

# The Algoritm

Octubre, 2024

Una revisión de la ética y la rendición de cuentas algorítmicas

## Enfoques

### Perspectivas

Un enfoque más científico y académico en el estudio de algoritmos.

### Consideraciones éticas

Perspectivas críticas

Dilemas éticos

Rendición de cuentas

La frontera entre ética y algoritmia

### La Justicia

Sistemas automatizados de toma de decisiones



## “Tono Formal”

Miroslava Kostova Karaboytcheva

### Índice:

1. Señalar los datos
2. Explicabilidad de los algoritmos
3. Responsabilidad y rendición de cuentas
4. Privacidad y uso ético de los datos
5. Desigualdad tecnológica
6. Interacción humana
7. Algoritmos adaptativos
8. Ética de la automatización en la toma de decisiones críticas
9. Transparencia frente a complejidad técnica
10. Responsabilidad compartida en ecosistemas algorítmicos
11. Efecto en la economía laboral y la distribución de poder
12. Desinformación y manipulación a gran escala
13. El desafío de la justicia algorítmica



# Introducción

## Retos para solucionar la gestión de inputs

Abrir los algoritmos y hacerlos transparentes en algunos casos puede facilitar el "juego" o manipulación por parte de los agentes que van a ser valorados por ellos. Eso no preocupa si los datos de entrada no son manipulables. Sería el caso de un algoritmo para decidir la asignación de viviendas sociales que depende de variables que son objetivas y demostrables.

Sin embargo, la cosa se complica con, por ejemplo, el algoritmo de búsqueda de Google. A todos nos parecería justo que un mecanismo distribuidor de reputación tan brutal como ese fuera más transparente, pero si supiéramos con precisión cómo decidir los rankings, además de descubrir que hay intereses comerciales opacos en sus reglas de decisión, sabríamos cómo optimizar nuestros propios resultados. En un escenario así, aumentan las probabilidades de que sobrevivan los más fuertes.

Un segundo reto, complicado de solucionar, es que, si se aceptara auditar los algoritmos,

- ¿quién controla al controlador?
- ¿Cómo podemos asegurarnos que el dictamen de un auditor es confiable e independiente y no está a su vez sesgado?

Lo que me queda claro, en todo caso, es que ningún sistema de toma de decisiones va a ser perfecto así que tendremos que elegir el menos malo. Tengo cosas favorables que decir de los algoritmos si se usan como lo que son:

- "Aproximaciones humildes de la realidad que pueden ayudar a prevenir decisiones arbitrarias -por eso de que "
- "los algoritmos no tienen amigos"

Pero esto es siempre y cuando se hagan acompañar de una segunda capa de interpretación humana que corrija posibles desviaciones que producen injusticias irreparables.

Además de estos puntos, hay otros aspectos clave que complican la gestión de datos y el uso de algoritmos.

## 1. Señalar los datos

El problema no solo está en la transparencia del algoritmo, sino también en los datos que lo alimentan.

Los algoritmos entrenados con datos sesgados (intencional o inadvertidamente) replicarán esos sesgos, perpetuando o incluso amplificando las desigualdades existentes.

Por ejemplo, si un algoritmo de contratación se entrena con datos históricos donde hay un sesgo hacia contratar a hombres en lugar de mujeres, el algoritmo aprenderá a replicar ese sesgo, independientemente de la transparencia del proceso.

Un reto adicional es identificar y corregir estos pasos sin que afecten la funcionalidad o la precisión del algoritmo.

---

## 1. Señalar los datos:

El problema no solo está en la transparencia del algoritmo, sino también en los datos que lo alimentan. Los algoritmos entrenados con datos sesgados (intencional o inadvertidamente) replicarán esos sesgos, perpetuando o incluso amplificando las desigualdades existentes. Por ejemplo, si un algoritmo de contratación se entrena con datos históricos donde hay un sesgo hacia contratar a hombres en lugar de mujeres, el algoritmo aprenderá a replicar ese sesgo, independientemente de la transparencia del proceso. Un reto adicional es identificar y corregir estos pasos sin que afecten la funcionalidad o la precisión del algoritmo.

## 2. Explicabilidad de los algoritmos:

Muchos algoritmos, especialmente los basados en redes neuronales o aprendizaje profundo, son efectivamente "cajas negras", lo que significa que incluso sus creadores no siempre pueden explicar por qué un algoritmo toma una decisión específica. Esto plantea serios problemas en áreas críticas, como la justicia penal o la salud, donde es importante que las decisiones algorítmicas sean comprensibles y justificables. Un gran reto es cómo hacer que estos sistemas sean explicables sin sacrificar su

## 3. Responsabilidad y rendición de cuentas:

¿Quién es responsable si un algoritmo comete un error o toma una decisión injusta? Este es un problema crucial en el ámbito legal y ético. Las empresas que desarrollan y despliegan algoritmos pueden argumentar que solo proporcionan herramientas, pero cuando estas herramientas se utilizan para tomar decisiones importantes (por ejemplo, sobre quién recibe un préstamo o quién es arrestado), alguien debe ser responsable. Establecer mecanismos claros de responsabilidad es complejo, ya que implica no solo a los desarrolladores de algoritmos, sino también a las organizaciones que los

## 4. Privacidad y uso ético de los datos:

El concepto de **\*\*responsabilidad compartida en ecosistemas algorítmicos\*\*** se refiere a la complejidad de asignar la responsabilidad de los resultados o impactos de un algoritmo cuando está integrado en un sistema que interactúa con otros algoritmos, actores humanos y tecnologías. En lugar de un único desarrollador o entidad a cargo, la responsabilidad por cualquier sesión, error o consecuencia negativa tiende a estar distribuida entre varias partes. **### Elementos clave: 1. \*\*Interdependencia entre algoritmos\*\***: Los algoritmos a menudo no operan de forma aislada, sino que dependen de datos procesados por otros sistemas o están integrados en flujos de trabajo más amplios. Por ejemplo, en una plataforma de comercio electrónico, un algoritmo de recomendación puede depender de datos generados por sistemas de personalización, marketing y comportamiento del usuario. Esto significa que un fallo en uno de estos sistemas puede afectar al resultado final, y no siempre está claro quién es el responsable.

5. Regulación y políticas públicas:

---

## 6. Desigualdad tecnológica:

La **desigualdad tecnológica** se refiere a la brecha entre diferentes grupos o regiones en el acceso, uso y beneficio de las tecnologías avanzadas, como la inteligencia artificial (IA), el big data y los algoritmos. Este concepto aborda las disparidades no solo en el acceso a las herramientas tecnológicas, sino también en la capacidad para desarrollar, implementar y aprovechar estas tecnologías de manera efectiva. **Elementos clave de la desigualdad tecnológica:** 1. **Acceso desigual a la infraestructura tecnológica:** El acceso a tecnologías avanzadas, como computación en la nube, grandes conjuntos de datos o hardware especializado para el procesamiento de IA, no está distribuido de manera equitativa. Países en desarrollo, pequeñas empresas o individuos sin los recursos económicos adecuados pueden verse excluidos de estas tecnologías. Las grandes empresas tecnológicas, especialmente en países desarrollados, tienen acceso a infraestructuras poderosas, mientras que en otras regiones del mundo el acceso a Internet de alta velocidad o servicios en la nube puede ser limitado o inalcanzable.

## 7. Interacción humana con algoritmos:

Además de los desafíos ya mencionados, hay algunas reflexiones adicionales que me parecen relevantes para ampliar el panorama sobre la gestión de datos y el diseño del algoritmo.

- La interacción humana con algoritmos es un aspecto crítico que amplía la discusión sobre la gestión de datos y el diseño algorítmico.

A continuación, se destacan algunos puntos clave que profundizan en este tema:

- **Comprensión y Transparencia;** Uno de los principales retos en la interacción humana con algoritmos es que los usuarios a menudo no entienden cómo funcionan los algoritmos, lo que genera desconfianza y preocupación. Si bien los algoritmos pueden procesar grandes volúmenes de datos de manera eficiente, su opacidad puede hacer que las decisiones sean difíciles de explicar, especialmente en casos donde impactan la vida de las personas (por ejemplo, en decisiones crediticias o en recomendaciones de contenido).
- **Desafío;** Lograr un equilibrio entre la complejidad técnica de los algoritmos y la necesidad de hacerlos comprensibles para los usuarios finales.

## 8. Algoritmos adaptativos y el problema del "feedback loop":

- Cuando los algoritmos se adaptan constantemente a nuevos datos, pueden generar lo que se conoce como un bucle de retroalimentación.

---

## 9. Ética de la automatización en toma de decisiones críticas:

La expansión del uso de algoritmos para tomar decisiones en campos sensibles como la justicia, la medicina o los recursos humanos plantea cuestiones éticas profundas. Aunque los algoritmos pueden ofrecer consistencia y reducir algunos tipos de sesgo humano, también existe la preocupación de deshumanizar las decisiones críticas, especialmente en contextos donde se requiere empatía o juicio moral.

Por ejemplo, en la justicia penal, el uso de algoritmos para predecir el riesgo de reincidencia puede ser útil para tomar decisiones de libertad condicional, pero si se confía demasiado en estos sistemas, se puede perder el matiz humano y la comprensión contextual que solo un juez puede aportar.

La clave es encontrar una integración ética y equilibrada entre la automatización y el juicio.

## 10. Transparencia frente a complejidad técnica:

En muchos casos, la complejidad técnica de los algoritmos modernos, especialmente aquellos basados en aprendizaje profundo o redes neuronales, hace que la transparencia sea difícil de lograr, incluso para los desarrolladores propios. Estos algoritmos son tan complejos que, aunque compartas su código o estructura, puede que no sea comprensible para una audiencia no técnica.

Esto plantea un reto doble:

- Por un lado, está el deseo de hacer que los sistemas sean más transparentes;
- por otro lado, hay una barrera inherente en la comprensión técnica que dificulta que los usuarios (o reguladores) puedan auditar realmente el comportamiento de estos sistemas.

Esto requiere una mayor inversión en la creación de herramientas y métodos que hagan la "explicabilidad" de los algoritmos más accesibles, sin sacrificar su capacidad operativa.

## 11. La responsabilidad compartida en ecosistemas algorítmicos.

Es un enfoque que reconoce que, en sistemas complejos donde los algoritmos están integrados en una red de actores humanos, tecnológicos y organizativos, la responsabilidad de los resultados generados por estos algoritmos no puede ser asignada a una sola entidad. En lugar de esto, la responsabilidad se distribuye entre varios actores que tienen diferentes roles en el ciclo de vida del algoritmo.

Este concepto aborda los desafíos éticos, legales y operativos que surgen cuando algoritmos y sistemas automatizados interactúan con diversas partes interesadas y generan impactos que afectan a la sociedad.

Elementos clave de la responsabilidad compartida en ecosistemas algorítmicos:

- 
1. Múltiples partes involucradas
  2. En un ecosistema algorítmico típico, hay muchas partes que contribuyen al desarrollo, implementación y uso de algoritmos.

Estas partes incluyen:

- Desarrolladores: Los que crean y diseñan el algoritmo.
- Proveedores de datos: Quienes proporcionan los datos necesarios para entrenar y ajustar el algoritmo.
- Empresas o instituciones que implementan el algoritmo en sus operaciones diarias.
- Usuarios finales que interactúan con los resultados o decisiones producidas por el algoritmo
- Reguladores que establecen normativas para asegurar la equidad y la transparencia en el uso de estos algoritmos.

## **12. Desinformación y manipulación a gran escala**

El equilibrio entre innovación y regulación es un reto crucial en el ámbito tecnológico, especialmente en el desarrollo y uso de algoritmos e inteligencia artificial (IA).

- Por un lado, la innovación impulsa el crecimiento económico, la competitividad y la resolución de problemas complejos.
- Por otro lado, la regulación es esencial para asegurar que estas tecnologías se utilicen de manera ética, segura y justa, protegiendo los derechos de los ciudadanos y mitigando los posibles daños.

Este equilibrio es difícil de lograr porque las normativas tienden a ir más despacio que el ritmo de la innovación tecnológica. Una regulación demasiado puede sofocar la creatividad y ralentizar el avance, mientras que una falta de regulación adecuada puede llevar a la explotación, la desigualdad y el mal uso de las tecnologías.

Factores clave en el equilibrio entre innovación y regulación:

1. Ritmo de la innovación tecnológica:
2. La tecnología, y en particular la IA, avanza a un ritmo muy rápido, lo que crea una presión para los reguladores. Los avances en algoritmos, machine learning y big data están ocurriendo mucho más rápido que la capacidad de los gobiernos para crear y actualizar leyes que abordan los nuevos desafíos que surgen con estas tecnologías. Si los reguladores imponen normativas demasiado rígidas, las empresas y desarrolladores pueden encontrar barreras para innovar y experimentar con nuevos enfoques tecnológicos.

## **13. Efecto en la economía laboral y la distribución de poder**

El efecto en la economía laboral y la distribución de poder causado por el avance de los algoritmos, la automatización y la inteligencia artificial (IA) es un tema de profundo impacto social y económico.

---

Estos avances tecnológicos están transformando la naturaleza del trabajo, cambiando las dinámicas de poder entre empresas, trabajadores y gobiernos, y generando nuevas formas de desigualdad. Aunque las innovaciones tecnológicas prometen mejorar la productividad y abrir nuevas oportunidades, también generan preocupaciones sobre la pérdida de empleos, la concentración del poder en manos de unas pocas corporaciones tecnológicas y el cambio en las relaciones laborales.

1. Desplazamiento de empleos: La automatización y la IA tienen el potencial de reemplazar trabajos repetitivos y rutinarios, especialmente en sectores como la manufactura, la logística, el comercio minorista y la atención al cliente.
2. Según algunos estudios, hasta un 47% de los empleos en las economías desarrolladas podrían estar en riesgo de ser automatizados. Esto genera una preocupación significativa sobre el desempleo tecnológico, donde millones de personas podrían perder sus empleos debido a la automatización, especialmente en trabajos de baja cualificación.

#### **14. Desinformación y manipulación a gran escala:**

La desinformación y manipulación a gran escala es uno de los retos más complejos que los algoritmos, las plataformas digitales y las redes sociales enfrentan en la actualidad. Con la capacidad de distribuir información rápidamente ya un público masivo, los algoritmos pueden ser usados tanto para difundir noticias veraces como para amplificar información falsa o manipulada. La propagación de desinformación a través de estas plataformas tiene profundas consecuencias políticas, sociales y económicas, y puede influir en procesos democráticos, polarizar sociedades y erosionar la confianza pública en las instituciones.

1. Efectos clave de la desinformación y manipulación:
  - Amplificación algorítmica de contenidos manipulados.
  - Las plataformas digitales, como Facebook, Twitter, YouTube y otras redes sociales, utilizan algoritmos para priorizar el contenido que los usuarios ven, calculando en sus preferencias, interacciones y comportamientos. Esto crea un "efecto de cámara de eco", donde los usuarios tienden a recibir contenido que refuerza sus creencias preexistentes. Los algoritmos tienden a promover contenido altamente emocional o sensacionalista, ya que este tipo de material genera más interacciones (clics, comentarios, me gusta), lo que favorece la viralidad de la desinformación.

#### **15. El desafío de la justicia algorítmica:**

El desafío de la justicia algorítmica se refiere a los problemas éticos y sociales que surgen cuando los algoritmos, que son utilizados para tomar decisiones en diversos contextos, producen resultados sesgados, discriminatorios o injustos.

A medida que las tecnologías basadas en algoritmos, como la inteligencia artificial (IA) y el aprendizaje automático (machine

---

learning), se utilizan cada vez más en áreas como el empleo, la justicia penal, la educación, la atención médica y el crédito. Es crucial garantizar que estos sistemas sean justos y equitativos para todos los usuarios.

1. Principales retos de la justicia algorítmica:

- Sesgo algorítmico.

Los sesgos algorítmicos ocurren cuando los algoritmos producen resultados que favorecen o perjudican injustamente a ciertos grupos de personas, a menudo debido a los datos en los que fueron entrenados.

Esto puede suceder por varias razones:

Las razones para abordar el “desafío de la justicia algorítmica” son diversas y se pueden clasificar en varias categorías:

- Impacto Social y Ético Equidad
- Es fundamental asegurar que todos los individuos tienen un trato justo, independientemente de su raza, género, origen étnico o estatus socioeconómico. Las decisiones algorítmicas sesgadas pueden perpetuar o agravar desigualdades existentes en la sociedad.

Y para terminar dejar aquí en el aire que hay más razones para abordar el “desafío de la justicia algorítmica”, ampliando el análisis desde diferentes perspectivas:

- Impacto Económico
- Reducción de Costos
- La implementación de algoritmos justos puede prevenir errores costosos asociados con decisiones sesgadas, como demandas legales o daños a la reputación. Esto, a largo plazo, puede reducir los costos operativos para las empresas.

Miroslava Kostova Karaboytcheva