

**Inteligencia artificial y sesgos.  
El caso de la predicción del embarazo adolescente en Salta \***

**Inteligência artificial e vieses.  
O caso da previsão de gravidez na adolescência em Salta**

***Artificial Intelligence and Biases:  
Prediction of Adolescent Pregnancy in Salta***

**Karina Pedace , Tobías Schleider  y Tomás Balmaceda\*\***

Frente a la retórica instalada en clave de una inteligencia artificial (IA) eficiente y libre de prejuicios, nuestra propuesta en este artículo es argumentar por qué esta visión tan popular de la IA es inadecuada y examinar algunas de sus derivas en el trazado de políticas públicas. En primera instancia, argumentaremos que es imposible –por principio– plantear una IA valorativamente neutra, a partir de lo señalado por algunos aportes filosóficos tanto canónicos como contemporáneos. En la segunda sección analizaremos un caso que ilustra de modo diáfano la carga valorativa que gravita en la IA. Esto nos llevará a indagar, en la tercera sección, los diferentes niveles en los cuales los sesgos pueden tener injerencia en la IA. Nuestro objetivo central es explicitar la paradoja consistente en cómo la aplicación de IA en políticas públicas y en la toma de decisiones en ámbitos privados, lejos de ser una herramienta para combatir desigualdades estructurales, termina consolidando escenarios perjudiciales para poblaciones vulnerables.

9

**Palabras clave:** inteligencia artificial; algoritmo de aprendizaje automatizado; sesgos; neutralidad valorativa; políticas públicas

---

\* Recepción del artículo: 30/06/2021. Entrega de la evaluación final: 04/01/2022.

\*\* *Karina Pedace*: profesora adjunta en la Universidad Nacional de La Matanza (UNLAM), jefa de trabajos prácticos en la Universidad de Buenos Aires (UBA) e investigadora en IIF-SADAF-CONICET, Argentina. Correo electrónico: karinapedace@gmail.com. *Tobías Schleider*: profesor titular e investigador en la Universidad Nacional del Sur (UNS) y jefe de trabajos prácticos en la Universidad Nacional de Mar del Plata (UNMDP), Argentina. Investigador en UNMDP, UNS y el Instituto Latinoamericano de Seguridad y Democracia (ILSED). Correo electrónico: tschleider@gmail.com. *Tomás Balmaceda*: jefe de trabajos prácticos en la Universidad de Buenos Aires (UBA), Argentina. Profesor asociado en la Universidad de Ciencias Empresariales y Sociales (UCES), Argentina. Investigador en IIF-SADAF-CONICET, Argentina. Correo electrónico: tomasbalmaceda@gmail.com.

Diante da retórica instalada em busca de uma inteligência artificial (IA) eficiente e livre de preconceitos, nossa proposta neste artigo é argumentar por qual motivo é inadequada essa visão tão popular da IA. Além disso, examinaremos alguns de seus desvios na elaboração de políticas públicas. Em um primeiro momento, argumentaremos que é impossível –por princípio– propor uma IA valorativamente neutra, com base no que foi indicado por algumas contribuições filosóficas tanto canônicas como contemporâneas. Na segunda parte, analisaremos um caso que ilustra de modo diáfano a carga de valor que gravita na IA. Isso nos levará a investigar, na terceira parte, os diferentes níveis nos quais os diversos vieses podem influenciar a IA. Nosso objetivo central é explicitar o seguinte paradoxo: a aplicação da IA nas políticas públicas e na tomada de decisões na esfera privada, longe de ser uma ferramenta de combate às desigualdades estruturais, acaba consolidando cenários prejudiciais para as populações vulneráveis.

**Palavras-chave:** inteligência artificial; algoritmo de aprendizado de máquina; vieses; neutralidade de valores; políticas públicas

*Faced with the rhetoric of an efficient and free of prejudice artificial intelligence (AI), this article discusses why this popular vision is inadequate and examines some of its shifts in the design of public policies. Firstly, we argue that it is impossible –in principle– to propose a value-neutral AI, based on what has been indicated by some canonical and contemporary philosophical contributions. In the second section we analyze a case that clearly illustrates the value burden that gravitates on AI. This will lead us to investigate, in the third section, the different levels at which biases can have an influence on AI. Our main objective is to unveil the paradox related to how the application of AI in public policies and private spheres, far from being a tool to combat structural inequalities, ends up consolidating harmful scenarios for vulnerable populations.*

10

**Keywords:** artificial intelligence; machine learning algorithm; biases; value neutrality; public policies

## Introducción

La gran mayoría de las aplicaciones de inteligencia artificial (IA) se basa en un método conocido como “aprendizaje automatizado” (*machine learning*), cuyos algoritmos están especializados en encontrar patrones en los datos. A pesar de la enorme utilidad de algunos de los usos de este enfoque, otros pueden tener un impacto muy negativo en la vida de las personas: pueden exacerbar (y aun perpetuar) desventajas sociales preexistentes que se reflejen, por ejemplo, en la injusticia de una contratación laboral, en el otorgamiento crediticio, en el acceso a la salud y a la vivienda, en el sistema de justicia penal y en la seguridad, entre otros ámbitos. Debemos estar atentos, entonces, a si los algoritmos conducen a resultados discriminatorios. Es lo que sucede, por ejemplo, cuando identifican correlaciones y hacen predicciones acerca de la conducta de un determinado grupo en una comunidad basados en atributos sensibles como género o grupo étnico.

Frente a la retórica instalada en clave de una IA eficiente y libre de prejuicios, nuestra propuesta en este trabajo es argumentar por qué esta visión tan popular de la IA es inadecuada y examinar algunas de sus derivas en el trazado de políticas públicas. En primera instancia, sostendremos que es imposible -por principio- plantear una IA valorativamente neutra, a partir de lo señalado por algunos aportes filosóficos tanto canónicos como contemporáneos. En la segunda sección analizaremos un caso en el que se ilustra de modo diáfano la carga valorativa que gravita en la IA. Esto nos llevará a explicitar, en la tercera sección, los diferentes niveles en los cuales los sesgos pueden tener injerencia en la IA. Nuestro objetivo central es explicitar la paradoja consistente en cómo la aplicación de IA en políticas públicas y en la toma de decisiones en ámbitos privados, lejos de ser una herramienta para combatir desigualdades estructurales a partir de su supuesta neutralidad y eficiencia, termina consolidando escenarios perjudiciales para poblaciones vulnerables, esta vez con una metodología que se presenta como “justa” e “imparcial” y que impide instancias de apelación o explicabilidad. Finalmente, haremos algunas recomendaciones, a modo de sugerencias, para mitigar la discriminación resultante de los sesgos en la toma de decisión algorítmica.

11

### 1. El sueño de una inteligencia artificial objetiva y neutral

Hasta hace unos pocos años solo escuchábamos la frase “inteligencia artificial” para hacer referencia a HAL 9000 de *2001 Odisea del espacio* o al androide Data de *Star Trek*, pero hoy pocos se sorprenden de su uso cotidiano. El consenso en los medios de comunicación y en cierta bibliografía académica es que nos encontramos frente a una de las revoluciones tecnológicas más importantes de la historia.

Sin embargo, son necesarias algunas aclaraciones. En primer lugar, cabe señalar que la expresión “inteligencia artificial” requiere elucidación, ya que suele ser utilizada como una noción paraguas bajo la que aparecen conceptos vinculados y muy cercanos, pero no sinónimos, como aprendizaje automatizado, aprendizaje profundo (*deep learning*) o computación cognitiva, entre otros. En el presente trabajo usaremos “inteligencia artificial” para hacer referencia a sistemas o modelos de algoritmos que

pueden procesar grandes volúmenes de información y datos a la vez que pueden “aprender” y mejorar en su tarea más allá de cómo fueron programados originalmente. Por ejemplo, es un caso de IA un algoritmo que, luego de procesar cientos de miles de fotos de gatos, puede extraer lo que necesita para poder reconocer a un gato en una nueva imagen, sin confundirlo con un peluche o un almohadón. A medida que se le entregan más y más fotografías, aprenderá más y cometerá menos errores.

Una segunda aclaración concierne al término “algoritmo”. Se trata de un vocablo que se repite indefectiblemente en la literatura, pero que está lejos de ser unívoco. En el hablar cotidiano se suele hablar del “algoritmo” de una determinada plataforma, como “el algoritmo de Facebook”, cuando en realidad nunca se trata de un único elemento, sino de sistemas compuestos por algoritmos numerosos. Y a la hora de pensar una definición sobre algoritmos ninguna parece capturar sus diversos usos, tanto en el discurso público como en las publicaciones especializadas. Para evitar confusiones, proponemos pensar en una distinción inspirada en diferentes autores (Mittelstadt *et al.*, 2016; Pasquinelli y Joler, 2021). Por un lado, entender algoritmo en un sentido estrecho, en tanto un constructo matemático que se selecciona por su eficacia pasada para resolver tareas similares (por ejemplo: redes neuronales profundas, redes bayesianas, cadenas de Markov, etc.). Por otro, algoritmo en un sentido extendido, en tanto un sistema tecnológico tripartito -que comprende datos de entrenamiento, el algoritmo de aprendizaje en sentido estrecho y un modelo estadístico como *output* final- diseñado, ensamblado e implementado para determinados propósitos relacionados con la resolución del problema práctico formulado inicialmente.

12

Esta segunda acepción, la de algoritmo extendido, es la que emplearemos en este trabajo, en tanto sistema tecnológico diseñado a partir de un algoritmo en sentido estrecho entrenado con datos que seleccionan quienes diseñan con la finalidad de resolver un cierto problema. Se trata de estructuras que están presentes en tecnologías que tenemos a mano, como el reconocimiento de voz o la comprensión de texto presente en aplicaciones de nuestros teléfonos, y también en proyectos más ambiciosos como los de automóviles que no necesitan conductor o los de estudios para conocer más sobre el cáncer. El campo de aplicación de estos desarrollos es vastísimo y afecta a muchas industrias y sectores de la sociedad, como la economía, con el auge de algoritmos que prometen sugerir las mejores inversiones en la bolsa de valores; la política, con campañas a favor o en contra de un candidato creadas para interpelar a distintos individuos en particular a partir de sus preferencias y conductas en la web, o en la cultura, con recomendaciones personalizadas en plataformas de *streaming* de series, películas o música.

Quizá por estos usos reales y sus potenciales beneficios, estamos viviendo un escenario donde las promesas de la IA están sobredimensionadas. En tanto producto de “la inteligencia humana”, la IA es presentada como el punto máximo de la actividad racional, libre de prejuicios, pasiones y cualquier error que hombres y mujeres podamos cometer. Existen, incluso, quienes creen que estos mecanismos “objetivos” y “neutrales” de la IA son el vehículo ideal para llevar adelante la propuesta del velo de ignorancia de Rawls (1971), escogiendo principios que sean justos para todos y todas desde una posición totalmente imparcial. Frente a ello, esgrimiremos razones -que abrevan, fundamentalmente, en algunos aportes de las epistemologías feministas y

del pragmatismo contemporáneo- para tornar explícito en qué sentido las pretendidas objetividad y neutralidad resultan imposibles.

Un primer punto para considerar es el inexorable impacto de nuestra condición humana sobre la IA. En efecto, el ciclo de vida del aprendizaje automatizado supone un vínculo inextricable entre la IA y una secuencia de decisiones humanas. Tal como veremos a lo largo de este artículo, allí donde el aprendizaje automatizado nos provee las ventajas del rápido procesamiento de una ingente cantidad de datos y la capacidad de extraer patrones a partir de ellos, hay una serie de instancias donde la injerencia humana no solo es posible, sino necesaria. De este modo, frente al mito de la -presunta- autonomía de la IA (Kelleher, 2018, p. 33), sobre la que reposa una extendida retórica acerca de su neutralidad valorativa y la objetividad que le sería concomitante, en lo que sigue pondremos el énfasis en la gravitación humana en diferentes etapas.

Una reflexión central que cabe subrayar es que esa objetividad pretendida se ha hecho recaer, en muchas ocasiones, en una concepción ingenua en torno de los datos. En efecto, se les ha dotado de una pátina de objetividad como punto de partida y piedra de toque neutral, asumiendo que estarían desprovistos de valores. Por el contrario, se impone advertir que los datos son generados por un proceso de abstracción y, en consecuencia, no son nunca una descripción objetiva del mundo, sino que son inexorablemente parciales y sesgados en la medida en que resultan de decisiones y elecciones humanas (Kelleher, 2018, p. 46). Así, por ejemplo, se decide incluir ciertos atributos y dejar de lado otros.

13

Correlativamente, la noción misma de predicción que surge de esta consideración de los datos dista mucho de tornarla un resultado impoluto y prístino del algoritmo. En este contexto, predecir equivale a estimar el valor de un atributo objetivo (*target*) para una instancia dada, con base en los valores de otros atributos seleccionados para esa instancia. Cabe que nos preguntemos más detenidamente, entonces, qué tipo de objetos culturales y técnicos son los conjuntos de datos que constituyen la fuente de la IA. Tal como sostienen Pasquinelli y Joler (2021), la calidad de los datos de entrenamiento es el factor más determinante de la así llamada “inteligencia” que extraen los algoritmos de aprendizaje automatizado. Los algoritmos (en el sentido estrecho de constructos matemáticos, al que nos referimos más arriba) son la segunda, en la medida en que son las máquinas las que computan ese valor e inteligencia en un modelo.

Sin embargo, nos interesa enfatizar que los datos de entrenamiento no aparecen nunca en bruto, ni son independientes e imparciales.<sup>1</sup> Veamos este punto más pormenorizadamente: el diseño, el formateo y la edición de los datos de entrenamiento constituyen una empresa laboriosa y delicada, que es, probablemente, más significativa para los resultados finales que los parámetros técnicos que controlan el algoritmo de

---

1. Incluso hay autores, como Gitelman (2013), que sostienen que los datos, en algún sentido, “son ellos mismos algorítmicos”.

aprendizaje (Pasquinelli y Joler, 2021). En efecto, la decisión de elegir una fuente de datos en lugar de otra es una marca ostensible que deja la intervención humana en el dominio de las inteligencias “artificiales”.

El conjunto de datos de entrenamiento es, pues, una construcción cultural, no solo técnica. Usualmente comprende datos de entrada que se asocian con datos ideales de salida, como imágenes con sus descripciones, llamadas “etiquetas” o “metadatos”. Un ejemplo canónico que ofrecen Pasquinelli y Joler (2021) es el de la colección de un museo y su archivo, en el cual las obras de arte están organizadas por metadatos tales como autor, año, medio, etc. El proceso semiótico de asignar un nombre o una categoría a una imagen o un objeto nunca es imparcial. Esta acción deja otra profunda huella humana en el resultado final.

Un conjunto de datos de entrenamiento para aprendizaje automatizado se compone, usualmente, de los siguientes pasos: 1) producción: trabajo o fenómenos que producen información; 2) captura: codificación de la información en un formato de datos producido por un instrumento; 3) formateo: organización de los datos en un conjunto; 4) etiquetado: en aprendizaje supervisado, la clasificación de los datos en categorías (metadatos).

La reflexión que ha hecho recaer en el “mito de los datos brutos” la retórica de la objetividad pretendida se revela, entonces, o bien como profundamente ingenua o bien como cómplice de intereses controvertidos (que prefieren no revelarse, sino esconderse tras la pátina de la neutralidad valorativa). Una cuestión clave para el uso gubernamental de la ciencia de datos -en general- y del aprendizaje automatizado -en particular- estriba, por lo tanto, en definir qué medir y cómo medirlo (Kelleher, 2018, p. 26). Esto, a partir de una definición clara y precisa del problema que se pretende abordar, instancia en la que la gravitación humana también parece resultar crucial.

Como es sabido, la concepción epistemológica que hemos heredado de la Modernidad, a la vera de la que se ha montado una visión abstracta, universal, neutral y objetiva de la ciencia, ha reposado en una serie de dicotomías que fueron puestas en zozobra desde distintas tradiciones y distintos abordajes del pensamiento filosófico contemporáneo. En efecto, la idea de una ciencia ayuna de valores se encaramó sobre dualismos tales como: subjetivo/objetivo, particular/universal, metafórico/literal, evaluación/descripción, valor/hecho, que han sido puestos en jaque por un amplio arco de reflexiones entre las que -sin pretensiones de exhaustividad- se puede citar a la sociología de la ciencia (por ejemplo, con los ya célebres aportes de Bruno Latour y Steve Woolgar, 1979), el pragmatismo norteamericano (en una línea que hunde sus raíces en las reflexiones de John Dewey, 1929, y que llega a nuestros días en la crítica pertinaz de Hilary Putnam: 1981, 2002), y los aportes críticos de las epistemologías feministas esgrimidos -*inter alia*- por Helen Longino (1990) y Donna Haraway (1985, 2019).

Una explicitación y ponderación crítica de las razones variopintas sostenidas por estas diversas vertientes contra las dicotomías mentadas excede los propósitos de este trabajo. No obstante, resulta posible y pertinente aislar algunos de sus núcleos argumentativos. Entre ellos, cabe destacar que se ha sostenido que no es posible

trazar distinciones tajantes, exclusivas y excluyentes, entre polos tales como subjetivo/objetivo y valor/hecho, puesto que, una vez que se abandona la categoría moderna de subjetividad *qua* autosuficiente y prescindente de todo vínculo con la alteridad, así como su antropocentrismo correlativo, se socavan los cimientos para sostener una ciencia abstracta, universal, neutral y objetiva.

Consideremos, a continuación, algunas de las críticas filosóficas a los supuestos en los que se hizo reposar la idea de una ciencia: abstracta, universal, neutral y objetiva. En efecto, esas cuatro “virtudes” cardinales que durante siglos fueron endilgadas al quehacer científico como epítome de la ausencia de valores y prejuicios tambalean al reparar en que:

- (i) los valores lo permean todo; no hay, por lo tanto, posibilidad de escindirlos -al menos, de la manera general que subyace a los planteos que nos resultan relevantes aquí- de los hechos (Dewey, 1929);
- (ii) el lenguaje ordinario revela, en este sentido, los diversos modos en que evaluación y descripción se hallan íntimamente entretejidos (Putnam, 2002);
- (iii) las “virtudes” invocadas resultan de una concepción errada de nuestra cognición, surgida al calor de valores vinculados con una visión androcéntrica de la ciencia que hizo pasar una mirada parcial hegemónica por patrón universal (Longino, 1990);
- (iv) de este modo, ha cundido la ceguera respecto de los diversos modos en que nuestro conocimiento es siempre encarnado, parcial y situado, en un sentido que nos desafía a resignificar y redimensionar la objetividad científica, y a abandonar el sueño roto de la absoluta neutralidad valorativa (Haraway, 1985, 2019).

15

Las cuatro objeciones que acabamos de reconstruir parecen echar por tierra las pretensiones de aquellas cuatro “virtudes” cardinales de la ciencia.<sup>2</sup>

A continuación, analizaremos un caso surgido en la Argentina que -frente a la retórica de una IA neutral y objetiva- ilustra de un modo patente la carga valorativa que gravita en ella. El caso se presenta alineado con el objetivo central de este trabajo, cual es, como fue referido, explicitar una paradoja que impone la necesidad acuciante de su debate extramuros académicos: cómo la aplicación de IA en el trazado de políticas públicas (y, también, en la toma de decisiones en ámbitos privados), lejos de mitigar desigualdades estructurales, las termina consolidando.

## **2. El caso disparador: algoritmos que predicen embarazos adolescentes en Salta**

En 2017, el entonces gobernador de Salta, provincia del noroeste de Argentina, anunció un convenio con la filial nacional de la empresa Microsoft para la aplicación de la IA “en la prevención de los problemas más urgentes” de la provincia. Entre estos

---

2. Cabe resaltar que estas objeciones no nos comprometen necesariamente con una salida relativista, cuyas versiones y complejidades, dado el alcance de nuestro trabajo, no discutiremos aquí.

problemas, en efecto, se destacaba el gran número de embarazos adolescentes. Según las estadísticas oficiales, en ese año más del 18% de los partos totales de la provincia habían sido de gestantes menores de 19 años: 4.914 hijos e hijas, a razón de más de 13 por día (DEIS, 2017). En la promoción de su iniciativa, el gobernador manifestó:

“Lanzamos un programa para prevenir el embarazo adolescente utilizando inteligencia artificial de la mano de una reconocidísima empresa de software del mundo. Con la tecnología vos podés prever cinco o seis años antes, con nombre, apellido y domicilio, cuál es la niña, futura adolescente, que está en un 86 por ciento predestinada a tener un embarazo adolescente” (Página 12, 2018).

16

Los anuncios de la empresa se refirieron a una “iniciativa innovadora, única en el país y un gran paso en el proceso de transformación digital de la provincia” (Microsoft, 2018). Y revelaron a un tercer actor en la coalición: la Fundación CONIN, presidida por Abel Albino, médico activista en contra de la legalización del aborto y del uso de preservativos (Albino, 2010). En esos anuncios se informó, además, parte de la metodología de trabajo. Por ejemplo, se comunicó que la información de base era “entregada voluntariamente por las personas” y posibilitaba “trabajar en la prevención del embarazo adolescente y la deserción escolar. Los algoritmos inteligentes permiten identificar características en las personas que podrían derivar en alguno de estos problemas y advierten al gobierno” (Microsoft, 2018). El coordinador de tecnología del Ministerio de Primera Infancia de Salta, Pablo Abeleira, declaró que “a nivel tecnológico, el modelo que desarrollamos tiene un nivel de precisión de casi un 90 % de una prueba piloto realizada en Salta capital” (Microsoft, 2018).

Por detrás de las manifestaciones grandilocuentes, que hoy -con el cambio de signo político en la provincia desde 2019- parecen acalladas, se agazapaba una propuesta con problemas graves de diversa índole. Los investigadores del Laboratorio de Inteligencia Artificial Aplicada (LIAA), del Instituto de Ciencias de la Computación (ICC) de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (UBA) realizaron un informe al respecto. Su conclusión es que la propuesta adolece de desaciertos metodológicos serios y se apoya en datos poco confiables. Además, plantean el riesgo de tomar medidas incorrectas a los responsables de políticas públicas, con estas palabras: “Las técnicas de inteligencia artificial son poderosas y demandan responsabilidad por parte de quienes las emplean; son solo una herramienta más, que debe complementarse con otras, y de ningún modo reemplazan el conocimiento o la inteligencia de un experto”, especialmente en un área tan sensible como la salud pública y cuando se involucran sectores vulnerables (LIAA, 2018).

Otros investigadores han llegado a conclusiones parecidas. Así, una revisión del sistema por parte de la World Web Foundation refiere que no existe información accesible sobre las bases de datos empleadas, ni sobre la hipótesis que sirve de apoyo para el diseño de los modelos, ni sobre el proceso de diseño de los modelos finales, criticando la opacidad del proceso (Ortiz Freuler e Iglesias, 2018). Además, se

ha sostenido que la iniciativa omite evaluar cuáles son las inequidades que se pueden producir, para prestar especial atención a los grupos minoritarios o vulnerables que pueden verse afectados, amén de las dificultades de trabajar con un rango de edad tan amplio en los relevamientos y el riesgo de discriminación o aun criminalización inherente al sistema (Martínez Elebi, 2020; Danesi, 2021).

En rigor, y de acuerdo con lo que se puede reconstruir por detrás de la falta de información relevante sobre las bases técnicas del proyecto, el procedimiento consistió en aprovechar datos reunidos por el gobierno provincial y organizaciones de la sociedad civil, y apoyo de Microsoft,<sup>3</sup> solamente en barrios de bajos ingresos de Salta Capital durante 2016 y 2017. La encuesta llegó a poco menos de 300.000 personas, de las cuales 12.692 fueron niñas y adolescentes entre 10 y 19 años. En el caso de las menores de edad, la información fue recogida luego de obtener el consentimiento de “los jefes de hogar” (Ramallo, 2018). Estos datos nutrieron un modelo de aprendizaje automatizado<sup>4</sup> que, según sus implementadores, permite predecir, con un grado cada vez mayor de precisión, qué niñas y adolescentes cursarán un embarazo en el futuro. Esto es: se obtiene un listado de personas a las cuales les fue asignada una probabilidad de embarazo. Los algoritmos, lejos de autoejecutar políticas, brindaban la información al Ministerio de Primera Infancia para que abordara los casos identificados. No se precisó desde el Estado salteño en qué consistiría ese abordaje, ni los protocolos empleados, ni las acciones de seguimiento previstas,<sup>5</sup> ni el impacto de las medidas aplicadas -ni si ese impacto se ha medido de algún modo-, ni los criterios de selección de los organismos no gubernamentales o fundaciones involucrados, ni el rol de la iglesia<sup>6</sup> (Ortiz Freuler e Iglesias, 2018; LIAA, 2018). En síntesis, aun si fuera posible (lo cual no parece ser el caso) predecir el embarazo adolescente, no queda claro para qué se haría: la prevención permanece ausente en todo el proceso. Considerar un riesgo alto de estigmatización de niñas y adolescentes por esta causa es, entonces, inevitable.

¿Qué queda de la promesa del entonces gobernador de predecir “cinco o seis años antes, con nombre, apellido y domicilio, cuál es la niña, futura adolescente, que está en un 86 por ciento predestinada a tener un embarazo adolescente”? Como veremos

---

3. Ante el señalamiento, desde organismos de la sociedad civil distintos a los participantes en el proyecto, de que los datos recabados se alojaban en servidores privados, ubicados fuera del territorio argentino, las autoridades salteñas se limitaron a responder que el procedimiento es “conforme a las leyes pertinentes” (Ortiz Freuler e Iglesias, 2018).

4. El modelo fue construido a partir de un proceso iterativo de aprendizaje automatizado conocido como “árbol de decisión potenciado” (*two-class boosted decision tree*), que permite crear un conjunto de árboles de regresión en el cual cada árbol corrige los errores del anterior. Pueden consultarse más detalles respecto de este método, por ejemplo, en el reporte técnico que presenta Microsoft para explicar sus algoritmos de base (Burgues, 2010).

5. En una entrevista, el responsable del proyecto indicó como políticas a emplearse “un fortalecimiento familiar” y el “abordaje humano”, sin otra precisión. Ante la pregunta concreta sobre si el programa incluye educación sexual y la entrega de anticonceptivos -en particular, preservativos-, el funcionario se limitó a afirmar: “Nosotros les damos los resultados a las áreas correspondientes” (Ramallo, 2018).

6. En un video promocional del Ministerio de Primera Infancia de Salta, rescatado por Ortiz Freuler e Iglesias (2018), en el cual participan el gobernador y el titular de la Fundación CONIN, se anuncia el rol de “las iglesias” como fundamental (disponible en: <https://www.youtube.com/watch?v=F8Lg9ytO9JY>, minuto 1.48). Este no parece un dato menor: el 91% de la población de esa provincia adhiere a la Iglesia Católica (Mallimaci *et al.*, 2019).

en la última sección de este artículo, no se trata solo de una promesa incumplida, sino de una inviable (además de éticamente cuestionable). Pero antes analicemos en qué instancias, desde la génesis de un algoritmo hasta su aplicación, intervienen sesgos.

### 3. Los sesgos en acción

Como señalamos en la primera sección de este artículo, vivimos en un momento histórico en el que se dan dos circunstancias: la adopción en tiempo récord de tecnologías basadas en IA y la creencia de que estos avances son beneficiosos por sí mismos, ya que estarían libres de prejuicios y de cualquier debilidad humana semejante. Lejos de las pantallas de cine y las páginas de los libros de ciencia ficción, la IA se encuentra a nuestro alrededor, incrustada de manera invisible en nuestras herramientas diarias y como parte de sistemas tecnológicos complejos (Boddington, 2017). Desde el mercadeo, el entretenimiento y la gestión de recursos laborales hasta ámbitos que históricamente han sido tradicionales y repelentes a los cambios -como la atención médica, las finanzas, la educación y las políticas públicas-, se incorporan en la actualidad sistemas de algoritmos de aprendizaje automatizado tanto para mecanizar pequeñas tareas como para obtener información clave a la hora de tomar decisiones. Los desarrollos actuales permiten no solo obtener imágenes de nuestro rostro sin nuestro consentimiento para identificarnos (Balmaceda y Schleider, 2021), sino que también están siendo programadas para leer nuestras emociones y recuperar todo tipo de información, a la vez que ya se aplican en el uso de armas autónomas que pueden matar sin intervención humana directa y en vehículos que circulan en algunas calles de Estados Unidos sin conductor.

18

El avance de esta tecnología, sin embargo, no está acompañado por una suficiente reflexión y una discusión abierta acerca de sus consecuencias no deseadas. Una vez más, parece prevalecer en la sociedad la noción de que la aplicación de algoritmos en diferentes ámbitos no solo es garantía de eficiencia y rapidez, sino también de la no intervención de los prejuicios humanos, que pueden “ensuciar” el prístino accionar de las matemáticas involucradas en el código que son el núcleo de los algoritmos. Así, se da por sentado que la IA fue creada para mejorar a la sociedad en su conjunto o, con más modestia, a ciertos procesos y productos. Sin embargo, no solo no se problematiza lo más elemental, esto es: para quién representaría esto una mejora, quién se beneficiaría, quién evalúa la mejora -¿los ciudadanos, el Estado, las empresas, las adolescentes salteñas, los adultos que abusaron de ellas?-. En cambio, no parece haber una real conciencia de la dimensión de su impacto social o de la necesidad de discutir si tal cambio es evitable. Las noticias continuas sobre la inserción de la IA en nuevos ámbitos son recibidas como un dato que no sorprende más que por lo novedoso, y que, inexorable como el avance del propio tiempo, no puede ser impedido o revisado. La creciente automatización de procesos que antes eran realizados por humanos -lo que Brynjolfsson y McAfee (2014) denominan “la segunda era de las máquinas”, en la cual, a diferencia de la Revolución Industrial, los artefactos no complementan a los trabajadores, sino que los sustituyen- puede generar alarma y desazón, pero no el interés por tratar de detenerlo o preguntarse cuál es el futuro del trabajo o el de las personas una vez que la IA se haga cargo de nuestras tareas laborales. Se impone una serie de preguntas que raramente se

formulan. ¿Es eso algo deseable? ¿Para qué sector de la sociedad? ¿Quién se beneficiará de esta transformación y quién saldrá perdiendo? ¿Qué podemos esperar de un futuro en el que gran parte de las ocupaciones tradicionales serán llevadas adelante por máquinas? No parece haber tiempo ni espacios para discutir sobre eso: la automatización simplemente ocurre.

Esta pasividad frente al continuo avance de la tecnología en nuestra vida privada, pública, laboral y cívica parece posible gracias a la confianza generada por la creencia de que estos desarrollos son “superiores” a lo que se podría conseguir a través del mero esfuerzo humano. De acuerdo con esto, ya que la IA es mucho más poderosa, es “inteligente” (la etiqueta *smart* hoy se aplica a teléfonos celulares, aspiradoras y cafeteras, entre otros objetos que harían sonrojar a Turing) y está libre de sesgos e intencionalidades. Sin embargo, como hemos señalado, esta retórica debe contrastarse con los desarrollos teóricos que muestran que es imposible, por principio, plantear una IA valorativamente neutra. Ahora analizaremos cómo los sesgos se dan, en efecto, en las diferentes instancias de su desarrollo.

Dicho en pocas palabras y de manera simple: los sesgos están presentes en todas las etapas del diseño, las pruebas y la aplicación de los algoritmos (Coeckelbergh, 2021) y, por eso mismo, son muy difíciles de identificar y más arduos aún de corregir (Hao, 2019). Sin embargo, esa es una tarea necesaria para desenmascarar su supuesto carácter aséptico, carente de los valores y errores humanos.

Tal como argumentamos en la primera sección, los valores están presentes desde el momento mismo en que se decide cuál será el rol del algoritmo a crear. Es un mito creer que la ciencia de datos es un proceso autónomo que se pone en funcionamiento para que encuentre por sí solo respuestas a nuestros problemas (Kelleher, 2018). Todo el proceso de puesta en marcha, prueba y funcionamiento de los algoritmos requiere de una supervisión humana calificada, pero parece ser una creencia arraigada en el discurso público que los problemas aparecen por generación espontánea y que simplemente hay que encontrarles respuestas. En realidad, la definición misma del problema es una actividad en la que intervienen humanos y en la que quedan patentes sus preocupaciones, creencias y sentimientos. A más de 80 años de la distinción de Hans Reichenbach (1947) entre contexto de descubrimiento y contexto de justificación, se puede recoger el señalamiento de este epistemólogo alemán de que es una fantasía sostener que puede sistematizarse el proceso de aparición de las ideas o hipótesis científicas y afirmar que la formulación de los problemas no surgirá nunca de un algoritmo porque no es posible que el análisis de datos baste para definir un problema. En su manual sobre ciencia de datos para profesionales del mercadeo, Linoff y Berry (2011, p. 3) lo dejan en claro: “La minería de datos permite a las computadoras hacer lo que mejor saben hacer: explorar muchos datos. Esto, a su vez, permite que las personas hagan lo que mejor saben hacer, que es plantear el problema y comprender los resultados”. En ese sentido, es útil trazar una distinción entre dos clases de sesgos que pueden gravitar en la definición misma del problema: el así llamado “sesgo por omisión de información”, según el cual faltan variables relevantes para la caracterización del problema, y el “sesgo histórico”, que la contamina con prejuicios socioculturales (Pasquinelli y Joler, 2021).

La siguiente instancia donde es posible identificar la presencia de sesgos es en el diseño del algoritmo que buscará dar respuesta al problema señalado, ya que cualquier tecnología surge en el contexto de instituciones, prácticas y actitudes sociales que tienen sus propios valores. Esto es lo que Friedman y Nissenbaum (1996) llaman “sesgos preexistentes” de las instituciones sociales, prácticas y actitudes de las que surge la tecnología. Se trata de valores que están en varios niveles: en la sociedad en general, en una organización en particular y en las personas involucradas, tanto en quienes tienen un rol preponderante en la organización o en el financiamiento del desarrollo como en quienes solo cumplen funciones menores. Los sesgos permean la tecnología tanto cuando hay decisiones explícitas y conscientes, como cuando se quiere apuntar a un determinado tipo de cliente en un desarrollo con fines comerciales. También, de forma implícita, a pesar de las mejores intenciones, como cuando un sistema que asesora sobre solicitudes de préstamos pondera determinadas ubicaciones como barrios de bajos ingresos o con altos índices de criminalidad como “indeseables”. Estos valores están tan imbricados en la sociedad que, en ocasiones, requiere mucho esfuerzo o una mirada externa para que se los pueda detectar, como cuando se asumen ciertas preferencias o conductas de acuerdo con el género.

En el diseño del algoritmo, una decisión fundamental es entender qué tipo de algoritmo de IA se utilizará. Existen muchos modelos diferentes y cada uno tiene una manera distinta para lograr generalizaciones a partir de un conjunto de datos. El tipo de generalización que codifica un algoritmo se conoce como “sesgo de aprendizaje” (*learning bias, modeling o selection bias*) y es clave a la hora de lograr resultados. De hecho, existe un principio informalmente conocido como *no free lunch theorem* (Wolpert y Macready, 1997) -algo así como “el teorema de la inexistencia de almuerzos gratis” que, españolizado, se ajustaría a un improbable “teorema del nada es gratis”- que señala que el sesgo de aprendizaje es inevitable a la hora de que un algoritmo aprenda algo y que, por lo tanto, no existe “el mejor algoritmo de aprendizaje automatizado”, sino que, en todos los casos, siempre es una elección en la que se privilegia algo y se pierde en algún aspecto. Por ejemplo, un algoritmo de regresión lineal codifica una generalización lineal a partir de los datos y, como resultado, ignora las relaciones no lineales que pueden ajustarse más a los datos (Kelleher, 2018). Una manera de evaluar qué conviene más es construir múltiples modelos usando diferentes algoritmos y comparar sus resultados para decidir cuál utilizar de acuerdo con el conjunto de datos y la tarea asignada.

El diseño y la funcionalidad de un algoritmo reflejan los valores de las personas que lo diseñaron y de sus usos pretendidos. Ningún desarrollo tecnológico es neutral porque no existe algo así como una elección objetivamente correcta en ninguna instancia del desarrollo, sino muchas posibles elecciones (Pedace, Balmaceda, Pérez, Lawler y Zeller, 2020; Johnson, 2006). En consecuencia, en muchas ocasiones es difícil detectar los sesgos latentes en los algoritmos y los modelos que producen. El resultado final es un modelo estadístico que, por definición, conlleva un cierto margen de error. De modo que aun el juicio experto acerca del algoritmo que se considerará el más apropiado involucra una decisión cargada de valores acerca de cuál es el margen de error que se está dispuesto a tolerar en el *output* final.

Existen también sesgos en el conjunto de datos que se utilizarán, ya que en su recolección habrá que determinar cuáles son considerados relevantes para el objetivo buscado, tal como desarrollamos en la primera sección de este trabajo. Si el conjunto de datos no es representativo del objetivo que se intenta lograr, el modelo que genera el algoritmo no será preciso. Por ejemplo, si quisiéramos desarrollar un modelo para predecir la posibilidad de problemas pulmonares en pacientes que atravesaron la enfermedad del COVID-19, pero solo utilizáramos información de clínicas privadas de la ciudad de Buenos Aires, sería poco probable que este modelo fuese exacto para predecir lo que sucederá con hombres y mujeres de toda América Latina, o incluso de otras partes de la Argentina, o de la Ciudad Autónoma de Buenos Aires en barrios en donde la atención de salud se centra en hospitales públicos. Los sesgos en los datos se trasladan a sus resultados y al análisis posterior que se pueda realizar. Los algoritmos están, por lo tanto, sujetos a la restricción de que el *output* nunca puede exceder al *input* (Pedace, Balmaceda, Pérez, Lawler y Zeller, 2020). En términos de la teoría matemática de la comunicación de Shannon (1948) y su lema *garbage in, garbage out*: las conclusiones solo pueden ser fiables en función de los datos sobre los cuales están basadas. Si la base de datos está viciada, conducirá a resultados sesgados y, en ocasiones, injustos e inequitativos. Aquí convergen, por tanto, cuestiones epistémicas y valorativas.

Los datos producen sesgos principalmente por dos vías: o bien los datos recabados no son representativos de la situación que se desea analizar, como cuando las aplicaciones de redes sociales no reconocen rostros con tonos de piel oscuro (al haber sido alimentadas con imágenes de rostros de hombres caucásicos), o bien cuando reflejan prejuicios ya existentes, como un algoritmo que ayuda a decidir si una persona puede tomar un crédito y privilegia a los hombres sobre las mujeres porque ellas suelen tener trabajos informales o realizar tareas domésticas o de cuidado que no son consideradas trabajo. En este último sentido, se trata de sesgos sociales enquistados en la base de datos elegida, que en ocasiones incluye etiquetas aplicadas por humanos y que reflejan sus preferencias e ideas.

La injerencia de los sesgos en la selección de datos no debe confundirse con la que opera en la etapa de la definición del problema. Es posible usar los mismos atributos para entrenar un modelo con objetivos muy diferentes o usar atributos muy distintos para entrenar un modelo con el mismo objetivo. En el caso de un algoritmo que determina la solvencia crediticia, por ejemplo, un atributo podría ser la edad del cliente, la zona de la ciudad donde vive, sus antecedentes penales o la cantidad de préstamos pagados. Elegir qué atributos se consideran o se ignoran puede influir significativamente en la precisión de la predicción de un modelo. Y, aunque resulta relativamente fácil medir su impacto en la precisión, medir sus sesgos es un proceso bastante complicado (Hao, 2019).

Una vez que se determinó el problema a resolver, se eligió el tipo de algoritmo que se utilizará, se lo diseñó y se escogió el conjunto de datos, sigue la etapa de testeo y, eventualmente, de implementación. En ambas instancias, de nuevo encontramos la presencia de decisiones humanas en las que gravitan valores. Si bien es cierto que nunca ha sido tan fácil acceder a la tecnología para realizar ciencia de datos, también lo es que nunca ha sido mayor el riesgo de hacerlo de manera equivocada, ya que

es necesario tanto el conocimiento apropiado como la experiencia con respecto a las propiedades de los datos y las suposiciones que sustentan los diferentes algoritmos de aprendizaje automatizado (Kelleher, 2018). Uno de los riesgos es que personas sin la correcta preparación se sientan intimidadas por la tecnología y elijan simplemente tomar los resultados que les presenta el *software* como verdades incontestables o reflejos de la realidad. Sin embargo, es posible que, inconscientemente, hayan enmarcado el problema de manera inadecuada, hayan ingresado los datos equivocados o hayan utilizado técnicas de análisis con supuestos inapropiados, lo que derivaría en resultados que son la respuesta a la pregunta incorrecta, se basen en datos incorrectos o en el resultado de un cálculo incorrecto. El proceso de evaluación del modelo debe seguir la regla de oro de que un modelo nunca debe evaluarse con los datos sobre los que fue entrenado, aunque la legislación de un país podría determinar que es necesario tener en cuenta otros parámetros como la transparencia y la facilidad de explicación de algunos modelos (Kelleher, 2018).

Es por eso que se impone una revisión constante y crítica de los resultados obtenidos. A esto se agrega, además, que, por definición, cualquier conjunto de datos es la fotografía fija de un momento en el tiempo. Y como los algoritmos de aprendizaje automatizado utilizan información del pasado para trazar en el presente patrones que puedan generalizarse al futuro, el factor temporal debe ser tenido en cuenta, ya que incluso los modelos más exitosos deben ser reentrenados o las bases de datos más efectivas deben ser actualizadas.

22

#### **4. El mito de una inteligencia artificial objetiva y neutral**

Los efectos de la retórica en torno a desarrollos de inteligencia artificial valorativamente neutrales y objetivos parecen desmoronarse cuando se los enfrenta a aquellas voces que aseguran mostrar su imposibilidad por principio, tal como señalamos en la primera sección, como cuando se exhibe la participación humana en numerosas etapas del desarrollo de los algoritmos. Son hombres y mujeres quienes determinan el problema al que hay que dar respuesta, quienes diseñan y preparan los datos, seleccionan qué algoritmos de aprendizaje automático son los más apropiados, interpretan críticamente los resultados del análisis y planifican la acción adecuada a tomar en función de las ideas que el análisis ha revelado.

El caso de los algoritmos que “predicen” embarazos adolescentes en Salta deja al descubierto la inviabilidad de la imagen de pretendida objetividad y neutralidad de la inteligencia artificial, ya que es un verdadero catálogo de los problemas de la aplicación descuidada de técnicas (concretas o aparentes) de aprendizaje automatizado para el apoyo a la toma de decisiones de políticas públicas. Como mencionamos, estos problemas son de índole diversa, tanto por principio como por desconocimiento de la participación humana en distintas etapas de su creación y aplicación. Pero existen, además, errores técnicos que nos gustaría señalar.

En primer lugar, existe una sobredimensión artificial de los resultados del sistema. Esto se atribuye a una contaminación sutil de los datos de evaluación, desde que los datos sobre los cuales se evalúa el sistema no son distintos de los datos que se usan

para entrenarlo. Así, se viola un principio fundamental del aprendizaje de máquinas (LIAA, 2018) que hemos visto que se asume como la regla de oro inquebrantable.

Por otra parte, los datos son inadecuados para el fin que se persigue. Estos datos provinieron de una encuesta a adolescentes que residían en la provincia de Salta, que indagaba información sobre su persona (edad, etnia, país de origen, etc.), su entorno (si contaba con agua caliente en su vivienda, con cuántas personas cohabitaba, etc.) y sobre si había cursado o estaba cursando un embarazo. Ahora bien, la pregunta que se intentaba responder a partir de estos datos actuales era si una adolescente podía cursar un embarazo en el futuro (LIAA, 2018), algo que, más que de una predicción, tiene el aspecto de una premonición. Cabe recordar, en este sentido, que en la sección anterior advertimos que como los algoritmos de aprendizaje automatizado emplean información del pasado para trazar en el presente patrones que puedan generalizarse al futuro, el factor temporal no debe ser soslayado, ya que todo modelo debe ser reentrenado y toda base de datos, actualizada, para que la predicción sea adecuada.

Por último, cabe enfatizar el carácter sesgado tanto de la definición del problema como de los datos. Por un lado, resulta menester controvertir el propio planteamiento del problema, en la medida en que asume al embarazo adolescente como una suerte de fenómeno natural dado, considerado unilateralmente (solo desde las niñas y adolescentes), y desdeñando, en consecuencia, la injerencia de otros factores de suma relevancia (tales como, por ejemplo, la responsabilidad de los varones involucrados). Por otro lado, los datos sobre embarazos adolescentes, por la sensibilidad inherente a esta clase de temas, tienden a estar incompletos o, directamente, sesgados. Esta información -parcial, y mediada por el sector de la sociedad al que pertenecen las encuestadas, su historia personal y familiar, etc.– es la que aprende el sistema:<sup>7</sup> “Así, incluso si la metodología usada para construir y evaluar los sistemas fuera correcta, las reglas estadísticas construidas sobre estos datos arrojarían conclusiones erradas”, que reflejarían sus distorsiones (LIAA, 2018).<sup>8</sup>

23

7. El informe de la World Web Foundation, aunque sin ser especialmente incisivo, destaca dos problemas centrales del sistema. El primero, que las bases de datos y el diseño del modelo no son accesibles al público, lo cual “limita la trazabilidad y posibilidad de que se realicen auditorías independientes”. El segundo, que no se involucró a las poblaciones afectadas, ni se tuvo en cuenta “la desigualdad estructural que padecen las mujeres y adolescentes que conforman la población objetivo” (Ortiz Freuler e Iglesias, 2018).

8. El Ministerio de Primera Infancia de Salta, organismo entonces responsable del proyecto, ensayó una respuesta a las críticas de la LIAA a requerimiento de los autores del informe de la World Web Foundation (que renovaron algunos cuestionamientos, pero no se refirieron al carácter eminentemente *ad hoc* de la pretendida explicación). En primer lugar, los funcionarios salteños indicaron que el modelo sobre el cual se basó el informe de la LIAA (publicado inicialmente en: <https://github.com/facundod/case-studies/blob/master/Prediccion%20de%20Embarazo%20Adolescente%20con%20Machine%20Learning.md>) reutilizó datos de entrenamiento como datos de evaluación, pero el modelo “actual” utilizó “bases independientes”. Por otro lado, calificaron de “apropiados para la pregunta que se quiere contestar” los datos, porque “el modelo se actualiza y reentrena de manera regular” y se trata de un modelo probabilístico, apoyado “en una serie de presunciones acerca del significado y relevancia de patrones y valores atípicos”. Finalmente, refirieron que no hubo un sesgo en la base de datos porque el objetivo del modelo es “realizar predicciones únicamente sobre poblaciones vulnerables”, debido a que ese “es el campo de acción” del Ministerio de Primera Infancia (Ortiz Freuler e Iglesias, 2018). Con todo, cabe consignar que el informe sobre el cual se basó el LIAA ha dejado de estar disponible. La correlación entre estos fenómenos no puede explicarse con certeza. No obstante, la publicación original del “caso de estudio” pudo recuperarse a través de Internet Archive Wayback Machine: <https://web.archive.org/web/20180412161447/https://github.com/facundod/case-studies/blob/master/Prediccion%20de%20Embarazo%20Adolescente%20con%20Machine%20Learning.md/>.

Resulta claro, pues, que el desarrollo de algoritmos no es neutral, sino que se realiza a partir de una decisión en medio de muchas elecciones posibles. En este sentido, como el diseño y la funcionalidad de un algoritmo reflejan los valores de sus diseñadores y de sus usos pretendidos, los algoritmos inexorablemente conducen a decisiones sesgadas. Como hemos argumentado, a la vera de la noción de algoritmo en sentido extendido que hemos propuesto, hay decisiones humanas en cada una de las siguientes etapas: la definición del problema, el diseño y la preparación de datos, la selección del tipo de algoritmo, la interpretación de los resultados y la planificación de acciones a partir de su análisis. Sin una supervisión humana calificada y activa, ningún proyecto de algoritmo de inteligencia artificial podrá lograr sus objetivos y ser exitoso. Los mejores resultados de la ciencia de datos se producen cuando la experiencia humana y la potencia de los algoritmos trabajan de forma conjunta. En este sentido, el caso de la predicción del embarazo adolescente en Salta revela el carácter acuciante de contar con un debate público que habilite instancias de profunda reflexión crítica.

Se trata, entonces, de ir más allá de los intramuros académicos para propiciar un diálogo que involucre a todos los actores relevantes: desde personas vinculadas con el diseño, desarrollo e implementación de IA y representantes de las corporaciones, así como investigadores procedentes de diversas áreas, la sociedad civil y el Estado.

Quizás sea cierto que tenemos la tecnología que nos merecemos. Pero también es cierto que podemos merecer algo mejor. Desde nuestras voces filosóficas hemos pretendido contribuir al debate polifónico que debemos darnos, a través del señalamiento y la ponderación crítica de los distintos tipos de sesgos que impactan en las diversas etapas que constituyen los algoritmos que moldean nuestras vidas y las sociedades que conformamos.

24

## Bibliografía

Albino, A (2010). *Gobernar es poblar. Criterios antropológicos y éticos para una correcta educación sexual*. Buenos Aires: Ediciones Logos.

Burgues, C. J. C. (2010). *From RankNet to LambdaRank to LambdaMART: An Overview*. Microsoft Research Technical Report MSR-TR-2010-82. Recuperado de: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/MSR-TR-2010-82.pdf>.

Carroll, W. (2001). *I, Toto: The Autobiography of Terry, the Dog who was Toto*. Nueva York: Abrams.

Danesi, C. C. (2021). *The Impact of Artificial Intelligence on Women's Rights: A Legal Point of View*. En K. Miller & K. Wendt (Eds.), *The Fourth Industrial Revolution and Its Impact on Ethics*. Sustainable Finance (267-282). Cham: Springer.

DEIS (2017). *Estadísticas vitales 2017*. Buenos Aires: Ministerio de Salud de la Nación. Recuperado de: <http://www.deis.msal.gov.ar/wp-content/uploads/2019/01/Serie5Nro61.pdf>.

Dewey, J. (1929). *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. Londres: Allen & Unwin Ltd.

Haraway, D. (1985). *Manifiesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s*. *Socialist Review*, 80, 65–108.

Haraway, D. (2019). *Seguir con el problema. Generar parentesco en el Chthuluceno*. Bilbao: Consonni.

Latour, B. & Woolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts*. Los Angeles: Sage.

LIAA (2018). *Sobre la predicción automática de embarazos adolescentes*. Laboratorio de Inteligencia Artificial Aplicada (LIAA). Buenos Aires: Universidad de Buenos Aires. Recuperado de: <https://liaa.dc.uba.ar/es/sobre-la-prediccion-automatica-de-embarazos-adolescentes/>.

Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Martínez Elebi, C. (2020). *Inteligencia artificial y salud*. Buenos Aires: Centro de Estudios en Tecnología y Sociedad (CETyS). Recuperado de: <https://cetys.lat/inteligencia-artificial-y-salud/>.

Mallimaci, F., Giménez Béliveau, V., Esquivel, J. C. E. & Irrazábal, G. (2019). *Sociedad y Religión en Movimiento. Segunda Encuesta Nacional sobre Creencias y Actitudes Religiosas en la Argentina. Informe de Investigación 25*. Buenos Aires: CEIL-CONICET. Recuperado de: <http://www.ceil-conicet.gov.ar/wp-content/uploads/2019/11/ii25-2encuestacreencias.pdf>.

25

Microsoft (2018). *Avanza el uso de la Inteligencia Artificial en la Argentina con experiencias en el sector público, privado y ONGs*. Blog News Center Microsoft Latinoamérica. Recuperado de: <https://news.microsoft.com/es-xl/avanza-el-uso-de-la-inteligencia-artificial-en-la-argentina-con-experiencias-en-el-sector-publico-privado-y-ongs/>.

Ortiz Freuler, J. & Iglesias, C. (2018). *Algoritmos e Inteligencia Artificial en Latinoamérica: un estudio de implementaciones por parte de Gobiernos en Argentina y Uruguay*. Washington: World Wide Web Foundation.

Página 12 (2018). *La inteligencia artificial de Urtubey*. 11 de abril. Recuperado de: <https://www.pagina12.com.ar/107412-la-inteligencia-artificial-de-urtubey>.

Pasquinelli, M. & Joler, V. (2021). *El Nooscopio de manifiesto*. *laFuga* 25, pp.1-20.

Pedace, K., Balmaceda, T. Pérez, D. Lawler, D. Zeller Echenique, M. (2020). *Caja de herramientas humanísticas*. Recuperado de: <https://proyectoguia.lat/wp-content/uploads/2020/05/Caja-de-herramientas-Humanistas.pdf>.

Putnam, H. (1981). *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Putnam, H. (2002). *The collapse of the Fact/Value Dichotomy and other essays*. Cambridge: Cambridge University Press.

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423.

Ramallo, F. (2018). Cómo funciona el sistema para “predecir embarazos adolescentes” de Salta. Entrevista a Pablo Abelira, Coordinador de tecnología del Ministerio de Primera Infancia de Salta. *Diario Uno*, 12 de abril. Recuperado de: [https://uno.com.ar/tecnologia/como-funciona-el-sistema-para-predecir-embarazos-adolescentes-de-salta-04122018\\_rJxfqbraiM](https://uno.com.ar/tecnologia/como-funciona-el-sistema-para-predecir-embarazos-adolescentes-de-salta-04122018_rJxfqbraiM).

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.

Reichenbach, H. (1947). *Elements of Symbolic Logic*. Nueva York: Macmillan Co.