

GOBERNANZA ALGORÍTMICA, EXPLICACIÓN POR MEDIO DEL DISEÑO Y JUSTICIA DEL DISEÑO *

Algorithmic Governance, Explanation by Design and Design Justice

PAOLA CANTARINI GUERRA **

Fecha de recepción: 22/08/2022

Fecha de aceptación: 02/09/2022

Anales de la Cátedra Francisco Suárez

ISSN: 0008-7750, núm. 57 (2023), 121-141

<http://dx.doi.org/10.30827/ACFS.v57i.25976>

RESUMEN A pesar de los innumerables beneficios, hay varias críticas y preocupaciones en relación con el uso de la Inteligencia Artificial (IA), especialmente en Brasil, donde no hay ninguna legislación federal en vigor hasta el momento. Las legislaciones estatales existentes y los proyectos de ley que se están votando en el Congreso, como el PL 21/20, son defectuosos e incompletos, especialmente en lo que respecta a algunas aplicaciones de la IA que implican altos riesgos para los derechos y libertades fundamentales, además de traer a menudo un sesgo potencial, duplicando o potenciando el racismo institucionalizado existente en la sociedad u otras formas de discriminación. Existe una inequidad codificada que propicia infraestructuras injustas, sin embargo, donde hay poder hay resistencia, y en este sentido es necesario señalar la importancia de las contra codificaciones basadas en la solidaridad y que repiensen la justicia, es decir, un enfoque emancipador de la tecnología. Este es el cambio paradigmático necesario, y siendo necesario, tiene que ser posible y debe buscarse con el máximo esfuerzo. Se trata de verdaderas formas de resistencia, como instrumentos abolicionistas para el "Nuevo Código Jim", nuevas formas de discriminación de raza, género y clase existentes a través del uso de las nuevas tecnologías, en el sentido de prácticas de diseño orientadas a la justicia.

Palabras clave: Compliance, Diseño, Justicia del Diseño, Inteligencia Artificial, Gobernanza.

ABSTRACT Despite the countless benefits, there are several criticisms and concerns regarding the use of Artificial Intelligence (AI), especially in Brazil, where so far there is no federal legislation in force, the existing state legislations and the PL currently being voted in Congress, such as PL 21/20, are flawed and incomplete, especially regarding some applications of AI that involve high risks to fundamental rights and freedoms, besides often bringing a potential bias, duplicating or potentiating the existing institutionalized racism in society or

* Para citar/citation: Cantarini Guerra, P. (2023). Gobernanza algorítmica, explicación por diseño y justicia por diseño. *Anales de la Cátedra Francisco Suárez* 57, pp. 121-141.

** Instituto Ethikai, Brasil. Correo electrónico: paolacantarini@gmail.com

other forms of discrimination. There is a codified inequity propitiating unjust infrastructures, however, where there is power there is resistance, and in this sense one has to point to the importance of counter-codifications based on solidarity and that rethink justice, that is, an emancipatory approach to technology. This is the necessary paradigmatic change, and being necessary, it has to be possible, and should be sought with the utmost effort. These are true forms of resistance, such as abolitionist instruments for the “New Jim Code”, new forms of race, gender and class discrimination existing through the use of new technologies, in the sense of justice-oriented design practices.

Keywords: Compliance, Design, Design Justice, Artificial Intelligence., Governance.

1. EL BIG DATA ANALYTICS Y LOS SEGOS

El tema de la inteligencia artificial está estrechamente relacionado con la protección de datos en el contexto de lo que se conoce como “big data”, ya que la mayoría de las aplicaciones de IA utilizan bases de datos.

Sin duda, son muchos los beneficios que se derivan del uso de la IA en diversos sectores, como la sanidad y la seguridad pública, ayudando a controlar la pandemia de Covid, creando servicios globales más eficientes y cómodos.

Así lo corrobora la Declaración del 29.06.2019 del G20 sobre Comercio y Economía Digital —formada por los países del G8 más los emergentes, en la que participan las 19 mayores economías del mundo—, en la que se afirma que la inteligencia artificial es capaz de promover el crecimiento inclusivo de la economía, aportando innumerables beneficios sociales, siempre que su uso sea responsable. Dicha declaración trae la recomendación de observar los principios de la OCDE —Organización para la Cooperación y el Desarrollo Económico—, con énfasis en: la sostenibilidad en el desarrollo, el crecimiento inclusivo y el bienestar, la imparcialidad y la centralidad en el control humano, la transparencia para justificar la lógica en el proceso de toma de decisiones, la solidez en cuanto a la seguridad y la justicia en la decisión automatizada, la responsabilidad y la rendición de cuentas.

A pesar de los numerosos beneficios, hay varias críticas y preocupaciones en relación con el uso de la IA, especialmente en Brasil, donde no hay legislación federal en vigor. Las leyes estatales existentes y los proyectos de ley que se votan en el Congreso como el PL 21/20 son defectuosos e incompletos, especialmente en lo que respecta a algunas aplicaciones de la IA que implican altos riesgos para los derechos y libertades fundamentales, además de traer a menudo un sesgo potencial, duplicando o aumentando el racismo institucionalizado existente en la sociedad u otras formas de discriminación.

Se crean perfiles publicitarios dirigidos a cada tipo de personalidad analizada, se elaboran puntuaciones y clasificaciones por parte de los sistemas de protección crediticia, además del uso de dichas tecnologías en la policía predictiva, así como en otros análisis de tendencias mediante la “big data analytics” [analítica de big data] para construir estrategias individualizadas, además del enorme potencial para influir e incluso manipular comportamientos individuales y colectivos.

Por ello, se produce el análisis del comportamiento humano por parte de la inteligencia artificial (biometría comportamental), perfilando conductas, que pueden llevar a situaciones de segmentación, imponiendo decisiones basadas en dichas clasificaciones, con una importante reducción de la libertad de elección individual, y una mayor homogeneización de las personas. Son los llamados filtros de burbuja, como señala Eli Pariser.

Como ejemplo paradigmático de esta práctica, destaca el gobierno chino, que realiza la elaboración de perfiles para la atribución de puntos en su sistema de crédito social, teniendo en cuenta todas las actividades sociales de los individuos, en un sistema de vigilancia panóptica total, que influirá incluso en el derecho a la libre circulación, ya que existen sanciones como, por ejemplo, la prohibición de la compra de billetes de avión y tren.

Al hablar de la IA, uno de los puntos principales es la gobernanza de los algoritmos, que debe enfrentarse a la cuestión de transformar e incorporar principios éticos y jurídicos abstractos en prácticas concretas. Entre las prácticas de gobernanza tenemos la elaboración de la «evaluación de riesgos», y el «diseño de la IA», y podemos destacar como ejemplos de Informes de Impacto de Algoritmos: la lista de evaluación del Grupo de Expertos de Alto Nivel sobre IA de la Comisión Europea para una IA digna de confianza, el Center for Information Policy Leadership (CIPL), la Plataforma para la Sociedad de la Información (Países Bajos), el Responsible AI Impact Assessment Toll (RAIIA) (International Technology Law Association). (Cantarini, 2020; Cantarini y Guerra, 2020)

La gobernanza de los algoritmos debe basarse en algunas preguntas esenciales para evitar el determinismo tecnológico, como, por ejemplo: ¿es posible identificar los riesgos que la tecnología elegida intenta abordar? ¿Es posible garantizar las entradas del sistema (exactitud de los datos e idoneidad de la metodología)? ¿Se pueden justificar los resultados del sistema y cómo repercutirán en las políticas, en la práctica y en las relaciones con la comunidad? ¿Puede probarse la tecnología, garantizando la responsabilidad y cierta medida de transparencia? ¿La política de uso de la tecnología respeta la autonomía de las personas a las que va a afectar?

En relación con esta cuestión, el Tribunal de Primera Instancia de La Haya emitió una decisión judicial el 20/02/2020, relativa al sistema deno-

minado SyRI en los Países Bajos (Systeem Risico Indicatie), con el fin de detectar fraudes como la evasión y la percepción indebida de prestaciones de la seguridad social, a partir de la construcción de perfiles de riesgo de los sospechosos, mediante el uso de bases de datos de comportamiento registradas en varios sistemas gubernamentales. Se entendía que el sistema no establecía las salvaguardias exigidas por el Convenio Europeo de Derechos Humanos (art. 8, §2), en cuanto a la ponderación entre los intereses sociales y los posibles daños a las libertades y derechos fundamentales. Se entendió que podría no cumplir los requisitos de proporcionalidad y transparencia, ya que no se hizo público el tipo de algoritmo utilizado en el modelo de clasificación de riesgos, ni se facilitó información sobre el método de análisis de riesgos utilizado.

La centralización de Internet, potenciada por el uso a gran escala de la IA, dio lugar a los fenómenos denominados plataformaización, datificación y performatividad algorítmica. Representa el dominio y la centralización del mercado por parte de los llamados “cinco grandes”, Google, Amazon, Facebook, Apple y Microsoft, lo que genera una centralización del control en una especie de nuevo panóptico digital, que pondría en jaque las ideas relacionadas con el origen de Internet, es decir, la libertad, la emancipación y la democratización del conocimiento.

Cada vez más, se producen decisiones automatizadas en varios sectores sociales, creando perfiles de comportamiento, una práctica conocida como “profiling”, que puede aplicarse tanto a individuos como a grupos, a partir de una enorme cantidad de datos personales obtenidos generalmente al utilizar Internet, sin un consentimiento válido, porque en este caso las políticas de privacidad y de cookies de varios sitios web son bastante problemáticas, con información fragmentada y en un lenguaje todavía difícil para la mayoría de la población, es decir, todavía lejos de la observancia de todos los principios relacionados con la legislación de protección de datos personales en lo que se refiere en particular a la transparencia y a los requisitos para un consentimiento libre, informado y válido.

Cuando se trata de decisiones automatizadas existe el riesgo de sesgo llamado “sesgo de máquina”, “sesgo algorítmico” o simplemente “sesgo”, y su eliminación es un campo de investigación en curso, que involucra el tema de la Inteligencia Artificial Explicable (XAI por sus siglas en inglés -Explainable Artificial Intelligence). Centra su atención en las auditorías de algoritmos y plataformas digitales, la construcción de conjuntos de datos, con el objetivo de dotar de una especie de control o regulación del uso y resultado de los algoritmos, equipos de revisión y supervisión independientes. Es necesario que los desarrolladores cuiden más sus prácticas respecto a la base de datos, y que realicen una selección detallada de los datos utiliza-

dos en la fase de aprendizaje (filtrado previo), prestando atención a la diversidad de los datos y a la diversidad en el equipo de desarrolladores, además de a la formación y las pruebas de las aplicaciones de IA, lo que implica un proceso continuo de “cumplimiento” u “observancia” [compliance].

Al hablar de la Inteligencia Artificial Explicable (XAI), en relación con la transparencia, el cumplimiento y la responsabilidad, así como los derechos de revisión, oposición e impugnación de una decisión automatizada, estamos hablando de la gobernanza algorítmica, es decir, de las formas de controlar las decisiones automatizadas. La transparencia está relacionada con el requisito de la confianza, es decir, la explicabilidad es un requisito de la fiabilidad. La confianza es una parte integral del concepto del modelo de Inteligencia Artificial Explicable. Además de la transparencia sobre la recogida y el tratamiento de los datos, se exige la transparencia del diseño tecnológico (el diseño técnico) y de los algoritmos utilizados en cada caso.

Sin embargo, la protección de los secretos comerciales es contraria al deber de divulgación. La transparencia involucra dos aspectos, la accesibilidad y la comprensibilidad, y la cuestión de romper el código fuente, como respuesta a la demanda de transparencia del algoritmo, estaría relacionada sólo con el primer punto, siendo insuficiente, según parte de la doctrina, para predecir el comportamiento de los algoritmos, ya que el código se limitaría a exponer el método de aprendizaje automático utilizado y no la regla de la decisión tomada, lo que llevaría a la “falacia de transparencia”. (Ferrari Becker, 2019)

Además de la explicabilidad “post-hoc”, sería imprescindible que la explicabilidad estuviera ya pensada desde la fase de diseño, implicando una explicación por medio del diseño o una explicación por defecto, teniendo siempre en cuenta al público al que va dirigido, y en particular, buscando la comprensión de un profano, ya que hasta ahora la mayoría de las implementaciones del requisito de transparencia, ligado al requisito de una IA de confianza, no van dirigidas al usuario final, sino a los ingenieros que lo utilizan para repensar el modelo. Por lo tanto, existe una brecha entre la explicabilidad en la práctica y el objetivo de la transparencia.

Uno de los principales usos de la IA en el ámbito social es en el ámbito del reclutamiento y la selección de personal, denominado “reclutamiento inteligente”. Se utilizan a menudo plataformas tecnológicas especializadas en el proceso de selección, las denominadas HR Techs (Tecnologías para Recursos Humanos –RRHH–), de la mano de empresas que aportan soluciones tecnológicas para el sector de los RRHH, con el objetivo de conseguir una mayor eficiencia en dicho procedimiento. Buscan que los RRHH sean más estratégicos, ágiles, y evitando criterios subjetivos o personales para la elección de los candidatos. Alegan que se reducen los errores en la

selección de los candidatos con más afinidad con la vacante, mediante el análisis de una gran cantidad de datos personales de los candidatos, es decir, con el uso de la tecnología “big data” (se trata de la metodología conocida como “people analytics”), y la creación de perfiles automatizados de los candidatos y potenciales trabajadores.

El uso de software y mapeo de perfiles de comportamiento de los candidatos indica los más afectos a la cultura empresarial, y uno de los principales fundamentos de su uso es la reducción de los sesgos en la adopción de decisiones y una adopción de decisiones más asertiva, a través de la personalización del proceso de selección, con mayor neutralidad y objetividad.

Además de los datos personales, también es posible obtener datos sensibles a partir de inferencias sobre el estado emocional de las personas analizadas, por ejemplo, a partir del análisis de los patrones de escritura, las expresiones faciales y el discurso de los candidatos, lo que hace posible trazar incluso el perfil y la orientación sexual, y verificar problemas de salud, entre otros aspectos.

2. ENFOQUE A TRAVÉS DE LOS DERECHOS FUNDAMENTALES Y EL CONSTITUCIONALISMO DIGITAL

Resulta especialmente urgente que en los temas relacionados con la IA se preste atención a los aportes del Constitucionalismo Digital, tal como señalan autores como Giovanni di Gregorio, de la Universidad Bocconi, Gilmar Mendes, Edoardo Celeste, Claudia Padovani y Mauro Santaniello. Enfatizan la necesidad de postular el uso de estructuras y mecanismos del constitucionalismo moderno para orientar la gobernanza del espacio digital, a fin de asegurar un mejor equilibrio en los ámbitos de las relaciones jurídicas y la protección de los derechos fundamentales.

En este sentido, Gunther Teubner (1989) llevaba a cabo una importante contribución al indicar que la formulación tradicional del enfoque de los derechos fundamentales, bajo una perspectiva individualista de equilibrio entre los derechos individuales de los actores privados en la esfera digital, estaría desfasada, ya no es suficiente, y se debe abordar la dimensión colectiva-institucional de los derechos fundamentales. Y es que las redes sociales, como Facebook con su Consejo de Supervisión o “Tribunal Supremo de Facebook”, estarían ejerciendo una verdadera función normativa.

Wolfgang Hoffmann-Riem (2021 y 2020) señala la importancia, en el ámbito de la Inteligencia Artificial, de los derechos fundamentales y la proporcionalidad para hacer compatible la protección aunque sin impedir la innovación, y sugiere la propuesta de “responsabilidad por la innovación”,

o “forzamiento de la innovación”. Se trata de la definición normativa de objetivos o normas que todavía no pueden cumplirse con el modelo de desarrollo actual, pero que es plausible que se cumplan en el futuro. Si no se produce dicha implantación en un determinado plazo, se debería abandonar el desarrollo y uso de la aplicación de IA en cuestión.

Así lo destaca también Laura Mendes: “El profesor Hoffmann-Riem nos enseña que la preocupación por la preservación y actualización de los derechos fundamentales debe ser constante, considerando el Derecho como un instrumento para limitar los poderes y regular la innovación, de acuerdo con los objetivos y valores establecidos en el ordenamiento jurídico, especialmente los principios constitucionales”.

Se destaca también la importancia de la transparencia para la rendición de cuentas, pues permite un control externo efectivo, fundamental para una co-rendición de cuentas democrática, y se habla del “debido proceso digital”. Por lo tanto, tenemos que pensar en el derecho a la explicación con una aplicación más amplia, con mayores implicaciones legales para el diseño, el prototipo, la investigación de campo, y más allá de su enfoque individual, como una forma colectiva de supervisión.

El constitucionalismo digital tiene un carácter pre- o proto-constitucional ya que se refiere a reacciones normativas difusas que no se limitan al ámbito del Estado-nación, y se centran en la protección de los derechos digitales, la limitación del ejercicio del poder en y a través de las obras de la red digital y la formalización de los principios de gobernanza para Internet. Aunque algunas regulaciones formales en Internet se sitúan en un plano infra constitucional, presentan una verdadera naturaleza “pre” o “protoconstitucional”, ya que establecen bloques de interpretación de las constituciones formales en el ámbito digital. Sería algo así como el constitucionalismo “societario” (social o de sociedad) de Sculli, que Teubner adopta y desarrolla.

En un sentido complementario, podrían tenerse en cuenta las aportaciones de Theodor Viehweg al reflexionar sobre la conjugación de la dogmática jurídica con la cetética pues amplían el contenido y la comprensión de lo que se considera ciencia jurídica. Ya no es suficiente pensar en una ciencia jurídica en sentido estricto (dogmática jurídica), sino que debe ampliarse su enfoque para pensar en una ciencia jurídica en sentido amplio, que abarque la cetética (Ferraz Júnior 2016, pp. 21 y ss.).

Desde tales consideraciones se llega al entendimiento de que el sistema jurídico no es cerrado y autosuficiente, sino un sistema abierto, salpicado de lagunas que serán llenadas ante el caso concreto, lo que subraya la importancia de los procedimientos para la concreción del Estado Democrático de Derecho, que depende de ellos, y la consagración, por lo que a Brasil

por ejemplo se refiere, de la Constitución Federal con carácter procesal. De este modo, tales procedimientos se muestran como la forma más adecuada de resolver los complejos e inéditos problemas de la posmodernidad. Vale la pena recordar que el sistema social autopoiético (en la formulación de Luhmann, a partir de la cual, además de Luhmann, autores como el mencionado Teubner y, en Brasil, Willis Santiago Guerra Filho y Marcelo Neves, entre otros, vienen pensando en el derecho), es considerado un sistema al mismo tiempo cerrado (normativamente) y abierto (cognitivamente).

3. EL DISEÑO TECNOLÓGICO Y EL ENFOQUE DE DERECHOS HUMANOS Y FUNDAMENTALES POR MEDIO DEL DISEÑO

La normativa más reciente de la Unión Europea supone un cambio de enfoque o paradigma: ya no se centra únicamente en la delimitación de una personalidad electrónica para algunas aplicaciones de IA, sino que aporta un enfoque basado en el riesgo. La propia protección de datos se caracteriza como una regulación del riesgo, implicando así el binomio gravedad y probabilidad. Se generaliza la defensa de la protección a través de la arquitectura de gestión de riesgos y de la regulación por medio de la tecnología, pasando de la autodeterminación informativa a los daños inducidos por la información.

En el ámbito de la protección de datos, uno de los fundamentos de este cambio de perspectiva es la fragilidad de una de las principales categorías jurídicas que autorizan el tratamiento de datos: el consentimiento. Se apunta a la ficción del consentimiento ante su fragilidad en la práctica. Así lo señala un importante informe de la Comisión Europea de 2017, en el que se destaca que el consentimiento se ha convertido en una pesada carga que hay que soportar y que no necesariamente proporciona la necesaria protección a los derechos fundamentales del interesado.

Junto al consentimiento (autodeterminación informativa, según la ley alemana, o libertad informática, según la ley española), tenemos también la responsabilidad digital, la ética digital y la rendición de cuentas, que implica el diseño responsable de los programas informáticos.

Es común recurrir a expresiones tales como el mito del consentimiento y la paradoja de la privacidad, o la fragilidad del consentimiento, debido a que nos hallamos ante contratos que solo implican la opción “todo o nada”, “tómalo o déjalo”. Aunque parece un procedimiento aparentemente inocuo o sólo una ficción, las consecuencias que pueden derivarse de él pueden ser poco claras y difíciles de identificar.

En este sentido, Juliana Abrusio (2020) señala la fragilidad del consentimiento como base jurídica para el tratamiento de datos en los casos de elaboración de perfiles y decisiones automatizadas, ya que dichas prácticas son poco transparentes. El consentimiento libre e informado resulta inviable, ya que no hay una comprensión real de los términos y las consecuencias jurídicas relacionadas. Se da incluso una verdadera “fatiga del consentimiento”, al ser prácticamente imposible el control y la autogestión de los datos personales.

Shoshana Zuboff (2020), por su parte, ha llamado a los acuerdos de términos de servicio “contratos de adhesión” porque imponen a los usuarios condiciones del tipo ya aludido “tómalo o déjalo”, y estos las aceptan, les gusten o no, y sin permitirse su modificación unilateral. A estos contratos opresivos se les denomina “click-wrap”, literalmente “envoltorio de clic”. Falsifican el consentimiento, que se presta sin ni siquiera leer el documento ni tener conocimiento del mismo, porque en muchos casos el simple hecho de navegar por una web ya conlleva la obligación de aceptar cláusulas contractuales. Se contraviene así el deber de transparencia y el derecho a la información, además de conllevar exenciones de responsabilidad de forma abusiva. Estos “contratos” forman parte de la dinámica y el sistema del capitalismo de datos, denominado por Zuboff como excedente de comportamiento.

El enfoque basado en el riesgo implica varios niveles de riesgo de las aplicaciones de la IA, como podemos observar en el Libro Blanco sobre la IA de la Unión Europea, de 19 de febrero de 2020, creado por la Comisión Europea, y en la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadoras en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos administrativos de la Unión, de 21 de abril de 2021.

La normativa europea (en especial el Reglamento General de Protección de Datos de 2016) también prevé la creación de instrumentos de regulación “ex ante”, como los códigos de conducta, las certificaciones, las auditorías independientes y la elaboración de documentos como los Informes de Evaluación de Impacto de Protección de Datos (EIPD o DPIA en sus siglas en inglés –Data Protection Impact Assessment–) y la Evaluación del Interés Legítimo para el tratamiento de datos (LIA por sus siglas en inglés –Legitimate Interests Assessment–). Las legislaciones deberían imponer la obligación de elaboración previa de dicho documento en el caso de riesgos altos o moderados para los derechos fundamentales, y también debería prever sus requisitos y procedimiento de elaboración. Esta es una debilidad encontrada tanto en la española Ley General de Protección de Datos (LGPD) como en el brasileño Proyecto de Ley (PL) 21/20, ya que prevén pocos instrumentos

de “cumplimiento”, además de incurrir en omisiones, especialmente en lo que se refiere a la obligación de elaboración previa del Informe de Impacto y su procedimentalización.

En el ámbito europeo, el Libro Blanco sobre la IA separa, a efectos normativos, entre IA de alto riesgo, de riesgo moderado y de bajo riesgo. En el caso de alto riesgo hay una serie de condiciones clave que deben observarse (solidez, precisión y supervisión humana y garantía de los derechos fundamentales). En el caso del bajo riesgo, existe el cumplimiento de las normas estándar, del etiquetado voluntario, una especie de certificación y etiquetado voluntario, o sello de calidad.

Por otro lado, la propuesta de Ley de Inteligencia Artificial también implica la llamada perspectiva de los derechos humanos por medio del diseño [del inglés “human rights by design”, o compromiso de diseñar tecnologías de IA respetuosas con los derechos humanos por defecto], los usos beneficiosos de la IA, la IA para el bien y la IA centrada en el ser humano, además de mantener la separación entre aplicaciones de riesgos inaceptables, con prohibición expresa, de alto, moderado y bajo riesgo, siguiendo la visión ya esbozada por el Reglamento General de Protección de Datos, de 2016 (RGPD), y la Comunicación de la Comisión “Inteligencia Artificial para Europa” de 25/04/2018 (COM/2018/237). Se buscaba por un lado, la protección de los derechos fundamentales, y por otro, no obstaculizar la innovación. A este respecto, una de las aplicaciones consideradas de riesgo inaceptable es la tecnología de vigilancia por reconocimiento facial, con la excepción del uso por parte de organismos gubernamentales para la práctica de la investigación de delitos graves.

Todos estos enfoques deben ser conjuntos para lograr una protección sistémica, a través de la hetero-regulación y la arquitectura técnica (“protección por medio del diseño”), teniendo en cuenta también la colectivización, es decir, desde el reconocimiento de la dimensión múltiple de los derechos fundamentales, que implica aspectos individuales, colectivos y sociales.

Así lo afirma por ejemplo Claudia Quelle, de la Universidad de Tübing, cuando señala que las metodologías en el caso de los enfoques basados en el riesgo estarán influenciadas por la teoría del equilibrio de los derechos fundamentales ante casos concretos. En este sentido, la Junta Europea de Protección de Datos (JEPD), al comentar los criterios de evaluación de riesgos para la elaboración de EIPD, en línea con el RGPD (considerando 84 y art. 35), señala la necesidad de observar un procedimiento que implique la evaluación de la necesidad y la proporcionalidad, al tratarse de posibles colisiones de normas de derechos fundamentales.

Uno de los principales temas objeto de reflexiones éticas y jurídicas en torno a las aplicaciones de la IA es la elevada posibilidad de que se produz-

can sesgos en el caso del reconocimiento facial y de las decisiones judiciales en el ámbito penal. A este respecto, varios estudios e investigaciones apuntan a la discriminación de raza, clase y género, es decir, discriminaciones múltiples, lo que precisa de un enfoque múltiple en términos de protección sistémica. Así se percibía ya en el enfoque clásico de Angela Davis en el libro *Mujeres, raza y clase*, cuando afirmaba que existe una intersección entre el feminismo, el antirracismo y la lucha de clases.

Angela Davis, a pesar de no haberse ocupado de las nuevas tecnologías, sí ha tratado en al menos dos libros —*Are Prisons Obsolete?* (2003) y *Abolition Democracy. Beyond Empire, Prisons, and Torture* (2011)— la cuestión del sistema de justicia penal estadounidense. Como activista contra la prisión, y habiendo sufrido ella misma las penurias del encarcelamiento, está comprometida con la eliminación de dicho sistema, lo que va más allá de la mera postulación de reformas en los establecimientos penitenciarios. Como señalaba la autora, hay un aumento exacerbado del número de prisiones, sobre todo de negros y latinos, y en 30 años, de 1960 a 1990, hay diez veces más presos en EE.UU., sin que, no obstante, haya una reducción correspondiente del nivel de delincuencia, es decir, su eficacia es cuestionable. Existe, pues, una industria del castigo, es decir, el encarcelamiento masivo genera beneficios y tiende a reproducir las condiciones que llevan a la gente a la cárcel.

Además de apostar por alternativas abolicionistas, como la desmilitarización de las escuelas y la revitalización de la educación, Davis justificaba la necesidad de una democracia de la abolición, que incluyera la eliminación de las prisiones. En la misma línea, Ruha Benjamin señalaba, usando la expresión “complejo industrial penitenciario” (PIC), que la libertad condicional y la vigilancia son la parte de la industria que más crece.

En este sentido, Ruha Benjamin (2019) se preguntaba al principio de su libro: “¿Qué hay en un nombre?”, al tiempo que señalaba la necesidad de reiniciar la solidaridad. Lo que está en cuestión implica una reflexión sobre el nombre de algo que engloba valores, creencias y narrativas, que transmite una historia; es decir, los nombres son códigos sociales y son culturales, no son neutrales sino que también están racializados y clasificados.

Corroboran esta afirmación las investigaciones que demuestran la repercusión de los nombres en la experiencia de las personas en el mercado laboral, ya que los solicitantes de empleo con nombres de pila que sonaban blancos acababan teniendo más oportunidades de ser llamados para el trabajo, en comparación con los solicitantes de empleo con nombres que sonaban negros. Así lo señala Latanya Sweeney en una línea similar, afirmando que existe una elevada asociación de las búsquedas en línea de

“nombres negros” con los registros de arrestos, en comparación con los “nombres blancos”.

Según Ruha Benjamin, además del etiquetado en el ámbito penal basado en las predicciones realizadas por los algoritmos de la IA, existen otras correlaciones a partir de clasificaciones y perfiles, como en los ámbitos del empleo, la educación, la salud y la vivienda. Esta circunstancia limita o restringe las oportunidades de las personas, por lo que los diseñadores técnicos, aunque posiblemente de forma inconsciente, están erigiendo un sistema de castas digital.

Es lo que la autora llama el “Nuevo Código Jim”, consistente en el empleo de nuevas tecnologías que reflejan y reproducen las desigualdades existentes, pero que se promueven y perciben como más objetivas y neutrales que los sistemas discriminatorios anteriores. El término “New Jim Code” se basa en el libro *The New Jim Crow: Mass Incarceration in the Age of Colorblindness* (2012), de Michelle Alexander. En él se argumenta cómo el sistema penitenciario estadounidense ha producido un “nuevo sistema de castas raciales” a través de la discriminación legalizada. Es decir, la codificación cultural racista o discriminatoria se incorpora a la codificación técnica de las aplicaciones de IA.

Como han venido señalando varios autores, lejos de ser neutrales y tener objetividad, los algoritmos de IA tienen sesgos. Esto se enfatiza en el libro de Frank Pasquale (2015), donde se cuestiona la protección de los “algoritmos secretos” a través de la protección de los secretos comerciales, considerados la mayoría de las veces de forma absoluta, incluso ante casos concretos de colisión con otros derechos fundamentales, y sin realizar la necesaria ponderación entre ellos.

En definitiva, es necesario “tomarse en serio” los derechos, especialmente los fundamentales, aun reconociendo que ninguno tiene carácter absoluto, y que han de considerarse en su relación con otros, si realmente queremos preservar la dignidad humana, y preservarla absolutamente, en la medida en que solo ella es absoluta; no es un derecho fundamental sino la razón de ser de todos ellos.

Esta situación también la señalan Cathy O’Neil (2021), Virginia Eubanks (2017), Safiya Umoja Noble (2018), y Latanya Sweeney, profesora de la Universidad de Harvard.

Por lo demás, existen varios casos de sesgo en los modelos de predicción policial que utilizan IA, como ha constatado el Grupo de Análisis de Datos sobre Derechos Humanos— HRDAG por sus siglas en inglés—, al referirse a la herramienta de software de predicción policial denominada “PredPol” [busca predecir dónde se van a producir delitos], la cual no cum-

ple los requisitos de imparcialidad y neutralidad, y duplica el racismo y la discriminación existentes en la sociedad.

Varios estados de EE.UU. también propician un castigo eterno para el ex convicto, al prever que no pueda votar, ni siquiera después de cumplir su condena; una pena extra injustificada y sin remisión, a lo que también se añade la casi imposibilidad de encontrar un nuevo trabajo, lo que a su vez le condena a la marginalidad y a un retorno perpetuo al sistema penitenciario¹.

Para Ruha Benjamin existe una paradoja en dicho sistema: la discriminación legalizada que proporciona el sistema penal estadounidense. La propia raza es una tecnología diseñada para separar, estratificar y santificar las numerosas formas de injusticia que sufren los miembros de los grupos racializados.

Es a lo que se refería Foucault (2012, 2010, 2005, 1972)² cuando señalaba el nacimiento del biopoder y la sociedad de la normalización, propios del siglo XIX y basados en el elemento de la norma. Esta va a circular entre el poder disciplinario y el poder normativo, yendo a aplicarse, del mismo modo, al cuerpo y a la población, lo que permite a la vez controlar el orden disciplinario del cuerpo y los acontecimientos aleatorios de una multiplicidad biológica. La norma, pues, puede aplicarse tanto a un cuerpo que se quiere disciplinar como a una población que se quiere regular.

Así aparece un nuevo tipo de racismo, ahora insertado en los mecanismos del Estado, debido al surgimiento del biopoder. Permite efectuar una división entre las personas, en busca de una vida pura y saludable, a través de la eliminación del peligro biológico, directamente ligado a la eliminación de la propia especie o de la raza que la representa. La raza, el racismo, es la condición de aceptabilidad de la vida en una sociedad de normalización, siendo indispensable para el biopoder.

El racismo moderno está intrínsecamente ligado al biopoder, y es, como ya se ha dicho, un fenómeno típico de finales del siglo XIX. Por lo

-
1. Cathy O'Neil, en su libro "Weapons of Math Destruction" (Armas de destrucción matemática), también comenta un sistema de IA llamado "Impact" (modelo de valor añadido) que implica la clasificación de los profesores de la escuela pública y su despido masivo, con un sesgo discriminatorio contra los profesores, por ejemplo, que prueban métodos de enseñanza alternativos y un incentivo para los profesores que tienen conductas que manipulan el sistema sólo para asegurar una buena evaluación y puntuación.
 2. Foucault señala: "Me parece que uno de los fenómenos fundamentales del siglo XIX fue, y lo que podría llamarse, la asunción de la vida por el poder: si se prefiere, una toma de poder sobre el hombre como ser vivo, una especie de estatización de lo biológico o, al menos, una cierta inclinación que conduce a lo que podría llamarse estatización de lo biológico" (2012, pp. 285 y ss.).

tanto, está vinculado al funcionamiento de un Estado que está obligado a utilizar la raza, la eliminación de las razas y la purificación de las mismas para ejercer su poder soberano. De este modo, se produce la generalización absoluta del biopoder, y este fenómeno estaría efectivamente inscrito en el funcionamiento de todos los Estados de la época contemporánea.

Hay pues una inequidad codificada que propicia infraestructuras injustas, pero que también provoca contra-codificaciones basadas en la solidaridad y que replantean la justicia: es decir, un enfoque emancipador de la tecnología. Este es el cambio paradigmático necesario, y siendo necesario, tiene que ser posible y debe perseguirse con el máximo esfuerzo.

Donde hay poder hay resistencia, como decía Foucault, y podemos considerar como formas de resistencia (Foucault, 2006, pp. 241 y ss.), como herramientas abolicionistas del Nuevo Código Jim —las formas de resistir y desafiar el Nuevo Código Jim—, y contribuir a un enfoque de la tecnología que vaya más allá del acceso a nuevos productos, para defender prácticas de diseño orientadas a la justicia.

Esto es lo que sostiene Ruha Benjamin, que destaca la iniciativa “Appolition” como una tecnología con un ethos emancipador, una herramienta de solidaridad que dirige las donaciones hacia el pago de fianzas para los presos que no tienen esa posibilidad, como los negros y los latinos.

La lucha debe ser conjunta, en primer lugar, para aliviar los marcos de desigualdad codificada, mirando el diseño de tal manera que se replantee para que deje de seguir simplemente la línea dominante, con el fin de lograr la “justicia del diseño”, abarcando en su metodología un enfoque inclusivo y democrático. No sólo en el sentido de un “diseño centrado en el ser humano”, sino centrado en los “derechos humanos y fundamentales por medio del diseño”; e incluso en una “IA centrada en el planeta”, ecológica, en los términos de Luciano Floridi, o, mejor aún, ecosófica, como propone Michel Maffesoli en su obra recientemente publicada entre nosotros, *Ecosofía* (2021)

Para desarrollar dicha metodología es fundamental observar las posibles vulneraciones de los derechos fundamentales en sus múltiples dimensiones, es decir, no sólo en el aspecto individual, sino también en el colectivo, social y ecológico, para obtener una protección sistémica. A fin de hacer posible este enfoque basado en los derechos humanos y fundamentales, es esencial que la metodología incorporada en el diseño técnico refleje una ponderación de los derechos fundamentales, en la línea de lo que elaborarían los instrumentos de cumplimiento [compliance], tales como la evaluación o el informe de impacto de la inteligencia artificial o algorítmica, centrados en dichos derechos.

A este respecto, el concepto de “justicia en el/del diseño” [design justice] fue desarrollado por Sasha Constanza-Chock, profesora del MIT, con colaboradores de la Allied Media Network, con el objetivo de repensar la dinámica del diseño de las tecnologías, a través de múltiples ejes de opresión, haciendo hincapié en la inclusión en el sentido de exigir la participación de las comunidades marginadas en todas las etapas del proceso de diseño tecnológico.

Un segundo punto fundamental de este enfoque de protección sistémica se refiere a la revisión de algunos documentos internacionales, en cuanto al enfoque de la protección de los derechos fundamentales.

Cuando observamos la mencionada propuesta de Ley de Inteligencia Artificial de la Unión Europea, a pesar de, como también se dijo, sus conexiones con un enfoque de derechos humanos por medio del diseño, los usos beneficiosos de la IA, una IA para el bien y una IA centrada en el ser humano, es decir, de un enfoque centrado en el ser humano, lo que sitúa en el eje a la persona y la dignidad humana, se detectan algunos fallos y omisiones. A pesar de su planteamiento en torno a los riesgos (tratando de no impedir la competencia internacional y la innovación, pero con prohibición expresa de las aplicaciones de IA de riesgo inaceptable, por atentar contra los valores de la Unión Europea), se comprueba que el uso de tecnologías de vigilancia (sistemas de identificación biométrica a distancia, en tiempo real, y en espacios accesibles al público), aunque prohibido, tiene algunas excepciones, como en el caso de la investigación de delitos graves. En otras palabras, el problema señalado por las autoras Angela Davis y Ruha Benjamin quedaría sin resolver.

Las excepciones se basan en razones de interés público y abarcan: la búsqueda de posibles víctimas de delitos, incluidos los niños desaparecidos, las amenazas a la vida o a la seguridad física de las personas físicas o las amenazas de atentados terroristas, y la detección, el rastreo, la identificación o el enjuiciamiento de delincuentes o de personas sospechosas de haber cometido un delito, tal y como se indica en la Decisión marco 2002/584/JAI del Consejo, de 13 de junio de 2002, relativa a la orden de detención europea y a los procedimientos de entrega entre Estados miembros, siempre que sean sancionables en el Estado miembro en cuestión con una pena o una medida de seguridad privativa de libertad por un período máximo de al menos tres años. Entre los delitos contemplados en la Decisión marco de 2002 están la participación en una organización delictiva, el terrorismo y la trata de seres humanos.

Aunque las solicitudes de usos de alto riesgo están sujetas al cumplimiento de normas y obligaciones estrictas, con especial atención a los requisitos relativos a la alta calidad de los datos, la documentación y la

trazabilidad, la transparencia, la supervisión humana, la precisión y la solidez, también se prevé en la propuesta de Ley de Inteligencia Artificial la preparación de una evaluación de conformidad “ex ante”. La fragilidad del documento queda reflejada cuando se permite que, por regla general, lo elabore el propio proveedor, salvo en el caso de los sistemas de IA destinados a la identificación biométrica remota de personas. La propia excepción prevé esta fragilidad, ya que, en el único caso de la excepción, exigiría como requisito la participación de un organismo notificado, que debe estar sujeto a una serie de requisitos, concretamente en términos de independencia, competencia y ausencia de conflictos de intereses.

Esto es lo que señala el Instituto Ada Love Lace, en el documento titulado *Algorithmic impact assessment: user guide*, de 8 de febrero de 2022, cuando afirma que es esencial respetar uno de los componentes constitutivos de una evaluación de impacto, y de necesaria inclusión en cualquiera de estas evaluaciones de impacto en la IA: a saber, la necesidad de una “fuente de legitimidad”, es decir, que dichas evaluaciones sean realizadas por otra estructura organizativa, institucional, como una agencia gubernamental. Es decir, una evaluación controlada y determinada por quienes toman las decisiones del proceso algorítmico carece de legitimidad, porque dicha evaluación también puede aportar sesgos.

Por lo tanto, además de un cambio de enfoque en el diseño tecnológico, y pensando en el largo plazo, de manera sostenible, tal vez en forma de un “diseño subversivo” en lugar del diseño dominante en el sentido de un proyecto colonizador, la regulación también debería revisar estos puntos de debilidad, llevando el foco a la protección sistémica. En este sentido es necesaria una reformulación de lo que se ha tratado como protección de los derechos fundamentales en el contexto de la protección de datos y la inteligencia artificial, por ignorar en muchos casos la construcción epistemológica y metodológica de la teoría de los derechos fundamentales como teoría fundamental del derecho, lo que lleva a errores en lo que se refiere a la naturaleza de los derechos fundamentales, y a la correcta resolución de la colisión de las normas de derechos fundamentales, a través del principio de proporcionalidad, que debe ser objeto de una mejor comprensión.

Aunque se ha producido un cambio en cuanto al marco de análisis de los instrumentos de gobernanza, es decir, la evaluación del impacto de los algoritmos, poniendo ahora el foco en los derechos y valores fundamentales en los mismos, y no sólo en las tecnologías sino también en los intereses difusos y colectivos, debería ampliarse dicha protección en el sentido de una adecuada y responsable ponderación de estos derechos fundamentales cuando entran en colisión.

Entre los documentos más importantes que han asumido este enfoque se encuentran *Unboxing AI: 10 steps to protect Human Rights* del Consejo de Europa; el Informe del Relator Especial de las Naciones Unidas para la Promoción y Protección del Derecho a la Libertad de Expresión y Opinión sobre la IA y su impacto en las libertades; y la Recomendación del Consejo sobre Inteligencia Artificial de la OCDE, que prevé una “diligencia debida” con la aplicación de la estrategia de “derechos humanos por medio del diseño”, en cada paso del desarrollo de la aplicación de la IA.

Algunos problemas plantea no obstante el documento titulado *Directrices éticas para una IA fiable*, elaborado por el grupo de expertos de alto nivel de la Comisión Europea en materia de Inteligencia Artificial, publicado en 2018, por incorporar un malentendido básico con respecto a los derechos fundamentales. El documento afirma, al comentar los conflictos entre principios (p. 16): “54) Cabe la posibilidad de que surjan tensiones entre los principios anteriores, y no existe una solución establecida para resolverlas”.

Y continúa:

En consonancia con el compromiso fundamental de la UE con la participación democrática, el respeto de las garantías procesales y la participación abierta en la esfera política, deberían establecerse métodos que posibiliten un debate sobre dichas tensiones. (...) Pese a que estos principios ofrecen ciertamente una orientación para la búsqueda de soluciones, no dejan de ser prescripciones éticas abstractas. (...) Pese a todo, pueden existir situaciones en las que no sea posible identificar compensaciones aceptables desde el punto de vista ético. Determinados derechos fundamentales y principios correlacionados son de carácter absoluto y no pueden ser objeto de un ejercicio de búsqueda de equilibrio (es el caso, por ejemplo, de la dignidad humana).

Sin embargo, no se trata de derechos fundamentales absolutos, ya que las normas que los consagran están en el mismo nivel de igualdad (en abstracto) que las normas constitucionales con naturaleza jurídica de principios constitucionales, en un estado de tensión mutua, lo que refleja precisamente una característica de los principios: su relatividad, a diferencia de los valores, que son absolutos. Por lo tanto, aunque el documento contempla principios en el sentido de directrices éticas, y no tiene fuerza normativa, lo cierto es que tales principios y valores éticos ya están consagrados en los países democráticos como derechos fundamentales, con eficacia.

Por lo tanto, no hay que hablar de inexistencia de solución en el caso de conflictos entre principios o derechos fundamentales, ya que existe una forma responsable de resolución, a través de la ponderación, que aporta un

procedimiento objetivo y racional, mediante la aplicación del principio de proporcionalidad, como en otros casos de colisiones entre normas de derechos fundamentales fuera del contexto digital. A través del procedimiento basado en la proporcionalidad, desarrollado a partir de la doctrina y la jurisprudencia del TEDH, ampliamente difundida en Europa y en el mundo, se desarrolla una aplicación contextualizada, adecuada para equiparar mejor los nuevos y singulares problemas que surgen en la posmodernidad, y en este sentido, se relaciona con un concepto dinámico de justicia.

No es posible sostener una solución para las colisiones de derechos fundamentales, en los casos actuales, fuera del entorno digital, a través de la aplicación de la ponderación, siguiendo la mejor doctrina y jurisprudencia, e ignorando tales aportes epistemológicos y metodológicos cuando se habla de protección de datos e inteligencia artificial. Hay buenas prácticas ya consolidadas en el tratamiento de otros problemas, todavía adecuadas para hacer frente a los nuevos.

Por lo demás, un importante documento internacional, que apunta en la dirección de ratificar las presentes notas, es el documento *Recomendación sobre la ética de la inteligencia artificial* publicado por la UNESCO³ en el que se establece que en caso de tensiones entre principios debe realizarse una evaluación contextual para gestionar las posibles tensiones, teniendo en cuenta el principio de proporcionalidad y de acuerdo con los derechos humanos y las libertades fundamentales.

4. OBSERVACIONES FINALES

Nos propusimos repensar los desafíos y las oportunidades para la gobernanza de los algoritmos de IA, no solo en el contexto brasileño, a largo plazo y de forma sostenible, para hacer compatible la innovación con una adecuada protección de los derechos fundamentales y humanos de toda la población. Para que se produzca una protección sistémica, a través de un sistema de protección proactivo, integral y sistémicamente seguro, es imprescindible reconocer las múltiples dimensiones de los derechos fundamentales, y otra perspectiva de los derechos humanos, vinculada a los conceptos de multiculturalidad, ética digital intercultural, tecnodiversidad y las epistemologías del sur, huyendo de visiones antropocéntricas, etnocéntricas o eurocéntricas, y reconociendo la diversidad cultural y social. Por lo tanto, hay que tener en cuenta una base normativa subyacente para

3. <https://en.unesco.org/artificial-intelligence/ethics>

el proyecto de gobernanza de la IA, con el fin de permitir la construcción de modelos orientados a la justicia social y a una visión sistémica, inclusiva y democrática, centrándose no sólo en la “IA centrada en el ser humano”, sino también en la “IA centrada en el planeta”, ya que también considera los impactos ambientales.

Se puede ver, pues, como ya se ha destacado, que efectivamente donde hay poder también hay resistencia, como en el caso de los contraargumentos sobre las dimensiones sociales y políticas del Nuevo Código Jim, instrumentos para combatir la inequidad codificada, llamados herramientas abolicionistas, fundamentados en un enfoque emancipatorio más holístico, y orientados hacia la justicia y la solidaridad, con los que nos alineamos y entendemos deben prevalecer, en beneficio de todos nosotros.

Una de las iniciativas consideradas como herramienta abolicionista es la Liga de la Justicia Algorítmica, que lanzó el “Compromiso de Seguridad Facial”, exigiendo a las empresas que adopten una postura pública” para mitigar el abuso de la tecnología de análisis de reconocimiento facial”. Con este compromiso, se limita el uso letal de la tecnología, el uso policial sin ley, además de exigir transparencia en cualquier uso por parte del gobierno, e incluir compromisos radicales como mostrar el valor de la vida, la dignidad y los derechos humanos.

Se trata de la necesidad de re-imaginar la ciencia y la tecnología con fines libertarios, reconociendo la interdependencia y, en este sentido, hablar de “co-liberación”, aplicando también dicha perspectiva al diseño tecnológico, en el sentido de lograr un diseño subversivo. Se trata de hallar imaginarios socialmente justos, que impliquen un enfoque socialmente consciente del desarrollo tecnológico, que exigiría dar prioridad a la equidad sobre la eficiencia, al bien social sobre los imperativos del mercado, y que se centra en la preocupación por la justicia, como parte del camino hacia una revolución de valores y un despertar democrático radical que la humanidad necesita desesperadamente para su supervivencia y la del planeta del que depende.

En este sentido de imaginar nuevas posibilidades para la relación hombre-algoritmo y de repensar la relación entre naturaleza y cultura, destacan la técnica y también la poética, abriéndose un campo de nuevas posibilidades para una poética negra, en su relación con la tecnología, como una poética del futuro.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abrusio, Juliana (2020). *Proteção de Dados na Cultura do Algoritmo*. Editora D'Plácido.
- Becker, Isabela Ferrari (2019). Accountability de Algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva, ITSRIO, <https://itsrio.org/wp-content/uploads/2019/03/Isabela-Ferrari.pdf>.
- Benjamin, Ruha (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press.
- Cantarini, Paola (2021). *O teatro Filosófico de Foucault e o Direito*. Tesis de Doctorado en Filosofía. PUCSP.
- Cantarini, Paola (2020). *Teoria Fundamental do Direito digital: uma análise filosófico-constitucional*. Clube de Autores.
- Cantarini, Paola y Guerra Filho, Willis S. (2020). *Teoria inclusiva dos direitos fundamentais e direito digital*. Clube de Autores.
- Davis, Angela (2011). *Are Prisons Obsolete?* Seven Stories Press.
- Davis, Angela (2005). *Abolition Democracy. Beyond Empire*. Prisons, and Torture, Seven Stories Press.
- Deleuze, G. (2013). *Conversações*, tradução de Peter Pál Pelbart, São Paulo: Editora 34.
- Eubanks, Virginia (2017). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- Ferraz Júnior (2016). *Tércio Sampaio. Teoria da Norma Jurídica*. Editora Atlas.
- Floridi, Luciano (2021). The end of an era: from self-regulation to hard law for the digital industry (November 9, 2021). Available at SSRN: <https://ssrn.com/abstract=3959766> or <http://dx.doi.org/10.2139/ssrn.3959766>.
- Foucault, M. (2012). *Em defesa da sociedade*. WMF Martins Fontes; 2.ª edição.
- Foucault, M. (2010). *Nascimento da biopolítica*. Edições 70.
- Foucault, M. (2006). *A hermenêutica do sujeito; cursos ministrados no Collège de France (1981-1982)*. São Paulo: Martins Fontes.
- Foucault, M. (2005). *História da sexualidade I: a vontade de saber*. Trad. e org. de Roberto Machado. 16. ed., Rio de Janeiro: Graal.
- Foucault, M. (1972). *A arqueologia do saber*. Petrópolis/Lisboa, Vozes/Centro do Livro Brasileiro.
- Jonas, Hans (2015). *O princípio responsabilidade: ensaio de uma ética para a civilização tecnológica*. Rio de Janeiro: Contraponto.
- Han, Byung-Chul (2022). *Non-things: Upheaval in the Lifeworld*. Polity.
- Hoffmann-Riem, Wolfgang (2021). *Teoria Geral do Direito Digital*. Forense.
- Hoffmann-Riem, Wolfgang (2020). Big Data e Inteligência Artificial: Desafios para o Direito, *Revista Estudos Institucionais*, v. 6, n. 2, p. 431-506.
- Maffesoli, Michel (2021). *Ecosofia*. Edições Sesc.
- Mittelstadt, Brent Daniel et al. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 1-21.

- Noble, Safiya Umoja (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- O’Neil, Cathy (2021). *Algoritmos de Destruição em Massa*, Editora Rua do Sabão.
- Pasquale, Frank (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Santos, Boaventura de Sousa (2003). *Reconhecer para libertar. Os caminhos do cosmopolitismo multicultural*. Rio de Janeiro: Editora Civilização Brasileira.
- Santos, Boaventura de Sousa y Meneses, Maria Paula (orgs.) (2010). *Epistemologias do Sul*. São Paulo: Editora Cortez.
- Teubner, Gunther (1989). *O direito como sistema autopoiético*. Tradução e prefácio de Jose Engrácia Antunes, Lisboa: Fundação Gulbenkian.
- Zuboff, Shoshana (2020). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

