

# LA ERA DE LOS ALGORITMOS ¿ENEMIGOS O ALIADOS?<sup>23</sup>

Ariadna Luján Martínez<sup>24</sup>

Francisca Ojeda Campos<sup>25</sup>

## RESUMEN

*Como sociedad estamos pisando terreno desconocido en lo que a nuevas tecnologías se refiere. Las potencialidades de la inteligencia artificial (IA) son innumerables, sin embargo, no son infalibles y los sesgos algorítmicos son uno de los grandes desafíos que presenta. En ese contexto, hemos desarrollado las bases de un protocolo para alcanzar una IA con sello ético y enfoque social.*

*El mayor reto es comprender que la IA, no solucionará automáticamente problemas estructurales. Los daños causados por algoritmos son de responsabilidad humana y la IA es una herramienta que se inserta en un determinado contexto y cuyo funcionamiento y éxito dependerá de las personas.*

*Una IA responsable es posible, pero requiere de educación, la fijación previa de parámetros de uso, y el desarrollo de marcos técnicos, de monitoreo de los sistemas, como también, una dimensión humana, que involucre a los distintos actores sociales que han de beneficiarse con ella.*

**Palabras clave:** Inteligencia artificial – Sesgos – Interdisciplinariedad – Descentralización – Colaboración.

---

23 Enlace al video <https://youtu.be/ESpnji3-kBg>

24 Estudiante avanzada de la carrera de Derecho de la Universidad Nacional de Cuyo. Email: [ariadnalujan41@gmail.com](mailto:ariadnalujan41@gmail.com)

25 Abogada de la Universidad Adolfo Ibáñez, Chile. Email: [fojedacampos@gmail.com](mailto:fojedacampos@gmail.com)

## INTRODUCCIÓN

Como sociedad estamos pisando terreno desconocido en lo que a nuevas tecnologías se refiere. Las potencialidades de la inteligencia artificial (IA) son innumerables, sin embargo, no son infalibles y los sesgos algorítmicos son uno de los grandes desafíos que presenta.

Los más fatalistas ven a la IA como una enemiga, creyendo que su implementación devendrá en el reemplazo de los humanos por las máquinas. Los más escépticos, piensan que aún se trata de ciencia ficción, del futuro. Por el contrario, nosotras creemos y buscamos promover una mirada optimista, que percibe a la IA como una herramienta que complementa y facilita labores humanas.

Hoy, el mayor reto es comprender que la IA, no solucionará automáticamente problemas estructurales. En ningún caso desconocemos los riesgos ni la posibilidad de errores en el desarrollo de la misma, pero el foco ha de dirigirse a los desafíos que se observan, y discutir las posibles soluciones.

Nuestro objetivo es plantear las bases de un protocolo para el uso de sistemas

algorítmicos basados en IA, inspirado en las indicaciones de la Comisión Europea, en las Recomendaciones de la Red Iberoamericana de Protección de Datos, las investigaciones de autores como Juan Corvalán y el trabajo realizado por científicos de datos que también creen posible una IA ética y responsable.

## DESARROLLO

*“La dificultad radica, no en las nuevas ideas, sino en escapar de las viejas”<sup>26</sup>.*

Previo a desarrollar nuestra propuesta de bases para un protocolo, creemos necesario precisar algunos conceptos.

### ¿Qué son los algoritmos?

Son un conjunto de instrucciones o reglas definidas, ordenadas que permiten solucionar un problema, realizar un cómputo, procesar datos y llevar a cabo tareas o actividades, siguiendo pasos sucesivos, llegando a un resultado final (obtener patrones). Estos algoritmos, son entrenados con datos, y tal como afirma Corvalán (2020) permite la vinculación entre algoritmos y Machine learning: El término *aprendizaje automático* se refiere

---

26

a la detección automatizada de patrones significativos en los datos.

Es importante remarcar que cuando hablamos de sesgos algorítmicos debemos saber que los mismos reflejan los valores de los humanos que están implicados en la codificación, recolección de datos para entrenar al algoritmo.

Por lo expuesto, es que creemos que cualquier desarrollo responsable de la inteligencia artificial, debe contemplar, un ámbito humano y un ámbito técnico.

Como mencionamos anteriormente, siguiendo los lineamientos del “Libro Blanco sobre Inteligencia Artificial” de la Comisión Europea y las Recomendaciones Generales de la Red iberoamericana de protección de datos personales, ambos documentos publicados durante el 2020, es que llegamos a las siguientes proposiciones:

### **Ámbito Humano o Ético**

Los sistemas son realizados por humanos, por tanto, el avance y desarrollo de la IA debe ir necesariamente acompañado de una formación tanto de quienes crean los sistemas, como de quienes toman las decisiones de implementación, y de los destinatarios o usuarios. Formar equipos

con un alto grado de conciencia en materia de ética es fundamental, de manera de fomentar el pensamiento crítico y cuestionador, y así poder crear sistemas que sean coherentes con los valores e ideas de la sociedad actual.

Por otro lado, los seres humanos naturalmente tenemos sesgos; para disminuir esto y poder avanzar en la educación y formación digital debemos formar un ecosistema de IA que acerque las ventajas de la tecnología a la sociedad, lo que debe hacerse desde distintos ámbitos y considerando los diversos actores que pueden estar involucrados.

Desde el sector público, se debe trabajar para asegurar el acceso de datos y a la infraestructura informática. Además es necesario un marco político que fomente la investigación y la innovación. Promover la educación digital desde los primeros niveles de educación hasta los universitarios, acompañado de regulación o protocolos que sean flexibles, ya que la necesidad de adaptación y re adecuación será constante con los avances que existen día a día.

Desde el sector privado, es necesario que el mismo pueda participar en la agenda de innovación e investigación (según sugiere

el Libro Blanco) lo cual creemos que es efectivamente necesario. Pero además contar con participación ciudadana, facilitando el acceso a la información y que la misma sea clara, promoviendo la transparencia de los algoritmos, la consulta pública a la hora de hacer regulaciones y la promoción de softwares de código abierto.

Interdisciplinariedad<sup>27</sup>: En palabras simples, significa que en el proceso de desarrollo de sistemas algorítmicos exista una “combinación de actores, elementos y valores de distintas áreas del conocimiento” (Cevasco, Corvalán y Fevre, 2019). Este es el factor clave de nuestra propuesta, sin colaboración, sin aportes de distintas áreas todo lo demás no será exitoso.

### Ámbito Técnico<sup>28</sup>

El ámbito humano debe necesariamente complementarse con protocolos de funcionamiento y revisión de los sistemas, lo que hemos denominado área técnica, es decir, a realizarse en el proceso de

<sup>27</sup> La Real Academia Española lo define como “Dicho de un estudio o de otra actividad: Que se realiza con la cooperación de varias disciplinas”.

<sup>28</sup> Como mencionamos anteriormente, el objetivo de este documento es hacer una breve reseña de nuestra disertación en el Congreso, por lo cual para ampliar conocimientos humildemente los dirigimos al canal de youtube “ElDerechoinformativo TV” donde podrán profundizar lo que aquí tratamos.

desarrollo del algoritmo o en el post procesamiento del mismo.

Elegimos dos sistemas de auditoría de modelos de machine learning, enfocados en reconocer y mitigar los sesgos que pueden introducirse en los sistemas de diversas formas y en distintas etapas del desarrollo del modelo.

#### a)

Es un paquete de Python (lenguaje de programación) desarrollado por Microsoft, que implementa algoritmos para detectar problemas de equidad de grupos de modelo de aprendizaje automático y a su vez mitigar injusticia. Se pueden utilizar algoritmos para mejorar el modelo durante el entrenamiento, pero también, algoritmos de post-procesamiento para mejorar modelos ya existentes. Funciona a través de un panel de evaluación para evaluar cómo las predicciones de un modelo afectan a distintos grupos y a su vez, utiliza algoritmos de mitigación para disminuir la injusticia en la clasificación. Además mide la dependencia de un modelo de sus entradas cambiandolas, si un cambio provoca drásticamente una modificación en la salida entonces el modelo es sensible a la característica.

<sup>29</sup> Más detalles en <https://fairlearn.github.io/>

## b) Modelo aequitas<sup>30</sup>:

Es una herramienta desarrollada por el Center for Data Science and Public Policy for Social Good Foundation, originalmente en la Universidad de Chicago, y actualmente en la Universidad de Carnegie Mellon en EE.UU. la misma es de código abierto<sup>31</sup> y permite auditar modelos de machine learning para detectar y mitigar sesgos, respecto de múltiples atributos (etnicidad, sexo, edad, etc.). Se basa en una definición de sesgo entendida como medida de disparidad entre grupos, en comparación con un grupo de referencia. A su vez, el grupo de referencia, se puede seleccionar utilizando varios criterios (los mencionados más arriba).

Lo distintivo de Aequitas es que tiene presente que sesgo y “fairness” no son conceptos absolutos y que están necesariamente vinculados al escenario al que se aplican, como asimismo al análisis y a la interpretación dentro de ese contexto. Por otro lado, pretende ser útil también a autoridades en general y no limitarse a ser entendido únicamente por personas con conocimiento técnico.

<sup>30</sup> La explicación que viene a continuación, se basa en el documento desarrollado por Saleiro, Kuester, Stevens, Anisfield, Hinksos, London, Ghani, (2018).

<sup>31</sup> Más detalles en: <https://github.com/dssg/aequitas>

Una explicación más detallada del funcionamiento tanto del modelo Fair ML como de Aequitas excede el objetivo de este documento, por lo cual, para aquellos que deseen profundizar en la materia, se recomienda acceder directamente a las fuentes. Lo importante es difundir su existencia y dar cuenta de que existe un interés y preocupación por desarrollar inteligencia artificial responsable con sello ético y en beneficio de las personas.

## CONCLUSIÓN

Es imperativo que la implementación de sistemas de IA vaya acompañada de discusiones y capacitaciones en el ámbito humano y de auditorías a los modelos, como alguno de los que mencionamos previamente, de forma de lograr una IA que sea a la vez útil, responsable y con enfoque social.

En este sentido, resulta acertado lo que expresa Kate Crawford, de Microsoft Research, “es hora de reconocer que los algoritmos son una creación humana que hereda nuestros prejuicios [...]: nuestra IA será tan buena como lo seamos nosotros”.

Creemos firmemente que esta afirmación va de la mano de educación ética y digital de los ciudadanos, y distintos actores sociales, como mencionamos, de estados

PROACTIVOS, que no se resistan a lo que ya está sucediendo sino que desde la interdisciplinariedad y la escucha de todos los sectores busquen soluciones de las que TODAS las personas participemos. Que la IA sea potencialmente destructiva o potencialmente beneficiosa, creemos está en manos de nosotros.

## BIBLIOGRAFÍA

- Cevasco, L., Corvalán, J., Le Fevre, E, "Inteligencia Artificial y Trabajo. Construyendo un nuevo paradigma de empleo". Astrea SRL, Buenos Aires, 2019.

- Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnología (Comisión Europea), "Libro Blanco sobre Inteligencia Artificial – Un Enfoque europeo orientado a la excelencia y la confianza", Bruselas, 2020.

- Corvalán, J, "Perfiles digitales humanos". La Ley, Buenos Aires, 2020.

- Red Iberoamericana de Protección de Datos – RIPD-, "Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial", 2020.

- Saleiro, P., Kuester, B., Stevens, A., Anisfeld, A., Hinksos, L., London, J., Ghani, R, "Aequitas: A Bias and Fairness Audit Toolkit". Center for Data Science and Public Policy, University of Chicago, Chicago, 2018.