



Intelixencia artificial

(Ética e normativa en materia de Educación)

Tina Campo Fernández - Fátima M^a García Doval



Nesgos e IA



Os nesgos son moldes de pensamento rápido, operativos pero, en ocasión errados:.

1. Todos os humanos temos nesgos, case que todo o tempo.
2. O primeiro nesgo é o nesgo do punto cego, o segundo, o nesgo de confirmación.
3. Correlación non implica causalidade.
4. A práctica totalidade das fontes de información humanas están nesgadas.
5. Certo tipo de investigacións, non todas, minimizan o impacto dos nesgos na información. Son poucas, pouco accesibles e pouco representadas.
6. **As IAs xerativas están alimentadas con coñecementos nesgados.**



As IAs xerativas seguen procesos probabilísticos que buscan asemellarse ao pensamento humano.

No primeiro no que o conseguen é na xeración de nesgos, alucinacións e outros erros de pensamento.





Eduardo Ruiz
@ejruizgarcia



Me ha parecido curiosa esta alucinación y la respuesta que me ha dado.
Contexto: le pedí a @GeminiApp una infografía, no me dio lo que necesitaba e iteré... terminó dándome esta imagen... 🙄



Eduardo Ruiz @ejruizgarcia · 11 ene.



Lo más curioso es que la calle de detrás bien podría ser una calle del barrio, pero el cartel del establecimiento que se ve por la ventana dice mucho... así que le pregunté por qué... y esta es la respuesta:



Eduardo Ruiz @ejruizgarcia · 11 ene.



"Es una observación muy aguda. La razón es técnica y tiene que ver con los metadatos de contexto que procesa el modelo.
El sistema recibe información sobre tu ubicación actual (Chiclana de la Frontera) para poder ofrecer respuestas útiles relacionadas con tu zona horaria...



Eduardo Ruiz @ejruizgarcia · 11 ene.



...o consultas locales. Al generar la imagen, la IA ha tomado ese dato del contexto y lo ha integrado creativamente en los elementos visuales del entorno (el cartel de la calle o del centro) para intentar dar mayor "realismo" y personalización a la escena del aula." 🤔



Eduardo Ruiz @ejruizgarcia · 11 ene.



Es un ejemplo que demuestra cómo los modelos generativos no solo escuchan el "prompt" explícito (lo que escribes), sino que también beben de las variables del entorno (ubicación, fecha, historial) para rellenar los huecos creativos de la imagen!





Cando unha persoa erra, a responsabilidade é da persoa.



Cando unha IA erra... a responsabilidade é da persoa que usou a IA!



ABRIL 2025

INTELIGENCIA ARTIFICIAL GENERATIVA Y DERECHOS DE AUTOR EN EL ÁMBITO CREATIVO

proyecto **451**

<https://www.gremieditors.cat/wp-content/uploads/2025/04/IA-Generativa-y-derechos-de-autor.pdf>



Guía práctica sobre derechos de autor en el ámbito educativo



<https://www.cedro.org/docs/default-source/0noticiasymas/guía-práctica-sobre-derechos-de-autor-en-el-ámbito-educativo.pdf>



Normativa da Unión Europea



Regulamentos

- Alcance: Aplicación xeral.
- Vincularidade xurídica: Xuridicamente vinculantes na súa integridade.
- Aplicación: Directamente aplicables (desde o momento da súa publicación no Diario Oficial da UE).
- Finalidade e obxectivo normativo: Harmonización total (aplicación uniforme en todos os Estados membros).



Directivas

- Alcance: Aplicación de regras/estándares mínimos.
- Vincularidade xurídica: Só vinculantes en canto ao resultado para alcanzar (a elección da forma e os medios para alcanzalo déixaselle aos Estados membros).
- Aplicación: Requírese de transposición (Estado membro debe adoptar unha norma nacional que recolle/implementa o contido e os obxectivos da Directiva).
- Finalidade e obxectivo normativo: Harmonización parcial (déixase certa marxe de discreción/manobra aos Estado para legislar como consideren, en función das súas particularidades xurídicas e sensibilidades nacionais).



Decisións

- Alcance: Sen aplicación xeral.
- Vincularidade xurídica: Xuridicamente vinculante aos seus destinatarios específicos (Estados, persoas xurídicas ou físicas).
- Aplicación: Directamente aplicables.
- Finalidade e obxectivo normativo: Crean obrigacións ou dereitos específicos, imponen sancións etc.



Recomendacións e ditames

- Alcance: Aplicación programática (*Soft law*).
- Vincularidade xurídica: Non son xuridicamente vinculantes (pero pode ter efectos xurídicos indirectamente).
- Aplicación: Non é precisa.
- Finalidade e obxectivo normativo: Orientar sobre políticas e lexislación, expresar opinións, convidar á adopción de normas de conduta etc.



Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial-RIA)



Considerandos 4, 5, 10, 44, 56 e 69



Definiciones 1, 34, 35, 36, 37, 49, 50, 51, 52, 60, 63 e 66



Considerandos

 **Considerando 4.** La IA es un conjunto de tecnologías en rápida evolución que contribuye a generar beneficios económicos, medioambientales y sociales muy diversos en todos los sectores económicos y las actividades sociales (...)

 **Considerando 5.** Al mismo tiempo, dependiendo de las circunstancias relativas a su aplicación, utilización y nivel de desarrollo tecnológico concretos, la IA puede generar riesgos y menoscabar los intereses públicos y los derechos fundamentales que protege el Derecho de la Unión.

Dicho menoscabo puede ser tangible o intangible e incluye los perjuicios físicos, psíquicos, sociales o económicos.



 **Considerando 10.** El derecho fundamental a la protección de los datos personales está garantizado, en particular, por los Reglamentos (UE) 2016/679 (11) y (UE) 2018/1725 (12) del Parlamento Europeo y del Consejo y la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo (13). Además, la Directiva 2002/58/CE del Parlamento Europeo y del Consejo (14) protege la vida privada y la confidencialidad de las comunicaciones, también estableciendo condiciones para cualquier almacenamiento de datos personales y no personales en los equipos terminales, y el acceso desde estos. Dichos actos legislativos de la Unión constituyen la base para un tratamiento de datos sostenible y responsable, también cuando los conjuntos de datos contengan una combinación de datos personales y no personales (...) Unas normas armonizadas para la introducción en el mercado, la puesta en servicio y la utilización de sistemas de IA establecidas en virtud del presente Reglamento deben facilitar la aplicación efectiva y permitir el ejercicio de los derechos y otras vías de recurso de los interesados garantizados por el Derecho de la Unión en materia de protección de datos personales, así como de otros derechos fundamentales.



Considerando 44. Existe una gran preocupación respecto a la base científica de los sistemas de IA que procuran detectar o deducir las emociones, especialmente porque la expresión de las emociones varía de forma considerable entre culturas y situaciones, e incluso en una misma persona. Algunas de las deficiencias principales de estos sistemas son la fiabilidad limitada, la falta de especificidad y la limitada posibilidad de generalizar. Por consiguiente, los sistemas de IA que detectan o deducen las emociones o las intenciones de las personas físicas a partir de sus datos biométricos pueden tener resultados discriminatorios y pueden invadir los derechos y las libertades de las personas afectadas. Teniendo en cuenta el desequilibrio de poder en el contexto laboral o educativo, unido al carácter intrusivo de estos sistemas, dichos sistemas podrían dar lugar a un trato perjudicial o desfavorable de determinadas personas físicas o colectivos enteros. Por tanto, debe prohibirse la introducción en el mercado, la puesta en servicio y el uso de sistemas de IA destinados a ser utilizados para detectar el estado emocional de las personas en situaciones relacionadas con el lugar de trabajo y el ámbito educativo. Dicha prohibición no debe aplicarse a los sistemas de IA introducidos en el mercado estrictamente con fines médicos o de seguridad, como los sistemas destinados a un uso terapéutico.



Considerando 56. El despliegue de sistemas de IA en el ámbito educativo es importante para fomentar una educación y formación digitales de alta calidad y para que todos los estudiantes y profesores puedan adquirir y compartir las capacidades y competencias digitales necesarias, incluidos la alfabetización mediática, y el pensamiento crítico, para participar activamente en la economía, la sociedad y los procesos democráticos. No obstante, deben clasificarse como de alto riesgo los sistemas de IA que se utilizan en la educación o la formación profesional, y en particular aquellos que determinan el acceso o la admisión, distribuyen a las personas entre distintas instituciones educativas y de formación profesional o programas de todos los niveles, evalúan los resultados del aprendizaje de las personas, evalúan el nivel apropiado de educación de una persona e influyen sustancialmente en el nivel de educación y formación que las personas recibirán o al que podrán acceder, o supervisan y detectan comportamientos prohibidos de los estudiantes durante las pruebas, ya que pueden decidir la trayectoria formativa y profesional de una persona y, en consecuencia, puede afectar a su capacidad para asegurar su subsistencia. Cuando no se diseñan y utilizan correctamente, estos sistemas pueden invadir especialmente y violar el derecho a la educación y la formación, y el derecho a no sufrir discriminación, además de perpetuar patrones históricos de discriminación, por ejemplo contra las mujeres, determinados grupos de edad, las personas con discapacidad o las personas de cierto origen racial o étnico o con una determinada orientación sexual.



Considerando 69. El derecho a la intimidad y a la protección de datos personales debe garantizarse a lo largo de todo el ciclo de vida del sistema de IA. A este respecto, los principios de minimización de datos y de protección de datos desde el diseño y por defecto, establecidos en el Derecho de la Unión en materia de protección de datos, son aplicables cuando se tratan datos personales. Las medidas adoptadas por los proveedores para garantizar el cumplimiento de estos principios podrán incluir no solo la anonimización y el cifrado, sino también el uso de una tecnología que permita llevar los algoritmos a los datos y el entrenamiento de los sistemas de IA sin que sea necesaria la transmisión entre las partes ni la copia de los datos brutos o estructurados, sin perjuicio de los requisitos en materia de gobernanza de datos establecidos en el presente Reglamento.



ART. 3-Definiciones

- [-] 35) **Identificación biométrica:** el reconocimiento automatizado de características humanas de tipo físico, fisiológico, conductual o psicológico para determinar la identidad de una persona física comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos;
- [-] 1) **Sistema de IA:** un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales;
- [-] 34) **Datos biométricos:** los datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física, como imágenes faciales o datos dactiloscópicos;
- [-] 36) **Verificación biométrica:** la verificación automatizada y uno-a-uno, incluida la autenticación, de la identidad de las personas físicas mediante la comparación de sus datos biométricos con los datos biométricos facilitados previamente;



- =** 37) **Categorías especiales de datos personales:** las categorías de datos personales a que se refieren el artículo 9, apartado 1, del Reglamento (UE) 2016/679, el artículo 10 de la Directiva (UE) 2016/680 y el artículo 10, apartado 1, del Reglamento (UE) 2018/1725;
- =** 49) **Incidente grave:** un incidente o defecto de funcionamiento de un sistema de IA que, directa o indirectamente, tenga alguna de las siguientes consecuencias:
 - a) el fallecimiento de una persona o un perjuicio grave para su salud;
 - b) una alteración grave e irreversible de la gestión o el funcionamiento de infraestructuras críticas;
 - c) el incumplimiento de obligaciones en virtud del Derecho de la Unión destinadas a proteger los derechos fundamentales;
 - d) daños graves a la propiedad o al medio ambiente;



- 50) **Datos personales:** los datos personales tal como se definen en el artículo 4, punto 1, del Reglamento (UE) 2016/679;
- 51) **Datos no personales:** los datos que no sean datos personales tal como se definen en el artículo 4, punto 1, del Reglamento (UE) 2016/679;
- 52) **Elaboración de perfiles:** la elaboración de perfiles tal como se define en el artículo 4, punto 4, del Reglamento (UE) 2016/679;
- 60) **Ultrasuplantación:** un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos;



- 63) **Modelo de IA de uso general:** un modelo de IA, también uno entrenado con un gran volumen de datos utilizando autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas, independientemente de la manera en que el modelo se introduzca en el mercado, y que puede integrarse en diversos sistemas o aplicaciones posteriores, excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su introducción en el mercado;
- 66) **Sistema de IA de uso general:** un sistema de IA basado en un modelo de IA de uso general y que puede servir para diversos fines, tanto para su uso directo como para su integración en otros sistemas de IA;



AXENCIA PARA A
MODERNIZACIÓN
TECNOLÓXICA DE GALICIA

Clasificación de los sistemas de IA





ART. 5–Prácticas e sistemas IA prohibidos

- ❌ Aqueles Sistemas IA que empreguen **técnicas subliminares** como poden ser estímulos de audio, imaxe ou vídeo que as persoas non poden percibir ou **técnicas manipuladoras ou enganosas** para alterar o comportamento humano, minguando a súa capacidade para tomar decisións e facéndolles tomar unha decisión que doutro xeito non tomarían, dunha maneira que probablemente lles cause un prexuízo considerable ou llo cause a outros.
- ❌ Os que **exploten vulnerabilidades** dunha persoa física ou grupo de persoas derivadas da súa idade ou discapacidade, ou dunha situación social ou económica específica, coa finalidade ou o efecto de alterar de maneira substancial o seu comportamento que provoque -ou poida provocar- prexuízos considerables.
- ❌ Os que se empreguen para **avaliar ou clasificar ás persoas** durante un período determinado de tempo **atendendo ao seu comportamento social ou as súas características persoais ou da súa personalidade** ("scoring social"), de forma que a puntuación cidadá provoque ás persoas un trato prexudicial ou desfavorable:
 - en contextos sociais que non garden relación cos contextos onde se xeraron ou solicitaron os datos, ou
 - que sexa inxustificado ou desproporcionado con respecto ao seu comportamento social ou a gravidade deste



- ❌ Aqueles que **avalíen ás persoas para valorar ou predicir o risco de que cometan un delito**, baseándose no seu perfil ou trazos de personalidade. Isto non afecta os sistemas que apoien a valoración humana da implicación dunha persoa nunha actividade delituosa, que se base en feitos obxectivos e verificables.
- ❌ Sistemas que **cren ou amplíen bases de datos de recoñecemento facial** mediante a extracción non selectiva de imaxes faciais da internet ou de circuítos pechados de televisión.
- ❌ Os que infiran as **emocións dunha persoa nos lugares de traballo e nos centros educativos**, agás que vaian destinados a fins médicos ou de seguridade.
- ❌ Os Sistemas IA de categorización biométrica que **clasifican ás persoas físicas baseándose nos seus datos biométricos** para deducir ou inferir determinados datos sensibles como a raza, opinións políticas, afiliación sindical, conviccións relixiosas ou filosóficas, ou sobre a vida sexual ou a orientación sexual. Isto non inclúe, por exemplo, os datos que se adquiriron de forma lícita, como poden ser imaxes.
- ❌ Os que utilicen unha **identificación biométrica remota en tempo real**, salvo que se considere estraitamente necesario e para determinados obxectivos moi concretos.



Prácticas de IA prohibidas



Técnicas Subliminales o Engañosas

Explotación de Vulnerabilidades

Sistemas de Puntuación Social

**Categorización biométrica para inferir
datos sensibles**

**Identificación Biométrica Remota 'en
Tiempo Real'**

**Evaluación de riesgos y perfiles
predictivos de Delincuencia**

**Creación o ampliación de bases de datos
de reconocimiento facial**

Inferencia de emociones





ART. 6–Sistemas IA de alto riesgo

Reglas de clasificación de los sistemas de IA de alto riesgo

- 1. Con independencia de si se ha introducido en el mercado o se ha puesto en servicio sin estar integrado en los productos que se mencionan en las letras a) y b), un sistema de IA se considerará **de alto riesgo** cuando **reúna las dos condiciones** que se indican a continuación:
 - a) que el sistema de IA esté destinado a ser utilizado como componente de seguridad de un producto que entre en el ámbito de aplicación de los actos legislativos de armonización de la Unión enumerados en el anexo I, o que el propio sistema de IA sea uno de dichos productos, y
 - b) que el producto del que el sistema de IA sea componente de seguridad con arreglo a la letra a), o el propio sistema de IA como producto, deba someterse a una evaluación de la conformidad de terceros para su introducción en el mercado o puesta en servicio con arreglo a los actos legislativos de armonización de la Unión enumerados en el anexo I.
- 2. Además de los **sistemas de IA** de alto riesgo a que se refiere el apartado 1, también se considerarán de alto riesgo los sistemas de IA **contemplados en el anexo III**.



ANEXO III

Sistemas de IA de alto riesgo a que se refiere el artículo 6, apartado 2

Los sistemas de IA de **alto riesgo** con arreglo al artículo 6, apartado 2, son los **sistemas de IA que formen parte de cualquiera de los ámbitos siguientes**:

- **3. Educación y formación profesional:**

- a) Sistemas de IA destinados a ser utilizados para determinar el acceso o la admisión de personas físicas a centros educativos y de formación profesional a todos los niveles o para distribuir a las personas físicas entre dichos centros
- b) Sistemas de IA destinados a ser utilizados para evaluar los resultados del aprendizaje, también cuando dichos resultados se utilicen para orientar el proceso de aprendizaje de las personas físicas en centros educativos y de formación profesional a todos los niveles
- c) Sistemas de IA destinados a ser utilizados para evaluar el nivel de educación adecuado que recibirá una persona o al que podrá acceder, en el contexto de los centros educativos y de formación profesional o dentro de estos a todos los niveles
- d) Sistemas de IA destinados a ser utilizados para el seguimiento y la detección de comportamientos prohibidos por parte de los estudiantes durante los exámenes en el contexto de los centros educativos y de formación profesional o dentro de estos a todos los niveles



Sistemas de IA de alto riesgo: Anexo III

Biometría:

1. **Identificación biométrica remota**
2. **Categorización biométrica:**
Clasificación por atributos sensibles.
3. **Reconocimiento de emociones:**
Interpretación potencialmente errónea de emociones.

Infraestructuras críticas:

1. Uso en sistemas de seguridad esenciales como energía y transporte, donde fallos pueden tener consecuencias graves.

Educación y formación profesional:

1. Admisión y distribución en centros educativos.
2. Evaluación de resultados de aprendizaje.
3. Seguimiento de comportamientos durante exámenes.

Empleo, gestión de trabajadores y acceso al autoempleo:

1. Sistemas para contratación y evaluación de empleados, impactando oportunidades laborales.

Acceso a servicios privados esenciales y públicos:

1. Evaluación para la admisión a servicios esenciales como salud y seguros.

Aplicación de la ley:

1. Herramientas para evaluar riesgos de criminalidad, afectando derechos civiles y potencialmente llevando a discriminación.



Migración, asilo y gestión del control fronterizo:

1. Evaluación de riesgos en la gestión de fronteras y asilo, impactando a individuos vulnerables.

Administración de justicia y procesos democráticos:

1. Influencia en la interpretación de leyes y hechos judiciales y en elecciones o referendos.



Autoridades nacionales competentes: AESIA

Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la AESIA:



- *tareas de supervisión, el asesoramiento, la concienciación y la formación a entidades de derecho público y privado para la adecuada implementación de toda la normativa nacional y europea IA.*
- *función de inspección, comprobación, sanción y demás que le atribuya la normativa europea.*
- *Autoridad responsable de la supervisión, y en su caso sanción, de los sistemas de IA con el objeto de eliminar o reducir los riesgos para la integridad, la intimidad, la igualdad de trato y la no discriminación, en particular entre mujeres y hombres, y demás derechos fundamentales que pueden verse afectados por el mal uso de los sistemas.*



RGPD

1. El presente Reglamento se aplica al tratamiento total o parcialmente automatizado de datos personales, así como al tratamiento no automatizado de datos personales contenidos o destinados a ser incluidos en un fichero.

1) «datos personales»: toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona;

1. Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar.

2. El apartado 1 no se aplicará si la decisión:

- a) es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento;
- b) está autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, o
- c) se basa en el consentimiento explícito del interesado.

3. En los casos a que se refiere el apartado 2, letras a) y c), el responsable del tratamiento adoptará las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión.

4. Las decisiones a que se refiere el apartado 2 no se basarán en las categorías especiales de datos personales contempladas en el artículo 9, apartado 1, salvo que se aplique el artículo 9, apartado 2, letra a) o g), y se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.