



Igualdad, inteligencia artificial, algoritmos y decisiones automatizadas: ¿Cómo detectar y abordar la discriminación en el contexto de la IA?

Dra. Ana Maria Corrêa
ana.correa@kuleuven.be
www.law.kuleuven.be/citip



Funded under the 'Citizens, Equality, Rights and Values programme 2021-2027' of the European Commission

Estructura

Parte I - Concepto jurídico de la IA

Parte II - Casos reales de presunta discriminación contra clases protegidas perpetrada por sistemas de IA

Parte III - Detección de la discriminación por IA

Parte IV - Lucha contra la discriminación en el ámbito de la IA

Parte I

Concepto jurídico de IA



Definición jurídica, IA

CoE, (Comité de Inteligencia Artificial) - CAI

Proyecto de Convención [Marco] sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho

Artículo 2.a - Definiciones

"**Sistema de inteligencia artificial**" significa **cualquier** sistema **algorítmico** o una **combinación** de tales sistemas que, tal como se define en el presente documento y en la legislación nacional de cada Parte, **utiliza métodos informáticos derivados de la estadística** u otras **técnicas matemáticas** para llevar a cabo funciones que comúnmente se **asocian** con la **inteligencia humana** o que, de otro modo, **requerirían inteligencia humana**, y que **ayuda** o **sustituye** el juicio de los responsables humanos de la toma de decisiones en la realización de dichas funciones. Dichas funciones incluyen, entre otras, la **predicción**, la **planificación**, la **clasificación**, el **reconocimiento de patrones**, la **organización**, la **percepción**, el **reconocimiento de voz/sonido/imagen**, la **generación de texto/sonido/imagen**, la **traducción de idiomas**, la **comunicación**, el **aprendizaje**, la **representación** y la **resolución de problemas**".

Definición jurídica, IA

UE

Definición de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas sobre la inteligencia artificial (**Ley sobre Inteligencia Artificial**) y se modifican determinados actos legislativos de la Unión

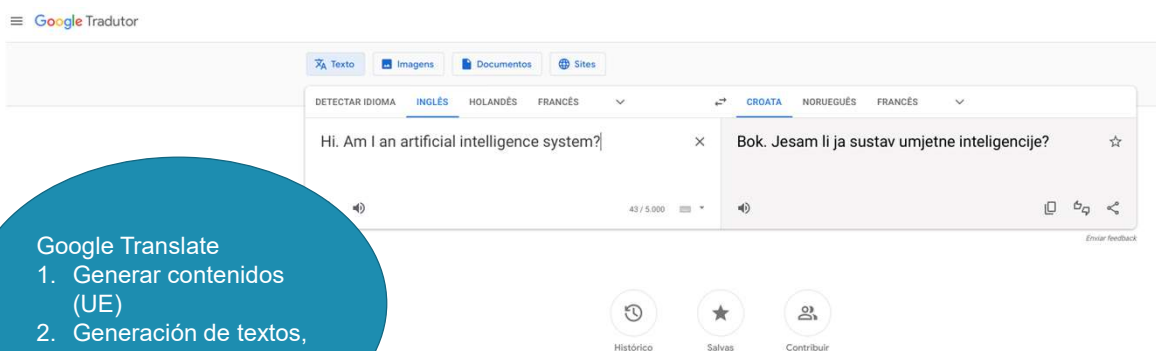
Artículo 3.1 - Definiciones

«Sistema de inteligencia artificial (sistema de IA)»: **el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa.**

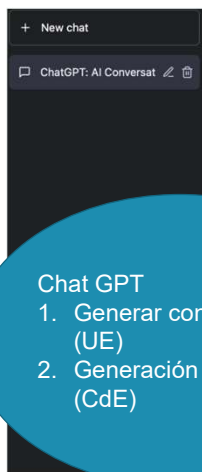
Anexo I Técnicas y enfoques de la inteligencia artificial

1. Estrategias de aprendizaje automático, incluidos el aprendizaje supervisado, el no supervisado y el realizado por refuerzo, que emplean una amplia variedad de métodos, entre ellos el aprendizaje profundo.
2. Estrategias basadas en la lógica y el conocimiento, especialmente la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencia y deducción, los sistemas expertos y de razonamiento (simbólico).
3. Estrategias estadísticas, estimación bayesiana, métodos de búsqueda y optimización.

Ejemplo, AI



Ejemplo, AI



Chat GPT

1. Generar contenidos (UE)
2. Generación de textos (CdE)

Hi. Why ChatGPT is an Artificial Intelligent System?

Hello! ChatGPT is an artificial intelligence system because it has been designed and programmed to simulate human-like conversation and provide intelligent responses to users. It uses advanced natural language processing (NLP) algorithms to understand and interpret user input, and then generates appropriate responses based on its understanding of the context and the knowledge it has been trained on.

The system has been trained on vast amounts of text data and is capable of recognizing patterns and generating appropriate responses. It can also learn and improve over time through machine learning techniques, allowing it to adapt to new situations and provide even more accurate and helpful responses.

Overall, ChatGPT is an example of the incredible progress that has been made in the field of artificial intelligence, and demonstrates how these technologies can be used to create systems that are capable of engaging in intelligent and meaningful conversations with humans.

Ejemplo, AI



Anuncios específicos en línea

1. Predicciones, recomendaciones(UE)
2. Reconocimiento de patrones(CdE)

Ejemplo, AI

Nethone Product Use cases Pricing Resources About Explore Book a call Try for free

AI-powered fraud detection software that puts your business first

Whether you want to take care of fraud prevention yourself or you wish to leave
fraud protection in the hands of professionals—we're here to help.



Evaluación de fraude

1. Predicción (UE)
2. Predicción (CdE)

9

Ejemplo, AI



Reconocimiento facial

1. Predicción (UE)
2. Predicción (CdE)

10

Parte II

Casos reales de supuesta discriminación de clases protegidas perpetrada por sistemas de IA



Sector público: Discriminación contra clases protegidas y elaboración de perfiles étnicos

- Detección del fraude a la Seguridad Social en los Países Bajos.
- El sistema algorítmico utilizó un modelo de clasificación de riesgos para estimar el riesgo de fraude e inexactitud de las prestaciones por cuidado de menores asignadas a las familias en los Países Bajos.
- En los Países Bajos, el gobierno cubre parte de los gastos de guardería. La cuantía se calcula en función de los ingresos de la familia. A las familias con ingresos bajos se les puede llegar a reembolsar hasta el 96% de los gastos de guardería, mientras que a las familias con ingresos altos se les reembolsa un 33% de los gastos de guardería.
- A grandes rasgos, el algoritmo tenía en cuenta la **nacionalidad** de las personas para crear **perfiles de riesgo** de los beneficiarios de la prestación
- Como consecuencia, miles de familias con rentas bajas y medias de origen extranjero fueron sometidas a un gran escrutinio, acusadas falsamente de fraude y obligadas a devolver las prestaciones que, en muchos casos, ascendían a decenas de miles de euros.
- Otras consecuencias para las familias: acabaron sumidas todavía más en la pobreza, se les pidió injustamente que devolvieran grandes sumas de dinero; algunas perdieron sus casas; más de 1.000 niños quedaron bajo custodia del Estado como consecuencia de las acusaciones.
- Dimisión del gobierno holandés en 2021.



XENOPHOBIC MACHINES

DISCRIMINATION THROUGH UNREGULATED USE OF ALGORITHMS IN THE DUTCH CHILDCARE BENEFITS SCANDAL



Sector privado: Discriminación contra clases protegidas



- Caso Employment Tribunals. (Reino Unido) Número de asunto - 3206212/2021.
- Uber utiliza el sistema de reconocimiento facial: "**Comprobación de identidad en tiempo real**".
- **Finalidad:** Garantizar la seguridad de los pasajeros y evitar que los conductores/transportistas compartan sus cuentas con personas no registradas en Uber.
- **Los conductores/transportistas de Uber pasan** por controles de antecedentes penales para obtener su permiso para trabajar en la plataforma.
- **Antes de empezar a trabajar**, los conductores/transportistas deben presentar un selfie.
- El software **comprueba la foto**. Si no coincide, el conductor/transportista queda fuera del sistema durante 24 horas.
- **Si se produce un error por segunda vez**, se desactiva la cuenta del conductor o transportista.
- Pa Edrissa Manjang, mensajero de Uber, presentó una demanda contra Uber, alegando que había sido discriminado por su raza. El reconocimiento facial y la comprobación humana habían fallado supuestamente en su caso. Por ello, su cuenta fue desactivada.
- "Las demandas son por **acoso** relacionado con la raza en virtud del artículo 26 de la Ley de Igualdad de 2010, **victimización** en virtud del artículo 27 de la Ley de Igualdad de 2010 y **discriminación racial indirecta** en virtud del artículo 19 de la Ley de Igualdad de 2021. El demandante es un hombre negro de ascendencia africana. Plantea sus quejas raciales haciendo referencia al color (tanto negro como no blanco) y al origen nacional/étnico (africano)."

Razones técnicas de la discriminación por IA

Formación de datos

- **Los conceptos de etiquetado** pueden dar lugar a discriminación cuando las etiquetas se correlacionan de algún modo con clases protegidas. Por ejemplo, en el sistema fiscal neerlandés, la nacionalidad se correlacionaba como factor de riesgo de fraude en las prestaciones sociales. Esto coloca a los ciudadanos no neerlandeses en una posición de desventaja frente a los neerlandeses.
- **Datos sesgados utilizados** para alimentar sistemas de IA. Por ejemplo, COMPAS, utilizado para tomar decisiones sobre la libertad provisional y la imposición de penas.
- **Identificación de sustitutos.** Cuando aspectos no protegidos, como la dirección, puedan servir de motivo de discriminación.

Parte III

Cómo detectar la discriminación por IA: Enfoques de la UE y el CdE

15

Instrumentos jurídicos y no jurídicos. Detección de la discriminación en la IA



Nivel de la UE.

1. Grupo de expertos sobre IA: **Directrices éticas para una IA digna de confianza**
2. Grupo de expertos sobre IA: **Lista de evaluación final sobre IA fiable**
3. **Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas sobre la inteligencia artificial (Ley sobre Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM/2021/206 final)**

16

Lista de evaluación de la IA fiable (ALTAI)

- Antes de evaluar la fiabilidad del sistema, debe realizarse una **Evaluación del Impacto sobre los Derechos Fundamentales (FRIA)**.

→ "Una **FRIA** podría incluir preguntas como las siguientes, basadas en artículos específicos de la Carta y del Convenio Europeo de Derechos Humanos (CEDH), sus protocolos y la Carta Social Europea.

1. El sistema de IA, ¿discrimina potencialmente de forma negativa a las personas por alguno de los siguientes motivos (no exhaustivos): sexo, raza, color, origen étnico o social, características genéticas, lengua, religión o convicciones o cualquier otra opinión, pertenencia a una minoría nacional, patrimonio, nacimiento, discapacidad, edad u orientación sexual? "¿Ha puesto en marcha procesos para comprobar y controlar la posible discriminación negativa (sesgo) durante las fases de desarrollo, despliegue y uso del sistema de IA?"

- La realización de la ALTAI puede implicar a un equipo multidisciplinar de personas: Diseñadores y desarrolladores de IA; científicos de datos; responsables o especialistas en contratación; personal de front-end que trabajará con el sistema de IA; **responsables jurídicos/de cumplimiento**; personal de gestión.

REQUISITO nº 4 Transparencia

1. Trazabilidad: por ejemplo, "¿Puede rastrear qué datos utilizó el sistema de IA para tomar una determinada decisión o recomendación?"; "¿Puede rastrear qué modelo o reglas de IA condujeron a la decisión o recomendaciones del sistema de IA?"

2. Explicabilidad: por ejemplo, "¿Explicó las decisiones del sistema de IA a los usuarios?"; "¿Consulta continuamente a los usuarios si entienden las decisiones del sistema de IA?"

REQUISITO nº 5 Diversidad, no discriminación e imparcialidad

1. Evitar sesgos injustos: por ejemplo, "¿Ha evaluado y puesto en marcha procesos para comprobar y supervisar posibles sesgos durante todo el ciclo de vida del sistema de IA? ¿Ha realizado pruebas con grupos destinatarios específicos o casos de uso problemáticos?"

REQUISITO Nº 7 Responsabilidad

1. Auditabilidad: por ejemplo, "¿ha establecido mecanismos que faciliten la auditabilidad del sistema de IA (por ejemplo, la trazabilidad del proceso de desarrollo, la obtención de datos de entrenamiento y el registro de los procesos, resultados e impacto positivo y negativo del sistema de IA?"

Fuente: [Comisión Europea - Modelar el futuro digital de Europa](#)

Parte IV

Lucha contra la discriminación en la IA



Cómo abordar la discriminación por IA: Discriminación indirecta

Discriminación indirecta

"existirá discriminación indirecta cuando una disposición, criterio o práctica aparentemente neutros sitúe a personas de [un motivo protegido] en desventaja particular con respecto a otras personas, salvo que dicha disposición, criterio o práctica pueda justificarse objetivamente con una finalidad legítima y salvo que los medios para la consecución de esta finalidad sean adecuados y necesarios".

Ref. Art. 2, 2 (b) Directiva 2000/43/CE; Art. 2, 2 (b); Directiva 2000/78/CE; Art 2,1(b) Directiva 2006/54/CE.

Enfoque: Al situar el sistema de IA como una disposición, criterio o práctica neutral, la atención se desplaza del funcionamiento técnico de la IA a su impacto.

Ventaja: Evitar explicaciones técnicas sobre el sistema de IA.

Desventaja: Justificación. Existen muchas más posibilidades de justificar la discriminación indirecta que la directa. La discriminación indirecta sólo tendrá consecuencias legales si el uso del sistema de IA no es un medio proporcionado para alcanzar un objetivo legítimo. Por ejemplo, necesidad empresarial.

Cómo abordar la discriminación por IA: Discriminación directa

Discriminación directa

"Existirá discriminación directa cuando una persona sea, haya sido o pudiera ser tratada de manera menos favorable que otra en situación comparable por razón de [un aspecto protegido]".

Ref. Art 2, 2 (a) 2000/43/CE; Art. 2, 2 (a); Directiva 2000/78/CE; Art 2,1(a) Directiva 2006/54/CE.

Enfoque: En la UE, **no se** requiere **ninguna intención** para la existencia legal de discriminación directa. En este caso, si se utilizan datos sesgados para alimentar un sistema de IA, por ejemplo, y, como consecuencia, una persona con un aspecto protegido recibe un trato menos favorable debido a su aspecto protegido, se produce una discriminación directa.

Ventaja: Muy excepcionalmente justificable. Por ejemplo, requisito profesional auténtico y determinante, siempre que sea proporcionado al objetivo legítimo.

KU Leuven Centre for IT & IP Law (CiTiP) - imec
Sint-Michielsstraat 6, apartado de correos 3443
BE-3000 Lovaina, Bélgica

<http://www.law.kuleuven.be/citip>

