

# Desafiando los Sesgos Algorítmicos: Un Análisis Crítico de las Implicaciones Éticas

Andrea Valentina Cubillos Pinto, Fundación Universitaria Konrad Lorenz, Facultad de Matemáticas e Ingenierías, Bogotá, Colombia  
*Andreav.cubillosp@konradlorenz.edu.co*

**Resumen:** El documento aborda una preocupación creciente en la era digital: los riesgos de perpetuar sesgos y discriminación a través del uso de algoritmos de inteligencia artificial (IA). A medida que estas tecnologías se vuelven más sofisticadas, existe un peligro latente de que amplíen y codifiquen los prejuicios y desigualdades sistémicas presentes en los datos utilizados para su entrenamiento, erosionando derechos fundamentales como la privacidad y la autodeterminación informativa.

Se analizan dos casos ilustrativos que ponen de manifiesto estos desafíos: el documental 'Coded Bias' expone cómo los sistemas de reconocimiento facial comerciales tienen dificultades para identificar correctamente a personas de piel oscura, especialmente mujeres, debido a los sesgos en sus datos de entrenamiento. Por otra parte, el caso legal Wisconsin State vs. Loomis cuestiona la imparcialidad del algoritmo predictivo COMPAS, utilizado para determinar sentencias penales, ante posibles sesgos raciales codificados.

Ambos ejemplos resaltan inquietudes éticas modulares respecto al desarrollo e implementación de la IA, como la falta de transparencia algorítmica, la integración de datos sesgados que perpetúan discriminaciones históricas y la posible transgresión de derechos fundamentales y garantías procesales.

El documento realiza un análisis ético exhaustivo, abordando aspectos cruciales como la transparencia y explicabilidad, los sesgos y la discriminación, los derechos humanos, el diseño ético desde las primeras etapas para mitigar sesgos e incorporar diversidad, y la necesidad imperiosa de regulación y supervisión.

Se concluye enfatizando la urgencia de establecer un marco regulatorio sólido y lineamientos éticos estrictos que rijan el desarrollo y uso de algoritmos predictivos, particularmente en ámbitos sensibles como el sistema judicial, a fin de garantizar equidad, transparencia y respeto irrestricto por los derechos humanos.

**Index Terms:** Sesgos algorítmicos, inteligencia artificial, reconocimiento facial, sistema penal, discriminación, ética, transparencia, derechos humanos, diseño ético, regulación.

## 1. Introducción

En la era actual, la inteligencia artificial (IA) y los algoritmos han permeado de manera omnipresente diversos ámbitos de nuestras vidas, desempeñando un papel fundamental en la toma de decisiones. Sin embargo, a medida que estas tecnologías se vuelven más sofisticadas y avanzadas, surge una preocupación creciente sobre los sesgos inherentes que pueden perpetuar y amplificar la discriminación existente en nuestra sociedad.

Por ello, la proliferación de tecnologías avanzadas plantea desafíos significativos en la protección de datos personales y la preservación de derechos fundamentales [1]. Es por ello que

---

la problemática de la dependencia algorítmica y los sesgos algorítmicos se vuelve crucial para garantizar la privacidad y la autodeterminación informativa de los individuos en un entorno cada vez más dataísta y algorítmocentrista.

Según [1], la protección de datos y la autodeterminación informativa son aspectos fundamentales que requieren una atención especial para garantizar la privacidad y los derechos de los individuos en un entorno tecnológico en constante evolución, por lo que es esencial realizar estudios detallados de los actores involucrados en estos procesos tecnológicos y establecer medidas que limiten la influencia de intereses incompatibles, evitando así la corrupción en la implementación de cambios tecnológicos disruptivos.

En este contexto, el documental 'Coded Bias' de 2020, dirigido por Shalini Kantayya, aborda esta problemática de manera contundente. La película expone cómo los algoritmos pueden reflejar y reforzar los sesgos históricos y sistémicos presentes en los datos utilizados para entrenarlos. Se centra principalmente en el trabajo de investigadores como Joy Buolamwini, quien descubrió que los sistemas de reconocimiento facial comerciales tenían dificultades para identificar correctamente a las personas de piel oscura, especialmente a las mujeres, un problema que se remonta a los conjuntos de datos sesgados utilizados para entrenar estos algoritmos.

Uno de los casos más emblemáticos que ilustra los riesgos de los sesgos algorítmicos es el de Wisconsin State vs Loomis . En este caso, Eric Loomis fue sentenciado en base a una evaluación de riesgo realizada por el algoritmo COMPAS (Corrective Offender Management Profiling for Alternative Sanctions), utilizado por el sistema de justicia penal. Loomis argumentó que el uso de este algoritmo violaba sus derechos constitucionales debido a los posibles sesgos raciales presentes en el mismo [2].

En definitiva, el caso Wisconsin State vs Loomis pone de manifiesto la relevancia y urgencia de abordar los sesgos algorítmicos, especialmente en contextos sensibles como el sistema de justicia penal, donde las decisiones pueden tener un impacto significativo en la vida de las personas. Por ende, a medida que avanzamos hacia un mundo cada vez más dependiente de la IA, es crucial desarrollar marcos éticos y regulatorios que garanticen la transparencia, la rendición de cuentas y la protección de los derechos fundamentales en el diseño y uso de estas tecnologías.

## 2. Desarrollo

El documental 'Coded Bias' destaca cómo los sistemas de reconocimiento facial pueden reflejar y perpetuar sesgos históricos y sistémicos presentes en los datos utilizados para entrenarlos. Por ello, la investigadora Joy Buolamwini pone de manifiesto una preocupación clave: **"los sistemas de reconocimiento facial basados en inteligencia artificial no son imparciales, sino que incorporan los sesgos y prejuicios de los desarrolladores que los crearon"**.

Esta afirmación subraya el hecho de que, a menos que se aborden activamente los sesgos presentes en los datos y los procesos de entrenamiento, los algoritmos de IA pueden terminar codificando y perpetuando las desigualdades y discriminaciones existentes en la sociedad, en lugar de ser herramientas verdaderamente objetivas e inclusivas.

En consecuencia, el documental expone una problemática fundamental: **cuando se emplean conjuntos de datos sesgados para entrenar sistemas de inteligencia artificial, el resultado inevitable es que dichos sistemas perpetuarán e incluso acentuarán los sesgos y desigualdades presentes en esos datos**. En consecuencia, estas tecnologías emergentes, lejos

de ser herramientas imparciales y equitativas, podrían tener un impacto discriminatorio al limitar significativamente su usabilidad y accesibilidad para ciertos grupos históricamente marginados o infrarrepresentados.

De esta manera, en lugar de mitigar las injusticias sociales, la IA mal diseñada corre el riesgo de exacerbar la exclusión sistemática que ya enfrentan algunas comunidades vulnerables. Es un llamado de atención sobre la urgente necesidad de abordar los sesgos algorítmicos desde su raíz para garantizar que la revolución tecnológica no deje a nadie atrás.

Por lo tanto, los riesgos de codificar sesgos discriminatorios a través del uso acrítico de algoritmos de inteligencia artificial han sido puestos de relieve tanto en documentales como en casos legales reales como se evidencio en el premiado documental 'Coded Bias' y el caso judicial Wisconsin State vs Loomis, los cuales, comparten una preocupación fundamental: **estos sistemas automatizados pueden proyectar una imagen engañosa de imparcialidad, ocultando sesgos y desventajas profundamente arraigados a menos que se desarrollen y examinen con sumo rigor.**

En 'Code Bias', se revelan las consecuencias potencialmente dañinas de los sesgos presentes en los datos que alimentan sistemas de reconocimiento facial, mientras que el caso Loomis cuestiona la imparcialidad de un algoritmo de evaluación de riesgos utilizados para determinar sentencias. Ambos ejemplos subrayan cómo estos procesos automatizados pueden vulnerar derechos fundamentales y perpetuar patrones de discriminación sistémica si no se los somete a un riguroso escrutinio público.'

En el mismo orden de ideas, uno de los desafíos que plantean es la falta de transparencia y rendición de cuentas que rodea a muchos de estos sistemas. Al tratarse de cajas negras.<sup>3</sup>Algorítmicas, resulta extremadamente difícil evaluar qué sesgos han sido codificados, dilucidar la lógica subyacente de las decisiones automatizadas, y determinar responsabilidades cuando se producen resultados discriminatorios. Esta opacidad socava principios básicos como el debido proceso y el derecho a un juicio justo.

Ambos casos presentan los desafíos de la inteligencia artificial (IA) en la experiencia de usuario (UI) y en el diseño de interfaces (UX) al abordar temas como la discriminación algorítmica, los sesgos en los datos y la falta de diversidad en el desarrollo de tecnologías. Si bien 'Coded Bias' y el caso Loomis abordan problemáticas diferentes, ambos ponen de manifiesto desafíos éticos y técnicos fundamentales en torno al uso de sistemas de inteligencia artificial.

No obstante, cada uno enfatiza aspectos distintos de esta problemática común. Mientras que 'Coded Bias' se enfoca en la urgente necesidad de promover una mayor diversidad en el desarrollo tecnológico para mitigar sesgos desde su raíz, el caso Loomis explora las implicaciones legales y éticas del uso de algoritmos opacos en el sistema de justicia penal.

En contraste, la película documenta los efectos discriminatorios potenciales de la IA sesgada en ámbitos tan variados como la vigilancia en las calles, la imposición del registro en la entrada y salida de ciertas viviendas, beneficios en las contrataciones de personal exclusivamente para los hombres y, demás situaciones que atentan contra los derechos de las personas. Por su parte, el caso judicial examina cómo un algoritmo predictivo específico podría vulnerar garantías constitucionales al utilizarse para determinar sentencias.

En esencia, ambos ejemplos ilustran los riesgos de implementar sistemas de IA sin las debidas salvaguardas técnicas y éticas. Abordan dimensiones complementarias de un mismo desafío crucial: **¿cómo desarrollar y desplegar estas poderosas tecnologías emergentes de manera**

responsable, equitativa y respetuosa en presencia de los derechos humanos fundamentales?

### 3. Análisis Ético

Por otra parte, se pueden identificar varias implicaciones éticas relevantes que afectan el diseño y la percepción de la tecnología, específicamente en el contexto del uso de algoritmos en el sistema penal y de reconocimiento facial. A continuación, se presenta un análisis ético detallado:

1. **Transparencia y explicabilidad:** La falta de transparencia en el funcionamiento de los algoritmos utilizados en el sistema penal y de reconocimiento facial es una preocupación ética fundamental. La opacidad en cómo se toman las decisiones automatizadas y qué variables influyen en ellas socava principios básicos como el debido proceso y el derecho a un juicio justo. Esto genera desconfianza en la legitimidad de estos sistemas y afecta negativamente la percepción pública sobre su uso.
2. **Sesgos y discriminación:** Los algoritmos de inteligencia artificial no son neutrales por naturaleza, ya que inevitablemente heredan los sesgos y prejuicios presentes en los datos utilizados para su entrenamiento. Cuando estos datos reflejan patrones históricos de discriminación y marginación hacia grupos minoritarios o vulnerables, existe un alto riesgo de que los algoritmos terminen codificando y perpetuando esas mismas inequidades de manera automatizada. Esta situación representa uno de los desafíos éticos más apremiantes en el desarrollo de sistemas de IA.
3. **Derechos humanos y garantías procesales:** El despliegue de algoritmos de inteligencia artificial en ámbitos tan sensibles como el sistema penal y las tecnologías de reconocimiento facial conlleva riesgos sustanciales de socavar derechos y libertades fundamentales consagrados en los principios de derechos humanos. Desde la libertad personal hasta la privacidad e intimidad, pasando por garantías básicas de debido proceso, el uso desregulado e inadecuadamente supervisado de estos sistemas automatizados amenaza con vulnerar un amplio espectro de prerrogativas ciudadanas esenciales.
4. **Diseño ético y responsabilidad:** Abordar las profundas implicaciones éticas que plantea el desarrollo de algoritmos predictivos y sistemas de reconocimiento facial requiere adoptar un enfoque ético integral desde las etapas iniciales de diseño de estas tecnologías disruptivas. Esta aproximación ética por diseño debe apuntar a múltiples frentes modulares: en primer lugar, implementar rigurosos protocolos para identificar y mitigar cualquier sesgo presente en los datos de entrenamiento, a fin de evitar la codificación de prejuicios y discriminaciones sistémicas. En segunda instancia, es imperativo garantizar altos estándares de transparencia algorítmica que permitan comprender la lógica subyacente del proceso de decisión automatizado y las variables que lo moldean. Asimismo, deben diseñarse mecanismos sólidos de rendición de cuentas y auditoría externa para dilucidar responsabilidades cuando se producen resultados dañinos o contrarios a la ética.

Finalmente, la incorporación de una multiplicidad de voces, perspectivas y experiencias en los equipos de desarrollo resulta indispensable para lograr soluciones tecnológicas verdaderamente inclusivas y equitativas. En esencia, la ética debe estar íntimamente entrelazada en el ADN de estos sistemas para asegurar el respeto a los derechos humanos de quienes se vean afectados por sus determinaciones. Sólo mediante un enfoque ético holístico se podrá revertir la desconfianza ciudadana y cultivar una percepción pública positiva en torno

al extraordinario potencial transformador de la inteligencia artificial.

5. **Regulación y supervisión:** Dada la naturaleza potencialmente dañina del uso indebido de algoritmos en contextos sensibles como el sistema penal y de reconocimiento facial, es necesario contar con un marco regulatorio claro y una supervisión adecuada. Esto implica establecer pautas y lineamientos éticos para el desarrollo y aplicación de estas tecnologías, así como mecanismos de control. Una regulación sólida puede mejorar la percepción pública y aumentar la confianza en el uso ético y responsable de estos sistemas.

## 4. Conclusiones

- El uso de algoritmos predictivos en el sistema penal, como el caso del algoritmo COMPAS evaluado en *Wisconsin State vs Loomis*, plantea serios riesgos éticos y legales si no se aborda adecuadamente la falta de transparencia y los sesgos inherentes a estos sistemas.
- Es imperativo establecer un marco regulatorio sólido y lineamientos éticos claros para el diseño y uso de algoritmos predictivos en el sistema judicial, a fin de garantizar la equidad y la transparencia en las decisiones automatizadas.
- Se debe promover un enfoque de 'ética por diseño' en el desarrollo de estos sistemas de IA, identificando y mitigando sesgos desde la raíz, asegurando la aplicabilidad algorítmica, e incorporando una diversidad de perspectivas en el proceso.
- Es necesario mantener un escepticismo saludable y un escrutinio constante sobre el uso de tecnología predictiva en sentencias judiciales, para evitar replicar injusticias sistémicas y preservar el discernimiento y valoración humana en la toma de decisiones críticas.
- El documental 'Coded Bias' expone cómo los algoritmos, como los sistemas de reconocimiento facial, pueden reflejar y reforzar los sesgos históricos y sistémicos presentes en los datos utilizados para entrenarlos, lo que puede resultar en discriminación hacia ciertos grupos, especialmente personas de piel oscura.
- Se requiere un enfoque ético integral desde las etapas iniciales del diseño de estas tecnologías disruptivas, implementando protocolos para identificar y mitigar sesgos.

---

## Referencias

- [1] J. C. Faliero, "Inteligencia artificial, sesgos y derecho al anonimato," *Nueva Sociedad*, no. 294, pp. 120–129, Aug. 2021. [Online]. Available: <https://biblat.unam.mx/hevila/Nuevasociedad/2021/no294/11.pdf>
- [2] M. del P. Roa Avella, J. E. Sanabria-Moyano, and K. Dinás-Hurtado, "Uso del algoritmo compas en el proceso penal y los riesgos a los derechos humanos," *Revista Brasileira de Direito Processual Penal*, vol. 8, no. 1, pp. 275–310, 2022.