

**REGULACIÓN O NO PARA LA IA? PROPUESTA DE REGULACIÓN HÍBRIDA DE
LA EN COLOMBIA****REGULATE OR NOT REGULATE AI? PROPOSAL FOR HYBRID REGULATION
OF AI IN COLOMBIA****Abg. Alex Bahamón ***, **Cient. Juan Pablo Barrero ******* Universidad de los Andes**, Facultad de derecho, Derecho.

Cra. 1 #18a 12, Bogotá, Colombia.

E-mail: ah.bahamon10@uniandes.edu.co.

**** Universidad de los Andes**, Facultad de Ciencias, Departamento de Física.

Cra. 1 #18a 12, Bogotá, Colombia.

E-mail: jp.barrero10@uniandes.edu.co.

Resumen: La inteligencia artificial es una tecnología que es cada vez más importante en el panorama científico e industrial en Colombia. Su implementación trae consigo empleo, eficiencia y otros beneficios para la economía del país, pero, al mismo tiempo genera riesgos para la privacidad (por ejemplo, en el uso de datos), el mercado y los consumidores, entre otros. Con esto en mente, a partir de parámetros éticos y estableciendo un conjunto de máximas se propone un marco ético que permita a los diferentes actores que hacen uso y desarrollan la inteligencia artificial autorregularse y que de una base sobre la cual se pueda desarrollar una legislación. De este modo se construyó un marco con el objetivo de permitir la innovación y el desarrollo de nuevas tecnologías y proteger los derechos de los individuos que se ven expuestos a ellas.

Palabras Clave: Inteligencia artificial, regulación, marco ético, uso de datos, autorregulación

Abstract: Artificial intelligence is a technology that is increasingly important in the scientific and industrial landscape in Colombia. Its implementation brings with it employment, efficiency and other benefits for the economy of the country, but at the same time it generates risks for privacy (e.g. the use of data), the market and consumers, among others. Based on ethical parameters and establishing a set of ethical postulates, we propose an ethical framework that allows the different actors that make use a develop artificial intelligence to self-regulate and that on a basis on which legislation can be developed. Thus, a framework was constructed with allowing innovation and the development of new technologies and protecting the rights of the individuals who are exposed to them as the main objectives.

Key words: Artificial intelligence, regulation, ethical framework, data usage, self-regulation

INTRODUCCIÓN

La inteligencia artificial es una de las tecnologías más prominentes del panorama económico, social y académico actual. Uno de los conceptos transversales para que una tecnología basada en algoritmos computacionales y modelos matemáticos se considere inteligente es que sea capaz de aprender a partir de datos históricos a realizar tareas que consideramos inteligentes cuando son realizadas por un humano. No obstante, los académicos no han logrado llegar a un consenso definitivo a la hora de definir qué constituye inteligencia artificial y qué no. No obstante, han desarrollado conceptos que son clave para tal efecto. Para poder plantear un marco ético que sirva como columna vertebral a las regulaciones que posteriormente se emitan en relación el desarrollo y uso de IA en Colombia, es necesario hacer la distinción entre las inteligencias artificiales generales y las específicas ya que, si bien ambas se consideran inteligencias, en este punto de la innovación tecnológica sólo es posible tratar con las últimas.

En un reporte del Consejo Nacional de Ciencia y Tecnología (NSTC, por sus siglas en inglés) de los Estados Unidos se define a la inteligencia artificial general como “un sistema de inteligencia artificial hipotético futuro que exhibe comportamientos aparentemente inteligentes, o por lo menos tan avanzados como los de una persona en el rango total de tareas cognitivas” (NSTC, 2016). Es evidente, como lo dice en la descripción, que aún no estamos cerca de conseguir una inteligencia artificial de este estilo por lo que este marco se limitará a expresarse sobre los principios éticos que deben regir la regulación de las inteligencias específicas, definidas en el mencionado reporte como inteligencias basadas en aprendizaje de máquinas (Machine learning) y aprendizaje profundo (Deep learning). Este último solo es un modelo en etapa de desarrollo que, si bien se ha implementado en usos comerciales o públicos, no es de uso extensivo, por lo que no será objeto de este marco ético.

En el caso de Machine learning, que es el objeto de este marco, se implementan algoritmos basados en modelos estadísticos que, a partir de un (i) entrenamiento hecho con datos históricos, (ii) aprenden a sacar conclusiones acerca de conjuntos de datos que tienen las mismas variables que los datos de entrenamiento (muchas veces comprobando con datos de prueba), (iii) para

predecir con cierta probabilidad que se dé un suceso. Esta es una de las herramientas más útiles y de mayor implementación en la industria y el sector público, pues permite resolver problemas que posiblemente no tendrían solución si se cuenta únicamente con supervisión humana (NSTC, 2016).

En los últimos años, de la mano con la tendencia de modernización por parte de particulares, empresas (como Facebook, Twitter y Google) y organizaciones gubernamentales (como la Fiscalía colombiana y la Policía de Nueva York), la IA es una buena herramienta para solucionar dichos problemas, puesto que se han desarrollado algoritmos de predicción basados en Machine learning que permiten agilizar tareas como el procesamiento de datos, la evaluación de peticiones y la identificación de personas facilitando muchos procesos que antes eran más dispendiosos (Elish y Hwan, 2017). Este es el caso de Colombia, país en el que tanto el sector privado como el público implementan sistemas de IA, no sólo optimizando ciertos procesos sino también generando una industria alrededor de dicha tecnología. Sin embargo, esta tecnología puede generar perjuicios de diferente índole para ciertos grupos poblacionales o la ciudadanía en general.

De este modo, surge la necesidad de crear un mecanismo que ayude al sector público y al sector privado a orientarse hacia un desarrollo e implementación sostenible y justo, donde considere los riesgos que enfrentan las poblaciones a las que se enfrenta esta tecnología. Además, es necesario permitir la experimentación e innovación (éticamente conscientes) sin restringirlas con una regulación legal, pensando en que la economía colombiana requiere entrar en este mercado para poder crecer y que, al mismo tiempo, los ciudadanos deben ser protegidos de alguna forma. Este marco se explicará a continuación bajo tres grandes ejes éticos que facilitarán la implementación de reglas mínimas morales que deberán cumplir los actores involucrados en el ambiente de la IA. Estos son la ética de los datos, la de los algoritmos y la de las prácticas (Buiten, 2019). Se tomarán estas tres etapas pues cada uno de ellos responde a diferentes problemáticas presentes en el contexto de la implementación y desarrollo de la IA (Buiten, 2019). (Plaza J, Ruiz M, Rosero C, Zapata L. 2017).

MARCO ÉTICO Y JUSTIFICACIÓN

La primera pregunta a contestar antes de cualquier tipo de intervención del Estado en la innovación y desarrollo tecnológico es: ¿Hay algún problema que justifique la intervención? En esta medida, estas preguntas no se pueden contestar con base en suposiciones o en el vacío, pues las implicaciones de intervenir son serias y reales, por lo que se debe evaluar con base en evidencia y conocimiento que permita tomar decisiones cuyos efectos se prevean (Gasser, 2017). Con esto en mente, en esta propuesta se buscan mitigar los riesgos inherentes al diseño e implementación de la IA en el contexto colombiano, partiendo de los problemas identificados a nivel mundial y con especial consideración en los que son característicos de un país en desarrollo como Colombia. En esta sección del documento se expondrán los principales problemas globales del diseño e implementación de la IA, con énfasis en algunos particulares de países del Sur Global; luego, se explicará cuál es el marco ético-filosófico que más se ajusta para solucionar estos problemas, donde se precisará en quiénes serán los sujetos de esta ética y, finalmente, se harán algunas precisiones con respecto a este marco.

En primer lugar, la literatura y la investigación a nivel mundial ha identificado problemas en cuanto a la IA en áreas que se refieren principalmente a los sesgos y la discriminación que producen estos sistemas, las vulnerabilidades, implicaciones de privacidad, la relevancia del ser humano como centro del diseño y la transparencia del desarrollo y de los sistemas son algunos de los problemas más relevantes. Estos serán estudiados a continuación:

1. Sesgos y discriminación (ética de los datos): estudios han arrojado que las aplicaciones de la IA en el sistema criminal, acceso al sistema financiero, moderación de contenido en línea y reclutamiento de recursos humanos pueden generar algunas formas de discriminación (Raso, Hilligos, Krishnamurthy, Bavitz y Kim, 2018). Por ejemplo, en el caso del sistema criminal, este estudio encontró que se podían permitir conductas discriminatorias con minorías históricamente víctimas de esta práctica, pues los datos que se usaban para el entrenamiento conservaban los sesgos estructurales del sistema penal (Raso et al, 2018).
2. Las vulnerabilidades (ética de los algoritmos): en 2016 Microsoft lanzó en Estados Unidos un chatbot que utilizaba un sistema de IA de Deep learning para brindar apoyo socio-afectivo a las personas, y se decidió a última hora que el canal de comunicación fuera Twitter. Esta IA ya había sido un éxito en China en conversaciones privadas a través de una app propia (Følstad y Bae, 2017). No obstante, luego de 16 horas de funcionamiento en Twitter, este sistema empezó a arrojar respuestas racistas y nazis, entre otros comentarios incendiarios. Esto se debió a que Tay realizó un aprendizaje de abundantes usuarios de Twitter con opiniones incendiarias que iniciaron conversaciones similares con la máquina de forma coordinada (Hunt, 2016). Por esto, el caso de Tay es un ejemplo de cómo se pueden explotar vulnerabilidad de las IA para generar resultados distintos a los previstos.
3. Implicaciones de privacidad (ética de las prácticas y de los datos): una firma de mercadeo político tuvo acceso a los datos de más de 87 millones de usuarios de Facebook, información con la cual se podían realizar perfilamientos, de tal forma que se podía generar contenido personalizado para los usuarios que reforzara algunas de sus creencias; y así favorecieron a un candidato en las elecciones presidenciales de 2016 en Estados Unidos (Isaak y Hanna, 2018). La violación a la privacidad de los datos que ocurrió en este caso generó consecuencias significativas en el desarrollo de la vida política de Estados Unidos, lo cual muestra el poder que otorgan los datos con los sistemas de procesamiento actuales. Así se puede ver como la IA tiene implicaciones severas en la protección de datos, pues de ellos puede surgir perfilamiento, segregación y dificultades para la autodeterminación (Scripa, 2017).
4. Ser humano como centro del diseño (ética de los algoritmos): a finales del siglo XIX y comienzos del XX se producían alrededor de 25 mil lesiones o muertes a empleados que trabajaban en los ferrocarriles de Estados Unidos acoplado vagones de carga, así que los legisladores expidieron el “Safety Appliance Act” donde ordenaban la instalación de acopladores automáticos (Leta, 2014). No obstante, las lesiones y las muertes no tuvieron reducciones significativas, pues los empleados seguían expuestos a los riesgos del

acoplamiento (Leta, 2014). Esto fue causado por un mal entendimiento de la interacción entre humanos y tecnología por lo que se generan afectaciones a la calidad de vida de los humanos al implementar sistemas (Leta, 2014). Además, es necesario considerar la sustitución de puestos de trabajo por máquinas como un gran reto que genera la IA en las sociedades contemporáneas (Chui, Mnyika y Miremadi, 2016).

5. Transparencia:

a. Desarrollo (ética de las prácticas): en la Provincia de Salta, en Argentina, el gobierno decidió implementar una IA para predecir embarazos entre población vulnerable. Para esto entregaron datos de más de 200 mil mujeres adolescentes en poblaciones vulnerables (Ortiz e Iglesias, 2018). Este fue un caso muy criticado por activistas, pues afirmaban que no solucionaba el problema y se estaba violando la privacidad de las menores (Ortiz e Iglesias, 2018). La IA no arrojó resultados esperados, se consideró un fracaso (Ortiz e Iglesias, 2018). En este caso hubo varias etapas en el diseño opacas, donde no hubo participación de la ciudadanía y que ocasionaron estos problemas de diseño (Ortiz e Iglesias, 2018).

b. Sistemas (ética de los algoritmos): cuando la “evidencia es inescrutable”, es decir, la conexión entre el dato como insumo de la IA y la conclusión de esta no tiene evidencia de cómo se produjo, se produce la violación de una expectativa legítima (Mittelstadt, Allo, Taddeo, Wachter y Floridi, 2016). Esto genera un problema epistémico alrededor de la IA, pues se sabe que produce resultados, pero no cómo los produce (Mittelstadt et al, 2016). Entonces, es necesario saber el proceso que llevó a la IA a tomar la decisión.

Con estos problemas éticos en cuenta, **se propone que el marco ético-filosófico que regirá este marco es híbrido**. Esto en la medida que acoge elementos de la ética deontológica, pues plantea principios inviolables, como el respeto a los derechos fundamentales y a los derechos humanos; asimismo, involucra elementos de la ética utilitaria,

puesto que ofrece la posibilidad de realizar evaluaciones de costo-beneficio para determinar el uso de la IA. La razón de que sea marco ético-filosófico híbrido radica en que los principios no son suficientes para mitigar adecuadamente los riesgos de la IA en Colombia. Esto se debe a que los principios generan problemas a la hora de su interpretación, pues pueden dar espacio a muchos entendimientos dada su generalidad. Así que la herramienta de analizar costo-beneficio es útil para darle entendimientos más inequívocos a estos principios. En consecuencia, los sujetos de la ética que se plantea son las personas involucradas en el diseño, entrenamiento e implementación de la IA, así como en los responsables de los datos que la alimentarán luego de su entrada en funcionamiento.

Es importante aclarar que este es un marco ético que no distingue entre las IA que usan personas de naturaleza pública o privada, pues eso dificultaría la interoperabilidad de los sistemas, lo cual no beneficiaría la innovación que se intenta promover con este marco. (García-León, RA, Solano, EF, Acevedo A. 2018).

OBJETIVOS

1. Fomentar la innovación y el avance tecnológico de las diferentes áreas de la Inteligencia Artificial en el país.
2. Incentivar el uso de la Inteligencia Artificial por parte de las entidades públicas y de las entidades privadas para optimizar costos y facilitar la vida de los ciudadanos.

Principios éticos para la IA en Colombia

1. Respeto a los derechos humanos y los derechos fundamentales (DDFF) consagrados en la Constitución Política de Colombia y los derechos humanos (DDHH) reconocidos internacionalmente.

Ética de Datos	Ética de Algoritmos	Ética de Prácticas
El proceso de recolección de la información para el desarrollo e	Las personas naturales o jurídicas, públicas o privadas, deben	El manejo que se dé a los datos obtenidos para el desarrollo e

implementación de sistemas de IA debe procurar el respeto y la plena satisfacción de los DDFF y los DDHH de los usuarios.	ajustar el proceso creativo, diseño, programación, implementación y uso de la IA a los DDFF y los DDHH que puedan verse vulnerados a partir de la información que suministran los ciudadanos y de las posibles respuestas suministradas por el sistema de IA.	implementación de sistemas de IA debe respetar los DDFF y los DDHH que puedan verse vulnerados a partir de la información que suministran los ciudadanos.
---	---	---

2. El uso de IA debe intentar una mayor efectividad, eficacia y eficiencia en los procesos que pueden ser implementados por el ser humano, pero siempre mediada por este.

Ética de Datos	Ética de Algoritmos	Ética de Prácticas
El proceso de obtención de información debe minimizar la recolección de datos, al punto de que se obtengan aquellos que son estrictamente necesarios para la implementación del sistema de IA.	El desarrollo de la IA debe estar orientado a facilitar actividades y/o procedimientos de los usuarios, en especial aquellos que, gestionados únicamente por el ser humano, implican mayores costos y tiempo para los usuarios.	Debe haber intervención humana para mitigar riesgos y efectuar mejoras en los problemas que pueda presentar un determinado sistema de IA en la práctica.

3. La implementación de la IA debe estar supeditada a unos fines específicos, conforme a la situación que pretende intervenir, de manera que sus responsables puedan prever las posibles contingencias de su implementación y así mitigar los riesgos de la misma.

Ética de Datos	Ética de Algoritmos	Ética de Prácticas
Cada sistema de IA debe enmarcarse dentro de los fines para lo cual fue creada, obteniendo únicamente los datos requeridos para las circunstancias particulares de cada caso y procura la mitigación de los riesgos que dicha implementación implique.	Los algoritmos deben ser desarrollados atendiendo a los posibles riesgos de su implementación, de manera que los mismos se vean mitigados.	

4. Publicidad y transparencia con los usuarios y las autoridades en relación con los fines, métodos, diseño, programación e implementación de los sistemas de IA.

Ética de Datos	Ética de Algoritmos	Ética de Prácticas
Los usuarios deben tener la posibilidad de conocer con exactitud los datos que han suministrado a quienes desarrollan, diseñan e implementan sistemas de IA, y tener control sobre los mismos.	Los elementos estructurales de cada sistema de IA deben ser públicos y puestos a disposición de los usuarios y el público en general, a través de un lenguaje de fácil comprensión para los mismos y de medios de fácil acceso.	

5. Responsabilidad de los desarrolladores, programadores, diseñadores e implementadores de los sistemas de IA frente a las consecuencias, implicaciones y efectos de la misma en relación con sus usuarios.

Ética de Datos	Ética de Algoritmos	Ética de Prácticas
		Los desarrolladores, programadores, diseñadores e implementadores

		de los sistemas de IA deben responder por todos los perjuicios causados a los usuarios y el público en general en los casos en que no haya mitigación de riesgos.
--	--	---

IMPLEMENTACIÓN

Teniendo en cuenta que uno de los objetivos es fomentar la innovación en IA e incentivar la adopción de esta tecnología en diferentes ámbitos, resulta conveniente considerar un modelo de regulación que sea armónico. Es importante reconocer que la regulación privada tiene beneficios que permiten el alcance de estos objetivos, pues es mucho más flexible, permite la libertad y la competencia en términos factibles para los privados (Perritt, 2001). No obstante, un sistema de pura autorregulación de los actores privados puede generar inconvenientes, en la medida que pueden minar los valores democráticos (Perritt, 2001). Así que la mejor forma de alcanzar los beneficios de la regulación privada y, al tiempo, se protege la rendición de cuentas pública que suele asociarse con la regulación y la acción estatal, es un modelo híbrido regulación (Perritt, 2001). Este es un modelo mediante el cual el derecho ofrece un marco amplio y la regulación privada se enfoca en los detalles (Perritt, 2001).

Además, un modelo que se basa en la confianza, como lo es un modelo de regulación privada bajo el marco público, genera los beneficios de una regulación 2.0 (Grossman, 2015). Pues este es un enfoque dinámico, que responde a las interacciones de los actores, y que permite implementar modelos de rendición de cuentas sobre los datos (Grossman, 2015).

CONCLUSIONES

En principio, se estableció que el enfoque de este marco ético de carácter híbrido, en donde el establecimiento de los principios éticos está en manos del sector público y la regulación de detalle está en el sector privado, donde se harán las evaluaciones de costo-beneficio respectiva en la

utilización de los principios. La regulación del sector público se debería ajustar a los principios expuestos antes, cuyos objetivos deberían ser el fomento del desarrollo de la IA en el país, así como incentivar el uso de esta tecnología a lo largo de la industria y el Estado con una base regulatoria flexible como los son los principios. Asimismo, estos principios recogen la experiencia de la IA en otras latitudes, como ocurre con la necesidad de respetar los derechos fundamentales y los derechos humanos, la mediación del ser humano en la IA, la lealtad de los objetivos de diseño en su implementación, la publicidad y la transparencia con los actores involucrados y, finalmente, la responsabilidad de los de los desarrolladores, programadores, diseñadores e implementadores de IA frente a los usuarios y la sociedad.

Como se expuso, es importante reconocer que cada principio sufre de entendimientos distintos de acuerdo a la etapa que se esté observando. Por ejemplo, el principio de respeto a los derechos humanos se puede llegar a entender diferente en una etapa de recolección de datos a como se entendería en una de implementación. Esto se debe a que cada etapa de desarrollo e implementación de IA representa retos distintos que es necesario evaluarlos de acuerdo a su naturaleza, pues de otra forma se puede frenar la innovación y el desarrollo.

REFERENCIAS

- Agwin, J.; Larson, J.; Mattu, S.; y Kirchner, L. (2016) Machine Bias There's software used across the country to predict future criminals. And it's biased against blacks. *Propublica*. Disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Buiten, M. (2019). Towards Intelligent Regulation of Artificial Intelligence. *European Journal of Risk Regulation*, 10(1), 41-59. Disponible en: doi:10.1017/err.2019.8
- Chui, M.; Mnyika, J. y Miremadi, M. (2016) Where machines could replace humans—and where they can't (yet). *McKinsey Quarterly*. Disponible en: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet>

- Ellish, M. y Hwan, T. (2017) An AI pattern language. *New York: Data & Society*. Disponible en: https://www.datasociety.net/pubs/ia/AI_Pattern_Language.pdf
- Følstad, A. y Bae, P. (2017) Chatbots and the New World of HCI. *Interactions*, 24(4), p. 38-42, doi: 10.1145/3085558.
- Isaak, J. y Hanna, M. (2018) User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *The Policy Corner, IEEE Computer Society*. Disponible en: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8436400>
- Hunt, E. (2016) Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*. Disponible en: <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>
- García-León, RA, Solano, EF, Acevedo A. (2018). Caracterización térmica de mezclas de arcillas utilizadas en la fabricación de productos de mampostería para la construcción. *Revista Tecnologías de Avanzada*, ISSN: 1692-7257
- Gasser, U. (2017) AI and the Law: Setting the Stage. *Medium*. Disponible en: <https://medium.com/berkman-klein-center/ai-and-the-law-setting-the-stage-48516fda1b11>
- Gasser, U. (2016) RECODING PRIVACY LAW: Reflections on the future relationship among law, technology, and privacy. *Harvard Law Review Forum*, 130, 2, pp. 60-70. Disponible en: http://harvardlawreview.org/wp-content/uploads/2016/12/Vol.130_Gasser.pdf
- Grossman, N. (2015) White Paper: Regulation, the Internet Way. *Data-Smart City Solutions*. Disponible en: <https://datasmart.ash.harvard.edu/news/article/white-paper-regulation-the-internet-way-660>
- Mittelstadt, B.; Allo, P.; Taddeo, M.; Wachter, S. y Floridi, L. (2016) The ethics of algorithms: Mapping the debate. *Big Data Society*. DOI: 10.1177/2053951716679679
- National Science and Technology Council. (2016). *PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE* (p. 58). Disponible en: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
- Ortiz, J. e Iglesias, C. (2018) Algoritmos e Inteligencia Artificial en Latinoamérica: Un Estudio de implementaciones por parte de Gobiernos en Argentina y Uruguay. *World Wide Web Foundation*. Disponible en: http://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Spanish_Screen_AW.pdf
- Plaza J, Ruiz M, Rosero C, Zapata L. (2017). Formación en competencias específicas para la industria del software colombiano. Experiencias del uso del aprendizaje basado en proyectos. *Revista Tecnologías de Avanzada*, ISSN: 1692-7257
- Perritt, H. (2001) Towards a Hybrid Regulatory Scheme for the Internet. *University of Chicago Legal Forum*, 2001, 1, 8. Disponible en: <http://chicagounbound.uchicago.edu/uclf/vol2001/iss1/8>
- Raso, F.; Hilligoss, H.; Krishnamurthy, V.; Bavitz, C.; Kim, L. (2018) Artificial Intelligence & Human Rights: Opportunities & Risks. *Berkman Klein Center*. Disponible en: <https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights>
- Scripa, A. (2017) Artificial intelligence as a digital privacy protector. *Harvard Journal of Law & Technology*, 31, 1, pp. 217-235. Disponible en: <https://jolt.law.harvard.edu/assets/articlePDFs/v31/31HarvJLTech217.pdf>