores de TI. Según lo visualizado, una de las limitaciones del estudio con los datos obtenidos es trazar el ámbito concreto de actuación, dado que la Tecnología de la Información es un sector muy amplio. Así, los resultados en la práctica pueden ser aún peores cuando concretamente son ubicados en los sectores relativos a la IA.

Las principales contribuciones de este trabajo, por lo tanto, residen en la constatación de que la responsabilidad civil de los actos ilícitos practicados por IA o durante el curso de su actuación, ejecución e implementación debe ser la modalidad subjetiva. De esa forma, se permite la mitigación de los daños según los costos de precaución efectivamente asignados para configuración de una IA ética, sólida y robusta. En cuanto a los aspectos de gobierno, se verifica que las políticas públicas deben ser diseñadas para la creación de informes de impacto de seguridad e informes continuos del uso de IA en los sistemas gubernamentales. Las medidas de gobierno interno que incorporan fases de implementación son esenciales para su realización. Por lo tanto, deben ser estipuladas funciones y responsabilidades claras para el uso ético de la IA. Además, se trata de una opción para identificar, corregir y lidiar con riesgos internos en cuanto al desarrollo de la IA.

Además, la EBIA propone un abordaje *Human-Centric AI*. En términos simples, ese abordaje coloca a IA como una herramienta en que agentes humanos deciden cuándo y cómo usarla. En ese punto, la implementación de políticas públicas puede seguir las tradicionales teorías de la administración con relación a la gestión de riesgos, o sea, la relación entre impacto x probabilidad. En lo que respecta a las políticas públicas del sector estatal, los informes de impacto de protección de datos pueden adquirir nuevas características e incrementar resultados concretos destinados a la promoción de la IA en conjunto con su responsabilidad y ética. Se trata de un diálogo social y multisectorial indispensable para impulsar las prácticas de *accountability* relacionadas a la IA en las organizaciones.

Ese abordaje puede implementarse mediante gestiones operativas en que todo el procedimiento es acompañado desde el inicio al fin, delimitándose su: origen, finalidad, recolección y tratamiento de datos, aplicación, decisión, modelos reglamentarios y modelos estructurales para aplicación. Es indispensable que el gestor público conozca a su público objetivo para elaborar términos de servicios adecuados y suficientemente explicables a los usuarios. En otras palabras, todos los procesos de gestión pública deben estar pautados en la explicabilidad para alcanzar una delimitación de como los algoritmos funcionan o incluso como el proceso de toma de decisiones fue realizado, así como en la transparencia, para que se pueda obtener informaciones de todo tipo de forma inteligible y dirigida al desarrollo continuo, rastreabilidad, estandarización y auditoría de las prácticas de IA. Por esa razón, es importante la comunicación y la interacción entre las partes involucradas. Informaciones más generales pueden ser elaboradas, siempre y cuando estén dirigidas a un primer contacto. A medida que el ciudadano se depara con el efectivo uso de la IA, deben ser elaborados documentos específicos conteniendo un paso a paso para su correcta comprensión, informando, también, los responsables por la ejecución.

Las consideraciones finales aquí presentadas de forma alguna son conclusiones. Son solo indicaciones ejemplificativas para la creación de políticas públicas y documentos dirigidos

especialmente al *feedback* del usuario final. Las tecnologías que involucran IA están constantemente evolucionando y ligadas al carácter ético, siendo que su gobierno no debe ni puede poseer parámetros fijos para su aplicación. En caso de que eso suceda, existiría en sí una contradicción en la propia estructura de gobierno, que requiere una revisión periódica para adecuar los sistemas a lo previsto por el usuario, por el responsable y por el público. Por lo tanto, son caminos iniciales a ser trazados y complementados por una estrategia que aún tendrá un largo y real camino a ser recorrido.

Referencias

- BARANIUK, Chris. The 'creepy Facebook AI' story that captivated the media. **BBC**. 2017. Disponível em: <https://www.bbc.com/news/technology-40790258>. Acesso em: 20 dez. 2021.
- BECKER, Daniel; FERRARI, Isabela. O direito à explicação sobre decisões automatizadas: uma análise comparativa entre a União Europeia e o Brasil. **Revista de Direito e as Novas Tecnologias**, São Paulo, vol. 1, n. 1, out./dez. 2018.
- BILEFSKY, Dan. Hackers Use New Tactic at Austrian Hotel: Locking the Doors. **The New York Times**. 2017. Disponível em: <https://www.nytimes.com/2017/01/30/world/europe/hotel-austria-bitcoin-ransom.html>. Acesso em: 20 dez. 2021.
- BURRELL, Jenna. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. **Big Data & Society**, [s.l.], v. 3, n. 1, p.1-12, jan. 2016.
- BRASIL. Estratégia Brasileira de Inteligência Artificial – EBIA. **MCTI**. 2021. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 30 jan. 2022.
- BRASIL. **Estratégia Brasileira para Transformação Digital (e-Digital)**. 2018a. Disponível em: <https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf>. Acesso em: 31 jan. 2022.
- BRASIL. Manual de Gestão de Riscos. **Tribunal de Contas da União**. 2018b. Disponível em: <https://portal.tcu.gov.br/planejamento-governanca-e-gestao/gestao-de-riscos/manual-de-gestao-de-riscos/>. Acesso em: 30 jan. 2022.
- BRASIL. **Startup Brasil**. 2022. Disponível em: <https://www.startupbrasil.org.br/startups/conheca-as-startups-do-programa/>. Acesso em: 30 jan. 2022.
- CANADÁ. **Pan-Canadian AI Strategy**. 2017. Disponível em: <https://cifar.ca/ai/>. Acesso em: 30 jan. 2022.
- CHINA. **A Next Generation Artificial Intelligence Development Plan**. 2017. Disponível em: <https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17.pdf>. Acesso em: 30 jan. 2022.
- CITRON, D. K.; PASQUALE, F. A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, Washington, v. 89, n. 1, p. 2-27, 2014.
- CONECTA STARTUP BRASIL. **Startups**. 2021. Disponível em: <https://conectastartupbrasil.org.br/startups>. Acesso em: 30 jan. 2022.
- COMISSÃO EUROPEIA. **Artificial Intelligence Act**. 2021. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>. Acesso em: 30 jan. 2022.
- DIAKOPOULOS, Nicholas. Algorithmic Accountability Reporting: On the Investigation of

Black Boxes. **Tow Center for Digital Journalism**, 2013.

DINAMARCA. **Strategy for Denmark's Digital Growth**. 2021. Disponível em: https://eng.em.dk/media/10566/digital-growth-strategy-report_uk_web-2.pdf. Acesso em: 30 jan. 2022

DIVINO, S. B. S. Desafios e benefícios da inteligência artificial para o Direito do Consumidor. **Revista Brasileira de Políticas Públicas**, Brasília, v. 11, n. 1, p. 655-689, 2021.

DIVINO, S. B. S.; MAGALHAES, R. A. Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade. **Economic Analysis of Law Review**, Brasília, v. 11, p. 72-89, 2020.

DOSHI-VELEZ, Finale; KORTZ, Mason. Accountability of AI under the law: the role of explanation. **Berkman Klein Center Working Group on Explanation and the Law**, 2017.

DUTTON, Tim. An Overview of National AI Strategies. **Medium**. 2018. Disponível em: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>. Acesso em: 30 jan. 2022.

EL NAQA, Issam; MURPHY, Martin J. What is machine learning?. In: **machine learning in radiation oncology**. Springer: Cham, 2015. p. 3-11.

FINLAND. **Finland's Age of Artificial Intelligence**. 2017. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/finland-ai-strategy-report_en. Acesso em: 30 jan. 2022.

FJELD, Jessica et al. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. **Berkman Klein Center Research Publication**, Cambridge, n. 2020-1, 2020.

FLORIDI L. Soft ethics, the governance of the digital and the General Data Protection Regulation. **Philosophical Transactions of the Royal Society**, London, v. 376, n. 2133, 2018. Disponível em: <http://dx.doi.org/10.1098/rsta.2018.0081>. Acesso em:

FRANCE. **AI for Humanity: French Strategy for Artificial Intelligence**. 2018. Disponível em: <https://super-ai.diascreative.net/ai-for-humanity-french-strategy-for-artificial-intelligence>. Acesso em: 30 jan. 2022.

GAROUPA, Nuno; PORTO, Antonio José Maristrello; FRANCO, Paulo Fernando De Mello. As indenizações pela perda do tempo útil do consumidor. Espera e custos de oportunidade. **Revista de Direito do Consumidor**, Brasília, v. 124, p. 263-293, jul./ago. 2019

GARRET, Filipe. O que é algoritmo? Entenda como funciona em apps e sites da Internet. **TechTudo**. 2020. Disponível em: <https://www.techtudo.com.br/listas/2020/05/o-que-e-algoritmo-entenda-como-funciona-em-apps-e-sites-da-internet.ghtml>. Acesso em: 31 jan. 2022.

GLAUBITZ, Alina; RAYMOND, Nathaniel. **How should liability be attributed for harms caused by biases in Artificial Intelligence?**. 2021. Disponível em: https://politicalscience.yale.edu/sites/default/files/glaubitiz_alina.pdf. Acesso em: 20 dez. 2021.

INDIA. **National Strategy For Artificial Intelligence**. 2018. Disponível em: <https://indiaai.gov.in/research-reports/national-strategy-for-artificial-intelligence>. Acesso em: 30 jan. 2022.

ITALY. **Strategia Nazionale per l'Intelligenza Artificiale**. 2020. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/italy-ai-strategy-report_en. Acesso em: 30 jan. 2022.

JAPAN. **Artificial Intelligence Technology Strategy**. 2017. Disponível em: https://ai-japan.s3-ap-northeast-1.amazonaws.com/7116/0377/5269/Artificial_Intelligence_Technology_StrategyMarch2017.pdf. Acesso em: 30 jan. 2022.

KEATING, Gregory C. The theory of enterprise liability and common law strict liability. **Vanderbilt Law Review**, Nashville, v. 54, p. 1285, 2001.

KOREA. **National Strategy for Artificial Intelligence**. 2019. Disponível em: <https://www.msit.go.kr/bbs/view.do?sCode=eng&mId=10&mPid=9&bbsSeqNo=46&nttSeqNo=9>. Acesso em: 30 jan. 2022.

LEMLEY, Mark A.; CASEY, Bryan. Remedies for robots. **The University of Chicago Law Review**, Chicago, v. 86, n. 5, p. 1311-1396, 2019.

MEXICO. **Artificial Intelligence Agenda MX**. 2018. Disponível em: <https://oecd.ai/en/dashboards/countries/Mexico>. Acesso em: 30 jan. 2022.

O GLOBO. Amazon coleta detalhes íntimos sobre seus usuários através da Alexa; veja quais. **IG Tecnologia**. 2021. Disponível em: <https://tecnologia.ig.com.br/2021-11-22/amazon-alexa-coleta-dados-intimos.html>. Acesso em: 20 dez. 2021.

OCDE. **Recommendation of the Council on Artificial intelligence**. 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em: 30 jan. 2022.

OSOBA, Osonde; WELSER IV, William. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. **RAND Corporation**, Santa Mônica, 2017.

PAÍSES NÓRDICOS. **Declaration on AI in the Nordic-Baltic Region**. 2018. Disponível em: https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf. Acesso em: 30 jan. 2022

PASQUALE, Frank. **The Black Box Society: The Secret Algorithms that Control Money and Information**. Cambridge: Harvard University Press, 2015. 320 p.

PEREZ, Sarah. Microsoft Silences Its New A.I. Bot Tay, After Twitter Users Teach It Racism, **TECHCRUNCH**. 2016. Disponível em: <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>. Acesso em: 20 dez. 2021.

POSNER, Richard. **Economic Analysis of Law**. 3. ed. Alphen aan den Rijn: Wolters Kluwer, 2012. 1026 p.

RACHUM-TWAIG, Omri. Whose Robot Is It Anyway? Liability for Artificial-Intelligence-Based Robots. **University of Illinois Law Review**, Illinois, v. 2020, p. 1141, 2020.

RUSSELL, Stuart. J.; NORVIG, Peter. **Artificial intelligence: a modern approach**. 3. ed. New Jersey: Pearson Education, 2010. 1115 p.

SANDVIG, Christian et al. Auditing algorithms: Research methods for detecting discrimination on internet platforms. **Data and discrimination: converting critical concerns into productive inquiry**, Seattle, v. 22, p. 4349-4357, 2014.

SERASA. **O que é Serasa Score 2.0?** Disponível em: <https://www.serasa.com.br/score/score-2-0/>. Acesso em: 03 set. 2021.

SHAVELL, Steven. **Economic analysis of accident law**. Cambridge: Harvard University Press, 2009. 352 p.

SINGAPORE. **Artificial Intelligence Governance Framework**. 2020. Disponível em: <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>. Acesso em: 30 jan. 2022.

SOUZA, Celina. **Políticas públicas: conceitos, tipologias e subáreas**. Trabalho elaborado para a Fundação Luís Eduardo Magalhães. São Paulo, 2002.

SWEDEN. **Nationell inriktning för artificiell intelligens**. 2018. Disponível em: <https://www.regeringen.se/informationsmaterial/2018/05/nationell-inriktning-for-artificiell-intelligens/>. Acesso em: 30 jan. 2022.

TAIWAN. **AI Taiwan Action Plan**. 2019. Disponível em: <https://english.ey.gov.tw/News3/9E5540D592A5FECDD/1dec0902-e02a-49c6-870d-e77208481667#:~:text=Capitalizing%20on%20this%20wave%2C%20the,greater%20momentum%20into%20Taiwan's%20industries>. Acesso em: 30 jan. 2022.

UNESCO. **UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence**. 2021. Disponível em: <https://en.unesco.org/news/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence>. Acesso em: 30 jan. 2022.

UNIÃO EUROPEIA. **Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC**. 2017. Disponível em: <https://www.eesc.europa.eu/en/news-media/press-releases/artificial-intelligence-europe-needs-take-human-command-approach-says-eesc>. Acesso em: 11 abr. 2020.

UNITED STATES. **Artificial Intelligence Index Report 2021**. Stanford University. 2021. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf. Acesso em: 30 jan. 2022.

UNITED KINGDOM. **AI Sector Deal**. 2019. Disponível em: <https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>. Acesso em: 30 jan. 2022.

VINCENT, J. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. **The Verge**. 2016. Disponível em: <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>. Acesso em: 20 dez. 2021.

WANG, Ge. **Humans in the Loop: The Design of Interactive AI Systems**. Human-centered Artificial Intelligence. Stanford University. 2019.

WIPO. **Índice Global de Inovação**. 2021. Disponível em: https://www.wipo.int/edocs/pubdocs/pt/wipo_pub_gii_2021_exec.pdf. Acesso em: 30 jan. 2022