LA OPACIDAD ALGORÍTMICA Y SUS IMPLICANCIAS BIOÉTICAS

Fecha de recepción: 22/05/2025
Fecha de aceptación: 11/07/2025

RAB. DR. FISHEL SZLAJEN

ORCID ID 0000-0002-0433-8506
Contacto: fszlajen@gmail.com

- Rabino (Yeshivá Maalé Gilboa)

  Master en Filosofía Judía (Bar Ilan University)

  Doctor en Filosofía (Universidad Empresarial, Costa Rica)

  Postdoctorado en Bioética (Pontificia Universidad Católica de Rio Grande do Sul, Brasil)

  Mandel Jerusalem Fellow (Mandel Leadership Institute, Israel)

  Scholar Fellow on Religion and The Rule of Law (University of Oxford).

  Miembro Titular de la Pontificia Academia para la Vida, Vaticano

  Miembro del Consejo Académico de Ética en Medicina en la Academia Nacional de Medicina, Argentina

  Miembro del International Consortium for Law and Religion Studies, Estados Unidos.

  Miembro del Consejo Argentino para la Libertad Religiosa.

  Profesor titular en la Universidad de Buenos Aires, Universidad Nacional de La Matanza y Universidad del Salvador, Argentina.

  Premiado por la Fundación Ashekor Internacional (2005); Galardonado por Consenso Salud (2017); Premiado con la Mención de Honor Domingo Sarmiento (Senado Nacional, 2018); Declarado Personalidad Destacada de la Cultura (Legislatura Porteña, 2019); Distinguido por el Ministerio de Asuntos para la Diáspora del Estado de Israel (2020) y por la Pontificia Academia para la Vida, Vaticano (2025).

Instituto de Bioética / UCA - Vida y Ética Año 26 Nº 1 Junio 2025
DOI: https://doi.org/10.46553/vye26.1.6

RESUMEN

La creciente utilización de algoritmos en contextos críticos como la salud, la justicia y la seguridad plantea serias cuestiones bioéticas debido a la opacidad que caracteriza muchos de estos sistemas. Este artículo analiza cómo dicha opacidad socava principios bioéticos fundamentales como la autonomía, la justicia y la no maleficencia, a través de casos concretos en medicina y derecho. Se destacan los riesgos de sesgos algorítmicos, falta de explicabilidad y ausencia de mecanismos de rendición de cuentas. Como respuesta, se propone un enfoque multinivel de transparencia algorítmica que contemple tanto el "qué" (resultados) como el "cómo" (procesos), balanceando la necesidad de información pública, la protección de la propiedad intelectual y el respeto a los derechos individuales. Finalmente, se enfatiza la urgencia de marcos normativos con auditorías independientes, comités interdisciplinarios y criterios pragmáticos de explicabilidad para garantizar un uso ético de la inteligencia artificial.

Palabras clave: Opacidad algorítmica, Bioética, Inteligencia artificial, Explicabilidad, Autonomía, Justicia, No maleficencia, Transparencia algorítmica, Sesgo, Rendición de cuentas.

ABSTRACT

The increasing use of algorithms in critical contexts such as healthcare, justice, and security raises serious bioethical concerns due to the opacity of many of these systems. This article examines how such opacity undermines core bioethical principles, including autonomy, justice, and non-maleficence, illustrated through specific cases in medicine and law. It highlights the dangers of algorithmic bias, lack of explainability, and insufficient accountability mechanisms. As a response, a multilayered approach to algorithmic transparency is proposed, addressing both the "what" (results) and the "how" (processes), while balancing public information needs, protection of intellectual property, and respect for individual rights. The article concludes by stressing the urgent need for regulatory frameworks, independent audits, interdisciplinary ethics committees, and pragmatic standards of explainability to ensure the ethical use of artificial intelligence.

Keywords: Keywords: Algorithmic opacity, Bioethics, Artificial intelligence, Explainability, Autonomy, Justice, Non-maleficence, Algorithmic transparency, Bias, Accountability.

INTRODUCCIÓN

La creciente dependencia de algoritmos en sectores como la salud y la justicia, por sobre todo para la toma de decisiones críticas plantea serias y fundamentales cuestiones éticas. A menudo, estos algoritmos operan como "cajas negras", con escasa o nula transparencia en su funcionamiento interno, lo que genera desafíos relacionados con la confianza, la rendición de cuentas y los derechos de los individuos afectados, creando además una brecha en el entendimiento público y también profesional sobre cómo se generan los resultados y las decisiones, lo que conlleva peligros potenciales. Este artículo explora cómo la opacidad algorítmica entra en conflicto con principios bioéticos fundamentales y cómo pueden mitigarse estos riesgos con mecanismos de control, transparencia y rendición de cuentas.

1. OPACIDAD ALGORÍTMICA Y EL PRINCIPIO DE AUTONOMÍA

El principio bioético de autonomía establece que las personas deben tener la capacidad de tomar decisiones informadas sobre su propia vida. Y esto resulta relevante, particularmente en el ámbito de la salud, porque los algoritmos utilizados en la medicina personalizada y los sistemas de diagnóstico asistido por IA tienen un impacto directo en la vida de los pacientes. Por ello, cuando estos algoritmos son opacos, no pudiendo entender el origen de los resultados ni comprender el mecanismo o proceso mediante el cual se llega a ellos y se toman las decisiones que afectan su bienestar, socava la capacidad de ejercer su autonomía.

Dos estudios publicados en The New England Journal of Medicine destacan cómo los sistemas de IA aplicados a la medicina han comenzado a sustituir gran parte del proceso de toma decisiones en contextos críticos, pero con poca explicación sobre su funcionamiento tanto para los médicos como para los pacientes. Los autores demuestran que la falta de transparencia en estos algoritmos erosiona la confianza en el sistema de salud, limitando la capacidad de los pacientes para dar un consentimiento verdaderamente informado (Obermeyer y Emanuel, 2016; Char, Shah y Magnus, 2018).

Un ejemplo claro de ello es el uso de la IA para la recomendación de tratamientos oncológicos, tal como ocurrió con el sistema IBM Watson for Oncology. El sistema fue criticado por proporcionar recomendaciones basadas en datos limitados y no siempre actualizados, lo que resultó en errores respecto de las sugerencias de tratamiento para ciertos tipos de cáncer. Los pacientes y médicos carecían de la claridad necesaria y suficiente respecto del método por el cual se generaban estas

recomendaciones, lo cual afectó negativamente la capacidad de los pacientes para ejercer su autonomía (Cavallo, 2019; Zou y otros, 2020; Taeyoung y otros, 2023).

Dicho algoritmo, entre otros similarmente opacos, socavan la autonomía de la persona, tanto del profesional médico como del paciente, y ello se debe a que los sistemas de IA no incluyen mecanismos que proporcionen explicaciones comprensibles para los usuarios finales, como los médicos y los propios pacientes, técnica conocida como "IA explicable" (XAI, por sus siglas en inglés). Dicha XAI, debería así estar regulada por normativas que exijan a los desarrolladores de algoritmos proporcionar tanto a los profesionales como a los pacientes información comprensible sobre cómo el sistema llega a sus conclusiones, permitiendo tomar decisiones más informadas.

2. OPACIDAD ALGORÍTMICA Y EL PRINCIPIO DE JUSTICIA

El principio de justicia se refiere a la distribución equitativa de los beneficios y riesgos, en este caso de las intervenciones médicas y tecnológicas. Sin embargo, los algoritmos opacos a menudo perpetúan sesgos estructurales, reforzando las desigualdades. Cuando los algoritmos operan sin transparencia, frecuentemente amplifican las desigualdades existentes, ya que tienden a reproducir los sesgos presentes en los datos con los que han sido entrenados.

Uno de los más importantes estudios en este respecto demuestra, desde hace años, que los algoritmos opacos tienden a castigar las poblaciones vulnerables, desde solicitantes de empleo hasta estudiantes y pacientes. Básicamente dichas investigaciones destacan que la falta de transparencia en los algoritmos no sólo impide que las personas afectadas comprendan cómo se toman las decisiones, sino que también refuerza las desigualdades estructurales (O'Neil, 2016). Esto es singularmente grave en el campo de la salud, donde los sesgos algorítmicos pueden tener consecuencias fatales, como la negación de tratamientos a pacientes pertenecientes a minorías.

En 2019, un algoritmo de salud diseñado por la compañía Optum y utilizado en los Estados Unidos para identificar a pacientes con mayores necesidades de atención médica y predecir sus necesidades, presentaba un sesgo racial significativo discriminando sistemáticamente contra pacientes afroamericanos. El algoritmo estaba diseñado para predecir qué pacientes requerirían cuidados adicionales, pero basaba sus decisiones en los costos previos de atención médica. Es decir, empleaba datos de costos médicos previos para predecir las necesidades futuras de atención. Y, dado que los pacientes afroamericanos habían recibido menor

atención médica en comparación con otros grupos, sus costos de salud eran menores. Estos datos, bajo dicho criterio de programación, hizo que el algoritmo determinará una menor necesidad de atención subestimando las necesidades de los pacientes afroamericanos. Esta falta de justicia, asignando sistemáticamente menos prioridad a pacientes afroamericanos con condiciones de salud similares a las de pacientes blancos, tuvo como resultado que millones de pacientes de color recibieran menos atención de la que necesitaban, reduciendo la equidad en la atención médica y poniendo en evidencia la necesidad de mejorar la transparencia y la ética en el diseño de algoritmos en el ámbito de la salud (Obermeyer, Powers, Vogeli y Mullainathan, 2019).

Otro caso relevante es el de los algoritmos de evaluación de riesgo utilizados en el sistema judicial de Estados Unidos, como el sistema COMPAS, el cual fue objeto de críticas debido a su opacidad y sesgo racial. Un análisis de ProPublica demostró que COMPAS sobrestimaba el riesgo de reincidencia entre personas de color subestimando el riesgo en blancos. Además, la falta de transparencia en el funcionamiento del algoritmo limitaba la posibilidad de cuestionar o refutar la forma en que se generaban las puntuaciones, causando graves consecuencias y, en algunos casos, irreparables para los afectados. (Angwin, Larson, Mattu, Kirchner, 2023).

Un mecanismo clave para evitar estos problemas es la implementación de regulares auditorías algorítmicas y la transparencia de datos, demostrando la fuente desde donde se nutre de información o se entrena dicho algoritmo. Estos algoritmos deben ser entrenados con datos diversos y evaluados continuamente para detectar sesgos, además de incluirse mecanismos de control que permitan identificar y corregir estos sesgos antes de que afecten decisiones de gran envergadura. La participación de terceros expertos independientes en estas auditorías ayudaría a garantizar que los algoritmos estén alineados con principios de equidad.

3. OPACIDAD ALGORÍTMICA Y EL PRINCIPIO DE NO MALEFICENCIA

El principio de no maleficencia exige que los desarrolladores de algoritmos no causen daño a los individuos. Sin embargo, la opacidad algorítmica hace que sea difícil identificar y corregir errores, lo que puede resultar en decisiones perjudiciales para los usuarios.

Recientes estudios sobre algoritmos de diagnóstico revelan frecuentes fallas en los sistemas, más la grave carencia de respuestas al demandar explicaciones claras cuando las recomendaciones son incorrectas, todo lo cual conlleva un serio riesgo de cometer errores médicos que comprometen la salud de los pacientes. Sus au-

tores argumentan que la opacidad algorítmica en dichos sistemas de diagnóstico es una amenaza directa a la seguridad del paciente, ya que las malas decisiones basadas en IA no se pueden corregir sin un entendimiento claro de los procesos subyacentes. Incluso para aquellos que cuestionan el valor real de la explicabilidad en IA clínica, lo hacen por su inestabilidad y su operatividad a nivel poblacional y no individual, limitando por ende su utilidad para decisiones clínicas. Por ello, sin descartar la explicabilidad, abogan como primer paso una validación rigurosa interna y externa de estos modelos. Es decir, en primera instancia evaluar su rendimiento en diferentes entornos, poblaciones y escenarios, como vía más confiable para garantizar seguridad y eficacia clínica, demostrando que el modelo sea sólido mediante validación poblacional y privada. Todo ello, mediante metodologías que ofrezcan transparencia real como evidencia de desempeño, auditorías independientes y supervisión continua del modelo, para así poder confiar en explicaciones que reflejen la lógica interna evitando sesgos (Ghassemi, Oakden-Rayner y Beam, 2021).

Así, frente a la antes mencionada XAI, buscando desarrollar sistemas de IA cuyos resultados puedan ser comprendidos e interpretados por los usuarios, el aumento del uso de algoritmos opacos de IA en áreas de alto riesgo, como la atención médica, ha acelerado la imperiosa necesidad de la categoría llamada "explicabilidad", para evitar graves riesgos y aumentar la confianza de los profesionales de salud y los pacientes en estos sistemas, por sobre todo en contextos críticos.

Para ello, y específicamente en el ámbito de la salud la más actualizada bibliografía demanda que la XAI debería enfocarse en la importancia de la transparencia y la interpretabilidad de los modelos algorítmicos categorizando su metodología en seis áreas: 1) Analizar la contribución de cada característica en las decisiones del modelo y su influencia en la predicción; 2) Explicar el comportamiento del modelo en conjunto; 3) Explicar la forma en que el algoritmo usa o aplica conceptos humanos; 4) Explicar la forma en que ciertos modelos simples se implementan como aproximación de modelos complejos; 5) Identificar las contribuciones en los datos provenientes de imágenes y 6) Explicar comprensiblemente todos los anteriores puntos para los usuarios no expertos, destacando su rol en decisiones clínicas críticas (Sadeghi y otros, 2024).

Estos seis lineamientos, como criterio de transparencia, permitirán a los profesionales de la salud comprender qué factores específicos, como síntomas o resultados de pruebas, contribuyen a un diagnóstico particular. Proporcionará también una visión general del modelo completo en lugar de centrarse en un solo caso, identificando patrones en poblaciones o subgrupos específicos, lo cual es útil para estudios poblacionales en salud o para comprender tendencias en grandes con-

juntos de datos clínicos. Ayudará además a identificar patrones clínicos complejos, como la sensibilidad de ciertas características a diagnósticos específicos. Hará más comprensible los árboles de decisión facilitando la interpretación sin renunciar a la precisión, pudiendo explicar decisiones individuales en sistemas complejos sin comprometer la fiabilidad de las predicciones. Se podrán identificar regiones específicas en imágenes médicas, como radiografías, ecografías o tomografías, que son relevantes para la decisión del modelo, permitiendo a los médicos validar visualmente los resultados, por ejemplo, al identificar áreas afectadas en imágenes de diagnóstico. Por último, facilitará al mismo tiempo la comprensión a usuarios no expertos, lo cual es esencial para la comunicación con pacientes y profesionales no técnicos. Estos seis lineamientos aplicados sistemática y metodológicamente, contribuyen a lograr que las decisiones de IA sean transparentes y de fácil interpretación para fomentar la confianza y la adopción en entornos médicos.

4. LA NECESIDAD DE TRANSPARENCIA ALGORÍTMICA CONTEXTUALIZADA Y SUS MECANISMOS

Con lo mencionado hasta aquí, resulta claro que la toma de decisiones algorítmicas no sólo influye en decisiones puntuales, sino que desencadena otras acciones y comportamientos que retroalimentan el sistema, creando una dinámica que amplifica sus efectos, generando consecuencias imprevistas si no se implementan efectivos mecanismos para lograr dicho monitoreo. Por ello, no basta con exigir la divulgación de la existencia y uso de un algoritmo para determinada aplicación, sino que es crucial considerar qué información se difunde, cómo se comunica y a quién está dirigida como método de transparencia. Y ello se aplica a toda utilización de sistemas algorítmicos, sea en medicina, seguridad, justicia, educación, entretenimiento o en cualquier otra plataforma de medios sociales o de recomendación de noticias, ya que tienen un impacto que va más allá de la simple selección de contenido. Estos sistemas pueden perpetuar la desinformación o fomentar la polarización social, al priorizar contenidos que generan más interacción sin que se tenga en cuenta la calidad y veracidad de la información. Esto muestra cómo las decisiones algorítmicas se propagan a través de un ecosistema de datos, decisiones y comportamientos de usuarios, impactando de manera acumulativa.

Pero en sectores críticos como la salud, se ha visto cómo un error algorítmico en la toma de decisiones puede no sólo afectar al paciente directamente, sino también influir en la asignación de recursos, en la priorización de tratamientos y el desarrollo de políticas de salud pública.

Es por ello que, la actual bibliografía especializada propone que además de las antes mencionadas seis categorizaciones para implementar la transparencia algorítmica, esta debe ser además entendida de manera diferenciada según el contexto en que se aplique, ya que no todas las audiencias necesitan el mismo tipo de información ni el mismo nivel de detalle. El enfoque debe ser pragmático, balanceando la necesidad de comprensión por parte del público con la viabilidad técnica de la divulgación. En el análisis sobre cualquiera de los algoritmos de salud y judiciales mencionados, la simple divulgación de que está siendo utilizado no es suficiente para cumplir con los principios de transparencia. Los médicos y pacientes, jueces, abogados y partes implicadas necesitan información detallada sobre los criterios utilizados por el algoritmo, para generar los resultados pretendidos y su margen de error. De lo contrario, las decisiones basadas en el algoritmo carecen de un sustento verificable, lo que compromete los principios bioéticos mencionados además de los propios deontológicos de cada profesión.

5. TRANSPARENCIA RESPECTO DEL "QUÉ" Y DEL "CÓMO" EN LOS ALGORITMOS.

Definida la transparencia como la disponibilidad de información respecto de un agente permitiendo a otros monitorear los trabajos o ejecuciones de aquel (Meijer, 2014; Fox, 2010), básicamente la transparencia se trata de información y su relación con el procedimiento y el resultado. Por ello, la transparencia algorítmica debe considerarse desde dos perspectivas complementarias: la transparencia de los resultados y la transparencia del proceso. Ambos tipos de transparencia responden también a diferentes necesidades y audiencias, y se enfrentan a desafíos distintos.

La transparencia de los resultados hace referencia a la difusión y comunicación de las decisiones, recomendaciones o predicciones que produce un sistema algorítmico. En los contextos mencionados como la salud o la justicia, el acceso a esta información permite a las partes afectadas conocer los efectos específicos de las decisiones algorítmicas sobre sus vidas y derechos. Sin una explicación clara del "cómo" o la forma en que se generan las puntuaciones, calificaciones o variables que resultan en un índice, el "qué" carecerá de sustento para proporcionar justicia o equidad.

Por ello, se deberá proporcionar a los usuarios una interfaz comprensible donde puedan ver de forma clara y detallada los resultados del algoritmo y cómo afectan su caso particular. Resultados técnicos que deberán traducirse en información

clara y útil, entendible para los usuarios finales, sean profesionales, pacientes o toda parte interesada.

La transparencia del proceso, por otro lado, implica comunicar y difundir la forma en que se diseña, desarrolla y opera el sistema algorítmico, incluyendo la lógica detrás de sus cálculos, el origen de los datos utilizados y los procedimientos de ajuste o entrenamiento. Este aspecto de la transparencia es crucial para una evaluación ética completa, ya que permite analizar si los métodos utilizados en el desarrollo y entrenamiento del algoritmo cumplen con los principios de justicia y no maleficencia.

La falta de transparencia sobre el proceso de entrenamiento del algoritmo ha suscitado frecuentemente críticas por utilizar bases de datos sesgadas o parciales, afectando desproporcionadamente a ciertos grupos sociales. Así, no sólo se debe atender a las decisiones que representan el "qué", sino también a la falta de transparencia en el proceso de desarrollo y entrenamiento, es decir, el "cómo", porque dificulta detectar y corregir estos sesgos. Para ello, las auditorías llevadas a cabo por terceros independientes, sobre los datos y el diseño del modelo algorítmico permiten identificar y mitigar los sesgos en las primeras etapas del desarrollo. Luego, las compañías de desarrollo de IA deberían regularse bajo normativas debiendo comunicar a las entidades encargadas de supervisión y control, los orígenes de los datos y los métodos empleados en el desarrollo del algoritmo, explicando cómo fueron seleccionados, depurados y procesados. Y ello, bajo una documentación exhaustiva del diseño, ajustes y actualizaciones del algoritmo, siendo accesible a profesionales capacitados en ética y auditoría tecnológica.

Se observa así que ambos enfoques de transparencia, sobre el qué y el cómo, son complementarios y necesarios para crear un marco de responsabilidad efectivo para luego posibilitar una rendición de cuentas. Mientras que la transparencia del "qué" permite a los usuarios conocer el impacto de las decisiones algorítmicas, la transparencia del "cómo" es fundamental para que reguladores, expertos y auditores comprendan y evalúen los principios y métodos empleados en el algoritmo.

Un sistema de IA que prediga la probabilidad de éxito de distintos tratamientos debe, por un lado, hacer accesibles los resultados a los médicos y pacientes para que tomen decisiones informadas, transparentando el "qué". Al mismo tiempo, debe ser auditado y documentado en cuanto a sus bases de datos, ajustes y modificaciones, para que se entienda el proceso de predicción, transparentando el "cómo". Este enfoque completo ayuda a mitigar riesgos, como los frecuentes sesgos o informaciones parciales o desactualizadas que resultan en serios errores en las recomendaciones de tratamiento.

De aquí, puede concluirse que la transparencia algorítmica no sólo debe referir a los resultados y a los procesos, sino que además no puede ser total ni absoluta, debido a la complejidad de los sistemas y la necesidad de también proteger la propiedad intelectual. Sin embargo, se pueden implementar mecanismos efectivos que logren un equilibrio entre la transparencia y los intereses comerciales. En particular, la transparencia del "qué" empodera a los usuarios afectados, mientras que la transparencia del "cómo" permite a los expertos evaluar y verificar los parámetros bioéticos y legales del algoritmo. Al adoptar ambos enfoques, se puede promover un uso responsable de los algoritmos que respete los derechos y la dignidad de los individuos.

6. SUGERENCIAS PARA UNA TRANSPARENCIA EFECTIVA MEDIANTE UN ENFOQUE MULTINIVEL

La sociedad debe saber qué sistemas algorítmicos se están utilizando y en qué contextos. Esta divulgación puede hacerse a través de portales de transparencia o informes anuales. Dicha información deberá contener criterios de explicabilidad dirigida a usuarios expertos, donde aquellos profesionales directamente involucrados necesitan explicaciones más detalladas sobre los factores que influyen en las decisiones algorítmicas y cómo se ponderan los diferentes datos. Un sistema de auditorías independientes, deberá tener acceso periódico a los detalles técnicos y códigos para verificar el comportamiento del algoritmo y su alineación con criterios bioéticos y legales.

Cabe destacar que la transparencia no implica que toda la información sobre el funcionamiento de los algoritmos debe ser divulgada al público en general, ya que hacer público excesivos detalles técnicos puede resultar contraproducente, dado que el público puede no estar capacitado para entender aspectos complejos de la programación algorítmica, confundiendo o malinterpretando la información. Además, se debe respetar el derecho a la propiedad intelectual del desarrollador de los algoritmos. Nicholas Diakopoulos explica que el límite para ello es el mismo que ocurre en las inspecciones de las cocinas de los restaurantes, donde no son los comensales sino los expertos delegados por parte de entidades gubernamentales a cargo de bromatología y salubridad, quienes evalúan la limpieza, la seguridad alimentaria y otros aspectos del establecimiento para asegurar que cumpla con los estándares establecidos para habilitar la actividad gastronómica. Sin embargo, dichos expertos no demandan de los restaurantes que les revelen sus recetas o secretos comerciales para demostrar que están cumpliendo con las normativas que regulan el servicio gastronómico. En este contexto, la transparencia se refie-

re a proporcionar suficiente información sobre las prácticas del restaurante tal como la higiene y la manipulación de alimentos, sin comprometer su propiedad intelectual.

De manera similar, la transparencia en los algoritmos de IA debería centrarse en la explicabilidad de cómo estos algoritmos toman decisiones y en los datos que utilizan, sin necesidad de revelar su "receta" o los detalles técnicos internos. Esto permitiría a los usuarios y reguladores comprender cómo funcionan los algoritmos y asegurar que operan dentro de una normativa que garantiza el cumplimiento de estándares legales, bioéticos y de responsabilidad, sin develar información que podría ser sensible o comercialmente competitiva. La idea es fomentar la confianza en los sistemas de IA mediante la implementación de estándares de transparencia que no comprometan la innovación o la privacidad. Aquella analogía con el restaurante ayuda a enfatizar la importancia de un equilibrio entre la transparencia necesaria para la rendición de cuentas y la protección de los intereses comerciales y técnicos en el ámbito de la inteligencia artificial (Diakopoulos, 2016; 2019).

Por ello, se sugiere la creación de comités de expertos independientes que, con un acceso privilegiado a los algoritmos, puedan evaluar su desempeño y emitir dictámenes que además ayuden tanto a los usuarios finales como a los responsables de la política pública a comprender los riesgos y limitaciones sin comprometer los secretos comerciales. De esta manera, la divulgación puede enfocarse en garantizar que los principios jurídicos, bioéticos y los derechos de los individuos sean respetados, sin comprometer la viabilidad comercial o la competitividad de las empresas tecnológicas.

Básicamente, el acento radica en que la transparencia algorítmica debe implementarse de manera pragmática, considerando los diferentes niveles de acceso y comprensión de los implicados. Un enfoque matizado de la transparencia podría facilitar un mejor equilibrio entre los principios bioéticos y las demandas comerciales y técnicas. El establecimiento de mecanismos de supervisión externos, la participación de auditores independientes y la divulgación de criterios clave para el funcionamiento del algoritmo, son pasos cruciales para mitigar los riesgos bioéticos asociados con la opacidad algorítmica.

Pero para todo ello, y dado que la transparencia afecta directamente la rendición de cuentas, debe implementarse un marco para evaluar la responsabilidad en la toma de decisiones automatizadas, mediante una legislación proactiva, desarrollando un marco normativo que obligue a las empresas tecnológicas de desarrollo y entidades usuarias u operadores a hacer públicos los modelos algorítmicos y los datos utilizados para entrenarlos, al menos en los sectores críticos como la salud y

seguridad. En el marco de una regulación de políticas públicas, bien cabe la recomendación específica para crear estándares para la transparencia algorítmica y la implementación de auditorías regulares de los sistemas algorítmicos, para garantizar que cumplan con los principios normalizados sometiéndose a evaluaciones de impacto antes de su implementación, asegurando que se consideren no sólo la integridad de los principios bioéticos mencionados sino también derechos como la no discriminación, la privacidad y la no deshumanización de los procesos de toma de decisiones. A estos efectos, las instituciones deberían establecer comités de ética compuestos por expertos en bioética, inteligencia artificial y derechos humanos, quienes supervisen la creación y uso de algoritmos en sectores como la medicina, la educación y el derecho (Kroll y otros, 2017).

CONCLUSIÓN

Para evitar los daños y riesgos mencionados, es crucial que los sistemas algorítmicos sean diseñados bajo un marco normativo con criterios claros de explicabilidad, responsabilidad y rendición de cuentas, todo lo cual tiene como base la transparencia. Ello coadyuva a un enfoque de intervención humana tan significativa como esencial, lo cual demanda que los sistemas automáticos no deben ser la única autoridad en la toma de decisiones críticas. Los humanos deben poder supervisar, cuestionar y revocar las decisiones algorítmicas cuando sea necesario. Además, es fundamental establecer sistemas de responsabilidad algorítmica, donde los desarrolladores y operadores del algoritmo puedan ser responsabilizados acorde a lo que les corresponda, por resultados y más aún por decisiones incorrectas.

REFERENCIAS BIBLIOGRÁFICAS

Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016). "Machine bias". ProPublica (Mayo, 23).

Cavallo, J. (2019). "Confronting the criticisms facing Watson for Oncology". The ASCO Post, 10 Sep.

Char, D. S., Shah, N. H., y Magnus, D. (2018). "Implementing machine learning in healthcare: Addressing ethical challenges". En The New England Journal of Medicine, 378(11), 981-983.

Diakopoulos, N. (2016). "Accountability in algorithmic decision-making". En Communication of the ACM 59, 2, pp. 56-62.

Diakopoulos, N. (2019). "Front Matter". En Diakopoulos N. (2019). *Automating the News: How Algorithms Are Rewriting the Media* (p. [i]-[vi]). Harvard University Press.

Fox, J. (2010). "The uncertain relationship between transparency and accountability". En Development in Practice 17 (4), 663-671.

Ghassemi, M., Oakden-Rayner, L., y Beam, A. (2021). "The false hope of current approaches to explainable artificial intelligence in health care". En The Lancet Digital Health 3 (11), e745-e750.

Kroll, J., Huey, J., Barocas, S., Felten, E., Reidenberg, J., Robinson, D. y YU, H. (2017). "Accountable algorithms". En University of Pennsylvania Law Review, 165, 633-705.

Meijer, A. (2014). "Transparency". En Bovens, M., Goodin, R., Shillemans, T. (2014). *The Oxford Handbook of Public Accountability*. OUP, pp. 507-524.

Obermeyer, Z., Powers, B., Vogeli, C., y Mullainathan, S. (2019). "Dissecting racial bias in an algorithm used to manage the health of populations". En Science, 366(6464), 447–453.

Obermeyer, Z., y Emanuel, E. J. (2016). "Predicting the future-big data, machine learning, and clinical medicine". En The New England Journal of Medicine, 375(13), 1216-1219.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.

Sadeghi, Z., Alizadehsani, R., Akif, M., Kausar, S., Rehman, R., Mahanta, P., Kumar Bora, P., Almasri, A., Alkhawaldeh, R. Hussain, S., Alatas, B., Shoeibi, S. Moosaei, H., Hladík, M., Nahavandi, N., Pardalos, P. (2024). "A review of explainable artificial intelligence in healthcare". En Computers and Electrical Engineering, 118 (A), 1-20.

Taeyoung P., Philip G., Chang-Hee K., Kwang Taek K., Kyung J. C., Tea Beom K., Han J., Sang Jin Y., Jin Kyu O. (2023). "Artificial intelligence in urologic oncology: the actual clinical practice results of IBM Watson for Oncology in South Korea". En Prostate International, 11(4), 218-221.

Zou FW, Tang YF, Liu CY, Ma JA, Hu CH. (2020). "Concordance study between IBM Watson for Oncology and real clinical practice for cervical cancer patients in China: A retrospective analysis". En Frontiers in Genetics, 11(200), 2-8.