

Desafíos filosóficos e
implicaciones morales de las
narrativas sobre la Inteligencia
Artificial

JONATHAN PIEDRA ALEGRÍA



DOCTORADO EN FILOSOFÍA
UNIVERSIDAD DE SALAMANCA

UNIVERSIDAD DE SALAMANCA
DOCTORADO EN FILOSOFÍA

Desafíos filosóficos e implicaciones morales de las
narrativas sobre la Inteligencia Artificial

Philosophical challenges and moral implications of
narratives on Artificial Intelligence



Compendio de publicaciones presentado como
TESIS DE DOCTORADO

por

JONATHAN PIEDRA ALEGRÍA

en el área de

«FILOSOFÍA MORAL Y POLÍTICA»

en cumplimiento de los requisitos establecidos por el Programa de
Filosofía para obtener el grado de

DOCTOR

por la

UNIVERSIDAD DE SALAMANCA

Vº Bº de la

Directora:

DRA. MARIA GONZÁLEZ NAVARRO



Salamanca, Junio del 2022

Cita recomendada:

Piedra, Jonathan (2022). *Desafíos filosóficos e implicaciones morales de las narrativas sobre la Inteligencia Artificial*. Tesis Doctoral. Universidad de Salamanca.

Recommended citation:

Piedra, Jonathan (2022). *Philosophical Challenges and Moral Implications of Narratives on Artificial Intelligence*. Tesis Doctoral. Universidad de Salamanca.

Dra. María G. Navarro. Profesor Contratado Doctor del Departamento de Historia del Derecho y Filosofía Jurídica, Moral y Política de la Universidad de Salamanca.

CERTIFICA

Que la tesis doctoral que lleva por título “Desafíos Filosóficos e implicaciones morales de las narrativas sobre Inteligencia Artificial” ha sido realizada por D. Jonathan Piedra Alegría para optar al título de Doctor por la Universidad de Salamanca en el marco del Programa de Doctorado en Filosofía por lo que autorizo su presentación mediante el procedimiento denominado “Presentación de la Tesis Doctoral en la USAL en el formato de compendio de Artículos/Publicaciones.”

Firmado digitalmente por GONZALEZ NAVARRO MARIA - 33528066T
Nombre de reconocimiento (DN): c=ES,
serialNumber=IDCES-33528066T, givenName=MARIA, sn=GONZALEZ
NAVARRO, cn=GONZALEZ NAVARRO MARIA - 33528066T
Fecha: 2022.06.10 13:34:12 +02'00'
Versión de Adobe Acrobat: 2022.001.20117

Fdo. Dra. María G. Navarro

Firmado por JONATHAN JONAS PIEDRA ALEGRIA (FIRMA)
PERSONA FISICA, CPF-01-1249-0290.
Fecha declarada: 12/06/2022 09:59 PM

Fdo. Doctorando Jonathan Piedra Alegría

La presente tesis doctoral está elaborada en el formato de compendio de publicaciones, según la normativa aprobada por la Comisión de Doctorado y Posgrado de Universidad de Salamanca. A continuación, se enlistan los cuatro artículos originales aportados en esta memoria.

ARTÍCULO I:

Piedra, Jonathan. (2021) “El Despertar de los Robots (sociales)”. *Luces en el camino: Filosofía y ciencias sociales en tiempos de desconcierto*. 1era Edición Dykinson S.L. Madrid. pp.29-51 ISBN 978-84-1377-322-3

Indicadores de calidad: Q1 según el SPI (Scholarly Publishers Indicators in Humanities and Social Sciences)

ARTÍCULO II:

Piedra, Jonathan. (2021). La Smartificación de la Administración Pública: apostillas críticas a partir del caso europeo. *Cuadernos Salmantinos de Filosofía* (Sección Monográfica: Filosofía e Inteligencia Artificial) Vol. 48, 235-250, ISSN: 0210-4857E-ISSN: 2660-9509 pp.235-250 DOI:10.36576/summa.144500

Indicadores de calidad: Q4 según SCIMAGO. Indexada en: CIRC, DICE, Dialnet, ERIHPLUS, INDICES CSIC, Latindex, MIAR, ProQuest, International Philosophy Bibliography, REDIB, Resh, UlrichswebTM, IBZ online, Dulcinea, REBIUN, ATLA, BASE (Bielefeld Academic Search Engine), Google Académico, Philpapers.org, HISPANA, OpenAIRE, OpenDOAR, RECOLECTA y WorldCat.

ARTÍCULO III:

Piedra, Jonathan. (2022). *Venditio fumi: Autorregulación Empresarial e Inteligencia Artificial*. *Sincronía*. Año XXVI / Número 81. Enero-junio DOI: 10.32870/sincronia.axxvi.n81.12a22

Indicadores de calidad: Indexada en Redalyc.org, DOAJ, AmeliCA, Latindex, Crue, REBIUN, FLACSO - Red Latinoamericana de Revistas Académicas en Ciencias Sociales y Humanidades, Latin REV, Dialnet, AURA, GenamicsTM, Revistaslatinoamericanas.org, Google Académico y WorldCat.

ARTÍCULO IV:

Piedra, Jonathan. (2022). Decolonizando la Ética de la IA. DILEMATA: Revista Internacional de Éticas Aplicadas. N° 38, mayo.

Indicadores de calidad: H Index: 9 según Índice H de las revistas científicas españolas en Google Scholar Metrics. Categoría ANEP: C.

ÍNDICE

RESUMEN EXTENDIDO	1
INTRODUCCIÓN	2
PROBLEMAS DE CONCEPTUALIZACIÓN Y DEMARCACIÓN	5
FALTA DE CLARIDAD SOBRE LAS PROPUESTAS REGULATORIAS PARA LA IA	8
AUSENCIA DE ALTERNATIVAS DESCOLONIZADORAS EN CUANTO A LOS MODELOS ÉTICOS PARA LA IA	10
BIBLIOGRAFÍA	15
ARTÍCULOS DE INVESTIGACIÓN QUE COMPONEN LA TESIS.	20

Resumen extendido

La investigación aborda los desafíos morales de las diferentes narrativas sobre la Inteligencia Artificial desde una perspectiva filosófica. Con esta finalidad se han seleccionado una serie de problemas paradigmáticos que representan los principales vacíos en los ámbitos de reflexión encontrados en la revisión de la literatura científica sobre el tema, a saber: (1) Un déficit en la conceptualización y demarcación de las nociones centrales sobre la Inteligencia Artificial; (2) Falta de claridad sobre las propuestas regulatorias para la IA y (3) Ausencia de alternativas descolonizadoras en cuanto a los modelos éticos para la IA. El trabajo realizado analiza estas situaciones a partir de la robótica social, la *smartificación* de gestión pública, la autogestión ética empresarial y por medio propuestas regulatorias para la IA. Mostrando como existen muchos desafíos morales y políticos que deben ser tomados en cuenta y que necesitan una concienzuda reflexión filosófica. Varios temas de gran relevancia se muestran en esta investigación, de entre ellos: Una crítica a la propuesta de autogestión empresarial, así como al principalismo ético de las Guías Éticas de la IA. En cuanto al primer punto, mostrando como estas propuestas son cortinas de humo que buscan convencer a los legisladores y políticos sobre que el autogobierno empresarial y los mecanismos autorregulatorios en la industria son suficientes, por lo que no se necesita ninguna regulación específica para reducir los posibles riesgos tecnológicos, eliminar los escenarios de abuso o impedir las constantes violaciones a los DDHH (en temas como privacidad, discriminación, libertad de expresión, sesgo algorítmico etc.). En cuanto al segundo punto, se ha mostrado como se deben replantear las bases de la Ética de la IA desde una praxis decolonial.

Desafíos filosóficos e implicaciones morales de las narrativas sobre la Inteligencia Artificial

Introducción

La Universidad Salamanca es referente en lo relacionado con la investigación, desarrollo y reflexión sobre la IA. Sus aportes van desde de la creación de aplicaciones específicas de IA (SalamancaAIDía.es, 2018) posicionando a la USAL como un referente nacional de investigación en IA o a través de la gestión y el desarrollo de actividades de gran nivel internacional sobre estos temas (i.e La XII Conferencia Internacional sobre Aplicaciones de Sistemas Agente y Multiagente). Este deseo de estar a la vanguardia generó que la USAL publicara la primera revista española sobre computación distribuida e inteligencia artificial. Recientemente, el Senado Español y la SAL (2021) acordaron trabajar conjuntamente para crear un equipo que tenga como objetivo el papel de la IA para la protección de los Derechos Humanos y evitar la discriminación. Lo que muestra que la USAL se preocupa no solo por los aspectos técnicos relacionados con la innovación y el desarrollo, sino también por los dilemas morales y jurídicos que aparecen con la implementación de la IA,

Es precisamente, dentro de este mismo espíritu de excelencia es que la presente tesis está conformada por cuatro artículos de investigación publicados en diferentes revistas académicas. Dicha investigación se encuentra ubicada en el área de “Filosofía Moral y política”, ya que todos los documentos tienen como eje transversal el análisis de los desafíos morales de las diferentes narrativas sobre la Inteligencia artificial (Kaplan 2016, Lee 2020, Coeckelbergh, 2021) desde una perspectiva filosófica. Los documentos que se presentan abordan el tema desde una perspectiva dilemática propia de las ciencias sociales, humanas y jurídicas. Con esta finalidad se han seleccionado una serie de problemas paradigmáticos que representan los principales vacíos en los ámbitos de reflexión encontrados en la revisión de la literatura científica sobre el tema, a saber: (1) Un déficit en la conceptualización y demarcación de las nociones centrales sobre la Inteligencia Artificial; (2) Falta de claridad sobre las propuestas regulatorias para la IA y (3) Ausencia de alternativas descolonizadoras en cuanto a los modelos éticos para la IA.

La selección de estas categorías no se debe únicamente al resultado de un análisis crítico de la literatura especializada, sino que se gestó como parte del conocimiento obtenido en

cursos y talleres especializados, así como a la participación en Congresos Internacionales. Durante el período del Doctorado, realicé cursos especializados en renombradas universidades internacionales, así como en organismos multilaterales, que me permitieron abordar el tema de los desafíos y controversias de la IA, desde puntos de vistas interdisciplinarios, así como desde posturas prácticas alejadas de la especulación teórica. Por mencionar algunos casos, pude participar en el curso de invierno «Winter School on The Regulation of Robotics & AI in Europe: Legal, Ethical and Economical Implications» en Sant’Anna, School of Advanced Studies que se llevó a cabo en Pisa, Italia. También el Seminario – «Inteligencia Artificial y su influencia en la libertad de expresión» realizado en la Universidad Autónoma de México. Otro curso de importancia fue «Gobernanza de la inteligencia artificial en la Administración Pública desde una perspectiva ética» del Centro Latinoamericano Administración para el Desarrollo (CLAD), entre muchos otros relacionados, así como los cursos de especialización brindados por la Universidad de Salamanca. Como podrá notarse, en el caso particular de los cursos mencionados cada uno responde de alguna manera a los temas que posteriormente fueron desarrollados en los documentos presentados como compilación para esta Tesis. No obstante, todas estas actividades académicas, solo fueron la primera parte de una labor mucho más amplia, la cual consistió en presentar los resultados de la investigación frente a un público experto y así tener la posibilidad validarlos o contrastarlos con la opinión de renombrados especialistas internacionales. Es así como teniendo como meta lo anterior, los elementos centrales de esta investigación fueron presentados en 12 Congresos Internacionales en países como: España, Portugal, Brasil, Chile, Colombia, México y Costa Rica. La participación este tipo de actividades fue de gran aprendizaje y utilidad, ya que me permitieron refinar algunos puntos, así como descartar otros. Por lo que la Tesis que se está presentando es el fruto de un amplio trabajo de investigación, difusión y retroalimentación del conocimiento, lo cual a mi modo de ver es uno de los puntos fuertes de esta investigación.

Es así como el objetivo de esta Tesis (en su modalidad por compendio de publicaciones) fue el análisis sistemático y acumulativo de los problemas filosófico-morales que se derivan de los vacíos encontrados en este amplio proceso, lo cual constituye, a su vez, el aporte principal de cada uno de los documentos. Es precisamente la intersección de los elementos anteriores, la clave para entender la originalidad y valor de cada uno de los artículos que aquí se presentan como resultado de las tareas de investigación. Los artículos

nos permiten reflexionar desde una «Historia conceptual» los elementos controvertidos, al tiempo que se utiliza la metodología propia del estudio de casos, para dimensionar, ya no solo filosóficamente, sino además desde un punto de vista social, las grandes transformaciones que se esperan de la IA. Esto nos permite ubicar en un contexto real (no-especulativo) los retos y dilemas morales de su utilización.

Este tipo de acercamiento metodológico se relaciona directamente con las características disruptivas de la IA. Esto se debe a que los problemas morales de esta tecnología tienen una fuerte dimensión valorativa y no solamente científica o técnica. Es claro que la dimensión del valor necesita una revisión sistemática por medio de interpretaciones y perspectivas morales. Por lo tanto, se requiere no solamente una aplicación mecánica de algunas posturas filosóficas, sino un equilibrio entre la información y datos técnicos, así como con algunas intuiciones éticas, que implican una interacción cuidadosa entre lo teórico y lo práctico.

Mucho se habla sobre la IA, pero ¿Qué se comunica realmente? Por lo general, los documentos relacionados con la IA utilizan una retórica excesivamente optimista que oculta las disyuntivas reales que se presentan con su utilización. En ocasiones, los documentos son demasiado generales o bien excesivamente detallados, generando una clara oposición en cuanto a los enfoques. En el primer caso, parten de un contexto poco delimitado y nada restringido, en el cual casi cualquier reflexión es posible sin entrar en mayores contradicciones discursivas. Mientras que cuando sucede lo opuesto, se debe a las características particularísimas de un texto, que debido al enfoque disciplinar o a la naturaleza especializada del problema en cuestión, impide un acercamiento básico, excepto si se poseen conocimientos excesivamente especializados. ¿Qué nos dice todo esto respecto a esta Tesis? En primer lugar, que cada «vacío» implica un análisis doble que se basa en evidencias (datos, casos, propuestas, etc.), así como en hipótesis sobre las posturas morales subyacentes a estas situaciones. En segundo lugar, que las conclusiones de los artículos implican un análisis sistemático que combina la reflexión filosófica con una lectura social (acotado al problema que se está trabajando en cada documento) sobre las implicaciones disruptivas en algún caso, mientras que, en otros, se dirigen a problemas concretos con características específicas.

Problemas de conceptualización y demarcación

Por las razones anteriormente indicadas resulta revelador que los problemas de conceptualización y demarcación del ámbito de investigación y aplicación de la IA comprometen al concepto y a las posibles definiciones de IA. No existe una definición precisa sobre este campo, y aun cuando se pueda trabajar operativamente con una idea general, en la mayoría de los documentos existe una polisemia que no ayuda mucho a una delimitación clara. Dependiendo de la orientación de la cual se parta, la IA puede ser vista como una ciencia o técnica exclusiva de especialistas que buscan crear máquinas o programas inteligentes para resolver problemas (McCarthy, 2018). Lo que vendría a suponer una instrumentalización utilitarista de la IA, por lo que su principal característica sería la resolución de dificultades, por medio de tecnología aplicada. Otros consideran que su núcleo se encuentra en la imitación del comportamiento inteligente (Franceschetti, 2018), lo que supondría, al menos en la teoría, algo más que el enfoque instrumental. Esto es así en tanto la definición de comportamiento inteligente implicaría ciertas interpretaciones sobre la cognición humana, que finalmente se aplicarían a dispositivos o programas inteligentes. Ciertamente, cada una de estas posiciones tiene muchos matices, y no son, propiamente, posturas homogéneas, aunque representan una línea muy asociada a la visión empresarial, ya que buscan la implementación práctica. En estos dos casos, la IA es vista como una tecnología prometedora por sus posibles aplicaciones concretas, así como por el posible crecimiento económico derivado de su uso e implementación, que a la postre generará una bonanza de empleos y riqueza (Kaplan), que dividirá al mundo entre quienes son potencias y quienes no son potencias de la IA (Lee, 2020). A la par de estas, existen interpretaciones más especulativas que se enfocan en el potencial disruptivo de la IA. Ya sea porque en un futuro la IA nos hará más que humanos (Kurzweil, 2012), llegará un punto que seremos casi como dioses transhumanos (Harari, 2016) con la ayuda de estas tecnologías o simplemente porque resignifican lo que se entiende actualmente por humano (Tegmar, 2018). Todo ello sin duda nos hace pensar en una «era artificial» (Velarde, 2021) o por decirlo más llanamente, en una «Era de la IA» (Mondal 2020), en la cual la inteligencia artificial será la tecnología de crecimiento más rápida.

Conceptos como «robot» o «autonomía» son igualmente confusos. Esta situación se puede notar más cuando se interrelacionan en la llamada «robótica social» un floreciente campo de la IA. Para comenzar, valga decir que con respecto a lo que es un robot, no

existe una definición global que pueda ser tomada como una base clara. Incluso autores como Arkin (2019) emplean definiciones separadas para robot y robot autónomo. Un robot sería «una máquina o vehículo automatizado, capaz de percepción, razonamiento y acción independientes» (pp.50). Mientras que un robot autónomo es «un robot que no requiere la participación humana directa, excepto para la misión de alto nivel; dicho robot puede tomar sus propias decisiones de acuerdo con su misión sin requerir autorización humana directa, incluidas las decisiones sobre el uso de la fuerza letal» (Arkin 2009, pp. 50-51). La Resolución del Parlamento Europeo (16 de febrero del 2017) con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica considera que un «robot inteligente tiene la capacidad de adquirir autonomía, autoaprendizaje (a partir de la experiencia y la interacción), la capacidad de adaptar su comportamiento y acciones al entorno, así como un soporte físico y la inexistencia de vida en sentido biológico» (pp.6). La principal característica de estos tipos de robots sería la autonomía, aun cuando, igualmente, no es claro qué significa la autonomía en este tipo de contextos. Cabe señalar que sobre la autonomía tampoco existe un consenso sobre lo que implica. Se la considera, con frecuencia, una idea elemental que subyace la capacidad para tomar decisiones propias (en oposición a la heteronomía), sin la influencia determinante de otra persona o entidad, pero ciertamente, las definiciones varían según el contexto. Muchas veces en los debates filosóficos sobre el tema, los criterios de autonomía pueden ser muy exigentes, mientras que, en otras áreas, como la de los robots con IA, la exigencia disminuye. Parece, además, que también tiene niveles. Moor (2006) menciona hasta cuatro tipos de «*ethical impact agents*», que varían en función de la idea autonomía y responsabilidad. Por otro lado, Hellström (2013) utiliza el concepto de «poder autónomo» (*autonomous power*) como un medio para evaluar las capacidades relativas de los agentes artificiales con diferentes grados de capacidad física, poder de decisión, y alcance para la interacción. Son esta clase de situaciones las que precisamente hacen que el tema de la autonomía sea relevante en lo relacionado con los desafíos morales de la IA.

El tema de la autonomía de los agentes con IA constituye una preocupación habitual en los Organismos Internacionales, así como para los Estados a la hora de generar políticas públicas. Verbigracia, el Parlamento Chileno con respecto a esto menciona que «Los sistemas de IA deberían diseñarse de forma tal que respeten las reglas de la ley, derechos humanos, valores democráticos y diversidad, y deben incluir salvaguardas apropiados»

(Comisión “Desafíos del Futuro, Ciencia, Tecnología e Innovación”. 2019, pág. 12). El Parlamento Europeo en una Resolución del 2017 también considera importante este tema, ya que la autonomía de los robots implica dejar de considerarlos como simples cosas o herramientas, lo que hace necesario revisar su estatus desde un punto de vista moral y legal, para pasar a pensar seriamente si las categorías filosóficas y jurídicas existentes son suficientemente consistentes o es necesario crear otras. Por ejemplo, la investigación de Toboso & Aparicio (2019) respecto a una robótica inclusiva plantea importantes reflexiones sobre los diseños universalmente accesibles. «Se trata no sólo de analizar cómo se pueden eliminar las barreras actuales o cómo ciertas aplicaciones tecnológicas pueden resolver los problemas de las personas con discapacidad, sino también de desarrollar estas aplicaciones a través de un enfoque prospectivo y participativo orientado a la creación de tecnologías inclusivas.» (pp.177). Un aspecto particularmente importante en el caso de los robots asistenciales y de cuidado. También es necesario, reflexionar la sugestiva aportación de Barrio Andrés (2019) que considera necesario hablar de un Derecho de los Robots. El término parece especialmente indicado pues hace notar la clase de problemas que se derivan de la dilemática imputación de responsabilidades; algo que resulta especialmente claro en lo que se refiere a la distribución de responsabilidades y el manejo de riesgos en los robots sociales. Igualmente sugerente es la propuesta de Aparicio *et al* (2019) especialmente en lo relativo a una nueva concepción sobre de la vulnerabilidad humana basada en la Convención sobre los Derechos de las Personas con discapacidad (CDPD) con la finalidad de crear una reflexión ético-política sobre la robótica asistencial. La hipótesis que presento en mi investigación se basa en la idea general subyacente en este documento, ya que igual que los autores, considero que la tecnología robótica exige una reflexión filosófica basada en la moral y la política. Mi propuesta parte de los argumentos desarrollados por Dumouchel & Damiano (2017) en tanto desarrollo la idea de los robots sociales como «sustitutos» y no como simples máquinas especializadas. Por lo tanto, en su «esencia» radicaría la característica de ser rol o una actividad y no un objeto propiamente. Un robot social busca sustituirnos en alguna tarea (i.e cuidado de niños, compañía de mayores, atención de personas con demencia etc.) pero sin tomar propiamente nuestro lugar. Como lo indican Navarro & Kovacic (2021) los robots sociales contemporáneos no pueden concebirse como objetos o instrumentos. Estos robots sociales «supone una reconsideración continua de la relación entre el ser humano y la máquina que acabara cada vez más definida en términos de relaciones interespecíficas.» (pp.50). Precisamente esta situación es la que provoca que el desarrollo

de este tipo de robots traiga no solo elementos positivos, sino también aspectos negativos. A pesar de esto, el debate sobre la necesidad de introducir normas legales y ético-jurídicas relacionadas con su funcionamiento se encuentra aún en proceso y no hay una perspectiva clara.

Los artículos presentan ejemplos del uso de la Inteligencia Artificial, que en un futuro próximo generará situaciones que no solo son un desafío legal, ético, sino también social en un sentido amplio. Cada uno de los artículos presenta algún reto de la Inteligencia Artificial, así como ejemplos de su uso. Temas como la ciudadanía digital, la *smartificación* de la Administración Pública son campos de estudio relativamente recientes, que adolecen de las consecuencias derivadas del escaso tratamiento analítico que han merecido hasta el momento, tal y como he mencionado líneas arriba. En los documentos que buscan la regulación de la IA (con algunas excepciones), estas carencias no solo son más visibles, sino en algunos casos (como las propuestas empresariales) son más intencionadas, con tal de mantener los debates en aspectos abstractos (Wagner, 2018)

Falta de claridad sobre las propuestas regulatorias para la IA

Independientemente de la posición desde la cual se conciba la IA o sus aplicaciones, es claro que el tema tiene una relevancia gigantesca en los debates especializados, reflexiones académicas, pronósticos futuristas, medios de comunicación o en las políticas de un país. Una de las principales limitaciones para desarrollar estos aspectos positivos es la falta de aspectos normativos que, junto con la ambigüedad ética (Morley *et al*, 2021) de algunos lineamientos o posiciones sobre su regulación, provocan un gran vacío en la adopción de medidas y directrices necesarias. De hecho, «Instrumentalized and emptied of its instrumental and intrinsic value, what we might have wanted to optimistically call “ethics” now appears trapped between meanings: objectionable instances of self-regulation, static and incomplete lists of guiding principles and other forms of narrow and conservative regulative “fixes”» (Bietti, 2019, pp.217-218). Esta situación, en ocasiones, se pretende resolver con resoluciones no-vinculantes de organismos internacionales o con dictámenes emitidos por grupos interdisciplinarios de expertos (Por ejemplo, el caso de la Unión Europea con las «Directrices Éticas para un IA fiable» o en los Estados Unidos con *Preparation for the Future of Artificial Intelligence*). Todas ellas con la intención de servir como guías que permitan crear lineamientos básicos sobre el tema. (Piedra, 2022a;

Piedra, 2022b) En América, por ejemplo, Berkman Klein del Center for Internet & Society, considera que debe existir una deliberación exhaustiva y a largo plazo del impacto derivado de la implementación de tecnologías relacionadas con la IA, y no únicamente una visión cortoplacista como la que usualmente se presentan en los discursos sobre este tipo de tecnologías. Por lo general, estos discursos se ubican en dos grandes grupos (Piedra, 2022b): los «progresistas económicos» que consideran a que la IA nos puede llevar a una bonanza de desarrollo económico y social, en el cual la IA en nuestra época sería una guía para la riqueza y el trabajo (Kaplan, 2016). Luego están aquellos que consideran que debería tenerse cautela con la implementación acrítica de esta tecnología. En el medio se encuentran quienes buscan la regulación, sobre todo porque es claro que existen infinitas interacciones de una IA con su entorno de manera tal que conduzca a resultados impredecibles. Esto plantea la necesidad de su regulación desde el punto de vista de las políticas públicas, así como desde el punto de vista ético. Sin embargo, las aproximaciones éticas son, en unos casos, un tanto imprecisas en cuanto los detalles técnicos y, en otras, posiciones bastante generales sobre postura ética que carecen de lo que se conoce como el «*on-the-ground approach*». De hecho, esto último es lo que usualmente reclaman las grandes compañías privadas que dominan el campo de la IA: la regulación es inadecuada y limita la innovación debido a esta falta de conocimiento de la realidad tecnológica. Situación que resulta ser un tanto paradójica ya que a pesar de que en el mundo existen algunas estrategias nacionales, esto no implica el establecimiento de una regulación particular, ni muchos menos un posicionamiento claro con respecto al tipo de regulación o postura ética deseada. Solamente existen ideas básicas sobre algunos puntos que deberían ser tratados. Una muestra de esto se puede encontrar en la OCDE. Este organismo creó una lista de cinco principios basados en valores para una administración responsable de la IA. De nuevo, lo que ahí se presenta, *grosso modo*, son premisas poco desarrolladas en cuanto a los sistemas de IA. Repitiendo *locus communes* tales como el respeto a las leyes nacionales, los Derechos Humanos (sin concretar ninguno en específico) o a la diversidad humana. Aspectos que desde luego son de gran importancia y siempre deberían tenerse en cuenta a la hora de regular alguna tecnología. No obstante, esto es interesante ya que en muchos países (i.e América) no existe una estrategia de política pública que intente abordar los desafíos ético-jurídicos que presentan estas tecnologías. Esto crea un peligroso un vacío que podría ser utilizado por aquellos que quieren utilizar de una manera poco ética las posibilidades de la IA.

Muchos organismos y centros Internacionales han reflexionado sobre estos desafíos. Por ejemplo, la Agencia Latinoamericana de Información menciona que «[u]n eje predominante de crítica y regulación de la IA son las normas internacionales de derechos humanos. Considerando el impacto sobre la privacidad, la libertad de expresión y la libertad de asociación, entre otros, el marco de los derechos humanos es un requisito mínimo al que deben adherirse los sistemas de IA.» (ALAI noviembre, 2019, pág. 2). Una línea similar es sostenida por el The Carr Center for Human Rights (Harvard Kennedy School) que sitúa el tema de los valores y derechos humanos en el centro de la discusión filosófica, revitalizando y brindándole un enfoque actualizado a temas clásicos de la filosofía ética y moral. Propuestas como las de Astobiza *et al* (2021) van un paso más allá, al incluir la Agenda 2030 para los Objetivos de Desarrollo Sostenible. Según ellos «A pesar de la gran cantidad de guías y enfoques basados en principios para la Ética de la IA, el único marco teórico y práctico base para la gobernanza global de la IA y, en particular, a su aplicación para los ODS son los Derechos Humanos (DH). DH deben limitar cualquier diseño o implementación de tecnologías de IA. En la era de la IA y los robots, los DH deben ocupar un lugar central.¹» (pp.3) Sin embargo, a pesar de que es habitual encontrar referencias generales a los Derechos Humanos en los documentos oficiales sobre estos temas, una parte importante de las propuestas se decantan hacia posiciones meramente orientadoras, dejando de lado los aspectos normativos.

Ausencia de alternativas descolonizadoras en cuanto a los modelos éticos para la IA

Las situaciones como las previamente mencionadas son elementos que justifican el hecho de que las propuestas éticas regulatorias sean, en la mayoría de los casos, un compendio de principios que se presuponen universales. En este sentido concuerdo con Ausín (2021) en lo que respecta a que « [s]e interpreta la ética de la IA como un recetario o conjunto de principios y buenas intenciones con respecto a su uso, renunciando al sentido primigenio y profundo de la reflexión ética que no es otro sino interrogar, analizar y cuestionar el sentido mismo de los desarrollos tecnocientíficos (no tanto su uso), por lo que la ética debe de estar presente desde el mismo diseño de la tecnología y en constante

¹ «Despite the large number of guidelines and principle-based approaches to AI ethics the only theoretical and practical framework on which to base a global governance of AI and, in particular, to its application for the SDGs is Human Rights (HR). HR should constrain any design or deployment of AI technologies. In the age of AI and robots the HR must take a central stage.»

diálogo con su desarrollo (*Ethics by Design*). La ética es básicamente pregunta.» (pp.13) Ciertamente eso así, pero además las preguntas siempre se hacen desde determinados espacios sociales, culturales y vitales. Por lo que existe todo un contexto histórico que justifica este reduccionismo. Mi investigación complementa la posición de Ausín, al trabajar un aspecto prácticamente ignorado en los debates sobre la Ética de la IA: Su fundamentación anglo-eurocéntrica. Para esto parto de una hipótesis desarrollada en la Filosofía Decolonial Latinoamericana según la cual la cultura occidental lleva mucho tiempo definiendo con base en su propia historia los criterios y valores con los que debe funcionar el mundo. Una perspectiva reduccionista (muchas veces dicotómica) centrada en el punto de vista del hombre occidental, blanco y heterosexual se busca imponer como el paradigma de conocimiento universal. Su posición hegemónica les permite crear «fronteras valorativas» (Cabezas, 2013) que hacen que muchas culturas y pueblos asimilen este pensamiento, a expensas de que sus propias experiencias hayan sido descartadas, excluidas e incluso eliminadas. Estas «fronteras valorativas, no son sino formas de discriminar. Efectivamente, dado que implican una valoración de lo real respecto del esquema dual previo, influyen en el tratamiento de los problemas morales y legales» (Cabezas, 2013, pp.2)

Mi investigación continúa una línea iniciada por Maldonado (2007, 2018) con respecto a la tecnología, según la cual «El discurso tecnofílico propone a las TICs como herramientas de desarrollo fundamental de las sociedades en el actual contexto de globalización, sin cuestionarse si su uso debe ser definido en base a necesidades y decisiones de los actores sociales». (Maldonado, 2018, pp.122). Los argumentos del autor se encuentran delimitados a las tecnologías de la información y comunicación, especialmente en lo relacionado con la pretensión de estas tecnologías para mejorar la capacidad y velocidad de distribución de la información, facilitando así el acceso al conocimiento. Mi propuesta complementa esta visión al incluir dentro del paradigma decolonial a una tecnología aún más disruptiva que las señaladas por Maldonado: la IA. La capacidad de los robots (Piedra, 2021), máquinas o sistemas para realizar tareas que tradicionalmente han sido ejecutadas por seres humanos, actualmente no se pone en duda. Debido a que IA combina diversas habilidades tales como el denominado aprendizaje autónomo, la capacidad de razonar, la resolución de problemas, e incluso la percepción y el reconocimiento (y uso) del lenguaje natural es necesario un debate no-reduccionistas sobre los modelos epistémicos y éticos sobre los cuales se debe regular esta tecnología.

Como indica Wagner (2015): «La diversidad étnico-cultural es una de las características más evidentes de las sociedades actuales (...) Esta pluralidad se refleja en diferentes costumbres, tradiciones, religiones, niveles educativos, estructuras sociales y relaciones de poder, que a su vez forman la base de las diferencias axiológicas, de una pluralidad de valores y normas, así como de formas morales-éticas de argumentación y conceptos básicos.»² (pp. 219). Sin embargo, la ausencia de una reflexión basada en las necesidades y condiciones de los actores sociales es particularmente notoria en el caso de las orientaciones regulatorias sobre la IA. En la actualidad, prácticamente la totalidad de las propuestas éticas regulatorias que tienen pretensión global se basan en la exclusión de las diferentes epistemologías, así como en una invisibilización de los conocimientos no-occidentalizados en lo relacionado con axiologías culturales alternativas. Todo con base en argumentos extremadamente positivos (o ingenuos) sobre el desarrollo y el crecimiento económico de los países. «Quienes se abanderan de este tipo de proclamas tecnodesarrollistas están diseñando una sociedad determinada por las tecnologías, desconociendo que los proyectos de sociedad que se engendran en las comunidades son diversos y, por ende, los usos tecnológicos responden a esas múltiples necesidades, lo cual implica poner en sospecha la redención que este tipo de ideología proclama en torno a los beneficios de las TICs como dispositivos de desarrollo». (Maldonado, 2018, pp.122)

Situación que es muy más fácil de observar en los «planteamientos regulatorios» provenientes de las grandes empresas de la IA. Es necesario decir, que actualmente las empresas tecnológicas son cada vez más coloniales en su accionar. Esto se puede observar claramente en la imagen social que pretende cultivar, mientras mantienen su soberanía económica y empresarial en la cultura dominante como arma de dominación tecnológica y social, utilizando como excusa el desarrollo de la IA como una excusa adecuada para la ausencia de reglas, la explotación, discriminación e incluso el irrespeto flagrante a los Derechos Humanos (Piedra, 2022a).

Muchas veces los «valores culturales» que se indican en estos documentos son solo aportes cosméticos o, en el mejor de los casos, elementos susceptibles de una ponderación económica. Tal y como indica Roldán (2020): «El mismo concepto de “valor cultural”

² «Ethnisch-kulturelle Vielfalt ist eines der augenfälligsten Merkmale der heutigen Gesellschaften (...) Diese Pluralität zeigt sich in unterschiedlichen Sitten, Traditionen, Religionen, Bildungsstandards, sozialen Strukturen und Machtverhältnissen, die wiederum die Grundlage bilden für axiologische Differenzen, für eine Pluralität der Werte und Normen sowie moralisch-ethischer Argumentationsformen und Grundbegriffe».

está cargado de una etimología económico laboral: lo que “vale” es tanto lo que cuesta conseguir como lo que adquiere “precio” en esa relación o transacción, pero más aun lo que quienes detentan el poder determinan como “valioso”». (pp.26) Pero este tipo de reflexión no se encuentra en las propuestas regulatorias éticas de la IA. En contraste, se omiten groseramente casi todos los elementos que se relacionen con una discusión sobre el sistema económico que subyace en estas tecnologías o las prácticas abusivas de las empresas de IA que dominan el mercado. Este tipo de ausencias son un elemento común en las propuestas éticas sobre la IA.

A pesar de que «En las sociedades moralmente pluralistas y democráticas, la ética no existe, existen en verdad diversas morales en el ámbito cotidiano (...) y distintas líneas desde la filosofía moral (...). Ninguna de ellas es hegemónica para toda la humanidad, ninguna de ellas es la moral o la ética» (Romero Muñoz, 2017, pp.48). Esto no es así en los documentos sobre las orientaciones éticas de la IA que hemos mencionado. De hecho, es todo lo contrario, en ellos se encuentra un discurso que parte de una base interpretativa supuestamente universal sobre el ser humano, que se muestra incuestionada en estos documentos. Pretendidamente neutras en cuanto a los elementos tecnológicos, así como a los modelos de IA, estas guías se basan en representaciones abstractas, generalizaciones apresuradas y simplificaciones de muchas realidades complejas, que se basan en criterios morales occidentales. Como indica O’Neil (2017) «[L]os modelos, a pesar de su reputación de imparcialidad, reflejan objetivos e ideología. [...] Nuestros propios valores y deseos influyen en nuestras elecciones, desde los datos que elegimos recopilar hasta las preguntas que hacemos. Los modelos son opiniones integradas en las matemáticas». Aspectos que desde luego no son abordados en ningunas de estas declaraciones éticas, las cuales, en su mayoría, responden a la misma lógica y enfoque.

Como ya he señalado, la IA es una tecnología con gran potencial disruptivo no solo en el ámbito económico, sino también desde un punto de vista más amplio, como lo es el social. Esta característica hace necesario una reflexión filosófica basada en la moral y la política que pueda generar un debate responsable sobre el modelo de IA que se quiere implementar y desarrollar, lo cual desde luego implica una meditación sobre el poder, la discriminación, exclusión o vulneración de derechos que la tecnología puede provocar. La propuesta que he desarrollado tiene como idea transversal un aspecto que Ausín (2021) menciona muy claramente, cuando escribe que «Una tecnología no solo debe ser evaluada

por la forma en que contribuye a la eficiencia y la productividad, sino también por la forma en que puede crear ciertas formas de poder y autoridad. Las tecnologías transforman los objetos, pero igualmente los hábitos, costumbres o relaciones» (pp.4) Precisamente por esto es por lo que la investigación plantea varias líneas de investigación innovadoras. El claro que existen profundos temas de conceptualización y demarcación que se encuentran en áreas como IA para la Administración y Gestión Pública o la robótica social que son necesarios que sean estudiados con tal poder definir enfoques adecuados frente a las implicaciones socio-jurídicas que generan estos temas.

El trabajo realizado aborda varios de estos problemas a partir de la robótica social y la *smartificación* de gestión pública. Mostrando exitosamente como existen muchos desafíos morales y políticos que deben ser tomados en cuenta y que necesitan una concienzuda reflexión filosófica. Por otro lado, la falta de claridad sobre las propuestas regulatorias para la IA es un área de investigación que necesita mucho trabajo. Dos temas de gran relevancia son las han sido abordadas en esta investigación: Una crítica a la propuesta de autogestión empresarial, así como al principialismo ético de las Guías Éticas de la IA. En cuanto al primer punto, mostrando como estas propuestas son cortinas de humo que buscan convencer a los legisladores y políticos sobre que el autogobierno empresarial y los mecanismos autorregulatorios en la industria son suficientes, por lo que no se necesita ninguna regulación específica para reducir los posibles riesgos tecnológicos, eliminar los escenarios de abuso o impedir las constantes violaciones a los DDHH (en temas como privacidad, discriminación, libertad de expresión, sesgo algorítmico etc.). En cuanto al segundo punto, se deben repensar las bases de la Ética de la IA desde una praxis decolonial. Práctica que no solo es atinente en cuanto metodología, sino además necesaria para replantear el verdadero impacto de este tipo de Guías. La descolonización la Ética de IA, implica repensar desde un marco intercultural, decolonial y pluralista conceptos como privacidad, autonomía o el manejo de datos, entre muchos otros tópicos. Es necesario legitimar y reconocer posiciones epistemológicas y morales diferentes a la perspectiva universalista anglo-eurocéntrica. Es ineludible una conversación verdaderamente global sobre la IA, en donde no se marginalice el conocimiento no-occidental. Situación, que sin duda alguna, es una tarea por realizar y una línea investigativa que debería tener mucho más desarrollo, al igual que las que hemos mencionado previamente.

Es claro que existen grandes desafíos que se deben abordar. Es necesaria una verdadera ética reflexiva frente a la inmediatez de los «solucionismos» (Roldán, 2021) que plantea la IA, así como al optimismo ingenuo que generalmente permea el debate sobre este tipo de tecnologías. Se necesita mucho trabajo interdisciplinario e investigación adicional en temas como la robótica inclusiva (Toboso & Aparicio, 2019), así como en la creación de un marco ético-político para la robótica asistencial (Aparicio *et al*, 2019). Tampoco se debe dejar de lado, lo relacionado con el tema de la perspectiva de género y la robótica, así como los cambios profundos que implican las interacciones humano-robot (Navarro & Kovacic, 2021). Por otro lado, el cuestionamiento del principalismo de las propuestas éticas de la IA, también es un asunto importante que debe debatirse, ya sea desde un enfoque basado en Derechos Humanos (Astobiza *et al*, 2021), a partir de un abordaje ético integral que esté presente desde el diseño de la tecnología (Ausín, 2021) o por medio de una metodología descolonizadora que visibilice nuevos marcos axiológicos y abordajes epistémicos (Piedra, 2022b). Ciertamente, existe mucha incertidumbre en esta área, lo demuestra que queda mucho trabajo por hacer. Trabajo que espero continuar y del cual esta Tesis ha servido para marcar futuras líneas de investigación.

Bibliografía

ALAI (noviembre, 2019): Inteligencia artificial: Ddhh, Justicia social y Desarrollo (43).

Aparicio Payá, M., Toboso Martín, M., Ausín Díez, T., Monasterio Astobiza, A., Morte Ferrer, R., & López Castro, D. (2019). Un marco ético-político para la robótica asistencial. *ArtefaCToS. Revista De Estudios Sobre La Ciencia Y La tecnología*, 8(1), 97–117. <https://doi.org/10.14201/art20198197117>

Arkin, R. (2009). *Governing Lethal Behavior in Autonomous systems*. Boca Raton, FL.: CRC Press.

Astobiza, M; Toboso, M; Aparicio, M; López, D (2021). *AI ethics for sustainable development goals*. *IEEE Technology and Society Magazine* 40 (2), 66-71 2021

Ausín, T. (2021). ¿Por qué ética para la Inteligencia Artificial? Lo viejo, lo nuevo y lo espurio. *SOCIOLOGÍA Y TECNOLOGÍA*, 11 Extra_2, (2021): 1-16. ISSN: 1989-8487. DOI 10.24197/st.Extra_2.2021.1-16

- Barrio Andrés, M. (Ed.) (2019). *Derecho de los Robots*. Wolters Kluwer España, S.A
- Bietti, E. (2019). From Ethics Washing to Ethics Bashing. A View on Tech Ethics from Within Moral Philosophy. In *Proceedings of ACM FAT* Conference (FAT* 2019)*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3351095.3372860>
- Cabezas, M. (2013). *Juicios morales y fronteras biológicas: más allá de la frontera razón / emoción*. *Arbor*, 189(762), a052. doi:10.3989/arbor.2013.762n4003
- Coeckelberg, M. (2021). *Ética de la Inteligencia Artificial*. Trad. Lucas Álvarez Canga. Catedra Teorema. España
- Coeckelberg, M. (2022). *The Political Philosophy of AI*. Polity Press. Cambridge. The UK.
- Comision “Desafíos del futuro, ciencia, tecnología e innovación”. (2019): *Inteligencia Artificial para Chile. La urgencia de desarrollar una estrategia*.
- Dumouchel, P, & Damiano, L (2017) *Living with Robots*, Harvard University Press.
- Franceschetti, D. (2008) *Principles of robotics & artificial intelligence*. H.W. Wilson Publishing. 2018
- G. Navarro, M., & Kovacic, M. (2021). Tecno-especies: la humanidad que se hace a sí misma y los desechables. *Bajo Palabra*, (27), 45–62.
- Harari, Y. N. (2016). *Homo Deus*. Harvill Secker.
- Hellström, Y. (2013). On the moral responsibility of military robots. *Ethics and information technology* 15(2):99–107.
- Kaplan, J (2016). *Abstenerse humanos. Guía para la riqueza y el trabajo en la era de la inteligencia artificial*. Teell Editorial S.L. España

Kurzweil, R. (2012). *La singularidad está cerca. Cuando los humanos trascendamos la biología*. Lola Books.

Lee, K. (2020) *Superpotencias de la inteligencia artificial*. Barcelona: Editorial Planeta.

Maldonado, C. (2018): *De-colonialidad en la era tecnomediática*. Ediciones CIESPAL.

Maldonado-Torres, N. (2007). Sobre la colonialidad del ser: contribuciones al desarrollo de un concepto. En S. Castro-Gómez & R. Grosfoguel (eds.) *El giro decolonial. Reflexiones para una diversidad epistémica más allá del capitalismo global* (pp. 127-167). Bogotá: Iesco-Pensar-Siglo del Hombre Editores.

Mondal, B. (2020): Artificial Intelligence: State of the Art. En: Valentina E. Balas, Raghvendra Kumar y Rajshree Srivastava (eds.): *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, t. 172. Cham: Springer International Publishing (Intelligent Systems Reference Library), pag. 389–425.

Moor, J. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems* 21(4): 18–21.

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). *Ethics as a Service: A Pragmatic Operationalisation of AI Ethics*. *Minds and Machines*, 31(2), 239–256. doi:10.1007/s11023-021-09563-w

O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. London: Penguin.

Parlamento Europeo (2017) Resolución del Parlamento Europeo con recomendaciones: Normas de Derecho civil sobre robótica.

Piedra, Jonathan. (2021) “El Despertar de los Robots (sociales)”. *Luces en el camino: Filosofía y ciencias sociales en tiempos de desconcierto*”. 1era Edición Dykinson S.L. Madrid. pp.29-51 ISBN 978-84-1377-322-3

Piedra, Jonathan. (2022a). *Venditio fumi*: Autorregulación Empresarial e Inteligencia Artificial. *Sincronía*. Año XXVI / Número 81. Enero-junio 2022. DOI: 10.32870/sincronia.axxvi.n81.12a22

Piedra, Jonathan. (2022b). Decolonizando la Ética de la IA. *DILEMATA: Revista Internacional de Éticas Aplicadas*. Nº 38, mayo.

Quadra-Salcedo y Fernández del Castillo, T; Pinar Manas, J; Barrio, Moises (2018): *Sociedad digital y derecho*. 1. ed. Madrid: Ministerio de Industria Comercio y Turismo; Red.es; Boletín Oficial del Estado.

Romero Muñoz, J. (2017). CiberÉtica como ética aplicada: una introducción. *Dilemata*, (24), 45–63. Recuperado a partir de <https://www.dilemata.net/revista/index.php/dilemata/article/view/412000100>

Roldán, C (2021). *Por una ética reflexiva frente a la inmediatez de los «solucionismos»* Editor: Fundació Víctor Grífols i Lucas: Citación: Salud Pública y COVID. Barcelona: Fundació Víctor Grífols

Roldán, C. (2020). “El papel de los valores culturales en la construcción de las relaciones económicas y sociales. Una reflexión desde la filosofía” en *Gaceta Sindical. Reflexión y Debate*. Número 34, junio. Confederación de comisiones obreras.

SalamancaALDIA.es (2018). Investigadores de la USAL desarrollan un sistema de inteligencia artificial para catar *jamón*. Retrieved 30 May 2022, from <https://salamancartvaldia.es/noticia/2018-06-08-investigadores-de-la-usal-desarrollan-un-sistema-de-inteligencia-artificial-para-catar-jamon-100331>

Tegmar, M (2018). *Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial*. Trad. Marcos Pérez Sánchez. Penguin Random House Grupo Editorial. Barcelona

Toboso, M. & Aparicio, M. (2019) “*Entornos de funcionamientos robotizados. ¿Es posible una robótica inclusiva?*”, en *DILEMATA: Revista Internacional de Éticas Aplicadas*, pp. 171-185

Velarde, G. (2021) *Era Artificial: Predicciones para ultrahumanos, robots y otros entes inteligentes*. Düsseldorf: PRICA.

Wagner, A (2015). Pluralität der Rationalitätsformen. Zur Irrationalität des interkulturellen Dialoges en Asmuth, C., & Neuffer, S. G. (Eds.). *Irrationalität: In Zusammenarbeit mit Kai Gregor*. Königshausen u. Neumann.

Wagner, B. (2018) “Ethics As An Escape From Regulation.: From ‘Ethics-Washing’ To Ethics-Shopping?” *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, edited by EMRE BAYAMLIOĞLU et al., Amsterdam University Press, Amsterdam, 2018, pp. 84–89. <https://doi.org/10.2307/j.ctvhrd092.18>

Artículos de investigación que componen la Tesis.

A continuación, se comentan brevemente los cuatro artículos de investigación que integran esta tesis. Como se ha indicado previamente, estos artículos tienen como línea transversal los desafíos morales de la IA, a partir de los vacíos (conceptuales, normativos etc.) encontrados en la revisión sistemática de la literatura respecto al tema.

Artículo I

Piedra, Jonathan. (2021) “El Despertar de los Robots (sociales)”. *Luces en el camino: Filosofía y ciencias sociales en tiempos de desconcierto*. 1era Edición Dykinson S.L. Madrid. pp.29-51 ISBN 978-84-1377-322-3

Planteamiento

A pesar de la gran cantidad de definiciones que podemos encontrar de IA, por lo general coinciden en la capacidad de los robots, máquinas o sistemas para realizar tareas que tradicionalmente han sido ejecutadas por seres humanos. Un área que se ha visto enriquecida por la IA ha sido la robótica. La unión entre estas disciplinas ha creado un cambio significativo en nuestra concepción habitual de lo que es una máquina, pasando de la idea tradicional de robot hacia la de robot social. El diseño de los robots sociales plantea grandes retos. Este tipo de robots tienen la habilidad de reconocimiento, la capacidad de expresar y percibir emociones, establecer y mantener relaciones sociales y usar señales naturales. Todos estos elementos los hacen parte de una serie de tecnologías socialmente disruptivas. En este sentido, a pesar de los grandes beneficios que pueden traer, plantean complejas dudas sobre la comunicación, las interacciones sociales y la forma en como nos relacionamos con otros seres no-humanos.

Discusión

La aceptación final de los robots sociales por parte de la sociedad dependerá de hasta qué punto son seguros y que tanto puedan asemejar nuestros códigos sociales de conducta, generando confianza y empatía hacia ellos. Los RS al estar diseñados para interactuar con seres humanos de una manera coherente con la psicología social, pueden servirnos para

aprender algo de nosotros mismos. Los robots (en general y más aún los RS) son un reflejo de nuestra propia humanidad y su desarrollo nos lleva a reflexionar muchos elementos que creemos fundamentales de la condición humana. Verbigracia, aspectos éticos y sociales que todavía no tienen una respuesta.

ARTÍCULO II:

Piedra, Jonathan. (2021). La Smartificación de la Administración Pública: apostillas críticas a partir del caso europeo. *Cuadernos Salmantinos de Filosofía* (Sección Monográfica: Filosofía e Inteligencia Artificial) Vol. 48, 2021, 235-250, ISSN: 0210-4857E-ISSN: 2660-9509 pp.235-250 DOI:10.36576/summa.144500

Planteamiento

La Inteligencia Artificial (IA) y las tecnologías derivadas están generando una gran cantidad de cambios en nuestro mundo. Estudios y publicaciones especializados muestran hasta qué punto las tecnologías de Inteligencia Artificial tendrán numerosas implicaciones que cambiarán de manera profunda y disruptiva nuestra sociedad. En el caso concreto de la Unión Europea, en el año 2018, 24 países miembros de la UE (más Noruega) firmaron la *Declaración de Cooperación sobre Inteligencia Artificial* en la cual se comprometían a trabajar juntos con la finalidad de estimular la adopción de la IA en Europa. En este artículo aludimos a algunos ejemplos europeos relacionados con el potencial de la IA en la gestión pública y presentamos argumentos acerca de por qué la IA para uso público es un campo de investigación teórico sin capacidad para ejercer, en la actualidad, el necesario impacto en los futuros diseños de la Administración pública.

Discusión

A pesar del potencial de la IA para el sector público y del número creciente de países que apuestan por esta vía para la modernización de la Administración, la IA para uso público es un campo de investigación en el que escasean los diagnósticos exhaustivos acerca de los desafíos ético-jurídicos asociados a la aplicación de estas sofisticadas tecnologías. En estas páginas se sostiene que este déficit prospectivo está estrechamente relacionado con la falta de análisis conceptual robusto y una comprensión política común sobre la IA. Asimismo, el marcado optimismo tecnológico y la poca claridad política e institucional con la que se desarrollan y aplican estas tecnologías, hace más difícil que tenga lugar un

acercamiento reflexivo en torno a cuál debería ser el modelo de aplicación de la IA en la Administración pública de los estados europeos. Contrasta con esta situación de indefinición en torno a cuáles habrían de ser los fundamentos ético-jurídicos y sociopolíticos que habrían de guiar a los distintos estados europeos, el hecho de que la aplicación de la IA en la gestión pública parece tener el poder de marcar un rumbo a los problemas de gobernanza, así como los relacionados con la ciudadanía digital.

ARTÍCULO III:

Piedra, Jonathan. (2022). *Venditio fumi: Autorregulación Empresarial e Inteligencia Artificial*. *Sincronía*. Año XXVI / Número 81. Enero-junio 2022. DOI: 10.32870/sincronia.axxvi.n81.12a22

Planteamiento

En la primera parte de este artículo se analizará la utilización reduccionista e instrumental de la ética en temas relacionados con la regulación privada de la Inteligencia Artificial. Principalmente en lo relacionado con la autorregulación ética empresarial. Se parte de la hipótesis de acuerdo con la cual, las propuestas éticas utilizadas por las grandes corporaciones de la Inteligencia Artificial son parte de una estrategia que busca solamente crear una apariencia de responsabilidad empresarial y social, que tiene como propósito, confundir a los legisladores y a los encargados de políticas nacionales, así como a los consumidores finales, con tal de mantener sus prácticas empresariales y ganancias intactas. Se proponen como ejemplos paradigmáticos los casos de Google y Amazon con tal de mostrar esta situación, para concluir como los desafíos éticos y legales de la implementación de la IA, son reinterpretados bajo una «gestión de riesgo éticos», en la cual se trivializa y se vacía de contenido el discurso ético.

Discusión

Existe un uso instrumental de la ética (al menos de las grandes corporaciones) que va más allá del *ethical washing*. Una gestión (ética) del riesgo que solo busca de mantener los beneficios económicos de las empresas. El discurso ético se vuelve una fachada para evitar la verdadera regulación. Pero incluso va más allá, la autorregulación ética, en concreto lo que busca es la desregulación normativa, promoviendo como la única solución la gobernanza impulsada por el mercado, y no por los Estados. Es una

instrumentalización del lenguaje ético que genera una apariencia de una empresa social y éticamente responsable.

Como resultado de esto, muchas de estas iniciativas, en particular las patrocinadas por la industria, no pasan de ser (en el mejor de los casos) señales de buenas intenciones y de conductas adecuadas con el único propósito de retrasar la regulación, centrar el debate en problemas abstractos (lo cual retrasa más aún la discusión) y evitar cualesquiera soluciones técnicas.

ARTÍCULO IV:

Piedra, Jonathan. (2022). Decolonizando la Ética de la IA. DILEMATA: Revista Internacional de Éticas Aplicadas. N° 38, mayo.

Planteamiento

Existe una base principialista (tomada del modelo bioético) en la Ética de la IA. Esto puede interpretarse como caso paradigmático de ética “colonizada” que invisibiliza los diferentes juicios morales y a las axiologías culturales alternativas. Partimos de la hipótesis de que existe la posibilidad de problematizar la finalidad y el verdadero aporte de esta ética aplicada, utilizando para ello una perspectiva dilemática, propia de las ciencias humanas y jurídicas con tal de mostrar los principales elementos anglo-eurocéntricos de las iniciativas públicas sobre la «Ética de la IA» que han surgido recientemente.

Discusión

Es necesario la descolonización epistémica de las propuestas teóricas con pretensión universal. Habilitar el conocimiento de otras subjetividades y otros espacios geográficos, ampliando así los relatos anglo-eurocéntricos respecto a la EIA que constantemente se reproducen como los únicos marcos cognoscitivos válidos. Un aspecto muy importante es el acceso y participación en las plataformas científicas con tal de poder visibilizar paradigmas alternativos

Descolonizar la EIA implica también, una reflexión sobre el poder. Se debe trabajar en la crítica de la base socioeconómica de estas propuestas, la tendencia al monopolio de las empresas y las condiciones materiales que implica el uso de la IA. La atención no debe

centrarse únicamente en el aspecto tecnológico. También hay que repensar las actuales condiciones estructurales, ambientales y económicas en la cuales se basan estas tecnologías.

EL DESPERTAR DE LOS ROBOTS (SOCIALES)

JONATHAN PIEDRA ALEGRÍA
Universidad Nacional, Costa Rica

RESUMEN

La IA, así como tecnologías derivadas están generando una gran cantidad de cambios en nuestra sociedad. A pesar de la gran cantidad de definiciones que podemos encontrar de IA, por lo general coinciden en la capacidad de los robots, máquinas o sistemas para realizar tareas que tradicionalmente han sido ejecutadas por seres humanos. Un área que se ha visto enriquecida por la IA ha sido la robótica. La unión entre estas disciplinas ha creado un cambio significativo en nuestra concepción habitual de lo que es una máquina, pasando de la idea tradicional de robot hacia la de robot social. El diseño de los robots sociales plantea grandes retos. Este tipo de robots tienen la habilidad de reconocimiento, la capacidad de expresar y percibir emociones, establecer y mantener relaciones sociales y usar señales naturales. Todos estos elementos los hacen parte de una serie de tecnologías socialmente disruptivas. En este sentido, a pesar de los grandes beneficios que pueden traer, plantean complejas dudas sobre la comunicación, las interacciones sociales y la forma en como nos relacionamos con otros seres no-humanos.

PALABRAS CLAVE

Robots sociales, Inteligencia Artificial, substitutos, relaciones sociales.

INTRODUCCIÓN

Los avances en la Inteligencia Artificial (IA) están transformando nuestra vida de manera incuestionable, ya sea desde aspectos cotidianos como comprar un boleto de avión, traducir un texto en un idioma desconocido e incluso la manera en cómo nos movilizamos (remodelación del transporte público y privado). Cambios en la salud y sanidad, la ciencia, en el manejo de datos o la privacidad, son espacios que tampoco han quedado fuera de las innovaciones que ha ido generando la IA. Sin duda alguna, también transformará el empleo e incluso la guerra (i.e. utilización de armamento autónomo). Según López de Mantaras algunos de los temas de investigación más importantes en el campo de la IA (y con un mayor posible impacto social) serán: el aprendizaje automático, los sistemas multiagente, el razonamiento espacial, la planificación de acciones, el razonamiento basado en la experiencia, la visión artificial, la comunicación multimodal persona-máquina y la robótica humanoide (Morte y Monasterio 2017, pág. 305), todas ellas con una infinidad de aplicaciones para la vida cotidiana. En este artículo nos centraremos en la robótica, un área que ha tenido un gran progreso en su interacción con la IA. Abordaremos específicamente el tema de los Robots sociales (RS) con tal de mostrar algunas características que puedan servir como un parámetro para su definición. Nuestra propuesta se basa en que un RS es mucho más que un objeto tecnológico especial, sino que adquiere las características de un sustituto. En un segundo momento trataremos algunos desafíos morales de su uso, haciendo mención concretamente al caso de los robots (sociales) para el cuidado y la asistencia.

1. LA IA COMO UNA TECNOLOGÍA DISRUPTIVA

La IA, así como tecnologías derivadas están generando una gran cantidad de cambios en nuestra sociedad. A pesar de la gran cantidad de definiciones que podemos encontrar de IA, por lo general coinciden en la capacidad de los robots, máquinas o sistemas para realizar tareas que tradicionalmente han sido ejecutadas por seres humanos. Su potencial disruptivo se presenta en sectores tan importantes como el educativo, las ciencias de la salud, las disciplinas jurídicas e incluso en la economía. Esto hace que algunos especialistas consideren que la IA es la tecnología

más floreciente de esta década, llamando, incluso, nuestro presente como la «Era de la IA» (Mondal 2020, pág. 424). Esta posición encuentra sustento en el hecho de que países como Japón, Estados Unidos, Brasil o México estén invirtiendo gran cantidad de recursos en investigaciones relacionadas con la IA. Solo China, por ejemplo, reúne el 73 por ciento de la inversión mundial, con el firme propósito de ser líder mundial en esta área en el año 2030. Según datos del año 2018 del *Allen Institute for Artificial Intelligence*, China superaba en número a los Estados Unidos en cuanto a publicación científica sobre la IA. Esto, entre otras cosas, hace pensar a Kai-Fu Lee (2020) que China se convertirá en la próxima superpotencia de la IA. Este tipo de datos nos muestran que los avances en IA tendrán consecuencias sociales masivas. Por ejemplo, solo la automoción autónoma podría, potencialmente, reemplazar millones de trabajos en los años por venir. Esto sin contar aspectos como reconstrucción de la infraestructura, la protección y seguridad de los vehículos y pasajeros o el resguardo de los datos. Camps (2020) considera que «Toda innovación tecnológica es positiva y puede aprovecharse orientándola al bienestar de la humanidad, pero con cautela. El principal inconveniente de esas precauciones es que hay que ir ejerciéndolas sobre la marcha.» (<https://tinyurl.com/y69gejem>). En América, por ejemplo, el *Berkman Klein Center for Internet & Society*, considera que, debido precisamente a estos nuevos desafíos disruptivos debería tenerse en cuenta, lo que ellos llaman «The principle of consideration of long term effects» que implica una deliberación a largo plazo de los impactos, y no únicamente una visión momentánea que es la que parece predominar en las reflexiones sobre la IA. Sobre todo, a partir de dos potenciales problemas: 1. Los diseñadores humanos de los sistemas de IA pueden incorporar sus propios valores en el diseño y 2. Las infinitas interacciones complejas. Un sistema de IA puede interactuar con su entorno de una manera que conduzca a resultados impredecibles¹ (esto es especialmente relevante en el tema de los RS). En principio, el gran avance de la IA en el mundo, así como la planificación estratégica-económica de

¹ Este es uno de los motivos por los cuales, en América «La Declaración de Montreal» sigue la misma línea al plantearle a los profesionales (tecnólogos, políticos y éticistas) anticipar los riesgos y proponer acciones, para mitigar los posibles efectos negativos.

un número cada vez mayor de países (i.e China, Estados Unidos o la UE) debería plantear un debate social sobre la necesidad de su regulación desde un punto de vista de políticas públicas, así como desde un enfoque ético². Es así como los principales desafíos se relacionan con el tema de los derechos humanos. Cortina (2020) señala con relación a las nuevas tecnologías (incluyendo la IA) que estas deben ser «transparente (sic) porque la tecnología debe buscar el bien común, promover un mundo inclusivo, no perjudicar a nadie, permitir la autonomía de las personas y la justicia» (<https://tinyurl.com/y3nl4j4q>) elementos que claramente hacen referencia al «espíritu» de los derechos humanos³ (en sentido amplio). The *Carr Center for Human Rights* ubica de manera más contundente el tema de los valores y Derechos Humanos en el centro de la discusión:

«So this is yet another case where a philosophical problem assumes new relevance: our philosophically preferred understanding of meta-ethics must enter to judge if we are comfortable putting human rights principles into the design of AI, or not.» (Risse 2018, pág. 10)

Aunque lo cierto del caso, es que frente a la falta de aspectos normativos (y éticos en general), se hacen necesarios aspectos normativos claros que vayan más allá de simple implementación de estas tecnologías o la forma en como promover su uso. Actualmente nos enfrentamos a desafíos

² En América Latina existen algunas estrategias nacionales (i.e Uruguay, Brasil o México han sido pioneros) aunque esto no implica el establecimiento de una regulación particular, ni muchos menos un posicionamiento claro con respecto al tipo de regulación o postura ética deseada. Solamente existen ideas básicas sobre algunos puntos que deberían ser tratados. Por mencionar un caso, la estrategia brasilera sobre la IA menciona que «os principais pontos de questionamento dizem respeito aos limites da aplicação da Inteligência Artificial, às implicações de seu uso em diferentes domínios econômicos e à necessidade de conjugar a tecnologia com o julgamento humano.» Un sentir similar lo encontramos en la OCDE con su lista de 5 principios basados en valores para una administración responsable que, *grosso modo*, menciona que los sistemas de IA deberían diseñarse en forma tal que respeten las leyes nacionales y los derechos humanos, sin operativizar ninguna forma concreta para esto.

³ La *Agencia Latinoamericana de Información* menciona que «Un eje predominante de crítica y regulación de la IA son las normas internacionales de derechos humanos. Considerando el impacto sobre la privacidad, la libertad de expresión y la libertad de asociación, entre otros, el marco de los derechos humanos es un requisito mínimo al que deben adherir los sistemas de IA.» ALAI noviembre, 2019, pág. 2.

éticos y políticos relacionados con la responsabilidad por los sistemas autónomos y semi-autónomos, su inclusión en ámbitos médicos-asis-tenciales o su interacción con los seres humanos (robots sociales), por ejemplo.

2. SOBRE EL CONCEPTO DE ROBOT

Para comenzar, es necesario indicar que no existe una definición unánimemente aceptada de lo que es un robot. Estas difieren según el contexto sociocultural o geográfico (i.e entre Japón o la UE). Además, debido a la especialización cada vez mayor de la robótica, encontramos un gran número de definiciones y taxonomías que varían según su función, aplicación, componentes etc. Aun dentro de esta complejidad podríamos decir que “los robots son dispositivos mecánicos dirigidos por circuitos electrónicos o programas informáticos que se caracterizan por su capacidad para desarrollar actuaciones físicas” (Quadra-Salcedo y Fernández del Castillo et al. 2018, pág. 208). Otra una definición razonablemente clara la encontramos en el ámbito industrial. En este sentido, el robot más común es el industrial (RI) o de producción. Según la *Robotics Industries Association* (RIA) un robot industrial⁴ es: «a programmable, mechanical device used in place of a person to perform dangerous or repetitive tasks with a high degree of accuracy.» (<https://tinyurl.com/y2n73bmx>)

⁴ Desde luego que existen subcategorías desde este marco de referencia. Por ejemplo: los robots cartesianos, GANTRY robots, SCARA robots o robots con brazo articulado.



Ejemplo de un robot industrial articulado
KUKA Roboter GmbH, Bachmann, Dominio Público, via Wikimedia Commons

Estos robots son manipuladores (procesan y modifican información contextual) multipropósito (i.e ensamblaje de partes, pintura, soldadura, revisión etc.) y reprogramables que ejecutan tareas rutinarias. Pueden ser controlados por un humano o ser totalmente automáticos. Esta clase de robots son los más utilizados por las empresas, debido a su versatilidad de funciones. Datos sobre Europa muestran que «Del 2010 al 2014 las ventas de robots aumentaron una media del 17 % por año y es la industria electrónica la que está generando el mayor número de ofertas de empleo» (*Normas de Derecho Civil sobre Robótica*). Asimismo, según el informe «World Robotics Industrial Robots 2019» en el resto del mundo en el año 2018 la venta de robots industriales creció un 6 por ciento y se esperaba que a partir del 2020 y hasta el 2022 tuviera un crecimiento del 12 por ciento anual.

Sin embargo, este tipo de robots, difieren bastante de aquellos que se encuentra enriquecidos con aplicaciones de IA (i.e reconocimiento de voz, reconocimiento facial, aprendizaje automático o la utilización de aprendizaje profundo) y que además están equipados con habilidades de autoaprendizaje e incluso iniciativas de acción (i.e autonomía). Es aquí

donde también encontramos algunas definiciones que están relacionadas con estas características. Es así como desde el tema de autonomía Arkin (2019) utiliza definiciones separadas para robot y robot autónomo. Según estas, un robot sería «una máquina o vehículo automatizado, capaz de percepción, razonamiento y acción independientes» (Arkin 2009, pp. 50-51). Mientras que un robot autónomo es

«un robot que no requiere la participación humana directa, excepto para la misión de alto nivel; dicho robot puede tomar sus propias decisiones de acuerdo con su misión sin requerir autorización humana directa, incluidas las decisiones sobre el uso de la fuerza letal» (Arkin 2009, pp. 50-51).

El Parlamento Europeo (*Normas de Derecho Civil sobre Robótica*) sigue esta línea de interpretación cuando se refiere como «robots inteligentes» aquellos que tienen la capacidad de adquirir autonomía, autoaprendizaje (a partir de la experiencia y la interacción), la capacidad de adaptar su comportamiento y acciones al entorno, así como un soporte físico y la inexistencia de vida en sentido biológico. Desde esta postura es claro que la principal característica de este tipo de robot es la autonomía⁵.

⁵ Es usual que, en los debates filosóficos sobre el tema, los criterios de autonomía pueden ser muy exigentes, mientras que, en otras áreas, como la de los robots con IA, la exigencia disminuya. Por ejemplo, ¿Un sistema es autónomo con respecto a qué tarea? ¿Es posible la «autonomía total» con respecto a todas las tareas? Y si esto fuera posible, ¿Un robot seguiría teniendo el mismo estatuto ontológico?



Ejemplo de un robot «mesero» en China
N509FZ, CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0/>>,
via Wikimedia Commons

Sobre este aspecto, tampoco existe un consenso sobre lo que implica la autonomía⁶, aunque es una preocupación central con respecto al

⁶ El Parlamento Europeo ha hecho un intento sugestivo, aunque no exento de dificultades, cuando menciona que: «Considerando que la autonomía de un robot puede definirse como la capacidad de tomar decisiones y aplicarlas en el mundo exterior, con independencia de todo control o influencia externos; que esa autonomía es puramente tecnológica y que será mayor cuanto mayor sea el grado de sofisticación con que se haya diseñado el robot para interactuar con su entorno; (...) Considerando que, cuanto más autónomos sean los robots, más difícil será considerarlos simples instrumentos en manos de otros agentes (como el fabricante, el operador, el propietario, el usuario, etc.); que esta circunstancia, a su vez, suscita la cuestión de si la normativa general sobre responsabilidad es suficiente o si se requieren normas y principios específicos que aporten claridad sobre la responsabilidad jurídica de los distintos agentes y su responsabilidad por los actos y omisiones de los robots cuya causa no pueda atribuirse a un agente humano concreto, y de si los actos u omisiones de los robots que han causado daños podrían haberse evitado». (*Normas de Derecho Civil sobre Robótica*)

desarrollo de este tipo de robots⁷. Una idea elemental que subyace es la capacidad de tomar decisiones propias (en oposición a la heteronomía), sin la influencia de otra persona o robot, pero las definiciones varían según el contexto y la postura filosófica-social de la que se parte⁸. El Parlamento Europeo en una Resolución del 2017⁹, continúa con el abordaje de este tema, ya que la autonomía de los robots implica dejar de considerarlos como simples cosas o herramientas, lo que hace necesario revisar su estatus desde un punto de vista moral y legal, para pasar a pensar seriamente si las categorías filosóficas y jurídicas existente son suficientes o es necesario crear otras. Todo esto debido al tema de la responsabilidad. Específicamente en cuanto a la distribución de responsabilidades y el manejo de riesgos en los robots sociales. Principalmente cuando se trata de aplicaciones prácticas, como en el caso de los RS, dedicados a la atención de adultos mayores o pacientes con algún tipo de demencia.

⁷ No únicamente para la UE. Verbigracia, el Parlamento Chileno pensando en este tema menciona que «Los sistemas de IA deberían diseñarse de forma tal que respeten las reglas de la ley, derechos humanos, valores democráticos y diversidad, y deben incluir salvaguardas apropiados» Comisión “Desafíos del Futuro, Ciencia, Tecnología e Innovación”. 2019, pág. 12.

⁸ Por ejemplo, Moor (2006) menciona 4 tipos de «ethical impact agents», que varían en función de la idea autonomía y responsabilidad. Por otro lado, Hellström (2013) utiliza el concepto de «poder autónomo» (autonomous power) para como un medio para evaluar las capacidades relativas de los agentes artificiales con diferentes grados de capacidad física, poder de decisión, y alcance para la interacción.

⁹ « (...) Considerando que, cuanto más autónomos sean los robots, más difícil será considerarlos simples instrumentos en manos de otros agentes (como el fabricante, el operador, el propietario, el usuario, etc.); que esta circunstancia, a su vez, suscita la cuestión de si la normativa general sobre responsabilidad es suficiente o si se requieren normas y principios específicos que aporten claridad sobre la responsabilidad jurídica de los distintos agentes y su responsabilidad por los actos y omisiones de los robots cuya causa no pueda atribuirse a un agente humano concreto, y de si los actos u omisiones de los robots que han causado daños podrían haberse evitado.» Resolución del Parlamento Europeo, del 16 de febrero de 2017.

3. ¿QUÉ ES UN ROBOT SOCIAL?

De manera similar con el caso anterior, el tema de los Robots sociales (RS) ha sido abordado, desde distintas posiciones. Dautenhahn y Billard (1999), por ejemplo, describen a los robots sociales como «agentes encarnados» que forman parte de un grupo (i.e una sociedad de robots o humanos). Además, tienen la habilidad de reconocimiento (mutuo) y de participar en interacciones sociales. Incluso, existe la posibilidad de que puedan poseer historias (perciben e interpretan el mundo en términos de su propia experiencia, aunque esta no es una condición necesaria). Así mismo se comunican explícitamente (palabras o gestos) y aprenden unos de otros. Fong et al, (2003) no hablan de «robots sociales», sino de «robots socialmente interactivos» (RSI) con tal de subrayar la capacidad del robot para la interacción social.



Imagen de Quori, un robot socialmente interactivo.

Mary Mark Ockerbloom, CC BY-SA 4.0

<<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Las características que pueden mostrar esos RSI se pueden resumir, aunque no de manera definitiva, en la capacidad de expresar y percibir emociones, establecer y mantener relaciones sociales y usar señales naturales.

No obstante, estas características son aún más complicadas de definir que en el caso de un robot-no social (i.e un RI). Sirva de ejemplo, el caso de la «comunicación», ¿Esta se refiere a transmitir información simplemente o algo más complejo? Jones (2017) considera que «robots cannot become authentic social agents unless they (somehow) possess dialogicality». (Jones 2017, pág. 558). Como vemos para ella, participar en una acción dialógica (y por lo tanto «tener algo que decir») es necesario para considerar un artefacto como auténticamente social. Por lo que es necesario una relación de igualdad (i.e reconocimiento de la otredad). Según Habermas (1999, 2002) para una relación dialógica se ocupa un entendimiento intersubjetivo (¿sería necesaria la intencionalidad el RS para esto?) para generar una racionalidad comunicativa. Lo cual es difícil de comprender en el caso de un RS, ya que no tienen subjetividad (aunque si pueden tener individualidad), al menos por ahora. Por otro lado, si partimos de la idea de Freire (1970) de que la acción dialógica promueve el entendimiento, la creación cultural y la liberación de los seres, entraría en juego el tema de las relaciones de poder. ¿Acaso los RS no deben obedecer nuestras ordenes? ¿Podría existir un RS en un contexto no-jerárquico? (i.e ontológico).

Otros autores introducen nuevos requisitos. Este es el caso de Nitsch y Popp (2014) que consideran que:

«In order for social robots to interact closely and intuitively with humans, the robots need to predict human intentions and actions and display behavior that is appropriate to the context. Moreover, intuitive interaction requires the development of new forms of communication between humans and robots» (Nitsch y Popp 2014, pág. 622)

Lo que implica la capacidad de reconocer emociones (o al menos de expresarlas) e interactuar de manera emocional en algún nivel con los seres humanos. Para esto ¿Será necesario que los RS tengan alguna forma antropomórfica (incluso humanoide -humanlike-) o zoomórfica? Posiblemente sí, en vista de los elementos empáticos que pueden generar este tipo similitudes¹⁰, aunque todo esto continua en un terreno

¹⁰ No todos los autores consideran que esto deba ser así. De hecho, también se podría hablar de robots sociales no humanoides (NH-SR) o zoomórficos. «NH-SR's have *spatial, physical,*

experimental y bastante hipotético. Es posible que las propuestas que buscan crear RS similares a humanos o animales para interactuar estén mediadas en algún sentido por la premisa de la similitud. Dicho en otras palabras, es fácil aceptar como una potencial «compañía» algo parecido a un humano o un animal. Los RS con cierta semejanza a humanos y animales nos muestran que la comunicación exitosa no solo se basa en la información o en las claves sociales, sino además en los elementos que se tienen en común, así como en los elementos comunes que se pueden obtener por medio de la interacción.

Lo que queda claro es que en todas estas interrogaciones sobre los criterios mínimos para que un robot sea social se muestra una pregunta sobre las condiciones mínimas para una interacción social. Lo cual queda fuera de la pretensión de este artículo. Para nuestros propósitos, definiremos a un RS con las siguientes características: (I) *Socialmente evocador*. Los RS evocan sentimientos sociales que más allá de la mera utilidad. El ser humano en algún sentido amplio se «involucra» con él (Breazeal 2002, 2003). (II) *Situación social*. Los RS están rodeados de un entorno social al que perciben y reaccionan, de manera distinta a un robot-no social (i.e RI). Esto implica a un «robot socialmente situado» con capacidad de distinguir entre otros agentes sociales y varios objetos en su entorno (Fong et al. 2003). (III) *Sociable*. Existe una relación proactivamente con los humanos para satisfacer objetivos sociales internos (i.e deseos o emociones) (Breazeal 2002, 2003). (IV) *Socialmente inteligente*: Los RS muestran algunos aspectos de la inteligencia social (Dautenhahn 1998) parecida a la humana (no es necesario que *tengan verdaderamente* inteligencia social, sino *solamente exhibirla*).

Como podemos ver, un RS no es simplemente un dispositivo mecánico que ejecuta una serie de tareas repetitivas, actuando de manera automática con su entorno. Posee conjuntamente, una serie de atributos que les

androbotic features to interact with or respond to users, either virtually or physically or using digitalphysical hybrid systems. The human-robotic interactions in this type of robotic environment create social dialogs in the form of lights, sounds, and movements, i.e., from the physical architectural level. Users, too, start to engage using their senses by talking, touching, moving or by using different types of body language, creating gestures and cues to activate and interact with these ubiquitous computing environments» (Korn 2019, pág. 16-17.)

permiten reconocer y participar en interacciones sociales con otros RS y con los humanos. Son precisamente estos atributos, lo que hacen que un RS no sea simplemente una máquina o un objeto tecnológico como cualquier otro.

3.1. LOS ROBOTS SOCIALES COMO SUBSTITUTOS

Si bien, un RS sigue siendo una máquina, este se convierte en un socio o compañero privilegiado de los humanos. Su interacción con todos los elementos de su entorno es parte de un proceso dinámico de interacciones que determinarán cómo están dispuestos hacia sus compañeros humanos y cómo son con los robots con los que interactúan. De hecho, cuando los robots están diseñados para encajar en un entorno social, ya no pueden ser considerados como simples objetos técnicos-autónomos, sino que deben ser vistos como elementos que forman parte de sistemas sociotécnicos que modifican y participan plenamente en los eventos que tienen lugar¹¹. Es así, como se debe tomar en consideración el acoplamiento entre el RS, las personas con las que interactúa, los otros objetos, el entorno y la organización social en la que se inserta. Esta caracterización hace que un RS no sea simplemente un «robot con funciones sociales» sino que son fundamentalmente «substitutos» (Dumouchel y Damiano, 2017). Por lo tanto, su «esencia» radica en ser un rol o una actividad y no un objeto propiamente. Un RS busca substituirnos en alguna tarea (i.e cuidado de niños, compañía de mayores, atención de personas con demencia) pero sin tomar propiamente nuestro lugar.

«To create an artificial nurse's aid, a robot companion for the lonely, or a mechanical nanny that takes care of children while their parents are away, indisputably amounts to introduce into our ordinary social relations artificial social agents that substitute for human (or animal agents)...» (Dumouchel y Damiano, 2017, pág.32)

Ser un sustituto, implica también cierto tipo de autoridad. No es simplemente cambiar una cosa por otra. O a una persona con un RS. No

¹¹ Esto implica que el ingeniero o el diseñador debería tener en cuenta esta inclusión de los robots (incluidos los aspectos éticos, legales, sociales y medioambientales) en el proceso de diseño de sus funcionalidades, fiabilidad y rendimiento.

es posible cambiar a un padre ausente, con otro ser humano, por ejemplo. Un docente, puede sustituir a un docente enfermo, cuando el primero tiene ciertas características, además de cierta legitimidad. Es decir, debe poder ejercer legítimamente su rol en virtud de una asignación, autoridad o condición especial. Un RS de cuidado puede sustituir adecuadamente la tarea de bañar a una persona anciana, que tiene un enfermero, debido a su diseño y características, pero no puede sustituir al enfermero en sí mismo. De hecho, según Dumouchel y Damiano (2017) un RS como sustituto posee 3 características importantes: (I) Aún cuando ser sustituto implica una función o rol, un RS no podría reducirse a ellas únicamente solamente. Es decir, un RS implica una capacidad de adaptación (que no es infinita) a una amplia gama de funciones. En la interacción con un ser humano, un RS de compañía debería ser capaz de dar compañía de muchas maneras y no solamente de una forma.¹² Un robot mascota de forma zoomorfica que acompaña a un niño durante su infancia puede necesitar mejorar sus habilidades para mantener el interés del niño¹³. (II) Tienen la habilidad de estar presentes. Esta habilidad constituye en sí mismo una acción, una afectación en el mundo de objetos físicos. Lo que implica que tienen una presencia social física. (III) Están dotados con una autoridad que pueden ejercer. Una madre tiene autoridad cuando le dice a su hijo que está haciendo berrinches que se calme o se vaya a otro cuarto. El niño reconoce la autoridad de su madre si efectivamente se calma o se va a otro cuarto. Por lo tanto, un RS (como sustituto) debe tener la capacidad de establecer algún tipo de autoridad con tal de cumplir adecuadamente su rol y tomar las decisiones necesarias para su función. En este punto, entran en juego las características que antes mencionábamos con la idea de que el ser humano se involucra de alguna

¹² Esto es un problema grave, ya que lo que un ser humano necesita no es lo mismo que otro busca. Lo que implica una dificultad (o imposibilidad del todo) para generalizar o trasladar el aprendizaje de un RS hacia otro.

¹³ Para esto se necesitan implementar modelos profundos de cognición humana y así como de la psicología social.

forma con el RS, así como sus capacidades dialógicas (o al menos comunicativas) o el aprendizaje profundo.

Todo esto implica que «human users tend to perceive social robots as interlocutors that break the object/subject divide. They tend to recognize these robots as a new category of interactors, with whom they can establish social relationships.» (Damiano y Dumouchel 2018, pág. 8), por lo que un RS no solo no es un objeto tecnológico común, sino que además se convierte en un verdadero agente social, que participa verdaderamente en las dinámicas sociales. En consecuencia, nosotros no usamos a los RS, nos relacionamos con ellos.

4. EL CASO DE LOS ROBOTS DE CUIDO Y ASISTENCIALES.

Las características de los robots sociales hacen que se acerquen mucho a nuestras vidas privadas. Debido a su interacción con nosotros surgen una serie de preguntas de gran importancia: ¿Queremos que ayuden a las personas mayores? ¿Queremos que nos ayuden a nosotros en nuestra vejez? ¿Aceptaríamos que nos ayuden para ir al baño?, ¿Estaríamos de acuerdo en que nos alimenten si sufrimos de algún accidente? ¿Estamos seguros en que deberían cuidar a un pariente con Alzheimer? A pesar de que las respuestas a estas preguntas difieren de una persona a otra, así como de la cultura y de las experiencias personales, el uso de estos robots ha encontrado un terreno fértil en los ámbitos médicos-asistenciales o de cuidado. Por mencionar un caso; Japón en un documento del año 2015 señaló como meta:

«Creating safe and stable work environments will be pushed forward by using robots that aid the aged to lead self-sustaining life by helping them move around when they go out so that they can live in a region to which they are accustomed even though they need care, and by introducing to medical sites the nursing robots that mitigate the physical burdens of care workers by making it easy for those who they look after to transfer. In addition, use of such robots for preventing people from needing care, rehabilitation, and health promotion will be propelled. Moreover, encouraging introduction of the robots that are applicable to medical institutions will also be considered. » (The Headquarters for Japan's Economic Revitalization 2015, pág. 64)

Si tenemos presente la gran cantidad de personas mayores en la sociedad, junto con los recientes avances en robótica, la implementación de los robots socialmente asistenciales y de cuidado (RsAC) podría ser beneficiosa. Un robot asistencial o de cuidado, es aquel que brinda ayuda, soporte o cuidado a un ser humano. En cambio, un RsAC no solo está diseñado para ayudar al ser humano a realizar tareas difíciles o rutinarias, sino que además «(...) but to give assistance through social interaction to achieve progress in, for example, convalescence, rehabilitation, and learning.» (Bemelmans et al. 2012, pág. 115).



Imagen de Paro. Robot terapéutico con forma zoomórfica (foca)

De Aaron Biggs, Flickr user ehjayb - <https://www.flickr.com/photos/ehjayb/21826369/>, CC BY-SA 2.0, <https://commons.wikimedia.org/w/index.php?curid=727258>

Proyecciones de la Naciones Unidas (2008) estiman que para el año 2030 con relación al aumento de la población anciana, Estados Unidos verá un aumento del 40 por ciento, Europa un 50 ciento y en Japón será del 100 por ciento. Al mismo tiempo, el número de personas mayores de 80 años aumentará en un 100 por ciento en todos los continentes. Igualmente, en las próximas décadas este crecimiento se acelerará, llegando el número de personas mayores de 60 años a alcanzar la cifra de 2.100 millones (ONU, 2015). Esto hace que la demanda del personal calificado para atender a esta población aumente. Este es uno de los 2

motivos principales (sin contar los posibles beneficios) que habitualmente se mencionan en la literatura por los que sería necesario el uso de RS asistenciales y de cuidado para atender a este tipo de población:

«First, it is expected that western countries will face a tremendous shortage on staff and qualified healthcare personnel in the near future¹. Second, people prefer more and more to live in their own homes as long as possible instead of being institutionalized in sheltered homes, or nursery homes when problems related to ageing appear.» (Broekens et al. 2009, pág. 95)

Las principales áreas en las que estos RsAC se insertarían serían: (I) La ayuda a los ancianos para monitorear su salud (II) La vigilancia de su seguridad y (III) Proporcionarles compañía. Todas estas funciones dentro de la finalidad de mejorar su bienestar general, autonomía e independencia, permitiéndoles vivir de forma independiente durante más tiempo. En teoría,

«socially assistive robotics seems to hold a potential of opportunities for the provision of care and enhancing quality of life, particularly against the background of increasing aging and limited availability of care providers, resulting in more people suffering from mental problems and psychological and social isolation.» (Bemelmans et al. 2012, pág. 117)

Aunque no todas son buenas noticias, varios problemas son mencionados regularmente en la literatura sobre el tema: (I) La potencial reducción del contacto humano (II) Sentimientos de objetivación, pérdida de control, engaño o «infantilización» (III) La pérdida de la privacidad (IV) Una posible pérdida de libertad personal y (V) La forma en cómo se deberían determinar las circunstancias en las que se debería permitir a las personas mayores o con demencia controlar a los RS.

El planteamiento detrás del problema (I) es que, si esta tecnología es accesible y eficiente, seguramente dedicaremos la tarea del cuidado de los ancianos a los RsAC, descuidando o cortando del todo los vínculos sociales y afectivos con la población mayor, lo que finalmente implicaría una deshumanización de este tipo de cuidado.¹⁴ Si partimos de la idea del

¹⁴ Cuestionamiento que también se encuentra en el Parlamento Europeo cuando señala que «el contacto humano es uno de los aspectos fundamentales de la atención a las personas;

RS como sustituto es claro que un RsAC nunca puede suplantar al ser humano, sino solo sustituirlo en una función concreta. En su interacción con los seres humanos, un RsAC no podría suplir el vínculo humano, ya que lo que hace es crear otro tipo de vínculo social. Nuestros vínculos con los animales, por más fuertes y hermosos que sean, no reemplazan nuestra interacción con los humanos. Pero interactuar con animales puede enriquecer nuestras vidas, así como también podría hacerlo un RS. Un ser humano tiene presente que un RS no es otro humano, pero esto no obsta que sea un compañero o que lo podamos ver como un soporte motivacional. En cuanto a la situación (II) es un problema que no es intrínseco a la interacción con un RsAC. De hecho, es un problema habitual que enfrentan las personas ancianas en buena parte del mundo occidental. A diferencia de Asia o en las tradiciones indígenas de los países americanos en que la población anciana representa el conocimiento y son motivo de respeto y reverencia, para Occidente (o para las culturas fuertemente occidentalizadas) los adultos mayores suelen ser un estorbo o un motivo de mucha incomodidad. Esto hace que se les cosifique, se aprovechen de ellos o se les «infantilice», limitándoles su capacidad de decisión o restringiendo su autonomía. Por lo que la suposición subyacente es que los RsAC mantendrán o perpetuarán los juicios humanos tradicionales sobre estos puntos. Un RsAC es un compañero de los seres humanos, que al igual que otros seres aprende (en este caso debido a la IA) de nuestros actos y conducta con los demás. Si nuestras interacciones con los demás humanos (los ejemplos con los que entrenamos a un RS) son discriminadoras, agresivas o cosificadoras, el RS posiblemente aprenderá que esta es la manera adecuada de actuar en un contexto específico. Estos aspectos negativos, lejos de toda duda, son un riesgo pero están relacionados con el diseño del RS, así como con

considera que sustituir el factor humano por robots podría deshumanizar la prestación de cuidados, pero, por otra parte, reconoce que los robots podrían realizar las tareas automatizadas de quienes prestan cuidados, aumentando la atención prestada por seres humanos y haciendo más selectivo el proceso de rehabilitación, lo que permitiría al personal médico y asistencial dedicar más tiempo al diagnóstico y a opciones de tratamiento mejor planificadas...» Resolución del Parlamento Europeo, de 16 de febrero de 2017.

la interacción misma humano-humano, no tanto con su implementación. Por lo que los roboticistas deberían actuar de manera interdisciplinaria con filósofos, psicólogos, sociólogos y juristas especialistas en derechos humanos para obtener en el proceso de diseño (como ya mencionamos antes) los aspectos éticos, jurídicos, sociales y medioambientales y que estos puedan implementarse dentro de las funcionalidades, fiabilidad y rendimiento del RsAC.

El tema (III) sobre la pérdida de la privacidad y (IV) Una posible pérdida de libertad personal, aun siendo distintos, se encuentran relacionados al menos desde el punto de vista de la interacción con un RsAC. Dentro del ámbito general de la IA, estos problemas se asocian a la recopilación y al manejo de datos personales. Estos asuntos toman relevancia sobre todo en aquellos RsAC que estén online o que envíen periódicamente la información una persona a un centro médico, por ejemplo. Ciertamente la privacidad es un valor fundamental que debe ser considerado en cada caso particular en el cual intervenga un RsAC. Habría que considerar quién está en control del RsAC, si es un servicio público o privado y quién asume los riesgos y los beneficios. Empero, en el caso de los RsAC este tema se relaciona más con la intrusión de un agente extraño en un espacio personal y de confianza. Nadie quiere que un extraño tenga el conocimiento de nuestras rutinas privadas y nuestros aspectos más íntimos, especialmente en caso de personas ancianas, con alguna discapacidad o en condición de vulnerabilidad. Menos aún que un objeto extraño nos vigile en cada uno de nuestros pasos y nos diga que podemos o no hacer, generando una sensación de incomodidad o incluso pérdida del control. Esta situación toma un matiz distinto, si partimos del hecho de que un RsAC no es un (objeto) extraño, ni se inserta de manera agresiva (al menos en la teoría) en un entorno social particular. Cuando un RsAC ingresa a la vida cotidiana de una persona, como en cualquier otra situación, hay un período de conocimiento y adaptación que puede resultar en unos casos bien y en otro mal. En principio, un RsAC busca colaborar con las personas como socios, construyendo una relación social. Obviamente para esto se necesita una relación altamente personalizada, adaptativa y a largo plazo a través del compromiso social y emocional. Lo cual siempre es un riesgo en

cualquier tipo de interacción social. En el caso de los RsAC es un tema que necesita mayor investigación empírica con tal de poder brindar un panorama más claro, sobre todo en los casos en que RsAC ocasione un daño involuntario o incluso voluntario.

Finalmente, el problema (V) se relaciona con el tema de la autoridad que mencionamos por primera vez cuando nos referimos a las características del sustituto. ¿Cuál sería la capacidad de decisión que debe tener un RsAC? La solución no es clara. Un acercamiento coherente implicaría que la respuesta a esto estaría estrictamente relacionada con la función o el rol que se substituye. Si un RS acompaña a una persona enferma y la ayuda a tomar sus medicamentos de manera adecuada, la autoridad del RsAC debería estar relacionada con esta actividad. Sin embargo, situaciones hipotéticas como estas se complican en casos con personas con demencia o alzheimer. Al igual que todos los ejemplos previos, se necesita mayor investigación empírica, así como el desarrollo de propuestas y acercamientos interdisciplinarios que permitan que los RS verdaderamente sean una vía de enriquecimiento humano y no un medio por el cual las relaciones sociales sufran un deterioro.

CONCLUSIONES

Ninguno de los argumentos previos pretende ser exhaustivo, ni mucho menos definitivo. La aceptación final de los RS por parte de la sociedad dependerá de hasta qué punto son seguros y que tanto puedan asemejar nuestros códigos sociales de conducta, generando confianza y empatía hacia ellos. Los RS al estar diseñados para interactuar con seres humanos de una manera coherente con la psicología social, pueden servirnos para aprender algo de nosotros mismos. Los robots (en general y más aún los RS) son un reflejo de nuestra propia humanidad y su desarrollo nos lleva a reflexionar muchos elementos que creemos fundamentales de la condición humana. Verbigracia, aspectos éticos y sociales que todavía no tienen una respuesta.

REFERENCIAS BIBLIOGRÁFICAS

- ALAI (noviembre, 2019): Inteligencia artificial: ddhh, justicia social y desarrollo (43).
- ARKIN, R. (2009). *Governing Lethal Behavior in Autonomous systems*. Boca Raton, FL.: CRC Press.
- BEMELMANS, ROGER; GELDERBLUM, GERT JAN; JONKER, PIETER; WITTE, LUC DE (2012): Socially assistive robots in elderly care: a systematic review into effects and effectiveness. En: *Journal of the American Medical Directors Association* 13 (2), 114-120.e1. DOI: 10.1016/j.jamda.2010.10.002.
- BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY OF HARVARD UNIVERSIT (2020): Coconet | Human Rights Concerns on Artificial Intelligence in Southeast Asia: An Overview. Recuperado de <https://tinyurl.com/y68nyow6>
- BREAZEAL, C. (2002) *Designing sociable robots*. Cambridge, MA: MIT Press.
- BREAZEAL, C. (2003) Towards sociable robots. *Robot. Auton. Syst.* 42, 167–175. DOI :10.1016/S0921-8890(02)00373-1)
- BREAZEAL, C. (2004) Social interaction in HRI: the robot view. *IEEE Trans. Syst. Man Cyber.: Part C* 34, 181–186. DOI:10.1109/TSMCC.2004.826268)
- BROEKENS, J.; HEERINK, M.; ROSENDAL, H. (2009): Assistive social robots in elderly care: a review. En: *Gerontechnology* 8 (2). DOI: 10.4017/gt.2009.08.02.002.00.
- CAMPS, V. (2020). Ser complacientes con la realidad no tiene nada que ver con la felicidad. *Ethic*. Recuperado de <https://tinyurl.com/y69gejem>
- COMISIÓN “DESAFÍOS DEL FUTURO, CIENCIA, TECNOLOGÍA E INNOVACIÓN”. (2019): *Inteligencia Artificial para Chile. La urgencia de desarrollar una estrategia*.
- CORTINA, A (2020): "La inteligencia artificial debe ser justa e inclusiva". Recuperado de <https://tinyurl.com/y3nl4j4q>.
- DAMIANO, L; DUMOUCHEL, P (2018): Anthropomorphism in Human-Robot Co-evolution. En: *Frontiers in psychology* 9, pág. 468. DOI: 10.3389/fpsyg.2018.00468.

- DAUTENHAHN, K. 1998 The art of designing socially intelligent agents— science, fiction, and the human in the loop. *Appl. Artif. Intell.* 12, 573–617.
- DAUTENHAHN, K.; BILLARD, A. (1999). Bringing up Robots or - The Psychology of Socially Intelligent Robots: From Theory to Implementation. Washington. Proceedings of the Third International Conference on Autonomous Agents. 3. 336-367
- DUMOUCHEL, P, AND DAMIANO, L (2017) *Living with Robots*, Harvard University Press.
- ESTRATÉGIA BRASILEIRA DE INTELIGÊNCIA ARTIFICIAL (2020). Recuperado de <https://tinyurl.com/yxwblcox>
- EUROPEAN COMMISSION. Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (26/4/2018)
- FONG, T., NOURBAKHSI, I. & DAUTENHAHN, K. (2003) A survey of socially interactive robots. *Robot. Auton. Syst.* 42, 143–166.
- FREIRE, P. (1970). *Pedagogía del Oprimido*. Madrid: Siglo: XXI.
- HABERMAS, J (1999): *Teoría de la acción comunicativa, II*. España. Taurus.
- HABERMAS, J (2002): *Teoría de la acción comunicativa, I*. México. Taurus.
- HELLSTRÖM, Y. (2013). On the moral responsibility of military robots. *Ethics and information technology* 15(2):99–107.
- JONES, R. (2017): What makes a robot 'social'? En: *Social studies of science* 47 (4), pág. 556–579. DOI: 10.1177/0306312717704722.
- KAI-FU, LEE (2020). *Superpotencias de la inteligencia artificial*. Barcelona: Editorial Planeta
- KORN, O. (ed.) (2019): *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*. Switzerland: Springer Nature.
- MONDAL, BHASKAR (2020): Artificial Intelligence: State of the Art. En: Valentina E. Balas, Raghendra Kumar y Rajshree Srivastava (eds.): *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, t. 172. Cham: Springer International Publishing (Intelligent Systems Reference Library), pág. 389–425.

- MOOR, J. (2006). The Nature, Importance and Difficulty of Machine Ethics. *IEEE Intelligent Systems* 21(4): 18–21.
- MORTE, R; MONASTERIO, A (9) 2017): Entrevista a Ramón López de Mántaras. En: *Dilemata* (24), pág. 301–309.
- NITSCH, V.; POPP, M. (2014): Emotions in Robot Psychology. En: *Biol Cybern* (108), pág. 621–629.
- RESOLUCIÓN DEL PARLAMENTO EUROPEO, de 16 de febrero de 2017, con recomendaciones: Normas de Derecho civil sobre robótica.
- RISSE, M. (2018): Human Rights and Artificial Intelligence. An Urgently Needed Agenda.
- QUADRA-SALCEDO Y FERNÁNDEZ DEL CASTILLO, T; PIÑAR MAÑAS, J; BARRIO, MOISÉS (2018): *Sociedad digital y derecho*. 1. ed. Madrid: Ministerio de Industria Comercio y Turismo; Red.es; Boletín Oficial del Estado.
- RISSE, M. (2018): Human Rights and Artificial Intelligence. An Urgently Needed Agenda.
- THE HEADQUARTERS FOR JAPAN’S ECONOMIC REVITALIZATION (2015): *New Robot Strategy*. Japan’s Robot Strategy.
- UNITED NATIONS (2008) Department of Economic and Social Affairs, Population Division, *World Population Prospects: The 2008 Revision*, New York.
- UNITED NATIONS (2015) *World population ageing 2015*. New York: United Nations

LA SMARTIFICACIÓN DE LA ADMINISTRACIÓN PÚBLICA: APOSTILLAS CRÍTICAS A PARTIR DEL CASO EUROPEO

*THE SMARTIFICATION OF PUBLIC ADMINISTRATION:
CRITICAL APOSTILLES FROM THE EUROPEAN CASE*

JONATHAN PIEDRA ALEGRÍA

Profesor de Filosofía
Escuela de Filosofía
Universidad Nacional
Heredia/Costa Rica
jonathan.piedra.alegria@una.cr
ORCID: 0000-0003-4532-4415

Recibido: 17/02/2021

Revisado: 21/08/2021

Aceptado: 6/09/2021

Resumen: La Inteligencia Artificial (IA) y las tecnologías derivadas están generando una gran cantidad de cambios en nuestro mundo. Estudios y publicaciones especializados muestran hasta qué punto las tecnologías de Inteligencia Artificial tendrán numerosas implicaciones que cambiarán de manera profunda y disruptiva nuestra sociedad. En el caso concreto de la Unión Europea, en el año 2018, 24 países miembros de la UE (más Noruega) firmaron la *Declaración de Cooperación sobre Inteligencia Artificial* en la cual se comprometían a trabajar juntos con la finalidad de estimular la adopción de la IA en Europa. En este artículo aludimos a algunos ejemplos europeos relacionados con el potencial de la IA en la gestión pública y presentamos argumentos acerca de por qué la IA para uso público es un campo de investigación teórico sin capacidad para ejercer, en la actualidad, el necesario impacto en los futuros diseños de la Administración pública.

Palabras clave: Administración pública, Europa, Inteligencia Artificial, Razonamiento moral.

Abstract: Artificial Intelligence (AI) and derived technologies are generating a great deal of change in our world. Specialized studies and publications show the extent to which Artificial Intelligence technologies will have numerous implications that will profoundly and disruptively change our society. In the specific case of the European Union, in 2018, 24 EU member countries (plus Norway) signed the Declaration of Cooperation on Artificial Intelligence in which they committed to working together to stimulate the adoption of AI in Europe. In this article, we refer to some European examples related to the potential of

AI in public management and present arguments about why AI for public use is a theoretical research field without the capacity to exert, at present, the necessary impact on future designs for public administration

Keywords: Public administration, Europe, Artificial Intelligence, Moral reasoning.

INTRODUCCIÓN

Entre los especialistas de las Ciencias Humanas, Sociales y Jurídicas que investigan en universidades y centros de investigación de Europa y América existe cierto consenso en torno al impacto que ejercerá la Inteligencia Artificial (IA) en el futuro de la burocracia y, en general, en las Administraciones públicas. La IA y otras tecnologías derivadas están generando una gran cantidad de cambios en nuestro mundo, y parece imponerse la previsión según la cual sus implicaciones conducirán a cambios profundos y disruptivos en nuestra sociedad. A pesar del número cada vez mayor de investigaciones relacionadas con la IA, su utilización en el ámbito público es relativamente joven. Tanto así que es común encontrar en la literatura, textos carentes de descripciones adecuadas, así como de sus posibles desafíos. Esta situación se debe, entre otros aspectos, a una ausencia de una definición común u homogénea de la IA y de algunos de sus elementos derivados. Es así, como previamente al desarrollo de los desafíos para su utilización en la Administración pública, es necesario establecer ciertos conceptos sobre la IA. Sobre esto, hay muchas definiciones de IA dependiendo del área o enfoque del cual se trata. Para los ingenieros informáticos, la IA se puede delimitar al desarrollo de programas que muestran o imitan un comportamiento inteligente (Franceschetti, 2018). Esta definición está en consonancia con la propuesta de John McCarthy, un padre fundador de la IA, cuando la define como la ciencia e ingeniería para hacer máquinas y programas inteligentes (McCarthy, 2018). Entendiendo la inteligencia como la capacidad de conseguir un objetivo. Este tipo de definiciones muestran indirectamente que la IA es más un campo de estudio que algo concreto que pueda describirse de manera sencilla. Incluso, es posible dividirla en varios campos como la robótica o el Big Data, por mencionar algunos casos. Verbigracia, la robótica es un campo interdisciplinario de la ingeniería que se ocupa se ocupa de la construcción y operación de los robots físicos (Franceschetti, 2018). A pesar de la falta de acuerdo universalmente aceptado sobre cómo definir un robot, se puede decir *grosso modo* que es una máquina programable con la capacidad de efectuar ciertas acciones de manera semiautónoma o autónoma. En realidad, la IA y la robótica son inicialmente dos campos diferentes, pero se entrelazan cuando ejemplo, se utiliza la técnica de *Machine Learning* para entrenar a

un robot para que ejecute una tarea. La integración de las tecnologías de la IA y la robótica se ha convertido en un tema de creciente interés para ambos campos.

Algo similar ocurre con el Big Data. Este término es utilizado para indicar una gran cantidad de información variada que llega en volúmenes crecientes y con una velocidad cada vez mayor. Estas características hacen que sea imposible para un ser humano realizar un análisis adecuado. No obstante, usando varias técnicas de IA es posible procesar todos esos datos con tal obtener información valiosa. Como se puede observar, existen relaciones significativas entre todos estos campos, aun cuando pueden ser diferenciables. Sin embargo, generalmente, todos coinciden en la capacidad de los robots, máquinas o sistemas para realizar tareas que tradicionalmente han sido ejecutadas por seres humanos, ya que la IA combina diversas habilidades tales como el denominado aprendizaje autónomo, la capacidad de razonar, la resolución de problemas, e incluso la percepción y el reconocimiento (y uso) del lenguaje natural. Su potencial disruptivo se presenta en sectores tan importantes para el diseño de políticas públicas como el energético, el educativo, el relativo a las ciencias de la salud, el transporte e incluso la administración de justicia.

Según Accenture (Mehr, 2017) para el año 2035 las aplicaciones de IA tendrán el potencial de duplicar el crecimiento económico global. En la actualidad, asistimos a una dura competencia para ver quién se convertirá en el líder planetario, es decir, quien será la superpotencia en este tema (Lee, 2020). China pretende invertir cerca de 147.800 millones de dólares americanos para convertirse en 2030 en el mayor innovador global en el campo de la IA (*State Council of the People's Republic of China*, 2017) y los Estados Unidos de América invirtieron en 2016 más de mil millones de dólares americanos en investigación y desarrollo en tecnologías relacionadas con la IA. Europa no se queda atrás, ya que en el año 2019 el gasto relacionado con tecnologías basadas en IA fue de cinco mil millones de dólares americanos, un 49 por ciento más que en el año 2018 (IDC, 2019). La competencia por el liderazgo a escala planetaria tiene un fuerte componente de promesa económica ya que, al cabo, la mayoría de los gobiernos reconocen abiertamente que dicha carrera se basa, sobre todo, en la percepción y el cálculo económico del potencial previamente asignado a la IA. En 2018 los veinticuatro países miembros de la Unión Europea (más Noruega) firmaron la *Declaración de Cooperación sobre Inteligencia Artificial* en la cual estos se comprometían a trabajar juntos con la finalidad de estimular la capacidad y adopción de la IA en Europa. Aunque el documento hace énfasis en los desafíos socioeconómicos y busca garantizar un marco ético y legal para poder ejercer un control democrático sobre la carrera global por el liderazgo en IA, en este artículo enumeraremos algunas de las dificultades y retos sociopolíticos que han destacado investigadores del área de las ciencias humanas, sociales y jurídicas con el objetivo

de dirimir si la sociedad civil europea está en disposición de abordar este debate en los próximos años y cuáles son los factores que incrementarán o, por el contrario, mermará la capacidad reflexiva de la ciudadanía digital.

1. LA AUTOMATIZACIÓN DE LA BUROCRACIA EUROPEA

Los avances tecnológicos basados en IA y el potencial económico asociado están ganando protagonismo en el contexto del sector público. Por ejemplo, Deloitte (2017) estimó que solo en Estados Unidos la automatización de las tareas de los empleados del Gobierno Federal podría llegar a suponer un ahorro de entre 96,7 millones y 1,2 mil millones de horas al año, lo que equivale a 3,3 mil millones y 41,1 mil millones de dólares americanos respectivamente. Si tenemos presente que las organizaciones y órganos del sector público constituyen los empleadores más importantes en la mayoría de los países (Organización Internacional del Trabajo), el ahorro pecuniario para la Administración pública, derivado únicamente de la automatización de algunas funciones (i.e. de carácter repetitivo), sería enorme. Pero el potencial en el sector público-administrativo no podría resumirse únicamente a esto, ya que además permitiría brindar una mejor calidad en los servicios públicos si se siguieran los pasos para dar lugar a lo que se ha denominado la *smartificación de la Administración pública* (Ramió, 2019). Esta expresión suele emplearse para describir el diseño de procesos encaminados a aumentar la inteligencia estratégica de las instituciones en el gobierno de redes públicas y privadas. Pero en los debates en torno a la *smartificación* de las administraciones públicas muchas veces se olvida que uno de los presupuestos objeto de deliberación es el de si, con estos procesos, se conseguirá aportar valor público a las actividades lideradas por administraciones capaces de interactuar con un modelo determinado de ciudadanía digital. Está fuera de toda duda que la implementación de la IA en el sector público ejercerá un impacto significativo sobre nuestras sociedades. Parece evidente que dichas implementaciones influirán sobre el diseño y la consecución de objetivos y políticas públicas. Tanto así, que no es extraño encontrar la introducción de este tipo de tecnología como una forma de garantizar la eficiencia en los procesos gubernamentales, logrando así una prestación efectiva de los servicios públicos en su totalidad. (Doorsamy *et al*, 2020). Sin embargo, todos los pronósticos tienen en común el hecho de que esto se realiza sin precisar qué modelo de ciudadanía digital se propugna y cuáles pueden ser las consecuencias sociopolíticas y económicas a escala global (en este caso europea) derivadas de las innovaciones disruptivas patrocinadas. No menos problemática es la tarea de dirimir si debiéramos calificar como positivos o negativos en relación con los intereses de la futura ciudadanía digital un escenario en el que

en el cual la IA (i.e *Machine learning*) deba manejar una cantidad masiva de datos (*Big Data*) porque o bien no existan suficientes expertos capaces de analizarlos o bien no se confíe en sus decisiones.

En principio, para la mayoría de las personas podría parecer evidente cuán positivo es el hecho de que una tecnología basada en IA pueda automatizar procesos burocráticos rutinarios en una escala de tiempo a la que no tiene acceso el ser humano en la consecución de tareas cognitivas equivalentes. Sea o no favorable la opinión de las personas ante el escenario de la automatización, lo cierto es que el primer fenómeno objeto de atención para el sentido común es la extrema reducción de cargas administrativas de la que quedarían liberados los funcionarios públicos junto a la opinión extendida de que existiría una correlación entre este factor y la obtención de beneficios en la resolución de la asignación de recursos para asumir tareas de elevada complejidad. ¿A qué se debe la tendencia de las personas a asumir que existe una correlación entre la automatización de procesos burocráticos, la liberación de tareas rutinarias entre el personal administrativo y la consiguiente obtención de beneficios derivados de este modelo de gestión? En el desarrollo y aplicación de la IA a las administraciones públicas no se encontrará la respuesta, ya que esta percepción se basa en el impacto ejercido por la cultura política en la que se hallan los ciudadanos, y en la asunción de un modelo de teoría de elección racional que no da cuenta de los efectos generados por la automatización de las tareas, ni de la digitalización de los procesos.

1.1. GESTIÓN 4.0

De hecho, si tenemos presente la velocidad con que se suceden los cambios tecnológicos resulta prácticamente inevitable que la Administración pública innove hasta tal punto en sus métodos tradicionales como para lograr no solo una mayor eficiencia en sus servicios sino, además, una mayor participación ciudadana, una mejor rendición de cuentas y un nivel de interoperabilidad constante entre los distintos niveles de gobernanza. Aun así, existe un país europeo que ha seguido esta vía, Estonia, considerado el país europeo número uno en ciberseguridad (Balví, 2017). El 98 por ciento de los ciudadanos poseen identidades digitales (ID Digital) y el 46,7 por ciento emite el voto *online*. En 1997 el país adoptó un modelo de digitalización que ha tenido como consecuencia que hoy el 99 por ciento de los servicios públicos estén digitalizados lo que haría de Estonia «el país más digital» del mundo. A este fenómeno de digitalización se le asocia el hecho de que en Estonia se ahorre un equivalente al dos por ciento del PIB en salarios. Incluso, en la página web oficial destinada a promocionarse, el país se presenta como «the world's most advanced digital society.»

El alto nivel de digitalización de Estonia fue determinante para que, en mayo del 2019, se iniciara un proceso de implementación de 16 soluciones de inteligencia artificial (o *kratts*) en el sector público (*Report of Estonia's AI Taskforce*, 2019). Para la mitología de Estonia, un Kratt es un ser mágico animado por un alma humana, compuesto de varios objetos (usualmente paja y elementos domésticos). En lo esencial, un Kratt era un sirviente diseñado para ayudar a los seres humanos. Es por esto por lo que la figura del Kratt en Estonia (no es una denominación que se haya internacionalizado) es utilizada como una analogía de la IA, así como de todas sus complejas relaciones con los seres humanos. En este contexto, un kratt es una aplicación práctica basada en IA completamente automática en ocasiones simplemente *software* y en otras instalada en *hardware* tales como vehículos autónomos, drones, robots etc. Estas nuevas tecnologías son parte de la estrategia nacional de IA de Estonia y del desarrollo de lo que se ha denominado E-State motivo por el cual nos encontramos ante un claro caso de una democracia robotizada (Moreno & Jimenez, 2018).

Otros países de la Unión Europea han tomado una posición menos activa, aunque siempre en la línea de la implementación de la IA. Italia, por ejemplo, ha desarrollado una estrategia digital que se enmarca en el contexto de la *Agencia Digital Europea*. El país entiende la transformación digital de la Administración pública como una búsqueda del crecimiento económico, promoción del empleo, calidad de vida y mayor participación democrática. En 2017 se aprobó un plan que buscaba el desarrollo de «cuatro pilares»: (1) ecosistemas digitales o áreas de política (e.g. ciudadanía digital, transporte, turismo, educación.); (2) infraestructuras inmateriales (e.g. manejo de datos de la Administración pública como datos de los bancos, sistema de pagos); (3) infraestructuras físicas (e.g. centros de datos, cloud storage, etc.); y (4) ciberseguridad. Esta búsqueda de la transformación digital del país ha dado lugar a diferentes proyectos de política pública tales como: *Electronic Health Record* (un expediente o registro médico digital), *E-billing* (sistema de facturación electrónica de uso obligatorio para la administración pública), *PagoAP* (plataforma para que ciudadanos y/o empresas pueden pagar a la Administración) y el *Equipo de Respuesta a Emergencias Informáticas de la Administración Pública* (CERT-PA) (una estructura que opera dentro de la *Agencia per l'Italia digitale* y es responsable del procesamiento de los problemas de seguridad en las administraciones públicas). Estas nuevas tecnologías basadas en IA lo que buscan es mejorar la calidad de vida de los ciudadanos italianos, permitiendo un mayor grado de accesibilidad a los servicios públicos, reduciendo el coste de determinados servicios públicos, así como del gasto social, un aspecto, este último que suele relacionarse con la supuesta reasignación de recursos. Estas políticas públicas a menudo se presentan bajo el lema de que la automatización y

gestión basada en IA ofrecerá a la ciudadanía la oportunidad de interactuar con el Estado de manera más rápida y personalizada.

El Gobierno de Suecia y el Nordic Council of Ministers emitieron en 2018 el documento *AI in the Nordic-Baltic región* en el que, en conformidad con el conjunto de países de la Unión Europea, se afirma que la aplicación de la IA en el diseño e implementación de políticas públicas conducirá a la resolución de muchos problemas sociales. Una de las principales diferencias de este texto con respecto a la *Declaración de Cooperación sobre Inteligencia Artificial* estriba en resaltar que los países nórdicos y la región báltica poseen una «alta madurez digital», así como la necesidad de evitar una regulación innecesaria que impediría precisamente las bondades de su desarrollo. Suele insistirse en que esta declaración es una hoja de ruta bastante general, que no pormenoriza ningún tema; no obstante, es importante comentar en detalle algunos de sus puntos. Los países nórdicos se posicionan como líderes digitales («digital front-runners») en el contexto mundial (no únicamente en el europeo), prueba de ello es la transformación digital operada en el sector empresarial nórdico, principalmente en las PYMES. Pero es importante resaltar el hecho de que a la hora de concebir y defender la digitalización de los servicios administrativos se asumió un enfoque comprometido con mejorar la vida de los residentes y no en la reducción de costes (Randall & Belina, 2019). La infraestructura digital ha sido en este país más relevante a nivel regional que estatal con el objetivo de evitar la brecha tecnológica e impedir disparidades entre el ámbito urbano y el rural. Prueba de ello es que se desarrollara una idea de «Smart cities» muy particular, según la cual lo «Smart» de una ciudad no provenía de la tecnología en sí misma, sino de las redes de cooperación y la capacidad de resolución de conflictos que abría la implementación tecnológica. Consiguientemente, la digitalización era percibida más un proceso humano que técnico (Randall & Belina, 2019), que podría mejorar la interacción y la comunicación entre los ciudadanos y la Administración pública.

En Noruega las municipalidades son organismos semiautónomos que se encuentran en el nivel administrativo noruego más elemental. Hay 356 municipalidades cuyo tamaño varía. Van desde las más pequeñas como Utsira que tiene 212 habitantes hasta el caso de Oslo con 647 676 habitantes. Las municipalidades asumen responsabilidades tales como la gestión y mantenimiento de los sistemas educativos o de las vías públicas locales, entre otras prestaciones sociales básicas. Servicios como sensores automatizados de tráfico, tecnologías para el bienestar o intercambios digitales de información, han ido tomando una relevancia mayor en las municipalidades. En el año 2018 se implementaron tres nuevas herramientas digitales y de automatización en los municipios bajo las siguientes denominaciones: *DigiSos*, (plataforma digital para solicitar asistencia social financiera), *DigiHelse* (sistema para la visualización de acuerdos alcanzados

y establecer diálogos con la Administración) y *Minside* (servicio de información sobre temas relativos a la municipalidad). Este tipo de digitalización es lo que ellos mismos llaman «alta madurez digital» algo que, a la postre, permitirá dar a Noruega el salto a las tecnologías de IA en la Administración pública. La existencia de todas estas iniciativas nos permite analizar resultados más concretos sobre la implementación actual de la IA en el servicio público noruego. Según Mikalef, Fjørtoft & Torvatn (2019) las tecnologías que más se han implementado en las municipalidades noruegas están relacionadas con la comunicación y la interacción con los ciudadanos. Un 28,9 por ciento utiliza agentes de interacción inteligente (i.e. *chatbots* y otros tipos de *Intelligent Conversational Agents*), un 21,1 por ciento ha puesto en funcionamiento traducciones en tiempo real (i.e. traducción voz a voz o voz a texto) y un 15,7 por ciento maneja de manera automatizada las solicitudes y el manejo de datos. Si bien no es un número muy alto, la situación global muestra un desarrollo continuando en la incorporación de la IA al servicio público y, en particular, en áreas como la salud, la ciberseguridad, la automoción autónoma, la investigación financiera y, en especial, la detección de fraudes, y, sin duda alguna, la automatización de todos los procesos burocráticos si bien, por encima de cualquiera de estos ámbitos, se mantiene un constante interés por la «interacción inteligente con la ciudadanía.»

Todos estos casos parecen atestiguar el potencial de la IA en la gestión pública, ya sea desde una perspectiva meramente económica —de eficacia en la gestión— como desde la perspectiva de las políticas de bienestar. Sin embargo, no todo es tan simple como parece. Aunque el uso de tecnologías inteligentes puede actuar como facilitador de la innovación, la sostenibilidad, la competitividad y la consulta ciudadana, lo cierto es que estas dimensiones se plantean por lo general asumiendo enfoques paradójicamente indeterminados, en los que no se problematizan los conceptos y concepciones de ciudadanía y estado digitales que se asumen incluso en los casos en los que existen rutas de acción concretas. La IA parece ser más una ideología que un aspecto tecnológico. En medio de este enamoramiento obsesivo por lo tecnología —*technological infatuation*— (Gauld, & Goldfinch. 2006) es a partir de donde se plantea una cultura de renovación administrativa por medio de IA en la gestión pública desde una perspectiva excesivamente optimista, que parece reducir los problemas de gobernanza a problemas técnicos.

En el contexto descrito, observamos cómo se promueve la implementación de la IA en el sector público sin antes mencionar cuál es el modelo de Administración pública subyacente, lo cual, en el mejor de los casos, impide visualizar correctamente las implicaciones del uso de la IA en la esfera de la gobernanza. En los documentos oficiales que hemos citado solo se alude a los aspectos tecnológicos de las aplicaciones de inteligencia artificial, lo que suena bastante vago sin un

modelo de Administración pública sólido que retrate las implicaciones para la gobernanza del estado administrativo. Por ello, algunos autores se preguntan si esta nueva cultura tecno-administrativa implica realmente, acaso un modelo de organización holocrático, es decir, un sistema más horizontal basado en la autogestión (Ramió, 2019). La respuesta a ese dilema no queda clara, precisamente porque se necesitan hacer explícitos y justificar desde un punto de vista ético-jurídico no solo conceptos y concepciones de la robotización de las administraciones públicas, también es preciso hacer explícito cuáles son las aspiraciones de bienestar humano y los compromisos asumidos con la sostenibilidad y el entorno ecosocial. En ausencia de todos estos debates, que exigen un tiempo de madurez no acelerado además del diseño de espacios para la deliberación pública, las políticas emprendidas por la Unión Europea y otros países han sido calificadas por algunos expertos como propias de democracias robotizadas. (Moreno y Jiménez, 2018)

En un marco como este, las vulnerabilidades que pueden crear una falla crítica o un grave problema son temas de gran relevancia. Ya en el 2014 un *Informe independiente sobre el voto electrónico en Estonia* (<https://estoniaevoting.org/>) mostraba serias preocupaciones sobre las garantías democráticas inherentes a esta forma digital de sufragio. Según este estudio, existirían graves brechas en la seguridad operativa y de procedimiento, junto con la arquitectura del sistema que expone el sistema de votación a ataques cibernéticos que podrían alterar el signo de los votos o incluso derivar en una situación de disputa los resultados de las elecciones. Problemas que eran bien conocidos desde hace años, debido a la prohibición del voto electrónico en los Países Bajos y Alemania. En el 2008, la votación electrónica fue suspendida en los Países Bajos. Todo comenzó cuando un grupo activista llamado: «No confiamos en las computadoras de votación» (*Wij vertrouwen stemcomputers niet*) demostró que los sistemas que se utilizaban en este tipo de votación podrían, en determinadas circunstancias, poner en peligro el secreto del voto. El primer problema que señalaron fue la facilidad para reemplazar los chips de programación, lo que permitía que un posible atacante cibernético tuviera acceso a la cuenta de la máquina pudiendo ejecutar cualquier tarea que quisiera. Lo más preocupante es que este tipo de ataques podían pasar desapercibidos debido a la falta de verificabilidad. Esto, generó una alerta adicional, en vista, de que el software utilizado no era de libre acceso (ni tampoco se liberó el código cuando se le solicitó a la empresa) y, por lo tanto, no era posible auditarlo. Para este grupo era claro, que cuando un proceso no puede ser verificado adecuadamente debido a las características de la tecnología utilizada (como la IA) no se debería implementar en áreas sensibles como los sistemas críticos para la seguridad (Yeung *et al*, 2020), o en las elecciones. Otro inconveniente importante que descubrieron es que era posible espiar a la máquina de votación por medio de un complejo ataque conocido como *Tempest* (Jacobs y Pieters, 2009).

Estos aspectos, junto a los problemas de seguridad en los almacenes donde se resguardaban las máquinas de votación, la imposibilidad de realizar un auditaje de manera efectiva porque no hay registro en papel, así como la dependencia del gobierno con los proveedores privados (empresa NEDAP) de las computadoras de votación (Oostveen, 2010), terminaron con 20 años del uso del voto electrónico en los Países Bajos.

Este escenario también tuvo resonancia en Alemania. Un grupo de hackers conocido como «Chaos Computer Club» (que trabajaba en colaboración con los activistas de los Países Bajos) solicitó la prohibición de la utilización de las máquinas de votación alemanas (CCC, 2006), ya que eran de la misma empresa proveedora (NEDAP) de su vecino país y por lo tanto corrían los mismos riesgos que los neerlandeses. Con esta finalidad, presentaron una demanda colectiva en la que solicitaban la suspensión del voto electrónico hasta que el gobierno pudiera demostrar la transparencia y verificabilidad de este tipo de sistemas. En la votación electrónica no es posible determinar cómo se procesan los votos dentro la computadora, por lo que el proceso se vuelve bastante opaco, reduciendo, en el mejor escenario posible, la capacidad de escrutinio y supervisión a la *expertise* de un grupo muy pequeño de técnicos especialistas. Como respuesta a esto, en el año 2009 la Corte Constitucional Federal Alemana (*Bundesverfassungsgericht*) declaró inconstitucional la votación electrónica debido a la falta de transparencia y verificabilidad de los ciudadanos, puesto que los votantes no tenían forma de saber que pasó realmente con su voto (se ocupaba conocimiento especializado) y por el contrario se les pedía una confianza ciega en las tecnologías de votación electrónica. Estas situaciones vulneraban el principio del carácter público de una elección, ya que no era posible garantizar de manera confiable los procesos de verificación en el conteo (o recuento de los votos), por lo que en este caso no se satisfacían las exigencias constitucionales fundamentales en un Estado de Derecho.

A pesar de que existe una amplia literatura sobre todos estos elementos, los entusiastas de este tipo de tecnologías suelen minimizarlos y con frecuencia se presentan estos problemas como si fueran exclusivamente técnicos cuando lo cierto es que afectan a la dimensión procedimental y por tanto a la legitimidad de democrática de los comicios.

Pero la votación es solo un tema más dentro de muchos otros, cuando hablamos sobre los riesgos. Tomemos, por ejemplo, las identidades digitales (ID Digital) en Estonia. Esta identidad digital es según Margit Ratnik (directora de la Oficina de Identidad y Estado de la Junta de Policía y Guardia Fronteriza) la «piedra angular de nuestra sociedad digital». Según datos oficiales (Politsei.ee, 2017) esta ID Digital es la base fundamental para el acceso a 1500 e-servicios públicos (i.e registros médicos o el sistema de votación), así como a 5000 servicios privados. Por números como estos es que los entusiastas de la automatización o digitalización de

la Administración pública consideran como un caso emblemático la situación de este país báltico. Sin embargo, ser «la sociedad digital más avanzada del mundo», también la convierte en la más dependiente tecnológicamente y vulnerable a ciberataques del mundo. En el año 2017 se encontró una grave vulnerabilidad en el cifrado de las Electronic ID Cards que afectó a 760000 personas (BBC, 2017), es decir a más de la mitad de la población del país. Este riesgo potencial permitía que la información pudiera ser robada o que un ciber atacante pudiera hacerse pasar por esa persona suplantando su identidad. El riesgo afectó a los chips utilizados en tarjetas de identificación, permisos de residencia e identificaciones digitales. Frente a esto, el gobierno estonio bloqueó estas ID cards, impendiéndole a miles de ciudadanos el acceso a los servicios gubernamentales. Como medida correctiva, los ciudadanos afectados debían actualizar de forma remota (o en su defecto en las oficinas policiales) el software de sus certificados para seguir utilizando los servicios electrónicos.

Algunas personas pasaron días (BBC, 2017) sin poder realizar la actualización electrónica y otras 18000 personas todo no pudieron realizar este cambio debido a problemas técnicos (News.err, 2017). Si bien algunos días, no parecen ser mucho tiempo, ciertamente lo son cuando este tipo de identificación es «la piedra angular de la sociedad». Durante este tiempo, estas personas no tenían participación real en el *E-State*. Todos sus derechos políticos y sociales se vieron comprometidos por una vulnerabilidad en los procesos de identificación. Aún cuando los elementos básicos de la interacción estatal se mantuvieron (en el mundo analógico) esta situación puso en entredicho la esfera de los derechos de más de la mitad ciudadanos estonios, de una manera inusitada en tiempos modernos. Aspectos como los previos, deberían ser suficientes para analizar con más cautela las supuestas ventajas de este tipo de procesos. En vista de que las características tecnológicas influyen no solo el proyecto mismo de automatización los servicios, sino en el concepto mismo de ciudadanía. Como muestran los casos que hemos citado, el uso de estas tecnologías puede generar una transformación radical en los procesos democráticos o incluso en las nociones tradicionales de participación ciudadana.

2. ¿UNA ADMINISTRACIÓN “INTELIGENTE”?

Si en la literatura en ciencia política es bien conocido el proceso de judicialización de la política, nos encontraríamos ante una amenaza distinta consistente en la nueva suplantación de la política por medio de su robotización. La automatización de nuestras sociedades no es un tema de ciencia ficción, ocurre actualmente y frente a la promesa económica de la IA, es muy probable que transfigure los

espacios políticos tradicionales, con el consiguiente impacto social y cultural, que genera dilemas éticos y jurídicos que van más allá de las descripciones tecnocráticas en torno a la mera implementación de la IA. Tomemos como ejemplo la «sociedad digital más avanzada del mundo» (Estonia) que se presenta a sí misma como una utopía tecnológica (al menos en cuanto a la gestión pública). Cuando nace un bebé este se convierte instantáneamente en un e-ciudadano. En el país en el que —en palabras Marten Kaevats, asesor nacional digital del Gobierno de Estonia— lo «digital es un estilo de vida» todos los ciudadanos se han convertido en «tecnociudadanos» (Echevarría & Almendros, 2019), seres que no forman parte de una nación, sino de un *E-State*. Su identidad no depende de alguna delimitación corporal o perspectiva subjetiva, sino que existe únicamente en virtud de procesos de identificación tecnológicos (ID Digital)¹. Los derechos políticos y sociales se reducen a la existencia de un usuario y a la interacción con un tecno-entorno despersonalizado que se ha convertido en el nuevo espacio público. Ahora el ágora es «Smart» y más eficiente. Si bien las formas constitutivas de interacción estatal siguen existiendo, ahora se trasladan a espacios más «inteligentes».

1 Una alternativa (al menos teóricamente), pero no exenta de problemas, podría ser la utilización de un Gemelo Digital para relacionarnos con la Administración pública. Ya que esto permitiría mantener la identidad física junto con la digital en nuestras interacciones con la Administración, sin llegar a transformarnos en meros usuarios tecno-despersonalizados, como parece que sucede en el caso anteriormente señalado. Un gemelo digital (*Digital Twin*) es una representación dinámica y actualizada de un objeto o sistema físico. Es básicamente un modelo virtual que interactúa con el sistema físico a lo largo de su ciclo de vida. (Grieves and Vickers, 2017). Por lo general, se utiliza en el contexto empresarial para realizar predicciones sobre el futuro de un sistema que se está probando. También se utiliza para valorar la interacción de un objeto o producto real en un ecosistema determinado con tal de optimizar o validar su funcionamiento (Semeraro *et al*, 2021). Sus principales usos se encuentran en la fabricación y manufactura, la gestión del ciclo de vida de un producto y en todo lo que se ha llamado como la industria 4.0. Sin embargo, su utilización para seres humanos es prácticamente inexistente hoy en día, en vista de que somos organismos vivos muy complejos (en los cuales el ciclo de vida completo es muy difícil de analizar), a diferencia de una entidad física usualmente inanimada. A pesar de esto, se ha planteado la posibilidad teórica de un «Human Digital Twin» (Shengli 2021) en ámbitos médicos-asistenciales con tal de resolver ciertas necesidades que los servicios médicos no pueden satisfacer. Este «Gemelo Digital Humano», sería una copia online de una persona real en nuestro mundo físico. Este «Gemelo» es principalmente una base de datos digital en tiempo real de una persona en una computadora o en el ciberespacio. Por lo tanto, cuando la información de esa persona cambia (esto depende de la manera en cómo se registren y sincronicen los datos), también se modificarían los datos registrados. Además, el «Gemelo» analizaría todos los datos utilizando diversas tecnologías (i.e Deep Learning) generando recomendaciones, predicciones o incluso diagnósticos. Sin embargo, a pesar de sus posibles beneficios teóricos, en la práctica su implementación en la Administración pública (o en cualquier otra área) es, como mínimo, una tarea que implicaría una cantidad descomunal de esfuerzo y trabajo especializado (sin contar la capacidad computacional) que presentaría muchos dilemas sobre todo en cuanto al tema de manejo y uso de datos personales, así como aspectos relacionados con la privacidad. Por lo que se hace necesario mayores estudios e investigaciones empíricas con tal concluir su factibilidad, no solo teórica, sino técnico-práctica.

Por consiguiente, se ha producido una transformación tecnológica de la ciudadanía que genera nuevos desafíos sociopolíticos y ético-jurídicos a la Administración pública. Esta situación tiene un impacto sobre la esfera de los derechos humanos la cual habría de obligar a los operadores políticos en general a analizar con más detenimiento las supuestas ventajas inherentes de la digitalización y robotización de los procedimientos democráticos y administrativos pues no es lo mismo garantizar el derecho a acceder a los servicios de la Administración pública por medio de una herramienta digital que tener la obligación de hacerlo únicamente por esa vía. Además, debido a su casi total grado de digitalización ya no existe una verdadera interacción humano-Administración pública. Según Sheryl Brahmam (2015), profesora en la Universidad Estatal de Missouri entre el 10 por ciento y el 50 por ciento de nuestras interacciones con agentes conversacionales (i.e. *chatbots*) son abusivas. ¿Qué quiere decir esto? Básicamente, que nuestra interacción con estos agentes no es precisamente placentera, en cierto modo debido a que, en la comunicación humana, no se pueden eliminar las expectativas y la necesidad de obtener explicaciones complejas (que nos llevan a realizar consultas), un aspecto humano determinante que debería tener presente la Administración en el diseño de su interacción con la ciudadanía. En este sentido, la automatización de respuestas puede generar experiencias frustrantes entre los usuarios de los servicios públicos. Si ya es difícil lidiar con un ser humano cuando solicitamos algo, la frustración rápidamente puede ir en aumento cuando tratamos con agentes que aparentan serlo. Se une a todo ello, el problema del diseño de algoritmos sesgados, que reproducen visiones estereotipadas desde el punto de vista de la etnia, el género, la clase social, etc., un tema que supo poner de manifiesto en una entrevista Brahmam (2015) cuando afirmó que es mucho más común encontrar comentarios sexistas hacia *chatbots* que se presentan como «mujeres» que entre quienes asumen rasgos «masculinos». ¿Debe este fenómeno obligar a la Administración a repensar el diseño de sus *chatbots*? ¿Es la Administración responsable de los efectos que determinados diseños tecnológicos tienen sobre los niveles de agresividad exhibidos por los usuarios? ¿Tienen esos niveles un impacto posterior sobre las relaciones interpersonales? ¿Debería clasificarse y reconocerse conforme a indicadores convalidados a nivel europeo el grado de agresividad generado en la interacción humano-máquina?

De momento, estas cuestiones no parecen ser problemas relevantes en los proyectos de *smartificar* la Administración pública.

CONCLUSIONES

A pesar del potencial de la IA para el sector público y del número creciente de países que apuestan por esta vía para la modernización de la Administración, la IA para uso público es un campo de investigación en el que escasean los diagnósticos exhaustivos acerca de los desafíos ético-jurídicos asociados a la aplicación de estas sofisticadas tecnologías. En estas páginas se sostiene que este déficit prospectivo está estrechamente relacionado con la falta de análisis conceptual robusto y una comprensión política común sobre la IA. Asimismo, el marcado optimismo tecnológico y la poca claridad política e institucional con la que se desarrollan y aplican estas tecnologías, hace más difícil que tenga lugar un acercamiento reflexivo en torno a cuál debería ser el modelo de aplicación de la IA en la Administración pública de los estados europeos. Contrasta con esta situación de indefinición en torno a cuáles habrían de ser los fundamentos ético-jurídicos y sociopolíticos que habrían de guiar a los distintos estados europeos, el hecho de que la aplicación de la IA en la gestión pública parece tener el poder de marcar un rumbo a los problemas de gobernanza, así como los relacionados con la ciudadanía digital.

*Agradezco todos los comentarios y aportes para la realización de este texto, realizados por la profesora María González Navarro de la Universidad de Salamanca. Un especial agradecimiento para los revisores anónimos que con sus comentarios mejoraron la versión definitiva de este documento.

REFERENCIAS BIBLIOGRÁFICAS

- BALVÍ, Muriel, *Los 7 secretos del país más digital del mundo*. Infobae. 27/11/2017. <https://www.infobae.com/tendencias/innovacion/2017/11/25/los-7-secretos-del-pais-mas-digital-del-mundo/>. [Consulta: 20/08/2020].
- BBC. Security flaw forces Estonia ID 'lockdown'. 2017 <https://www.bbc.com/news/technology-41858583> [Consulta : 16/08/2021]
- BRAHNAM, Sheryl, "Episode #121" in *Techemergence*. Nov 1, 2015. Podcast, 0h32m27s, <https://emerj.com/ai-podcast-interviews/calling-siri-names-youre-not-alone-a-closer-look-at-misuse-of-ai-agents/>
- CCC | CHAOS COMPUTER CLUB. Chaos Computer Club fordert Vervot von Wahlcomputern in Deutschland,". 2006. <https://www.ccc.de/updates/2006/wahlcomputer> [Consulta: 01/08/2021]
- DOORSAMY, Wesley y BABU SENA PAUL, Tshilidzi Marwala, *The disruptive fourth industrial revolution : technology, society and beyond*. Cham, Switzerland: Springer, 2020.
- ECHEVARRÍA, Javier y LOLA, S. Almendros, *Tecnopersonas. Cómo las tecnologías nos transforman*. Gijón: Ediciones Trea S.L., 2020.

- EGGERS, William, VIECHNICKI, Peter y SCHATSKY, David, *AI-augmented government*. United States. Deloitte, 2017.
- FRANCESCHETTI, Donald, *Principles of robotics & artificial intelligence*. H. W. Wilson Publishing, 2018.
- GOVERNMENT OFFICE AND MINISTRY OF ECONOMIC AFFAIRS AND COMMUNICATIONS. *Report of Estonia's AI Taskforce*. Estonia, 2019.
- GRIEVES, Michael y VICKERS, John, "Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems". In: *Transdisciplinary Perspectives on Complex Systems*. Springer, 2017: 85–113.
- IDC, *Automation and Customer Experience Needs Will Drive AI Investment to \$5 Billion by 2019 Across European Industries*. IDC. 01/04/2019. <https://www.idc.com/getdoc.jsp?containerId=prEMEA44978619> [Consulta: 20/08/2020].
- JACOBS, Bart y WOLTER, Pieters, "Electronic Voting in the Netherlands: From Early Adoption to Early Abolishment". En: Alessandro Aldini, Gilles Barthe y Roberto Gorrieri (eds.), *Foundations of Security Analysis and Design V*, t. 5705. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science). 2009: 21–144.
- LEE, Kai-Fu, *Superpotencias de la inteligencia artificial*. Barcelona: Editorial Planeta, 2020.
- MCCARTHY, John, "What Is AI? / Basic Questions". 2018. <http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html>. [Consulta: 01/08/2021]
- MEHR, Hila, *Artificial Intelligence for Citizen Services and Government*. United States: Harvard Ash Center Technology & Democracy Fellow, 2016.
- MIKALEF, Patrick, OLSEN FJØRTOFT, Siw y Y. TORVATN, Hans, "Artificial Intelligence in the Public Sector: A Study of Challenges and Opportunities for Norwegian Municipalities". In *Digital Transformation for a Sustainable Society in the 21st Century*. Lecture Notes in Computer Science, edited By Pappas, I. O., Mikalef, P., Dwivedi, Y. K., Jaccheri, L., Krogstie, J., & Mäntymäki, M. Norway: Springer, 2019: 267-277.
- MORENO, Luis y JIMÉNEZ, Raúl, *Democracias Robotizadas. Escenarios futuros en Estados Unidos y la Unión Europea*. Madrid: Los libros de la Catarata, 2018.
- NEWS. EE., 35,000 people have renewed ID card certificates. 2017. <https://news.err.ee/640156/35-000-people-have-renewed-id-card-certificates> [Consulta : 16/08/2021]
- NORDIC COUNCIL OF MINISTERS, AI in the Nordic-Baltic region. Sweden. 2018
- OOSTVEEN, Anne-Marie, "Outsourcing Democracy: Losing Control of e-Voting in the Netherlands." En: *Policy & Internet* 2 (4), 2010: 196–215. DOI: 10.2202/1944-2866.1065.
- POLITSEI. EE., *Estonia resolves its ID-card crisis*. 2017. <https://www2.politsei.ee/en/uudised/uudis.dot?id=801245> [Consulta: 16/08/2021]
- RAMIÓ, Carles, *Inteligencia Artificial y Administración Pública. Robots y Humanos compartiendo el servicio público*. Madrid: Los libros de la Catarata, 2019.
- RANDALL, Linda y BERLINA, Anna, *Governing The Digital Transition in Nordic Regions: The Human Element*. Sweden: Nordregio Report, 2019.

- SEMERARO, Concetta, LEZOCHÉ Mario, PANETTO, Hervé y DASSISTI, Michele, “Digital twin paradigm : A systematic literature review”. *Computers in Industry*, 130, 103469. 2021. <https://doi.org/10.1016/J.COMPIND.2021.103469>
- SHENGLI, Wei, “Is Human Digital Twin possible ?” *Computer Methods and Programs in Biomedicine Update*, 1, 100014. 2021. <https://doi.org/10.1016/J.CMPBUP.2021.100014>.
- YEUNG, Karen, HOWES, Andrew y POGREBNA, Ganna, “AI Governance by Human Rights–Centered Design, Deliberation, and Oversight: An End to Ethics Washing”. En Dubber, M., Pasquale, F. & Das, S., *The Oxford handbook of ethics of AI*. New York, NY: Oxford University Press, 2020.



***Venditio fumi*: Autorregulación Empresarial e Inteligencia Artificial.**

Venditio fumi: Business Self-regulation and Artificial Intelligence.

DOI: 10.32870/sincronia.axxvi.n81.12a22

Jonathan Piedra Alegría

Universidad Nacional de Costa Rica. (COSTA RICA)

CE: jonathan.piedra.alegría@una.cr / ID ORCID: 0000-0003-4532-4415

Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/)

Recibido: 24/09/2021

Revisado: 05/10/2021

Aprobado: 03/11/2021

RESUMEN

En la primera parte de este artículo se analizará la utilización reduccionista e instrumental de la ética en temas relacionados con la regulación privada de la Inteligencia Artificial. Principalmente en lo relacionado con la autorregulación ética empresarial. Se parte de la hipótesis de acuerdo con cual, las propuestas éticas utilizadas por las grandes corporaciones de la Inteligencia Artificial son parte de una estrategia que busca solamente crear una apariencia de responsabilidad empresarial y social, que tiene como propósito, confundir a los legisladores y a los encargados de políticas nacionales, así como a los consumidores finales, con tal de mantener sus prácticas empresariales y ganancias intactas. Se proponen como ejemplos paradigmáticos los casos de Google y Amazon con tal de mostrar esta situación, para concluir como los desafíos éticos y legales de la implementación de la IA, son reinterpretados bajo una "gestión de riesgo éticos", en la cual se trivializa y se vacía de contenido el discurso ético.

Palabras clave: Inteligencia artificial. Autorregulación. Ética. Riesgo ético.

ABSTRACT

In the first part of this article, we will analyze the reductionist and instrumental use of ethics in issues related to the private regulation of Artificial Intelligence. Mainly to ethical business



self-regulation. It is based on the hypothesis according to which, the ethical proposals used by the large Artificial Intelligence corporations are part of a strategy that seeks only to create an appearance of business and social responsibility, whose purpose is to confuse legislators, as well as well as end consumers, to keep their business practices and profits intact. We will use the cases of Google and Amazon as paradigmatic examples to show this situation. This part concludes how companies reinterpret the ethical and legal challenges of the implementation of AI, towards an "ethical risk management" that trivializes and empties the ethical discourse of content.

Keywords: Artificial Intelligence. self-regulation. Ethics. Ethical risk.

Introducción

El avance y desarrollo en la Inteligencia Artificial (IA) transformará nuestra vida de manera incuestionable. Asistentes virtuales *-reconocimiento de lenguaje natural-*, chatbots *-procesamiento de lenguaje escrito-* o sistemas que nos recomiendan una película *-algoritmos de recomendación-* ya se han vuelto parte de nuestras vidas. Cambios profundos están ocurriendo en áreas como la salud y la sanidad, la educación o en la movilidad ciudadana. Su impacto potencial es tan grande que algunos académicos consideran que prácticamente todos los temas de la vida cotidiana podrían ser mejorados por la IA (Morte & Monasterio, 2017), llegando inclusive a resolver (en algún momento) los problemas mundiales más apremiantes. Esto hace creer a ciertos especialistas que la IA es la tecnología más floreciente de esta década, denominando así el tiempo actual como la "Era de la IA" (Mondal 2020). Esta capacidad transformativa, así como el ritmo creciente con que se está implementando en las actividades humanas, no dejan duda respecto a que los avances en IA tendrán consecuencias sociales de gran alcance. Los posibles beneficios económicos de su implementación han hecho que tanto el sector privado como el público se interesen cada vez más en su investigación y puesta en funcionamiento.

Un estudio de *Price Waterhouse Coopers* citado por el BID (2020) señala que para el año 2030 la adopción de la IA podría hacer crecer el PIB mundial cerca de un 14 por ciento (equivalente a 16,5 trillones de dólares americanos), debido, fundamentalmente, al aumento de la productividad



(i.e robots industriales, vehículos autónomos o inteligencia asistida). Por mencionar un ejemplo, algunos datos sobre Europa muestran que "Del 2010 al 2014 las ventas de robots aumentaron una media del 17 por ciento por año y es la industria electrónica la que está generando el mayor número de ofertas de empleo" (*Normas de Derecho Civil sobre Robótica*).

Dado este contexto, así como la amplia gama de usos potenciales de la IA (en particular el aprendizaje automático *-machine learning-*, la computación cognitiva o uso de los macrodatos *-Big data*) la conciencia sobre una serie de problemas sobre su uso ha aumentado. Temas como la privacidad, el uso ético de los datos, la responsabilidad, autonomía, la sostenibilidad ambiental o la transparencia ahora no solo se discuten en espacios especializados sino también en los principales medios de comunicación. Un sentimiento de desconfianza y preocupación nace no solo de los riesgos inherentes de este tipo de la tecnología, sino además de la incuestionable acumulación de poder (económico, tecnológico y social) que las empresas líderes en IA están logrando.

Esto plantea dudas sobre el control y su posible regulación. Cuestiones como: ¿Cuál sería el punto de partida para la regulación? O ¿Qué se debería tener en cuenta para esto? Son preguntas que las empresas, así como los Estados y comunidades supranacionales están considerando actualmente. Aspectos éticos y políticos relacionados con la aplicación de ciertas tecnologías de IA son comunes los debates sobre este tema.

A saber, el tema de la autonomía (ie. responsabilidad imputable a los sistemas autónomos y semiautónomos, así como su inclusión en ámbitos médico-asistenciales o su interacción con los seres humanos *-robots sociales-*) es uno de los focos de discusión en los documentos que buscan su regulación. Otros temas como el reemplazo del trabajo humano por robots con IA, el manejo de los datos o la creación de algoritmos sesgados o discriminatorios tampoco son dejados de lado, siendo estos los principales riesgos y temores que genera la expansión y utilización de la IA. El Parlamento Europeo expresó que:

Considerando que, ahora que la humanidad se encuentra a las puertas de una era en la que robots, bots, androides y otras formas de inteligencia artificial cada vez más sofisticadas parecen dispuestas a desencadenar una nueva revolución industrial —que probablemente



afecte a todos los estratos de la sociedad—, resulta de vital importancia que el legislador pondere las consecuencias jurídicas y éticas, sin obstaculizar con ello la innovación. (EU Commission, 2017).

Como cabría esperar, estos aspectos plantean cuestiones importantes. Si tenemos presente que los sistemas y las tecnologías de IA son desarrolladas fundamentalmente por empresas privadas (Kai-Fu Lee, 2020), y que, además, no necesariamente basan su operación en algún sustento ético; ¿Cómo definir el *approach* más idóneo para que se "no obstaculice la innovación"? o ¿Cuál sería el acercamiento más adecuado para ponderar estas consecuencias?

¿Regular o no regular?

Las respuestas a estas preguntas se pueden ubicar, grosso modo, en dos grandes grupos: Podemos encontrar en primer lugar (I) a los "progresistas económicos" que consideran que la IA nos puede llevar a una bonanza económica y social, en el cual esta tecnología se convertiría en una guía para la riqueza y el trabajo (Kaplan, 2016). Así pues, lo más adecuado sería fomentar la investigación, el desarrollo e implementación de la IA, manteniendo las regulaciones estatales al mínimo necesario. Este vendría a ser, concretamente, el enfoque estadounidense sobre la IA. Este se centra en la innovación y en el desarrollo de tecnologías (que no podrían darse si hay regulación) más que en la defensa de las personas o la protección de los Derechos Humanos.

Esta perspectiva no implica el establecimiento de una regulación particular, ni un posicionamiento claro con respecto al tipo de marco normativo necesario y menos aún con postura ética deseada, aunque si destaca la necesidad de ciertos mínimos que deben ser respetados por las empresas por la vía de la autorregulación.

Luego encontramos aquellos que plantean que debería tenerse cautela con la puesta en marcha acríticamente de esta tecnología. En el mismo Estados Unidos, el *Berkman Klein del Center for Internet & Society*, considera que debe existir una deliberación exhaustiva y a largo plazo de los impactos de la utilización de tecnologías relacionadas con la IA, y no únicamente una visión



cortoplacista como la que comúnmente se presentan en los discursos sobre este tipo de tecnologías. Usualmente estas narrativas se dirigen hacia (II) la creación de un marco legal que permita regular de algún modo las complejas interacciones de una IA con su entorno.

Esta posición contempla la necesidad de su regulación desde el punto de vista de las políticas públicas, así como desde la creación de normas, leyes y reglamentos claros que definan el espacio de acción de las tecnologías de IA. Una crítica importante a esta perspectiva proviene de las grandes compañías privadas que dominan el campo de la IA. Muchas de ellas consideran que la regulación es inadecuada (carece del "*on-the-ground approach*"), limitando la innovación debido a una falta de conocimiento de la realidad tecnológica y de la dinámica del mercado actual.

Un punto medio.

Precisamente esto ha provocado que en muchos casos se opte por un punto de vista intermedio: Las *Guías éticas* sobre la IA. Según algunos estudios, solo en el año 2019 podían encontrarse 89 documentos de este tipo (Jobin et al, 2019). Llegando casi duplicarse (más de 160 *Guías éticas* sobre IA) su número en el año 2020 (Algorithm Watch, 2020). Estas *Guías* son resoluciones no-vinculantes de organismos internacionales o dictámenes emitidos por grupos interdisciplinarios de expertos (tanto públicos como privados). Todas ellas tienen la intención de servir como pautas que permitan crear lineamientos básicos sobre el tema, así como situar algunos de los dilemas de la IA en un marco de análisis crítico, al mismo tiempo que presentan algunas alternativas éticas frente a estos desafíos.

No obstante, tampoco escapan a las críticas. Los inconvenientes más importantes de estas *Guías* son varios: Para comenzar, son bastante imprecisas en cuanto a los detalles técnicos y conceptuales. También es normal que carezcan de mecanismos de implementación o supervisión y, por el contrario, presenten una gran cantidad de descripciones ambiguas, confusas o claramente problemáticas. Por lo general se basan en una lista muy limitada de principios, que comúnmente se pueden reducir a cinco de ellos: transparencia, justicia y equidad, un uso no-dañino, responsabilidad e integridad (Jobin et al, 2019). Principios que pocas veces son operativizados y que usualmente



carecen de elementos para su aplicación real. Tampoco es extraño encontrar referencias vagas a lugares comunes como el respeto al derecho civil o a las normas del ordenamiento jurídico correspondiente.

Estos documentos presentan pautas no-obligatorias que pueden servir como un paso inicial para la posterior regulación. Desde esta perspectiva, es común encontrar propuestas que buscan la *autorregulación ética-normativa* de las empresas (i.e Google, Amazon, Microsoft) frente a los retos sociales y políticos de la IA. En términos generales es complicado discernir claramente como este tipo de propuestas aportan a la discusión sobre la regulación, en vista de que no existe ninguna pauta obligatoria, postura ética específica o un sistema de valores que pueda orientar el debate¹. Por lo que son necesarios más estudios sobre este tema con tal de dimensionarlas adecuadamente.

Aun así, los gobiernos de diferentes países, organismos supranacionales y las mismas empresas están impulsando propuestas en este sentido. Tomemos como caso de referencia la propuesta de la OCDE a este respecto. Este organismo creó una lista de cinco principios (y cinco recomendaciones)² basados en valores éticos para una administración responsable de la IA (OECD, 2019). Lo que ahí se presenta, *grosso modo*, son premisas poco desarrolladas en cuanto a los sistemas de IA. Repitiendo *locus communes* tales como el respeto a las leyes nacionales, los DDHH o a la diversidad humana. Aspectos que desde luego son de gran importancia y sin duda alguna, deberían tenerse en cuenta, pero no tienen ni un solo elemento de política pública que intente abordar realmente los desafíos ético-jurídicos que presentan estas tecnologías.

Lo mismo sucede con el tema de los DDHH. A pesar de que se menciona -con mayor o menor énfasis- el respeto a estos derechos (bajo la premisa del cumplimiento de los instrumentos internacionales), en casi todos los documentos, lo cierto es que siempre aparecen sin ninguna referencia directa a alguno de ellos -más allá del respecto a la dignidad o la no-discriminación- y en las raras ocasiones que sí lo hacen siempre están totalmente carentes de contenido normativo. Por

¹ Esto no quiere decir que la Guías éticas no sean importantes. Sin duda alguna son un paso inicial para una futura regulación. Nuestra crítica se dirige hacia ciertos elementos de estos documentos, como la autorregulación empresarial, así como un acercamiento *light* en ciertos temas de la IA.

² Hoy 42 países han adoptado estos principios.



el contrario, se omiten groseramente los derechos sociales, laborales o los relacionados con el ambiente. Manteniendo incuestionable la estructura de poder o las desigualdades estructurales propias del sistema económico en el cual lucran. Es decir, este tipo de propuestas crean un ambiguo espacio de interpretación ético-normativo que podría ser utilizado por aquellos que quieren utilizar de las *Guías éticas* como una forma de evadir la regulación o simplemente como un *ethical washing* (Floridi, 2019).

La narrativa ético-regulatoria en la Inteligencia Artificial

Para nadie es un secreto que los principales avances en el ámbito de la IA son producto de las empresas privadas. Más específicamente, son dos grandes grupos de empresas quienes manejan el desarrollo y la innovación de la IA: El grupo conocido como GAFAM: Alphabet, Amazon, Facebook, Apple y Microsoft, (Prisma European Network, 2021) así como el conjunto de empresas chinas incluidas bajo el acrónimo de BATX: Baidu, Alibaba, Tencent y Xiaomi (Prisma European Network, 2021) controlan el mercado de la IA. Esta concentración del poder económico y tecnológico no solo crea una mayor desigualdad social (Kai-Fu Lee, 2020), sino que además hace más notorias las transgresiones a la ley, a los derechos fundamentales -o incluso en temas de DDHH³- por parte de estas empresas.

Como respuesta a estos problemas, muchas de estas empresas han creado "Comités Éticos" o "Guías éticas de comportamientos"⁴ sobre la IA con la finalidad de autorregularse en su prácticas empresariales y tecnológicas. Sin embargo, esta práctica difiere mucho de lo que se esperaría.

AI for Good?

³ El caso DeepMind de Google en Reino Unido, Cambridge Analytica y Facebook o el software de reconocimiento facial discriminatorio en Amazon o Microsoft, son una muestra directamente relacionada con la IA. Pero no se reducen a esto. Por ejemplo, la manera en Amazon abusa y explota laboralmente a sus empleados (Adler & Schneider, 2020) o las prácticas anti-sindicales de Google (Scheiber & Wakabayashi, 2019) son bien conocidas

⁴ Vid. "Perspectives on Issues in AI Governance" de Google (2019), "Operationalizing responsible AI" de Microsoft o la propuesta "Partnership on AI" (PAI) (2018).



En los primeros meses del año 2019, Google anunció de una forma llamativa la creación de un *Comité Asesor a nivel global sobre temas de IA* (Advanced Technology External Advisory Council). Con esto, la empresa californiana se unía a un grupo selecto de empresas "comprometidas" (como Amazon o Microsoft) con el "bienestar social" que también habían creado grupos asesores para fines parecidos. Este comité estaba formado por especialistas en tecnología, ética digital y personas con experiencia en políticas públicas. El objetivo principal de este grupo era proporcionar recomendaciones de carácter ético (en general) para el uso de IA en Google, así como para los investigadores que trabajan en áreas como el software de reconocimiento facial, manejo de datos o temas relacionados con la privacidad. Sin embargo, una semana después de su conformación el *Comité* fue disuelto, debido a los insistentes reclamos de sus miembros para fuera excluida una de sus integrantes, en vista de que era bien conocida su postura en contra de la población LGBTI y anti-migrantes (Joshi, 2019).

Este comité surgió, en alguna medida, como una respuesta al *Proyecto Maven -2018-*, en el cual Google le vendió al Ejército de los Estados Unidos un software de IA para los drones no tripulados. Cuando los empleados de Google se enteraron, muchos de ellos se pronunciaron en contra del proyecto, llegando incluso a renunciar a la empresa. Esta oposición finalmente llevó a Google a no renovar con el Ejército este contrato⁵.

A raíz de esto, el CEO de Google emitió en ese año -2018- una especie de lista de principios éticos⁶ en los que se mencionaba que la empresa no diseñaría o desplegaría IA para tecnologías que «puedan causar un daño general», para armas, para vigilancia o para tecnologías "whose purpose

⁵ No parece que esto fuera del todo cierto. En vista de un enigmático Tweet (cerca de un año después) de Donald Trump (presidente de Estados Unidos en ese momento) en el cual mencionaba, como luego de una reunión con Pinchai, tenía claro que Google estaba comprometido con el ejército de USA. "Just met with @SundarPichai, President of @Google, who is obviously doing quite well. He stated strongly that he is totally committed to the U.S. Military, not the Chinese Military..." — Donald J. Trump (@realDonaldTrump) March 27, 2019

⁶ Microsoft también emitió una serie de principios ("Operationalizing responsible AI") aún más generales que Google (<https://www.microsoft.com/en-us/ai/our-approach?activetab=pivot1:primaryr5>) en ese mismo año. De igual manera en el 2021, publicó una entrada en su blog oficial en la que se hace referencia a "The building blocks of Microsoft's responsible AI program" pero si precisar ningún aspecto en particular. (<https://blogs.microsoft.com/on-the-issues/2021/01/19/microsoft-responsible-ai-program/>).



contravenes widely accepted principles of international law and human rights". (Pichai, 2018). Lugares comunes, que como ya hemos indicado, se repiten a lo largo de la mayoría de las *Guías Éticas sobre la IA*. A pesar de esta publicación, Google, en ese mismo año, reconoció haber probado una versión de su motor de búsqueda diseñado únicamente para cumplir con la censura que China aplica a sus habitantes (Simonite, 2020) en abierta oposición con todos con los principios que ellos indicaron.

Frente a esto, debemos decir que este tipo de prácticas no son necesariamente contradictorias. Esto se debe, ante todo, a que los *Principios* de Google son un texto bastante superficial que no aborda ningún problema concreto. En este documento no se menciona ni un solo mecanismo o alguna forma para rendirle cuentas al Estado, inversionistas o público-usuario. Solo aparecen unos cuantos principios genéricos - privacidad, la responsabilidad, un uso no-dañino o la seguridad, entre otros- que no llegan a ningún lado. Lo mismo sucede con *Perspectives on Issues in AI Governance* (Google, 2019) un texto que se publicó en enero del año 2019 cuya finalidad consiste en informar a la opinión pública y a los gobiernos -al estilo de un *White Paper*- sobre cuál sería la «ruta» idónea respecto a la IA. Es necesario decir, que estas *Perspectivas* están planteadas de una manera más formal que los *Principios* del 2018. Lo que hace que su propuesta sea considerablemente más seria y amplia que las pautas previas, aunque ni así su finalidad.

Su contenido principal se resume al desarrollo de cinco áreas en las que los gobiernos (no necesariamente las empresas) pueden (no deben) trabajar en conjunto con la sociedad civil y los profesionales de la IA para proporcionar una guía importante sobre el desarrollo y uso responsable de la IA. Si bien el documento señala algunas consideraciones ético-políticas, no podría decirse que haya un desarrollo de ningún problema concreto. Por el contrario, una parte importante de este *White paper* es un proto panfleto que busca mostrar la importancia de la IA para nuestras vidas, así como lo inoportuno de una regulación estricta. El texto se toma la molestia, incluso, de utilizar toda una sección para señalar lo inadecuado de un enfoque regulatorio como el de la UE⁷, ya que una regulación como esa, provocaría una grave "inseguridad jurídica" a las empresas, creando una

⁷ Por ejemplo, The European Commission Staff Working Document on liability for emerging digital technologies.



«responsabilidad fuera de lugar», "sobrecargando" a los fabricantes de sistemas de IA, lo que crearía sin lugar a duda "un efecto paralizador en la innovación y la competencia".

Claramente este tipo de propuestas éticas lo único que buscan es evitar la verdadera regulación, que se muestra como inoportuna, innecesaria, ignorante de las dinámicas tecnológicas o simplemente equivocada. Para evitar esto, la solución más adecuada es la autorregulación. Prácticas de autoevaluación o enfoques regulatorios propios son la clave para dinamizar el sector y beneficiar a todos. "A self-regulatory or co-regulatory set of international governance norms that could be applied flexibly and adaptively would enable policy safeguards while preserving the space for continued beneficial innovation" (Google, 2019). No obstante, todo indica que el propósito final de estas posiciones es evadir un compromiso real con la regulación de las actividades comerciales de estas empresas⁸.

Un tema insignificante

Una situación similar ocurre con Amazon. El del 22 de mayo del año 2018, diferentes medios de comunicación reportaron que *Amazon Web Services*, estaba vendiendo su tecnología de reconocimiento facial, llamada *Rekognition* a una serie de departamentos policiales⁹ en Estados Unidos (Oregon y Florida¹⁰).

Rekognition puede rastrear a personas en tiempo real en la transmisión de un video de lugares públicos o de multitudes. Además, es capaz de buscar en bases de datos con millones de imágenes e identificar 100 rostros (personas) en una sola imagen (Ng, 2018). Amazon promovía el uso de esta tecnología para utilizarla en las cámaras corporales que llevan los policías (*American Civil Liberties Union Foundations of California*, 2018), lo cual transformaría totalmente el concepto

⁸ Lo mismo podría decirse de iniciativas como "Partnership on AI" (2018) que agrupa a empresas líderes en la IA como Amazon, Apple, Baidu, Facebook, Google, IBM o Intel, en la cual se busca resaltar más la afiliación a esta asociación que brindar alguna solución concreta o una propuesta de regulación.

⁹ No obstante, se viene utilizando por distintas agencias de Gobierno de los Estados Unidos desde el año 2016. Por ejemplo, ya era utilizada por la U.S. Immigration and Customs Enforcement (ICE).

¹⁰ A raíz de esto muchos otros Estados también estaban interesados. Como fue el caso de California y Arizona.



de intrusión a la vida privada, vigilancia policial y responsabilidad del Estado¹¹. Por si esto fuera poco, este software tiene un sesgo discriminatorio contra las personas no-caucásicas (i.e población afrodescendiente o latina).

En una prueba realizada por la *American Civil Liberties Union* (UCLU) utilizando el servicio de Amazon para el reconocimiento facial, esta organización descubrió que el software coincidía incorrectamente los rostros de 28 miembros del Congreso de los Estados Unidos (congresistas con diferentes tonalidades de piel morena) con las fotos de una serie de delincuentes arrestados (Armasu, 2018).

En el año siguiente, un estudio del *Massachusetts Institute of Technology* (Singer, 2018) mostró que este software identificó de manera errónea a las mujeres de piel oscura (darker-skinned women) como hombres el 31 por ciento de las veces, pero no cometió ningún error con los hombres de piel clara (light-skinned men)¹². Frente a esto, Amazon comentó que los miedos derivados de estos resultados eran "an insignificant public policy issue." (Pasternack, 2019), restándole importancia al tema.

Esta situación no es una cuestión pública banal. Su uso abiertamente discriminatorio hacia ciertos grupos poblacionales, junto a su implementación en las fuerzas policiales, generaban un reclamo legítimo frente a las posibles violaciones de los DDHH derivadas de su utilización. Un miedo que no es nada insignificante. La respuesta de Amazon fue lo que podríamos esperar: Discursos vacíos que evadían cualquier tipo de responsabilidad. La empresa se empeñó en criticar en primer lugar a la UCLU por utilizar una "configuración incorrecta" o diferente a la recomendada de *Rekognition* (Singer, 2019), Incluso, criticaron a los Gobiernos por el mal uso de su software, ya que Amazon posee una «Política de uso aceptable» que prohíbe el uso de sus servicios para actividades ilegales, que violen los derechos de los demás o que puedan ser dañinos en general.

¹¹ Un control y vigilancia de esta clase solo lo hemos visto en la serie de televisión británica *Black Mirror* o en las peores distopías literarias cyberpunk.

¹² Este tipo de sesgos raciales y por género, ya habrían sido previamente mostrados por las investigadoras del MIT, Joy Buolamwini y Timnit Gebru en el caso de los softwares de IBM y Microsoft. Así como la plataforma Face++ de la compañía china Megvii (Vid. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*)



Estas declaraciones nunca hicieron referencia al algoritmo en sí mismo, ni a los ejemplos discriminatorios utilizados por *Rekognition* para entrenarse, a los sesgos por género o raciales encontrados en los estudios, ni muchos menos a la responsabilidad social de Amazon cuando vende su tecnología para fines represivos o bélicos. La culpa es de quién usa la tecnología, no de ellos. Con esta alusión (ya clásica) a la neutralidad tecnológica no solo evaden su responsabilidad social, sino que además le atribuyen totalmente a los Estados los posibles malos usos, así los problemas éticos y legales correspondientes.

Para sustentar esta tesis, la empresa a través de su Gerente General de IA indicó en una publicación oficial -junio del 2018- del Blog de Amazon que no habían recibido ni un solo reporte sobre algún tipo de abuso policial con la utilización de *Rekognition*. A pesar de esto, también en una entrada del mismo Blog – febrero del 2019- el Vicepresidente de políticas públicas globales de *Amazon Web Services* escribió que "We support the calls for an appropriate national legislative framework that protects individual civil rights and ensures that governments are transparent in their use of facial recognition technology." y para esto señaló -de nuevo- una especie de mini-guía ética a manera de manual sobre la IA en general (Punke, 2019), centrada en los softwares de reconocimiento facial.

En esa entrada se indicaron un conjunto de principios como los siguientes: En el caso del reconocimiento facial siempre debía existir revisión humana. Utilizar softwares con un umbral de 99 por ciento de confianza cuando se use para fines que pueden lesionar los DDHH (¿Contradicción con la *Política de uso aceptable?*) y en los siguientes puntos más de lo mismo: Respeto a la ley. Transparencia por parte de los Gobiernos y avisos para que los ciudadanos sepan que estas tecnologías se están utilizando en espacios públicos o comerciales. Orientaciones que, aun siendo válidas en general, son intencionadamente superficiales, vagas y sin ninguna referencia a la responsabilidad empresarial, al desarrollo de tecnologías intrusivas, las lesiones presentes a los derechos (cometidas actualmente por Amazon), mecanismos de corrección, políticas concretas o formas específicas para impedir su uso «ilegítimo».



Inyectar ética.

Como podemos ver esto es parte de una instrumentalización que distorsiona el discurso ético para sus propios fines. Tomemos como ejemplo el caso de Reid Blackman. Un filósofo estadounidense, fundador y director ejecutivo de *Virtue*, una consultora de riesgos éticos (gestión del riesgo ético) que trabaja con empresas líderes en IA¹³ para integrar la ética y la mitigación de riesgos éticos en la cultura empresarial, así como el desarrollo, despliegue y adquisición de productos de tecnología emergente.

Un riesgo ético es una consecuencia negativa inesperada, resultado de acciones no éticas. En este caso, estos riesgos se asocian a problemas reputacionales, disminución de capital de inversión, detrimento de credibilidad pública y obviamente, pérdida de clientes. Es por esto por lo que en la página web oficial destinada a promocionarse, presenta el slogan: "How do you inject ethics into culture and product?"

¿Qué quiere decir exactamente inyectar ética en un producto? No queda claro. Lo cierto es que según Blackman, los consumidores quieren comprar productos o servicios en "empresas éticas", por lo que las consecuencias de no seguir un "enfoque ético empresarial" son graves.

These companies are investing in answers to once esoteric ethical questions because they've realized one simple truth: [...] Missing the mark can expose companies to reputational, regulatory, and legal risks [...] leads to wasted resources, inefficiencies in product development and deployment, and even an inability to use data to train AI models at all. (Blackman, 2020).

Posiblemente estas "preguntas éticas esotéricas" a las cuales se refiere Blackman, son preguntas básicas sobre la responsabilidad empresarial, justicia social o respeto a los derechos humanos. Preguntas que, a pesar de lo que el estadounidense pueda decir, aún ahora son ignoradas, en vista de que una propuesta como esta (*inject ethics into culture and product*), lo único que busca es

¹³ Una de las empresas con las que colaboró Blackman fue Amazon, tal y como lo indica en su página web. (<https://reidblackman.com>)



establecer una apariencia de corrección y responsabilidad empresarial y social, que tiene el propósito de confundir a los legisladores y a los encargados de políticas nacionales, así como a los consumidores finales, con tal de mantener las prácticas empresariales intactas, en las cuales se mantiene incuestionado el sistema económico subyacente, las malas prácticas empresariales con sus empleados o las transgresiones flagrantes a los DDHH. Todo esto cubierto con un hábito de conciencia y preocupación social.

Blackman señala que si bien, introducir declaraciones o planteamientos éticos no es necesario (un argumento bastante dudoso) para la operación de las empresas, debería hacerse ya que son "herramientas útiles para lograr sus objetivos" (Tardif, 2021). Es decir, desde esta lógica utilitarista, la única intención de las propuestas éticas es la obtención de fines comerciales (basados en el lucro y la optimización de recursos). Por lo tanto, evadir la regulación o evitar una disminución en su reputación, no es en sí mismo algo éticamente indebido.

Es así como esta "inyección de ética" viene acompañada de conferencias, conformación de grupos de expertos o con la creación de centros de investigación sobre Ética de la IA -como el caso de Facebook en Munich (Quiñonero, 2019)- que nunca llegan a ningún punto. Tampoco es extraño encontrar reflexiones sobre el futuro de la IA, que lo único que hacen es evadir las consideraciones sobre los temas urgentes en el presente, manteniendo el *status quo* empresarial intacto.

Estas propuestas simplemente cortinas de humo que sirven para exponer a los legisladores que el autogobierno empresarial y los mecanismos autorregulatorios en la industria son suficientes, por lo que no se necesita ninguna regulación específica para reducir los posibles riesgos tecnológicos, eliminar los escenarios de abuso o impedir las constantes violaciones a los DDHH (en temas como privacidad, discriminación, libertad de expresión etc.). A pesar de en algún caso concreto, estas empresas puedan solicitar leyes más concretas sobre los sistemas de IA, estas peticiones son vagas y superficiales.

Esto nos plantea un escenario muy distinto al que nos podríamos imaginar cuando una empresa quiere resolver los problemas éticos que se le presentan. En este contexto el debate sobre una IA ética va en paralelo con una resistencia a toda regulación. Cualquier idea de (auto)



regulación empresarial de las tecnologías tiende a dejar de lado un papel activo de Estado y, en cambio, busca aumentar la influencia del sector privado, llegando incluso a ver a las leyes como «un simple obstáculo» para la innovación tecnológica. Según Calo "Ethics guidelines of the AI industry serve to suggest to legislators that internal self-governance in science and industry is sufficient, and that no specific laws are necessary to mitigate possible technological risks and to eliminate scenarios of abuse" (Hagendorff, 2020). La finalidad última no es aplicar principios éticos para crear una tecnología menos intrusiva, más accesible o para erradicar los sesgos discriminatorios de sus algoritmos, sino para la mitigación de riesgos éticos que puedan traducirse en perjuicios empresariales.

Desde luego, esto no quiere decir que la ética no sea importante en las empresas, sino solamente que existe una tendencia que utiliza el discurso ético de manera floja, simplona o que lo instrumentaliza como una gestión de riesgo que solo busca seguir generando beneficios económicos. Si, además, a esto le agregamos la imposibilidad de crear una regulación homogénea para el uso de las tecnologías (mucho menos aún para la IA, debido a sus aspectos cambiantes y disruptivos) se crea un terreno fértil para que las corporaciones y empresas utilicen a su antojo la ética, seleccionando aquellos principios que más les conviene (*ethical shopping*), vaciándola de contenido o trivializándola (*ethical washing*). Generando, de manera creciente, la apariencia de que la toma de decisiones políticas y legales sobre estos temas están descontextualizadas o carentes de asidero tecnológico.

Conclusiones

En estas páginas se sostiene que existe un uso instrumental de la ética (al menos de las grandes corporaciones) que va más allá del *ethical washing*. Una gestión (ética) del riesgo que solo busca de mantener los beneficios económicos de las empresas. El discurso ético se vuelve una fachada para evitar la verdadera regulación. Pero incluso va más allá, la autorregulación ética, en concreto lo que busca es la desregulación normativa, promoviendo como la única solución la gobernanza impulsada por el mercado, y no por los Estados. Es una instrumentalización del lenguaje ético que genera una



aparición de una empresa social y éticamente responsable. Simplificando el verdadero trabajo ético, creando comités solo para la opinión pública o contratando a filósofos especialistas en ética, sin ningún poder real en modificar las políticas de empresa o sin siquiera tener una participación en el diseño de propuestas. El uso autorregulatorio de la ética es solo parte de una estrategia comunicativa que busca crear y consolidar una imagen de responsabilidad en los clientes y en la sociedad. En el mejor de los casos la ética se convierte en un *risk managing* en el que no importa la justicia social, ni la protección o resguardo de los seres humanos. Pareciera que este tipo de empresas no solo buscan el control sobre los consumidores, sino además sobre los políticos. Como resultado de esto, muchas de estas iniciativas, en particular las patrocinadas por la industria, no pasan de ser (en el mejor de los casos) señales de buenas intenciones y de conductas adecuadas con el único propósito de retrasar la regulación, centrar el debate en problemas abstractos (lo cual retrasa más aún la discusión) y evitar cualesquiera soluciones técnicas.

Referencias

- Adler, D. & Schneider, J. (2020). Amazon workers are fighting for their rights. This holiday season, think of them. Retrieved 6 May 2021, from <https://www.theguardian.com/commentisfree/2020/dec/01/amazon-workers-fighting-for-their-rights>
- Algorithm Watch. (2020, April 30). AI Ethics Guidelines Global Inventory. *Algorithm Watch*. <https://inventory.algorithmwatch.org/>
- Armasu, L. (2018). Amazon 'Rekognition' Falsely Identifies 28 Congress Members as Criminals (Updated). *Tom's Hardware*. (2021). <https://www.tomshardware.com/news/amazon-rekognition-congress-members-criminals,37515.html>
- Blackman, R. (2020). A Practical Guide to Building Ethical AI. *Harvard Business Review*. <https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai>
- Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*.



- Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law and Security Review*. <https://doi.org/10.1016/j.clsr.2019.04.007>
- Coeckelbergh, M. (2019) Artificial Intelligence: Some ethical issues and regulatory challenges. *Technology and Regulation*, 31-34.
- EU Commission (2017) Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL))
- Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology*, s13347-019-00354-x. <https://doi.org/10.1007/s13347-019-00354-x>
- Google. (2019). *Perspectives on issues in AI governance* (pp. 1–34). <https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf>
- Hagendorff, T. (2020). *The Ethics of AI Ethics: An Evaluation of Guidelines*. *Minds and Machines*. doi: 10.1007/s11023-020-09517-8
- Jobin, A., Ienca, M., & Vayena, E. (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9): 389-399.
- Joshi, A. (2019) Google dissolves its Advanced Technology External Advisory Council in a week after repeat criticism on selection of members. *Packt*. <https://hub.packtpub.com/google-dissolves-its-advanced-technology-external-advisory-council-in-a-week-after-repeat-criticism-on-selection-of-members/>
- Morte, R & Monasterio, A (2017). Entrevista a Ramón López de Mántaras. *Dilemata* (24), 301–309.
- Ng, A. (2018) Amazon is selling facial recognition technology to law enforcement. *C/Net*. <https://www.cnet.com/news/amazon-is-selling-facial-recognition-technology-to-law-enforcement/>
- OECD. (2019). *AI Principles*. <https://oecd.ai/ai-principles>
- Partnership on AI. (2018). *About us*. <https://www.partnershiponai.org/about/>.



- Pasternack, A. (2019) Amazon says face recognition fears are "insignificant." The SEC disagrees. *Fast Company*. <https://www.fastcompany.com/90329464/amazon-cant-block-investor-vote-on-face-recognition-says-sec>
- Pichai, S. (2018). *AI at google: Our principles*. <https://www.blog.google/technology/ai/aiprinciples/>
- Prisma European Network. (2021). *Digital Sovereignty: GAFAM, BATX, and ... the European Union?* <https://prisma-network.eu/our-work/DIGITAL-SOVEREIGNTY-GAFAM-BATX-EUROPEAN-UNION>
- Punke, M. (2019) Some Thoughts on Facial Recognition Legislation. *Amazon Web Services*. <https://aws.amazon.com/es/blogs/machine-learning/some-thoughts-on-facial-recognition-legislation/>
- Scheiber, N. & Wakabayashi, D (2019) Google Hires Firm Known for Anti-Union Efforts. *New York Times*. Tecnología. <https://www.nytimes.com/2019/11/20/technology/Google-union-consultant.html>
- Simonite, T. (2020). Google Offers to Help Others With the Tricky Ethics of AI. *Wired*. <https://www.wired.com/story/google-help-others-tricky-ethics-ai/>
- Singer, N. (27 de julio 2018). ¿La tecnología de reconocimiento facial de Amazon puede ser racista? *New York Times*. Tecnología. <https://www.nytimes.com/es/2018/07/27/espanol/amazon-rekogniton-aclu.html>
- Singer, N. (2019). Amazon is pushing facial technology that a study says could be biased. *MIT Media Lab*. from <https://www.media.mit.edu/articles/amazon-is-pushing-facial-technology-that-a-study-says-could-be-biased/>
- Tardif, A. (2021). Reid Blackman, Ph.D, Founder and CEO of Virtue Consultants. *Interview Series*. <https://www.unite.ai/reid-blackman-ph-d-founder-and-ceo-of-virtue-consultants-interview-series/>

Descolonizando la “Ética de la Inteligencia Artificial”*

Decolonizing the “Ethics of Artificial Intelligence”

Jonathan Piedra Alegría

Universidad Nacional, Costa Rica
Universidad de Salamanca, España
jonathan.piedra.alegria@una.cr

ISSN 1989-7022

RESUMEN: En la primera parte de este artículo se contextualiza el debate sobre la regulación ética de la Inteligencia Artificial, para posteriormente realizar un repaso general sobre los principales elementos anglo-eurocéntricos de las iniciativas públicas sobre la “Ética de la IA” que han surgido recientemente. Esto con el propósito ulterior de analizar la base principialista de la Ética de la IA. Nos referiremos a esto con el objetivo específico de presentarla en los términos de un caso paradigmático de ética “colonizada” que invisibiliza los diferentes juicios morales y a las axiologías culturales alternativas. Partimos de la hipótesis de que existe la posibilidad de problematizar la finalidad y el verdadero aporte de esta ética aplicada, utilizando para ello una perspectiva dilemática, propia de las ciencias humanas y jurídicas. Para llevar a cabo este tipo de reflexión, muy necesaria para entender cuál es el futuro de estas propuestas y como pueden afectar la regulación de la IA, se llevará a cabo un análisis acotado a la literatura contemporánea procedente del área de la Decolonialidad.

ABSTRACT: The first part of this paper contextualizes the debate on the ethical regulation of Artificial Intelligence, and then reviews the main Anglo-Eurocentric elements of public initiatives about the ‘Ethics of AI’ that have recently emerged. All this with the ultimate purpose of analyzing the basis of principlism of the Ethics of AI. We refer to this with the specific purpose of presenting it in terms of a paradigmatic case of ‘colonized’ ethics that hides the different moral judgments and the alternative cultural axiologies. We begin with the hypothesis that there is a possibility of questioning the real purpose and contribution of this applied ethics, using a dilemmatic perspective specific to human and legal sciences. In order to execute this type of analysis, which is highly required to understand what the future of these proposals is and how it can affect the regulation of AI, an analysis related to the contemporary literature from the Decoloniality field will be performed.

PALABRAS CLAVE: Decolonialidad, ética, principialismo, inteligencia artificial

KEYWORDS: Decoloniality, ethics, principlism, artificial intelligence

1. Contextualizando el debate

Ordinariamente, existen dos propuestas sobre cómo se deberían abordar los problemas procedentes de la implementación de la IA: (I) Un enfoque regulatorio duro (*hard law*) que busca la creación de un marco legal adecuado a los problemas propios de la IA, en cuanto a su investigación, desarrollo y utilización. Desde esta orientación es indispensable la creación de normas, leyes o reglamentos claros que definan el espacio de acción de las tecnologías de IA, así como la responsabilidad derivada de los posibles daños que puedan ocasionar. Para esto es necesario la participación del Estado (y organismos supranacionales) en la promoción de políticas públicas, así como de su efectivo cumplimiento. Un ejemplo de esta posición se encuentra en la propuesta china sobre IA del año 2017 —“New Generation Artificial Intelligence Development Plan” (AIDP)—. Teniendo como base este plan, dos años después —2019—, 19 municipalidades en China formularon 26 estrategias y políticas relacionadas con AI, con tal de cumplir con los objetivos propios de cada municipio (*China New Generation Artificial Intelligence Development Report —2019—*), pero siempre en consonancia con la propuesta general del país.

Frente a esta posición *dura*, se encuentra (II) un enfoque suave (*soft law*) basado en recomendaciones, declaraciones, manifiestos o propuestas que no tienen la fuerza vinculante, ni coactiva de la Ley, pero que pueden servir como elementos para mantener abierto el debate sobre la regulación y generar orientaciones que puedan ubicar los

* Agradezco todos los comentarios y aportes para la realización de este texto, realizados por la profesora María G. Navarro de la Universidad de Salamanca. Un especial agradecimiento para los revisores anónimos que con sus comentarios mejoraron la versión definitiva de este documento.



Received: 04/04/2022
Accepted: 28/05/2022

debates en ciertos aspectos de interés. Este rumbo es que ha tomado la Unión Europea (i.e. *Directrices Éticas para un IA fiable*), los Estados Unidos (i.e. *Preparation for the Future of Artificial Intelligence*) o Rusia (*Кодекс этики в сфере*—Código Ético de la Inteligencia Artificial—) por mencionar algunos casos.

El primer enfoque (I) es una discusión básicamente legal, dirigida por abogados, juristas o por *policymakers*. Desde esta posición, los dilemas y problemas de las tecnologías disruptivas son un asunto que se puede resolver con el cumplimiento de las normas legales existentes y en caso de que no fuera así, con la creación de nuevas. Mientras que la segunda postura (II) sería fundamentalmente un acercamiento ético conducido por filósofos especialistas en esta área, filosofía política e incluso académicos en general. Esta visión ha mostrado ser en los últimos años la que ha tenido mayor desarrollo. Desde esta posición, los problemas normativos de la IA no son únicamente aspectos jurídicos que se deban resolver exclusivamente con las herramientas de los sistemas legales. Es necesario indicar, que esta perspectiva no implica el desarrollo de alguna propuesta regulatoria concreta, ni un posicionamiento claro con respecto al tipo de marco normativo necesario, sin embargo, esto no ha impedido que bajo el nombre de "*Ethical guidelines*" en el año 2019 se pudieran encontrar 89 documentos de este tipo (Jobin et al. 2019). Llegando casi a duplicarse (más de 167 Guías Éticas sobre IA) en el año 2020 (Algorithm Watch 2020). Esto podría deberse a dos motivos: En primer lugar, debido al lobby e influencia de las grandes compañías privadas que dominan el campo de la IA. Es bastante conocido que muchas de ellas consideran que la regulación es inadecuada, en vista de que limita la innovación debido a una falta del conocimiento de la realidad tecnológica y de las dinámicas mercantiles. Por lo que la *regulación dura* se presenta como un obstáculo. Un ejemplo de esto, lo podemos encontrar en las propuestas empresariales, las cuales se basan en la idea de la autorregulación. Es decir, en que las mismas empresas se impongan lineamientos éticos que sirvan como base para de sus prácticas comerciales. En un contexto práctico, la autorregulación empresarial, es solo una forma de distracción que tiene como finalidad aparentar preocupación social (Piedra 2021a), pero con poca o ninguna incidencia real en los problemas que pretenden solucionar. Con el agravante de que retrasan la verdadera regulación, ubican los debates éticos en problemas generalísimos y en los casos que no es así, no existen soluciones técnicas concretas o viables.

El otro motivo que podría explicar la "explosión" de estas *Ethical guidelines* se encuentra en la base de un problema jurídico conocido como el "*Regulatorisches Trilemma*" (Teubner 1984). El *Trilema* es un modelo teórico socio-jurídico utilizado para identificar disfunciones sistémicas derivadas de la falta de armonía en la articulación estructural entre las esferas del Derecho, la Política y la Sociedad. Teniendo como resultado fallas regulatorias. Una de las premisas de este modelo es que la legitimación de las normas legales se consigue en una etapa reflexiva de los ordenamientos jurídicos, en donde por medio de una regulación indirecta (basada en la auto referencialidad de los sistemas sociales) se logra la regulación de las actividades de esta misma sociedad. A partir de esto, se infiere que una regulación excesiva, puede traer efectos inesperados creando nuevas disfunciones sistémicas.

Este sería precisamente uno de los principales problemas de los enfoques *hard law*. La regulación de la IA necesita una comprensión que va más allá de las formales clásicas de la legisla-

ción tradicional. Un acercamiento únicamente formal, podría destruir el sistema intervenido debido a una regulación excesiva. En el caso de la IA, este problema implica que la aplicación de una "normativa excesiva" podría provocar efectos secundarios inesperados debido a su carácter obligatorio. Una extensiva y detallada regulación de esta figura (i.e regulación estatutaria) podría no resultar ser adecuada para una tecnología como la IA, en donde los cambios ocurren muy rápidamente o en los cuales se requiere un alto grado de creatividad para su desarrollo. Desde esta posición, se puede concluir que la solución más adecuada para los problemas normativos de la IA se encuentra en normas menos directas (i.e normas éticas), ya que no es adecuado introducir normas rígidas en un área que se encuentra en constante cambio y transformación. Otra conclusión derivada de esta argumentación es que son las mismas empresas y compañías tecnológicas las que deberían dentro de su propia lógica solucionar los problemas de la IA, dejando el papel fuerte estatal en un plano secundario. Este acercamiento, se parece, *mutatis mutandis*, a la propuesta autorregulatoria de las grandes compañías. Aunque a diferencia del enfoque empresarial, el papel estatal no es innecesario ni un obstáculo en sí mismo, pero no debería ser la primera opción para corregir las contrariedades de la IA. Es así, como muchas de las orientaciones o *Guías éticas* (GE) sobre la IA pueden entenderse a partir de este contexto.

2. Una ética colonizada: Crítica al modelo principialista.

Este problema explica en parte la enorme cantidad de iniciativas públicas sobre la "Ética de la IA" (EIA) que han surgido recientemente. La otra parte de la explicación se encuentra en el deseo de las empresas privadas por evadir la verdadera regulación (Floridi 2019), centrando la discusión en soluciones técnicas que no abordan el fondo de los problemas o dirigiendo la atención sobre problemas abstractos (Wagner 2018) que parecen orientarse más a encontrar motivos para evitar la regulación, que servir como una verdadera guía o lineamientos sobre estos temas. En ambos casos, muchas de estas GE se caracterizan por una serie de propuestas vagas e indeterminadas que poco aportan a la creación de un verdadero marco de gobernanza.

Nuestro argumento se centra en que esto se debe, primordialmente, a que estos documentos se encuentran llenos de declaraciones generales, basadas en un enfoque principialista que sigue la tradición clásica anglosajona de la bioética (Floridi et al. 2018). Si bien, este modelo puede tomarse como la versión más célebre de una ética aplicada, no necesariamente es la más adecuada, ni tampoco la más conveniente para un "contexto global" en la IA.

Para comenzar, es necesario decir que los principios clásicos de la propuesta de Beauchamp y Childress redujeron la Bioética al ámbito biomédico exclusivamente. Particularmente la propuesta inicial de Potter (1971) quedó reducida a una serie de temas morales relacionados con la investigación biomédica, ajustados a lo que luego se conoció como el "Georgetown mantra", en vista de la repetición mecánica (Clouser y Gert 1990) de estos cuatro principios, para resolver cualquier conflicto. Sustituyendo las teorías éticas e incluso invisibilizando propuestas morales alternativas. Una situación similar sucede en la EIA. En vista del éxito de este modelo, prácticamente todas las GE sobre la IA, se basan en este sistema orientado a unos pocos principios.

Los principios propiamente éticos que se mencionan en estas GE¹ son, por lo general, muy pocos (debido a la influencia del modelo bioético) y cuando no lo son, es porque se presentan como versiones reiteradas de un mismo elemento, así como aclaraciones y/o extensiones de una idea o un tema central, por lo que aun cuando en un primer acercamiento parezca que se muestran una amplia gama de principios², lo cierto es que se resumen a una pequeña lista que incluye: respeto por la autonomía (humana), prevención (o reducción) del daño, justicia y explicabilidad. En este último caso, la “explicabilidad” se muestra como un principio independiente, a pesar de que es habitual encontrarlo en una relación de dependencia con los principios de responsabilidad o la transparencia. La “transparencia” aparece como un quinto principio ético, aunque en este caso, lo que significa propiamente difiere según el documento en el que aparezca. Tampoco es extraño encontrar a la “privacidad” como un sexto principio. No obstante, en este caso, y aún a pesar de su supuesta autonomía frente al resto de principios, se encuentra intrínsecamente relacionado con la justicia y la prevención del daño.

Generalmente, estas GE implican una breve presentación de conceptos y algunos usos de la IA de manera superficial con tal de “explicar” estos principios. La mayoría de estas GE tiene estructuras muy definidas a partir de las cuales describen los temas. Por ejemplo, la mayoría se basan en un *contenido probatorio*, el cual tiene como objetivo principal mostrar los beneficios que brinda el uso concreto IA (i.e el documento francés “*How can humans keep the upper hand? Report on the ethical matters raised by AI algorithms*”). En ocasiones también presentan algunos problemas que podrían ocurrir o han ocurrido debido al mal uso de la IA (i.e *Artificial Intelligence Ethics and Principles, and toolkit for implementation* de los Emiratos Árabes Unidos). Finalmente, un número mucho menor de GE tiene un *contenido resolutivo*. La característica de este tipo de contenido se basa en la mejora real de un servicio o actividad a través de la IA, pero también ciertas limitaciones para tener en cuenta a la hora de desarrollar la IA. (i.e “*A guide to using Artificial Intelligence in the public sector*” del Reino Unido). A pesar de esto, ciertamente en la gran mayoría de estas GE solo encontramos una exposición vaga e indeterminada de algunos cuantos principios que poco aportan a la creación de un verdadero marco de gobernanza.

Famosas iniciativas privadas como *AI at Google: our principles (2018)*, *Perspectives on Issues in AI Governance* (también de Google), *Some Thoughts on Facial Recognition Legislation (2019)* de Amazon, *Microsoft AI principles (2019)*, *IA Ethics* de IBM, *AI Principles (2008)* de GE Healthcare o *Guidelines for Artificial Intelligence (2019)* de Deutsche Telekom, siguen esta línea que hemos expuesto. Lo mismo sucede con otra serie de propuestas públicas. La *National Strategy for Artificial Intelligence (2018)* de la India, la *Australia’s Ethics Framework (2019)*, *Ethical Principles for Artificial Intelligence (2020)* del Departamento de Defensa de los Estados Unidos, *Principles on Artificial Intelligence (2019)* de la OECD o la propuesta de un grupo de expertos de alto nivel sobre IA (constituido por la Comisión Europea en el año 2018) se basan en un sistema de principios para una «IA Fiable». En Rusia, documentos derivados de alianzas público-privadas como el “Código Ético de la Inteligencia Artificial” (*Кодекс этики в сфере искусственного интеллекта*) plantean el mismo enfoque principialista que estamos mencionando.

Al igual que con la propuesta de Beauchamp y Childress, el “principialismo de la IA” carece de los elementos necesarios para resolver de manera adecuada las situaciones éticamente complejas que resultan de la interacción de la IA y la sociedad. Especialmente los dilemas

relacionados con los Derechos Humanos o las estructuras económicas subyacentes a estas tecnologías. Es particularmente extraño como para una tecnología tan dinámica y cambiante se promueve una lista bastante limitada de principios éticos estáticos, sin criterios operativos o sin siquiera elementos contextualizadores. Como señala Mittelstadt (2019) a diferencia de la Bioética, la EIA de no posee objetivos comunes (basados por ejemplo en un ideal médico), un código deontológico para los profesionales de la IA, alguna metodología para traducir los principios a la práctica o mecanismos legales efectivos de responsabilidad profesional. No obstante, este principialismo parece ser la solución para todos los problemas relacionados con la IA, a pesar de que no existe ninguna fundamentación teórica clara en estas propuestas con la cual se puedan ubicar las soluciones en un marco conceptual que permita abordar los dilemas morales adecuadamente.

Hay que tomar en cuenta que cada principio refleja un ideal y, por lo tanto, cada principio implica un valor determinado. En el caso de la Bioética, estos principios, fueron creados a partir de una base moral común que reflejaba un aspecto particular de la moral de un país concreto: el pragmatismo estadounidense. Holm (1995) considera que este principialismo se encuentra influenciado por lo que él llamó *American common-morality*, es decir por la moral y la cultura estadounidense de ese momento. Esto mismo sucede en el caso de la EIA. La base de esta ética es una narrativa moral centrada en occidente. Con esto nos referimos a un perspectiva moral y cultural anglo-eurocéntrica basada en criterio liberal del ser humano y en una propuesta ética utilitarista, en la cual subyace una estructura de racionalidad tecno-instrumental consolidada en el sistema capitalista moderno.

Es en esta *narrativa* en donde se muestra el *colonialismo ético* de estas propuestas. Por un lado, es bastante pretencioso considerar que una serie de normas (i.e la propuesta europea o estadounidense) sirvan para regular todos los desarrollos mundiales en esta área³, sin tomar en cuenta las características de los problemas que pueden aparecer en otros países. Aquí es donde opera precisamente la *colonialidad del saber*, ya que esta EIA —universal— funciona como un dispositivo de dominación legitimado por unas dinámicas de control en el ámbito técnico, así como en el aspecto epistemológico. La colonialidad es “[...] es uno de los elementos constitutivos y específicos del patrón mundial de poder capitalista. Se funda en la imposición de una clasificación racial/étnica de la población del mundo como piedra angular de dicho patrón de poder, y opera en cada uno de los planos, ámbitos y dimensiones, materiales y subjetivas, de la existencia cotidiana y a escala social”. (Quijano 2007, 93). Si tomamos al principialismo como una categoría de análisis de construcción occidental se hace claro que la EIA, responde a ciertas reglas sistémicas que articulan de una manera sutil las relaciones entre los centros hegemónicos del conocimiento tecnocientífico y las periferias. Sirva como ejemplo de esto, el número de artículos publicados sobre la IA. Según el *Artificial Intelligence Index Report* del año 2019, el 86 por ciento de los artículos presentados en conferencias de IA en el año 2018 fueron de autores de Norteamérica (Estados Unidos y Canadá), Europa o Asia oriental. Con respecto a las otras regiones del mundo, menos del 10 por ciento provenían de lugares distintos⁴. Esto es una muestra de la desigualdad de poder entre el centro hegemónico y periferia en el conocimiento. Es a partir de los centros hegemónicos del conocimiento desde donde se produce una realidad científica y ética, basada como hemos mencionado en el contexto histórico-geográfico del colonialismo/colonialidad (Quijano 2014). Situaciones como estas se agravan en vista de que la mayoría de las editoriales internacionales reconocidas —por este mismo sistema excluyente— se encuentran en América del Norte (principal-

mente Estados Unidos) y Europa, propagando las creencias, modos culturales, formas morales e ideales centrados en Occidente (Chimakonam 2017). Esto genera, como es de suponer, una serie de distinciones que sirven de líneas limítrofes en cuanto al conocimiento científico. Santos (2010) llama a esto *pensamiento abismal*. El pensamiento occidental moderno se basa en una serie de distinciones abismales que crean dos universos diferentes: Occidente/Otros-Mundos. Los *Otros-Mundos* son todas las realidades culturales-morales que son excluidas y que son obligadas a ajustarse a las reglas del mundo occidental. Desde este marco de interpretación, la esfera ética se podría entender desde la imposición de valores anglo-eurocéntricos que se presentan como la única manera de conocer e interpretar la realidad. Esta colonialidad del saber en la IA niega la producción intelectual de otros contextos culturales (femeninos, indígenas, afrodescendientes etc.) como verdadero conocimiento (Walsh 2007). Si bien, en el colonialismo tradicional la negación se daba únicamente por el color de la piel o el lugar de origen, ahora el rechazo es más sutil (pero igualmente xenófobo y racista). Actualmente se presenta bajo el argumento de un saber neutral, objetivo y universal. La ontología nacional e identitaria propia de estas narrativas centradas en occidente toman primacía frente a las demás culturas (Maldonado-Torres 2007) que son simplemente oscurecidas sobre la base de un proyecto ético como el que señalamos. Esto genera una amplia desigualdad en la producción del conocimiento que “[...]se rige por mecanismos epistémicos que le otorgan validación al tipo de sistema social que busca ser implementado como el molde a reproducir en el orbe.” (Maldonado 2018, 41). En el caso de la ética esta distribución desigual del poder obstaculiza y entorpece una participación verdadera de los países de la periferia en las discusiones sobre ética, particularmente con respecto a la IA. Los debates “globales” sobre la ética de esta tecnología, parten de un marco epistemológico limitado en el cual no existe una conversación con otras lógicas culturales que permitan repensar otras alternativas frente a este modelo epistemológico colonial.

3. Éticas-otras

Esto lo podemos ver en el caso del principio de autonomía. Desde la perspectiva kantiana, la autonomía tiene que ver con la capacidad humana de darse sus propias leyes. Esto quiere decir una capacidad de auto legislarse sin someterse a algún elemento externo. Los seres humanos como seres racionales no se encuentran obligados por el cumplimiento de leyes morales impuestas externamente, sino que, por el contrario, “está sometido a su propia legislación” (Kant 2000, §432). En las sociedades liberales-utilitaristas occidentales esto se traduce en una serie de “libertades personales” que priorizan los derechos individuales y la utilidad sobre las relaciones sociales. Sin embargo, esto no sucede necesariamente en las culturas africanas, asiáticas o latinoamericanas. Como ya lo demostró *The Moral Machine Experiment* (Awad *et al.* 2018) los valores y elementos en consideración a la hora de tomar una decisión moral difieren fuertemente entre la *ética individualista* de las sociedades anglo-europeas y de la *ética colectivista* de los países de Medio Oriente, Asia y África. Desde una ética colectivista, las decisiones no se miden por el interés personal o la maximización de la utilidad, sino que se basan en las necesidades y características del grupo o colectivo. Lo mismo puede decirse del contexto latinoamericano, particularmente desde los valores y tradiciones ancestrales de la población indígena. Tomemos como referencia un caso concreto. En la ética comunitaria del

Sumak Kawsay (una teoría moral de Ecuador y Bolivia), lo que se busca es “pensar bien, sentir bien para hacer bien con el objetivo de conseguir la armonía con la comunidad, la familia, la naturaleza y el cosmos” (Lajo 2008). El ser humano no es el centro de todas las decisiones, ni se espera que actúe conforme a una moralidad objetiva (como implica la perspectiva kantiana). La comunidad y las relaciones sociales existen antes que la idea de individuo. La ética del *Sumak Kawsay* está vinculada a una virtud comunitaria y no a una serie de valores individualistas. El ser humano es un elemento más de la Madre Tierra —*Pachamama*—. Este modelo busca el *buen vivir* de los demás y el bienestar de la comunidad, junto con un equilibrio de la naturaleza. Dentro de esta perspectiva, la autonomía incluso toma un matiz regional, ya que la base territorial es un elemento sustancial de la cultura. No podría existir la autonomía sin este elemento étnico-cultural. La perspectiva de una autonomía abstracta de un agente racional —basada en principios— no tiene cabida en un planteamiento como este.

Algo similar sucedería con el concepto de privacidad. Este concepto se usa de manera vaga para referirse a muchas ideas diferentes. No obstante, en el tema de la IA se asocia, *grosso modo*, a tres temas fundamentales (Tucker 2019) relacionados con los datos personales: La creación y persistencia de los datos en el tiempo, la utilización de estos datos y el derrame de datos. Como vemos, la privacidad está asociada a una esfera personal e individual del ser humano, específicamente en cuanto a su identificación. Esto implicaría que la privacidad es una derivación lógica de la individualidad. Una suposición razonable de esto sería que la autonomía se protege también por medio de la privacidad. Situación que es coherente con una moral individualista. Sin embargo, la noción de privacidad cambia según las distintas tradiciones y contextos sociales. Para las culturas indígenas latinoamericanas tampoco aplica la idea individualista de la privacidad. En este caso, sería más apropiado referirnos a una “privacidad colectiva”. Esto en vista, de que el ser humano se concibe como una totalidad y no como una individualidad aislada. El *Todo* existe antes que la idea individualidad. “Se trata de un tiempo y de un espacio subjetivado, o sea privado, que se refiere a un hábitat vital, en donde *nuestro* tiempo y *nuestro* espacio se funden en el hecho puro de vivir aquí y ahora cuando involucra el tiempo de *mi* vida, *mi* oficio, *mi* familia, y en este lugar, de *mi* comunidad”. (Kusch 1999, 422) El sentido de identidad se construye desde una base comunitaria y a partir de un sentido de pertenencia originaria con la naturaleza. La vida solo se puede comprender desde la colectividad —*sumak kamaña*—. Esto implica que la privacidad debe entenderse como un modo de ser social, y particularmente como un fenómeno relacionado con la forma en como los seres (humanos y no humanos) comparten el mundo entre sí.

Otra diferencia entre la posición individualista y la colectiva la encontramos en la visión instrumentalista de la naturaleza presente en la primera postura. En lo que respecta a los documentos de éticos de la IA, el ser humano se presenta como el núcleo del desarrollo de esta tecnología —*human-centric approach*— (i.e. *Código Ético de la Inteligencia Artificial* de Rusia). Un argumento categórico, pero peligrosamente ambiguo, ya que no es solo evidente que todas las tecnologías deben estar basadas en los intereses y valores humanos, sino además porque casi siempre «“humancentric” seems to be synonymous with “anthropocentric”» (Floridi 2021, 218). Este tipo de pensamiento continua una línea de explotación que solo considera los aspectos ambientales desde el punto de vista de la utilidad. Un antropocentrismo que solo valora a la naturaleza solo como una reserva de materias primas. Esta manera de pensar es la que nos ha metido en todos los problemas ambientales del Antropoceno. La naturaleza se concibe desde un “extractivismo ontológico” (Grosfoguel 2016). Un modo ser propio de

estas narrativas centradas en occidente. Una forma de *estar en el mundo* que solo busca la apropiación y el beneficio. Como claramente lo indican Harle *et al* (2018, 15) hay “Evident in the progression of modern Western technologies is the drive for power and exploitation.”. Una faceta más del colonialismo propio de la cultura occidental. Cualquier oposición a esta forma del ver el mundo es concebida como “anti-progresista”, y por lo tanto invisibilizada, ridiculizada o lo que es peor, eliminada. En lo referente a la EIA, la naturaleza se aborda desde una instrumentalización que despolitiza y descontextualiza a las diferentes culturas por medio de propuestas —como la principalista— que eliminan los diferentes sentidos y formas de *ser* o en mejor de los casos, asimilándolas a la matriz cultural moral anglo-europea.

Frente a este instrumentalismo y mercantilización de la naturaleza, podemos utilizar como ejemplo, de nuevo, el *Sumak Kawsay* como una herramienta decolonial que ha sido utilizada por algunos países sudamericanos como una alternativa al modelo económico capitalista y a su extractivismo ontológico. El buen vivir parte de un desarrollo económico que respeta y que tiene como límite a la Pachamama. En este contexto la naturaleza no es objeto de protección, sino que es sujeto de derechos. Si el clásico antropocentrismo supone una superioridad destructiva del ser humano frente a todo lo demás, el paradigma ético del *Sumak Kawsay* busca una relación de equilibrio y plenitud con la naturaleza. Una forma armónica de vivir colectivamente basada en la responsabilidad social y no en la explotación del medio ambiente o en el beneficio propio.

La visión ética de estos pueblos revela un sistema unificado de valores, así como un enfoque que muestra una cosmovisión basada en el buen vivir, la cohesión social, la sostenibilidad ambiental, así como el respeto por las tradiciones culturales y la espiritualidad comunitaria. Todos estos elementos deberían ser considerados a la hora de estimar las tensiones normativas y políticas subyacentes en conceptos clave para la IA como la autonomía o la privacidad, por ejemplo. Sin embargo, las iniciativas de la EIA se basan en declaraciones generales en las que aparecen una serie de principios vagos que prometen ser orientaciones para la acción, pero que en la práctica brindan pocas recomendaciones específicas. Si a esto le sumamos que este conjunto de principios está basado en narrativas centradas en modos de vida occidentales (que invisibilizan las posiciones éticas alternativas), la EIA se reduce a una mera selección mecánica y repetitiva de principios que poco aportan en un nivel práctico. Una propedéutica que contribuye muy poco al análisis y comprensión de los actuales fenómenos que se desarrollan en el campo de la IA.

Las perspectivas descolonizadoras nos brindan una mejor comprensión de los aspectos morales y jurídicos que se presentan en la implementación y desarrollo de la IA. Ya que nos permiten reflexionar críticamente sobre los problemas y controversias éticas, políticas, económicas y sociales de los sistemas de conocimientos que se muestran como universales, objetivos o neutros. La incorporación de la producción intelectual de otros contextos culturales, así como la inclusión de diferentes epistemologías y de conocimientos no-occidentalizados nos brindan una gran oportunidad para afrontar los desafíos globales contemporáneos relacionados con la IA. Desafíos que en la actualidad parecen encontrar algunas respuestas (teóricas y abstractas) en un principalismo reduccionista. Es necesario considerar seriamente todo lo que se puede aprender de distintos marcos epistémicos y morales con tal de plantear verdaderas soluciones que incorporen, por ejemplo, a la sustentabilidad social y ambiental en el debate de la EIA. La protección a la cultura (i.e los derechos culturales) o la identidad colecti-

va son temas que igualmente deben ser considerados no solo para las GE, sino para la EIA en general. Superar estos marcos occidentales de pensamiento e incorporar otras perspectivas éticas, es necesario para contrarrestar todas las prácticas coloniales, los supuestos culturales y los prejuicios dentro de los sistemas occidentales de pensamiento (en este caso el ético). La descolonización de la EIA es un reto más a la narrativa excluyente (anglo-eurocéntrica) que busca invisibilizar, reprimir o silenciar todos conocimientos que no se ajustan a su propio marco epistémico, moral o económico.

4. Conclusiones

Se necesitan más estudios sobre las propuestas éticas de la IA con tal de mostrar las implicaciones políticas y sociales de este tipo documentos (en países latinoamericanos, por ejemplo). No obstante, hemos mostrado un problema importante de estas GE: La invisibilización y exclusión de diferentes epistemologías, así como de conocimientos no-occidentalizados en lo relacionado con axiologías culturales alternativas. La EIA se encuentra atascada en una ruta estrecha que tiene poco espacio para la diversidad de pensamiento o el debate ético basado en conocimientos y cosmovisiones alternativas. La propuesta de la autorregulación empresarial tampoco es la solución, ya que no busca un cambio real y lo único que hace es postergar un auténtico debate sobre cuestiones primordiales (Piedra 2021a).

El verdadero potencial de las GE desaparece en un discurso que intenta imitar el principialismo bioético, sin tomar en cuenta sus limitaciones, o sin siquiera considerar la características sociales, culturales o estructurales de los países en los cuales se busca desarrollar e implementar las tecnologías de IA.

Las implicaciones de esto son claras. Se deben repensar desde un marco decolonial conceptos como privacidad, autonomía o el manejo de datos, entre muchos otros tópicos. Temas como la robótica social, por ejemplo, plantean desafíos particulares (Piedra 2021b) que ocupan mucha investigación empírica adicional, así como un debate que tenga presente y valore múltiples contextos que sirvan para integrar diversos puntos de vistas, así como diferentes propuestas éticas. Es necesario legitimar y reconocer posiciones epistemológicas y morales diferentes a la perspectiva universalista anglo-eurocéntrica. Es ineludible una conversación verdaderamente global sobre la IA, en donde no se marginalice el conocimiento no-occidental. Para esto es preciso que las recomendaciones o GE tengan en cuenta los contextos y las peculiaridades de los diferentes grupos sociales. Esto solo será posible si en los comités y órganos que debaten estos temas, se conformen grupos interdisciplinarios con personas de muy distintos orígenes y condiciones vitales.

Esto propone un reto doble para la EIA: En primer lugar, es necesario la descolonización epistémica de las propuestas teóricas con pretensión universal. Habilitar el conocimiento de otras subjetividades y otros espacios geográficos, ampliando así los relatos anglo-eurocéntricos respecto a la EIA que constantemente se reproducen como los únicos marcos cognoscitivos válidos. Un aspecto muy importante es el acceso y participación en las plataformas científicas con tal de poder visibilizar paradigmas alternativos. Los investigadores de la "periferia epistémica" con tal visibilizar sus ideas se ven obligados a publicar en editoriales internacionales

(que solo publican en inglés) que posean altas “tasas de impacto”⁵, pero poco espacio para posiciones distintas a las anglo-eurocéntricas o que de hecho no tienen ningún interés por los saberes alternativos. Como muy acertadamente lo mostraron Halfffman y Radder (2015), este tipo de prácticas (aun dentro del mismo “centro epistémico”) ni siquiera garantizan la calidad y lo que, de hecho, provocan es que “They ignore and destroy the variety of knowledge forms and practices in various fields of study” (Halfffman y Radder 2015, 167). En América Latina, por ejemplo, a pesar de que existen muchos medios alternativos de excelente nivel, estos no están legitimados por la institucionalidad, lo que produce una especie de crisis en el conocimiento. Es necesario dedicarles más recursos financieros a estos medios alternativos, así como a las investigaciones provenientes de estos espacios excluidos.

Esto nos lleva al segundo reto: la dimensión praxeológica (Maldonado 2018). Descolonizar la EIA implica también, una reflexión sobre el poder. Se debe trabajar en la crítica de la base socioeconómica de estas propuestas, la tendencia al monopolio de las empresas y las condiciones materiales que implica el uso de la IA. La atención no debe centrarse únicamente en el aspecto tecnológico. También hay que repensar las actuales condiciones estructurales, ambientales y económicas en la cuales se basan estas tecnologías.

Finalmente, es claro que una serie de principios (cualesquiera que sean) no pueden garantizar por sí solos la EIA. Frente a esto, los acercamientos no reduccionistas, interculturales y decoloniales serían beneficiosos para generar soluciones o paradigmas alternativos que promuevan contribuciones epistémicas diferentes. Generando un verdadero debate global sobre todos estos temas.

Referencias

- Algorithm Watch. (2020). AI Ethics Guidelines Global Inventory. *Algorithm Watch*. <https://inventory.algorithm-watch.org/>
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., et al. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64. <https://doi.org/10.1038/s41586-018-0637-6>
- Brussels Bureau (2021). US-EU need to write the rules for new technologies ‘before China’. Retrieved 18 June 2021, from <https://www.euronews.com/2021/06/15/us-and-eu-need-to-write-the-rules-for-new-technologies-before-china-says-us-secretary-raim>
- CEPAL. Comisión Económica para América Latina y el Caribe. (2004). Retrieved 25 May 2022, from <https://www.cepal.org/es/comunicados/america-latina-caribe-rezagada-investigacion-desarrollo>
- Chimakonam, J. O. (2017). African philosophy and global epistemic injustice. *Journal of Global Ethics*, 13(2), 120-137. <https://doi.org/10.1080/17449626.2017.1364660>
- Clouser KD, Gert B. (1990) A critique of principlism. *Journal of Medicine and Philosophy* 15 (2):219-236 <https://doi.org/10.1093/jmp/15.2.219>
- Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy y Madhulika Srikumar. (2020) “Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI.” Berkman Klein Center for Internet & Society. <https://doi.org/10.2139/ssrn.3518482>
- Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology*, s13347-019-00354-x. <https://doi.org/10.1007/s13347-019-00354-x>

- Floridi, L. (2021) The European Legislation on AI: a Brief Analysis of its Philosophical Approach. *Philosophy & Technology*. :1-8 DOI: [10.1007/s13347-021-00460-9](https://doi.org/10.1007/s13347-021-00460-9)
- Floridi, L., Cowls, J., Beltrametti, M. *et al.* (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Harle, J; Abdilia, A y Newman, A (eds) (2008). *Decolonising the Digital: Technology as Cultural Practice*. Old Ways, New. Sidney. Australia.
- Grosfoguel, R. (2016). Del extractivismo económico al extractivismo epistémico y al extractivismo ontológico: una forma destructiva de conocer, ser y estar en el mundo. *Tabula Rasa*, nº 24, pp. 123-143.
- Halffman, W., y Radder, H. (2015). The Academic Manifesto: From an Occupied to a Public University. *Minerva*, 53(2), 165-187. doi:[10.1007/s11024-015-9270-9](https://doi.org/10.1007/s11024-015-9270-9)
- Holm, S. (1995). Not just autonomy--the principles of American biomedical ethics. *Journal of Medical Ethics*, 21(6), 332-338. doi:[10.1136/jme.21.6.332](https://doi.org/10.1136/jme.21.6.332)
- Jobin, A., Ienca, M., y Vayena, E. (2019) "The global landscape of AI ethics guidelines." *Nature Machine Intelligence*, 1(9): 389-399. DOI:[10.1038/S42256-019-0088-2](https://doi.org/10.1038/S42256-019-0088-2)
- Jonsen AR. (2003) *The birth of bioethics*. New York, Oxford University Press.
- Kant, I. (2000). *Lógica. Un manual de lecciones*. Madrid, Akal.
- Kusch, Rodolfo (1999). *Obras completas II*. Rosario, Santa Fe: Editorial Fundación Ross.
- Lajo, J (2008) ¿Qué dice el Sumaj Kawsay? La escuela indígena de "Qhapaj Ñan". ¿Qué dice el Sumaj Kawsay? La escuela indígena de "Qhapaj Ñan". Retrieved 14 June 2021, from <https://sumakkawsay.tieneblog.net/?p=1988>
- Maldonado, C. (2018) *De-colonialidad en la era tecnomediática*. Ediciones CIESPAL.
- Maldonado-Torres, N. (2007). Sobre la colonialidad del ser: contribuciones al desarrollo de un concepto. En S. Castro-Gómez y R. Grosfoguel (eds.) *El giro decolonial. Reflexiones para una diversidad epistémica más allá del capitalismo global* (pp. 127-167). Bogotá: Iesco-Pensar-Siglo del Hombre Editores.
- Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat Mach Intell* 1, 501-507 (2019). <https://doi.org/10.1038/s42256-019-0114-4>
- ONU (2018). El racismo, un obstáculo para el desarrollo en América Latina. (2018). Retrieved 25 May 2022, from <https://news.un.org/es/story/2018/08/14406>
- Piedra, J. (2021a) Venditio fumi: Autorregulación Empresarial e Inteligencia Artificial. *Sincronía*. Año XXVI / Número 81. Enero-junio. DOI: [10.32870/sincronia.axxvi.n81.12a22](https://doi.org/10.32870/sincronia.axxvi.n81.12a22)
- Piedra, J. (2021b). El despertar de los robots (sociales). En: *Lucas en el camino: filosofía y ciencias sociales en tiempos de desconcierto*. España. Editorial Dykinson S.L. ISBN 978-84-1377-322-3
- Potter, V.(1971). *Bioethics: bridge to the future*. New Jersey: Prentice-Hall
- Quijano, A. (2000). Colonialidad del poder, eurocentrismo y América Latina. En E. Lander (ed.) *La Colonialidad del saber: Eurocentrismo y Ciencias Sociales. Perspectivas Latinoamericanas* (pp. 201-245). Caracas: CLACSO.
- Quijano, A. (2007). Colonialidad del poder y clasificación social. En: S. Castro-Gómez & R. Grosfoguel (eds.) *El giro decolonial. Reflexiones para una diversidad epistémica más allá del capitalismo global* (pp. 93-126). Bogotá: Iesco-Pensar-Siglo del Hombre Editores.
- Quijano, A. (2007). Colonialidad del poder y clasificación social. En: S. Castro-Gómez y R. Grosfoguel (eds.) *El giro decolonial. Reflexiones para una diversidad epistémica más allá del capitalismo global* (pp. 93-126). Bogotá: Iesco-Pensar-Siglo del Hombre Editores.
- Quijano, A. (2014). *Cuestiones y horizontes: de la dependencia histórico estructural a la colonialidad/descolonialidad del poder*. CLACSO.

- Santos, B. (2010). *Para descolonizar Occidente. Más allá del pensamiento abismal*. Buenos Aires, CLACSO.
- Springboard (2020). Tech Companies' Investment in R&D Pushes the U.S. Economy Forward - (2020). Retrieved 25 May 2022, from <https://springboardccia.com/2020/03/26/tech-companies-investment-in-rd-pushes-the-u-s-economy-forward/>
- Steering Committee (2019). Artificial Intelligence Index Report 2019 Steering Committee https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf
- Teuber, G. (1984). Das regulatorische Trilemma. Zur Diskussion um post-instrumentale Rechtsmodelle. *Quaderni fiorentini per la storia del pensiero giuridico moderno*. Vol. 13, N°. 1. 109-149
- Tucker, C (2018). "Privacy, Algorithms, and Artificial Intelligence," NBER Chapters, in: The Economics of Artificial Intelligence: An Agenda, pages 423-437, *National Bureau of Economic Research*, Inc.
- Wagner, B. (2018). Ethics As an Escape From Regulation.: From "Ethics-Washing" To Ethics-Shopping? In Bayamlioglu E., baraliuc I., Janssens L., y Hildebrandt M. (Eds.), *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen* (pp. 84-89). Amsterdam: Amsterdam University Press. doi:10.2307/j.ctvhrd092.18
- Walsh, Catherine (2007). ¿Son Posibles Unas Ciencias Sociales/culturales Otras? Reflexiones En Torno a Las Epistemologías Decoloniales. *Nómadas* (26): 102-113.

Notas al final

1. Información obtenida de "AI Ethics Guidelines Global Inventory" de AlgorithmWatch. Hasta la fecha, al menos 145 de estas iniciativas de EIA han creado documentos que describen principios éticos, valores o posturas generales para el desarrollo y la implementación de la IA.
2. Por ejemplo, el "principio de no discriminación" está muy relacionado con el principio de justicia. Es normal encontrarlos en una relación de género y especie.
3. Tal como y lo dijo la secretaria de Estado Unidos, Gina Raimondo (Brussels Bureau 2021).
4. Ciertamente, esto se debe a una gran variedad de factores. Quizá, el más evidente es que estas regiones poseen más recursos económicos, lo que les permite esta acumulación de poder y conocimiento. Por ejemplo, según un informe del año 2004 de la Comisión Económica para América Latina y el Caribe (CEPAL) en Latinoamérica solo se invierten en I+D el 0,5 por ciento del su producto interno bruto (PIB). Frente a un 2,5 y 3 por ciento del PIB de Estados Unidos o Japón. Así como los 2 puntos porcentuales que le dedica la Unión Europea a estos asuntos (CEPAL 2004). Estos montos se quedan cortos frente a los 65.000 millones de dólares americanos que solamente en el año 2018, las empresas Amazon, Alphabet, Microsoft, Apple e Intel, invirtieron en investigación y desarrollo (Springboard 2020). Pero si bien esto es un factor importante, no es el único. Los países de la "periferia del conocimiento" son países con una larga tradición colonial que pervive aún hoy. El racismo sistémico, la xenofobia, la violencia estructural y social, las condiciones de empobrecimiento (que entre otras cosas provocan una gigantesca "fuga de cerebros"), las migraciones forzadas o la presencia de una cultura fuertemente machista son temas que influyen en perpetuar estas relaciones coloniales entre los centros hegemónicos del conocimiento tecnocientífico y su relación con las periferias. Por ejemplo, en América Latina "Debido a su exclusión y discriminación los afrodescendientes tienen casi tres veces más posibilidades de vivir en la pobreza, tienen menos acceso a la educación y al empleo y están poco representados en cargos de toma de decisiones." (ONU 2018). Estos problemas sociales y políticos no pueden desvincularse de los motivos por los cuales existe una concentración en el conocimiento tecnocientífico.
5. Esto no solo pasa en la periferia, sino también en los países del "centro" del conocimiento. Por ejemplo, "In the Netherlands, counting output started off with the number of publications, then international publications, after which only English-language publications counted, thereafter articles in high-impact journals, and eventually often-cited publications (leading to a high 'h-index')." (Halfman y Radder 2015, 167).