

**03.**

Doctrina

Inteligencia artificial y sesgos

**Revista de la Escuela Judicial:** ISSN 2796-874X

**Año:** 02/Nº 3 - Noviembre 2022

**Recibido:** 6/10/2022

**Aprobado:** 20/11/2022

# Inteligencia artificial y sesgos

*Artificial Intelligence and Bias*

**Por Noelia Giselle Bertocchi<sup>1</sup>**

Universidad de Buenos Aires

**Resumen:** En este artículo se plantea cómo funciona la inteligencia artificial (IA) y su aplicación en el ámbito judicial. Se hace mención del colapso del Poder Judicial y cómo lograr que la IA colabore en la mejora del acceso a la justicia. Se hace enfoque en el problema de los sesgos en la IA, cómo impactan, cómo identificarlos y cómo evitar que se trasladen a las resoluciones judiciales.

**Palabras clave:** Inteligencia artificial – Sesgo algorítmico – Resoluciones judiciales.

---

1 Abogada (Universidad de Buenos Aires). Especialista en Administración de Justicia. Auxiliar letrada en Juzgado de Familia N° 2 del Departamento Judicial de Moreno - General Rodríguez. Egresada (eximida) de la Escuela Judicial de la provincia de Buenos Aires. Correo electrónico: noeliabertocchi@gmail.com. Identificador ORCID: <https://orcid.org/0000-0001-6842-6022>.

**Abstract:** *In this article we focus on how Artificial Intelligence (AI) works and its application in the judiciary system. This article provides a brief overview of how The Judiciary Branch is collapsed, and how to achieve Artificial Intelligence to collaborate in improving the access of Justice. A debate about biases in IA, how they affect the way it works; how to identify biases, how to address and remedy them in order to prevent them from being transferred to judicial resolutions.*

**Keywords:** *Artificial Intelligence – Algorithmic bias – Judicial Resolutions.*

## Qué es la inteligencia artificial

En términos simples, la inteligencia artificial (IA) se refiere a sistemas o máquinas que imitan la inteligencia humana para realizar tareas y pueden mejorar iterativamente a partir de la información que recopilan.

Muchos tenemos un concepto más ligado a la ciencia ficción de lo que es la IA, con robots humanoides que reemplazan en prácticamente todas las áreas al humano, e incluso complotan para apoderarse del mundo. Sin embargo, lejos está de eso. Su objetivo es mejorar significativamente las capacidades y contribuciones humanas.

Podría decirse entonces, dentro del amplio espectro de lo que abarca la IA, que es un programa computacional que aprende, a partir de datos que se le incorporan de manera repetitiva (iteración), a crear patrones. Es decir, un sistema de aprendizaje automático. A partir de estos patrones, genera respuestas o soluciones a distintas cuestiones que se le plantea al sistema. Mientras más datos se le incorporen, más aprende, lo que a su vez permite que realice predicciones más acertadas.

Estos datos pueden ser de todo tipo (textos, sonidos, imágenes, etcétera). Es información que le incorporamos los humanos a la máquina, información que consideramos necesaria para que el sistema de IA pueda resolver el problema o situación que le planteemos. También se le irán incorporando nuevas variables a medida que el sistema trabaje. Con estas nuevas variables, los datos ya cargados y

los patrones creados, sus predicciones serán más acertadas. Y estas predicciones a su vez funcionan como datos que el sistema almacena para futuras predicciones. Es una rueda de conocimiento, aprendizaje y mejoría constante.

La IA o sistema de aprendizaje automático resulta una herramienta muy tentadora para implementar en el Poder Judicial, dado que hoy en día nuestra Justicia no da abasto para resolver en tiempo y forma la enorme cantidad de causas en trámite. Nos planteamos si una alternativa que incorpore un sistema inteligente que reduzca tiempos y mejore la eficiencia de los resultados obtenidos puede ser implementada y sostenida en el tiempo en la Justicia.

Analizaré brevemente el impacto de la inteligencia artificial en el ámbito judicial, sus beneficios y dificultades a tener en cuenta al momento de su aplicación. Específicamente me centraré en los sesgos y cómo estos pueden surgir a partir de los datos que se le incorporen al sistema de aprendizaje automático; cómo detectarlos y evitar que se trasladen a la resolución de causas por medio de IA.

## **Colapso en la Justicia**

Como decía en el apartado anterior, no es ninguna novedad que actualmente el acceso a la justicia se ve mayoritariamente perjudicado por el gran caudal de causas en trámite. Esto hace que el Poder Judicial y el trabajo de los agentes judiciales sea muy criticado y poco valorado por la falta de eficiencia al momento de resolver.

En el ámbito judicial la IA puede ser utilizada para mejorar el servicio. De hecho, ya se han creado y puesto a prueba diversos sistemas con la finalidad de automatizar tareas de índole más mecánica. Un ejemplo es el sistema Prometea, creado en 2017 y puesto a prueba por la Fiscalía de la Ciudad Autónoma de Buenos Aires.

Prometea aplica IA para preparar automáticamente dictámenes judiciales. La herramienta le ha permitido a la Fiscalía incrementar la eficiencia de sus procesos de manera significativa: una reducción de noventa minutos a un minuto (99 %) para la resolución de un pliego de contrataciones; de 167 a 38 días (77 %) para procesos de requerimiento a juicio; de 190 a 42 días (78 %) para amparos habitacionales con citación de terceros, entre otras. Esta ganancia permitió que los empleados y funcionarios dedicados a realizar las tareas automatizadas pudieran dedicar más tiempo a casos más complejos que requieren un análisis más profundo, mejorando la calidad de sus dictámenes (Estevez, Fillottrani & Linares Lejarraga, 2020).

Una de las posibilidades que brinda la inteligencia artificial es la de automatizar determinados procesos a fin de que en última instancia solo sea necesaria una revisión superficial por parte del humano. Por ejemplo, al permitir que una máquina tome el rol principal de proveer determinadas causas cuyas resoluciones son de índole mecánica y no requieren un desarrollo o enfoque profundo, sino que conllevan un proceso rutinario. Existen causas cuyo proceso se resuelve de manera automatizada aun a nivel humano y no precisa de un estudio tan riguroso. Hoy en día, en los juzgados y tribunales los agentes judiciales son quienes llevan causas en su totalidad, sean de la materia que sean –acorde al fuero, claro-. Si bien hay materias que precisan de una mirada mucho más atenta

y tiempo y dedicación, sobre todo en temas de gran sensibilidad, como las cuestiones familia, hay determinados trámites judiciales que no y son resueltos por los mismos agentes, lo que hace que estos deban destinar valioso tiempo en algo que podría ser proyectado o resuelto por un sistema de aprendizaje automático.

Al ahorrar tiempo y permitir que haya más personal judicial disponible para cuestiones más delicadas o sensibles que además requieren una respuesta rápida, se logra que, cuando se llegue a un fallo, el justiciable tenga verdadero acceso a la justicia.

Lograr esto implica estudiar el modo en que se trabaja una causa, los mecanismos de proveimiento a partir de determinadas pautas, cómo un juez toma las decisiones, cómo resuelve causas, e identificar cuáles son los datos necesarios para incorporar a una máquina que le permitan actuar de manera adecuada y así llegar a una resolución eficaz. Es decir que hace falta un trabajo muy minucioso que requiere que haya interdisciplinariamente, entre letrados, magistrados, programadores, especialistas en la materia que se pretende aplicar a la máquina, un debate profundo sobre la información que se le pretende incorporar al sistema de aprendizaje automático. No basta solamente con darle Códigos, la Constitución nacional y algunas leyes. La máquina debe cumplir con ciertos recaudos para poder resolver conforme a derecho. Para esto será necesario también que pueda dilucidarse qué valores se pretende que pueda aprender y generar patrones con ellos. Así, su decisión no solo será hacer valer lo que dice la norma, sino también que haya un análisis ético detrás. ¿Será posible esto en una máquina?

Antes de pasar al siguiente apartado, y si bien no es objeto de este artículo, creo importante mencionar que más allá de los beneficios y ventajas que trae aparejada la IA en el Poder Judicial, existe parte de la sociedad en general y de agentes judiciales en particular que aún hoy desconfían de esta tecnología. Por desconocimiento o aprehensión, opinan que una máquina no podría suplantar a un ser humano dictaminando una sentencia, cuando en realidad el objeto del sistema de IA, al menos hoy, es básicamente colaborar con la Justicia reduciendo los tiempos, agilizando tareas y facilitando el trabajo de los juzgados o tribunales.

## **Datos y sesgos en la inteligencia artificial**

Para funcionar, la IA se nutre de datos a partir de los cuales crea patrones para luego utilizarlos. Los datos que se incorporen a estas máquinas o programas inteligentes serán la fuente a partir de la cual el sistema generará respuestas o predicciones. El sistema de IA aprende a través de las iteraciones, o sea, de las repeticiones de tareas que hará a partir de los datos que se le incorporen, de los patrones que genera, que irá puliendo con la información generada y mejorará sus respuestas; ergo, sus predicciones serán más acertadas.

Ahora bien, quiero hacer especial hincapié en que estos datos a partir de los cuales el sistema trabaja son incorporados, de hecho, por humanos. En otras palabras, el humano es quien arbitrariamente selecciona la información administrada a la máquina para que elabore un sistema automatizado. Y aquí es cuando surge un debate muy

interesante acerca de un elemento que puede ocasionar muchos problemas: el sesgo.

Los sesgos pueden ser definidos como una orientación o un error sistemático en el que se puede incurrir deliberadamente favoreciendo o perjudicando determinadas respuestas respecto de otras. Pueden identificarse en los sistemas de IA y es el humano quien los incorpora. ¿De qué manera lo hace? A través de los datos que les da, ya que por medio de los mismos puede transferir ciertos valores, orientaciones, tendencias, modas u opiniones que directa o indirectamente son potenciales amenazas ya que pueden crear respuestas discriminatorias, parciales o tendenciosas. Algo alejado de lo pretendido por el sistema de IA, lo cual es su gran paradoja.

Por ejemplo, uno podría tener una mirada simplista y pensar en principio que un futuro juez robot sería totalmente imparcial puesto que no tendría sentimientos, valores humanos o antecedentes familiares, o no podría sufrir ningún tipo de crisis o angustia personal que pudiera eventualmente hacer que sus decisiones fueran sesgadas o tendientes a resolver de tal o cual modo. Sin embargo, los sistemas de aprendizaje automático tienden a reflejar sesgos.

Existen sesgos de todo tipo. En este artículo nos enfocaremos en los cognitivos y los algorítmicos.

El concepto de sesgo cognitivo fue introducido por los psicólogos israelíes Daniel Kahneman y Amos Tversky en 1972. Este tipo de sesgos puede influir en la forma que vemos el mundo. "Están determinados por implicaciones culturales, influencia social, motivaciones emocionales o éticas, atajos en el procesamiento de la información,

o distorsiones en la recuperación de los recuerdos y la memoria, entre muchos otros”.<sup>2</sup>

Se trata de sesgos implícitos propios del procesamiento de información, tanto sensorial como conceptual, que afectan a todos los seres humanos en todos los ámbitos de la vida, incluso cuando toman decisiones difíciles e importantes (Conlisk, 1996), incluyendo las decisiones judiciales (Guthrie et al., 2001). (Páez, 2021, p. 1)

Por su parte, los sesgos algorítmicos se dan cuando un sistema de aprendizaje automático refleja los valores de las personas que lo desarrollaron o entrenaron.

Friedman y Nissenbaum (1996) fueron pioneros en el análisis del sesgo en los sistemas informáticos, bajo el argumento de que dichos sistemas presentan sesgos si “discriminan de forma sistemática e injusta a ciertas personas en favor de otras [negando] la oportunidad de obtener un bien o [asignando] un resultado indeseable a una persona o grupo de personas por motivos irracionales o inapropiados. (Simon, Wong & Rieder, 2022, p. 6)

¿Y cómo identificamos estos sesgos? A través de los datos. Estos

juegan un rol esencial en el entrenamiento de sistemas por medio de aprendizaje automático, dado que son la fuente de información que le indicará al sistema cuándo ha llegado a conclu-

---

2. Disponible en: <https://www.braininvestigations.com/neurociencia/sesgo-cognitivo-negocios/#:~:text=Un%20sesgo%20cognitivo%20es%20una,Kahneman%20y%20Tversky%20en%201972>.

siones correctas y cuándo no. Algo que resulta fundamental en este proceso, y que no siempre es tenido en cuenta, es que un sistema raramente se construye para realizar predicciones con los datos con que fue entrenado. Por el contrario, se espera que los modelos puedan sacar conclusiones acertadas sobre datos nunca vistos durante el “aprendizaje” –los datos de prueba– y cuyas etiquetas no se conocen.<sup>3</sup>

Entonces, sin perjuicio de los datos ingresados, el entrenamiento realizado con la incorporación de nuevos datos puede potencialmente provocar que el sistema realice predicciones sesgadas. En consecuencia, encontrar la manera de identificar estos sesgos y eliminarlos es parte del proceso de introducción de los sistemas de aprendizaje automático a la Justicia.

Los sistemas de inteligencia artificial están diseñados por personas con sus propias visiones del mundo, prejuicios, valoraciones de los hechos y sesgos adquiridos a lo largo de su experiencia de vida, que pueden filtrarse en el diseño y la definición de criterios de evaluación para estos modelos. Si esos grupos de trabajo no son lo suficientemente diversos como para reflejar una amplia variedad de visiones, muy probablemente no lleguen siquiera a darse cuenta de la existencia de los sesgos, y por tanto a corregirlos.<sup>4</sup>

Los jueces, fiscales, letrados, y toda persona humana en el mundo tienen un bagaje cultural, experiencias de vida, opiniones formadas,

---

3. Disponible en: <https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>.

4. *Ibid.*

educación e instrucción determinada que les ha dado la vida misma, sus familias, sus amigos, parientes, trabajos y demás. Todo su contexto y estilo de vida, vivencias, experiencias, son parte de los sesgos cognitivos.

## **¿Cómo combatir los sesgos?**

Los sesgos se presentan ante la incorporación de determinados valores implícitos en los datos que se trasladan al sistema de IA. Estos datos crean patrones que replican en el sistema y reflejan resultados que pueden ser tendenciosos o discriminatorios.

Veamos. Al sistema de IA se le incorporarán variables –datos– para que resuelva una causa determinada. El sesgo aparece ante la imposibilidad del sistema o mecanismo de aprendizaje automático de plasmar un resultado equitativo. Esto es porque interpreta de manera errada la relación entre las variables que se le incorporan y la realidad. Es decir que la solución a la que arriba lejos está de ser justa, sino que está parcializada. La medida en la que un modelo verdaderamente refleja la relación de las variables en una población indica qué tan sesgado está.

Pero antes de poder hallar respuestas a cómo combatir estos sesgos debemos saber cómo identificarlos.

A simple vista, pareciera que todo se reduce a identificar aquellos resultados que no se corresponderían a nivel jurídico con una resolución imparcial. Esto conlleva obviamente un control humano de cada resolución que se dictamina. Pero en muchos casos los sesgos

son tan sutiles que resulta una tarea mucho más engorrosa de lo que creemos. Por ello es necesario hacer tareas de control o monitoreo regular, no de los resultados obtenidos, sino de los datos cargados para que el sistema funcione. Estos deberían ser relevados por un equipo interdisciplinario que rastrear e identificara los que pudieran trasladar sesgos al sistema que trajeran aparejadas decisiones discriminatorias o parciales.

Mejor aún sería que los mismos grupos de trabajo que programasen los sistemas de aprendizaje automático se encontrasen capacitados para identificar potenciales sesgos en los datos que se utilizan para la creación de patrones.

La supervisión que se realice del sistema de IA tiene que ser integral. Es vital que haya diversidad en los equipos de trabajo que desarrollen y posteriormente supervisen los sistemas. Lamentablemente, hoy en día sigue habiendo poca representatividad de distintas minorías, lo que provoca que no siempre sea tarea fácil identificar sesgos en los sistemas que se producen; incluso que se adviertan formas de evitarlos.

Más allá del control que realice el creador del sistema, nosotros como sociedad también estaremos obligados a supervisarlos. Por muy tentador que suene delegar en una máquina tareas aburridas o mecánicas, no podemos dejar de controlar o monitorear el –buen– funcionamiento de estos sistemas.

Muchos ignoran la velocidad con la que la IA avanza en nuestra vida en diversos ámbitos. Es nuestro deber capacitarnos e informarnos

y ser parte del proceso. Pero, por sobre todas las cosas, aceptar que la IA llegó para quedarse.

## **Consideraciones finales**

Los sesgos son parte de la inteligencia artificial. Como vimos, el sistema de IA nace, se desarrolla y aprende a partir de datos, y estos datos provienen de humanos, quienes a través de variables los incorporan para que el sistema resuelva. Pero estas resoluciones no son infalibles.

Ya dijimos que las predicciones que brinda la IA tienen que ser revisadas y supervisadas periódicamente, ya que los sesgos pueden resultar peligrosos. Por ejemplo, dando resultados discriminatorios o tendenciosos.

Siendo que la IA y su implementación en distintas áreas de trabajo son relativamente nuevas en nuestro país, particularmente en el ámbito judicial, resulta importante que se plantee lo antes posible legislar sobre las mismas. Que se creen marcos regulatorios acordes, que se planteen y debatan las diversas problemáticas que puedan surgir. Sobre todo, plantear la cuestión de los sesgos. Debatir y concientizar sobre los beneficios que la IA trae a nuestra sociedad y cómo debemos asumir el rol de “supervisores” de su funcionamiento.

Este proceso es continuo. Corregir los sesgos que se identifiquen es una tarea necesaria a fin de evitar que se llegue a resoluciones discriminatorias.

Sin perjuicio de que sea probable que las nuevas tecnologías y el afán de evitar “perder tiempo” en la IA hagan que se creen nuevos sistemas que se autorregulen, es importante no perder el contacto humano a fin de que se pueda vigilar que el autocontrol que se está llevando a cabo es eficiente en identificar potenciales sesgos.

Asimismo será importante que quienes creen estos sistemas estén capacitados lo suficiente como para saber separar sus propios sesgos de lo que programen. Si bien será prácticamente imposible eliminar por completo lo que se incorpore inconscientemente a estas creaciones, que se pueda cambiar la perspectiva e identificar los potenciales sesgos o errores o tendencias para que no sean trasladados a lo programado será mucho más fácil si nos capacitamos para ello.

Me pareció interesante el enfoque abordado por Simon, Wong y Rieder (2022) cuando hacen hincapié en el diseño de algoritmos de manera consciente. Esto es, abordándolos de manera responsable y con atención a los valores humanos, teniendo presente que “las técnicas algorítmicas pueden ‘afectar la suerte de grupos enteros de personas de forma sistemáticamente desfavorable’ (Barocas y Selbst, 2016, p. 673)” (p. 3).

Si bien no vamos a profundizar en el tema, en líneas generales lo que propone el diseño sensible al valor es definir los valores humanos “como ‘lo que las personas consideran importante en su vida, con énfasis en la ética y la moral’ (Friedman y Hendry, 2019, p. 24)” (p. 4). Asimismo, que se haga valer la presencia de grupos de personas que se suelen omitir por no ser usuarios en el diseño de la tecnología pero que sin embargo se ven afectados por ella.

Este sistema de diseño propone una herramienta analítica que abra el debate respecto de la valoración en el diseño y desarrollo de la tecnología que se incorpora a los sistemas de inteligencia artificial, que tienden a ser ignorados o descuidados. Asimismo, proporciona una herramienta constructiva que permite y apoya la realización de valores específicos deseados en el diseño y desarrollo de nuevas tecnologías. Para explicarlo de manera más simple: la idea sería que cuando se diseñase un sistema de aprendizaje automático no fuera solo con miras a mejorar la eficiencia del sistema judicial, sino que también se considerasen las cuestiones que tienden a ser omitidas. La omisión se suele dar porque los equipos de trabajo que diseñan estos sistemas están compuestos por personas que no consideran determinados valores humanos necesarios para que el sistema propenda a que las resoluciones judiciales que dicte sean ajustadas a derecho. De ahí que se pretenda que sean interdisciplinarios y haya más heterogeneidad.

A esto último sumémosle que

no solo es necesario ampliar el alcance de las perspectivas disciplinarias, sino también el del propio objeto de investigación. Es decir, en lugar de limitarse a la equidad, la responsabilidad y la transparencia en el aprendizaje automático, la investigación sobre el sesgo algorítmico también debería considerar el sistema sociotécnico más amplio en el que se inscriben las tecnologías y [...] las diferentes lógicas y órdenes que estas tecnologías algorítmicas producen y generan. (Simon, Wong & Rieder, 2022, p. 11)

Para ir terminando, quiero dejar aclarado que los sesgos requieren un enfoque mucho más profundo del que se les dio aquí, porque

abarcan una enorme cantidad de información y perspectivas que no plasmó en su totalidad este espacio. Creo que merecen un estudio pormenorizado que no se circunscriba a una mirada superficial, más si tenemos en cuenta que son responsables de lo limitado o ilimitado, imparcial o parcial que pueda ser el sistema de inteligencia artificial.

La incorporación de inteligencia artificial al derecho es objeto constante de debate. Por supuesto que un enfoque interdisciplinario es necesario para mejorar la manera en que se pretende incorporar un sistema automatizado propiamente dicho a la Justicia. Estamos en constante estudio del modo en que queremos que estos sistemas aprendan, evolucionen y trasladen a un sistema automático lo que queremos que sea la Justicia al momento de dictar una resolución. El modo en que el humano desarrolle y mejore estos sistemas será reflejo de la Justicia del mañana.

## Bibliografía

- DANESI, C.** (2021). “Procesamiento de datos, sesgos e Inteligencia Artificial”. Recuperado de: <https://www.youtube.com/watch?v=-fS6d4Xnnz3A>.
- ESTEVEZ, E., P. FILLOTTORANI & S. LINARES LEJARRAGA** (2020). “PROMETEA: Transformando la administración de justicia con herramientas de inteligencia artificial”. Recuperado de: <http://dx.doi.org/10.18235/0002378>.
- FERRANTE, E.** (2021). “Inteligencia artificial y sesgos algorítmicos. ¿Por qué deberían importarnos?”. En: *Nueva Sociedad*, N° 294. Recuperado de: <https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>.
- KAHNEMAN, D. & A. TVERSKY** (1972). “Subjective probability: A judgment of representativeness”. Recuperado de: [https://doi.org/10.1016/0010-0285\(72\)90016-3](https://doi.org/10.1016/0010-0285(72)90016-3).
- PAEZ, A.** (2021). “Los sesgos cognitivos y la legitimidad racional de las decisiones judiciales”. En ARENA, F. J., P. LUQUE & D. MORENO CRUZ (eds.), *Razonamiento Jurídico y Ciencias Cognitivas*. Bogotá: Universidad Externado de Colombia. Recuperado de: <http://dx.doi.org/10.2139/ssrn.3956986>.
- SIMON, J., P.-H.WONG & G. RIEDER** (2022). “El sesgo algorítmico y el enfoque del diseño sensible al valor”. En: *Revista Latinoamericana de Economía Y Sociedad Digital*. Recuperado de: <https://doi.org/10.53857/tzvn9229>.