

***TAREA 50***

***Inteligencia Artificial Ética***



INDICE

[1 Introducción 1](#_Toc71938410)

[2 Iniciativas relacionadas 1](#_Toc71938411)

[2.1 Moral Machine de MIT Media Lab [1] 1](#_Toc71938412)

[2.2 Principios rectores de Microsoft [2] [3] 2](#_Toc71938413)

[2.3 Unión Europea 4](#_Toc71938414)

[2.4 Manifiesto ético del dato del Gobierno de Navarra [6] 5](#_Toc71938415)

[2.5 Leyes de la robótica de Isaac Asimov [7] 5](#_Toc71938416)

[3 Propuesta de normas de uso 5](#_Toc71938417)

[4 Referencias 7](#_Toc71938418)

# Introducción

En este documento se describe una reflexión sobre principios éticos aplicados a la Inteligencia Artificial (IA) que se plasma en una serie de principios. El impacto de la IA en nuestro día a día cada vez es mayor y se aplica en ámbitos muy diferentes. El beneficio que crean aplicaciones positivas de IA es inmenso. Sin embargo, en general las tecnologías son una moneda con doble cara, y el ser humano se las ingenia para utilizarlas tanto para hacer el bien como el mal.

En estos momentos en los que las aplicaciones basadas en IA se están expandiendo, se está generando una conciencia sobre la necesidad de establecer una ética e incluso legislación que marque el rumbo y las reglas a seguir por las aplicaciones de IA.

Tal y como menciona Max Tegmark en el documental “Límites éticos para la inteligencia artificia” en ámbitos como la biología se ha conseguido limitar su aplicación en aplicaciones “perversas”, como las armas biológicas. Sin embargo, en el caso de los físicos la aplicación de los avances físicos derivó en la bomba nuclear. En el caso de la AI, en estos momentos estamos a tiempo de influir en las aplicaciones futuras de esta tecnología.

Tanto las empresas como los gobiernos, a nivel local, nacional, europeo e internacional, comienzan a emitir recomendaciones y legislación sobre la IA. La complejidad técnica de la IA y la rapidez a la que esta evolucionando hace que esta sea una tarea complicada. La importancia de los datos de entrada y entrenamiento y validación en los desarrollos de las aplicaciones de IA complican de forma adicional esta tarea y hacen que la revisión del alcance ético de las aplicaciones de IA tenga que llevarse a cabo de forma continúa.

En este documento en primer lugar se describen brevemente algunas iniciativas relacionadas al uso ético de la IA. A continuación, se proponen una serie de principios básicos para mejorar el nivel ético de estas aplicaciones, que debieran ir acompañados de una legislación que regule, busque y sancione su uso indebido.

# Iniciativas relacionadas

## Moral Machine de MIT Media Lab [1]

Se trata de un juego de coches autónomos que imita el llamado dilema del ferrocarril. Se plantean nueve situaciones en las que hay vidas en peligro y se analiza cómo se toma la decisión de salvar a unas personas sobre otras y qué criterios utilizan. El estudio se hizo viral y en tres años se obtuvieron más de cuarenta millones de respuestas de hasta 233 países.

Los resultados mostraron como la percepción de lo correcto o ético era dependiente de la cultura. En función de la cultura se considera más ético salvar la vida de personas mayores o jóvenes. Así mismo, el problema y la percepción de lo ético en estas situaciones está muy influenciado por el anonimato de las personas. Por ejemplo, si un padre conduce un coche en el que va su hijo, es ético evitar un choque mortal aunque se atropelle a un grupo de personas, o lo ético sería sacrificarse con su propio hijo en el coche y no evitar el choque?



Figura 1 Preferencias identificadas en las más de cuarenta millones de respuestas recibidas en el experimento Moral Machine.

## Principios rectores de Microsoft [2] [3]

En Microsoft han identificado seis principios que, en su opinión, deberían regir el desarrollo y el uso de la IA: imparcialidad, confiabilidad y seguridad, privacidad y seguridad, inclusión, transparencia y responsabilidad. La descripción de estos principios se encuentra en el curso relacionado de Microsoft y se incluye a continuación.



Figura 2 6 principios rectores de Microsoft en la IA

* Imparcialidad: Los sistemas de inteligencia artificial deben tratar a todo el mundo de forma imparcial y evitar afectar a grupos de personas en situaciones similares de maneras diferentes.
* Confiabilidad y seguridad: Los sistemas de inteligencia artificial funcionen de forma confiable, segura y coherente en circunstancias normales y en situaciones inesperadas. Estos sistemas deben ser capaces de funcionar según su diseño original, responder de forma segura a situaciones inesperadas y resistir la manipulación perjudicial. También es importante poder comprobar que estos sistemas se comportan según lo previsto en condiciones de funcionamiento reales.
* Privacidad y seguridad:  Los sistemas de inteligencia artificial deben cumplir las leyes de privacidad que exigen transparencia sobre la recopilación, el uso y el almacenamiento de datos y obligan a los consumidores a establecer controles adecuados para elegir cómo se usan sus datos.
* Inclusión: Todas las personas deben beneficiarse de la tecnología inteligente, lo que significa que debe incorporar y abarcar una amplia gama de necesidades y experiencias humanas.
* Transparencia: Cuando se usan sistemas de inteligencia artificial para ayudar a tomar decisiones que tienen un impacto enorme en la vida de las personas, es fundamental que se entienda cómo se han tomado esas decisiones.
* Responsabilidad: Las personas que diseñan e implementan sistemas de inteligencia artificial deben ser responsables de su funcionamiento. Las organizaciones deben basarse en estándares del sector para desarrollar normas de responsabilidad. Estas normas pueden garantizar que los sistemas de inteligencia artificial no sean la autoridad final sobre cualquier decisión que afecte a la vida de las personas y que los seres humanos mantengan un control significativo sobre sistemas de inteligencia artificial que, por otra parte, son altamente autónomos.

## Unión Europea

La Unión Europea ha publicado recientemente (2021) [4] un comunicado con sus directrices para generar una legislación específica para la IA. Aunque no cubre el ámbito militar, la recomendación marca una clasificación de las aplicaciones de IA en función de su potencial de riesgo. Ejemplos de aplicaciones de alto riesgo son sistemas para contratar a personas, generar clasificaciones de riesgo crediticio, o decisiones judiciales.

Estos sistemas debieran basarse en datasets de calidad, establecer documentación y trazabilidad, compartir información adecuada con los usuarios, incluir medidas de evaluación y control manual por parte de humanos, y aplicar los estándares más elevados de robustez, seguridad, ciberseguridad y efectividad. Estos sistemas tendrían que certificar que cumplen estos requerimientos antes de comenzar su utilización.

En el 2019 un grupo de expertos de alto nivel en el campo de la IA publicó el informe “Directrices éticas para una IA fiable” [5]. Entre las directrices, destacan las siguientes:

* Desarrollar, desplegar y utilizar los sistemas de IA respetando los principios éticos de: respeto de la autonomía humana, prevención del daño, equidad y explicabilidad
* Prestar una atención especial a las situaciones que afecten a los grupos más vulnerables, como los niños o las personas con discapacidad, así como a las situaciones caracterizadas por asimetrías de poder o de información, como las que pueden producirse entre empresarios y trabajadores o entre empresas y consumidores
* Garantizar que el desarrollo, despliegue y utilización de los sistemas de IA cumpla los requisitos para una IA fiable: 1) acción y supervisión humanas, 2) solidez técnica y seguridad, 3) gestión de la privacidad y de los datos, 4) transparencia, 5) diversidad, no discriminación y equidad, 6) bienestar ambiental y social, y 7) rendición de cuentas.

## Manifiesto ético del dato del Gobierno de Navarra [6]

Este manifiesto se dirige a formular los principios rectores de la gestión de datos tanto personales como públicos, así como del uso de algoritmos de inteligencia artificial (IA). Recoge seis principios generales:

1. Protección de los datos
2. Apertura de los datos y reutilización
3. Gobernanza de los datos
4. Una IA para la ciudadanía y el bienestar social y ambiental
5. Promover la innovación gracias a los datos y la IA
6. Rendir cuentas de los sistemas de IA

El contenido de este manifiesto distingue entre tres tipos de enunciados:

1. Obligaciones exigibles al Gobierno de Navarra.
2. Derechos de la ciudadanía y las organizaciones.
3. Compromisos de avance que se adquieren

## Leyes de la robótica de Isaac Asimov [7]

Por último, se hace necesario recordar las tres leyes formuladas por Asimov en 1942 para los robots, que se pueden aplicar para la IA y que sirven de base para muchas de las reflexiones y movimientos de legislación de la IA actuales:

1. Un robot no hará daño a un ser humano ni, por su inacción, permitirá que un ser humano sufra daño
2. Un robot debe obedecer las órdenes dadas por los seres humanos, excepto si estas entran en conflicto con la primera ley.
3. Un robot debe proteger su propia existencia en la medida en que no entre en conflicto con la primera o la segunda ley

# Propuesta de normas de uso

En este apartado recojo una propuesta sobre el uso ético de la IA. Esta propuesta se basa en varias visiones personales que comparto con varios organismos:

* Tal y como propone la UE, al menos en las aplicaciones de alto riesgo habría que legislar el uso de la IA, incluyendo responsabilidades y sanciones y excluyendo o restringiendo su uso para ciertas aplicaciones
* Cuando un humano interactúa con una IA, la IA se tendría que identificar
* En todos los sistemas gobernados por una IA, un humano tendría que tener la opción de lograr el control y tomar una decisión contraria a la IA
* Cualquier sistema de IA que interactúe o tome decisiones relacionadas con personas, antes de implantarse tendría que superar una auditoría (interna o externa en función de la clasificación de su riesgo o impacto) para comprobar que aspectos éticos críticos se han considerado.
* La IA está formada por dos componentes principales, la algoritmia y los datos. Aunque hay opiniones sobre que los datos son neutros y los algoritmos y su aplicación éticamente positivos o negativos, considero que hay que valorar el conjunto.

La importancia y el impacto de la IA en nuestras vidas requiere que se garantice el uso ético de los datos y los algoritmos. Una aplicación de IA éticamente correcta, por ejemplo detectar si alguien se está ahogando en la playa de forma automática, puede ser contraproducente si sólo se ha entrenado con personas de ciertos rangos de edad o razas y etnias, ya que no detectaría que personas no representadas en el dataset de entrenamiento se están ahogando.

* Los usos no éticos de la IA, al menos en aplicaciones de alto riesgo, deberían estar penalizados a través de una legislación específica

Tras esta reflexión, reflejo una serie de principios sobre el uso ético de la IA, muy ligados al establecimiento de una legislación asociada:

1. El uso de la IA se debe prohibir en aplicaciones que atenten contra los derechos fundamentales, que busquen provocar daño físico o psíquico a las personas, o modificar el comportamiento de las personas o la sociedad de forma subliminal o fines comerciales y políticos
   1. En el caso de su aplicación por las fuerzas de seguridad para la monitorización de individuos, se deberá contar con autorización judicial
2. Cualquier aplicación de IA debe cumplir la legislación vigente. En caso contrario, la persona responsable de la aplicación de IA deberá responder frente a la justicia
3. Cualquier aplicación de IA que interactúe con humanos y no sea evidente (incluso para personas con alguna diversidad funcional) que se trate de una aplicación de IA se debe identificar como tal
4. Cualquier aplicación basada en IA debe someterse a una auditoría interna o externa para verificar que cumple con ciertos criterios establecidos, como por ejemplo los 6 seguidos por Microsoft. Esta auditoría se debe repetir de forma periódica para adaptarse a posibles malos usos, deficiencias o evoluciones tecnológicas de relevancia en la aplicación
5. Cualquier aplicación de IA debe poder delegar su decisión a un humano, independientemente de su dominio de uso y nivel de riesgo.
6. Cualquier aplicación de IA de alto impacto o riesgo debe ser capaz de dar explicaciones sobre su funcionamiento
7. Los datos utilizados para el entrenamiento, el desarrollo y la validación de la IA se deben de trazar y deben estar anonimizados.
8. La validación del algoritmo se debe hacer de forma periódica para asegurar que su rendimiento es aceptable a lo largo del tiempo.
9. En las aplicaciones de alto riesgo o impacto, los datos y los algoritmos de base de las aplicaciones deben estar visibles, para toda la sociedad en el caso de aplicaciones públicas y al menos para los auditores en las aplicaciones de empresas privadas
10. Las aplicaciones de alto riesgo o impacto se deberán registrar en un registro público. Cada registro tendrá asociado un responsable de la aplicación, que tendría que asumir las penalizaciones asociadas a un uso no ético de la aplicación de IA

# Referencias

|  |  |
| --- | --- |
| [1] | Quint Group, “Responsabilidad ética e inteligencia artificial,” 2021. [Online]. Available: https://www.quintgroup.com/es-es/insights/responsabilidad-etica-inteligencia-artificial/. |
| [2] | “Curso Microsoft Identificación de los principios y las prácticas para una IA responsable,” 2021. [Online]. Available: https://docs.microsoft.com/es-es/learn/paths/responsible-ai-business-principles/?wt.mc\_id=msftsource\_issue24link2\_email\_gdc&?ocid=eml\_pg272283\_gdc\_comm\_az&mkt\_tok=MTU. |
| [3] | Microsoft, “The Future Computed: Artificial Intelligence and its role in society,” 2018. [Online]. Available: https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed\_2.8.18.pdf. |
| [4] | Europen Union, “Fostering a European approach to artificial intelligence,” 2021. [Online]. Available: https://ec.europa.eu/newsroom/dae/redirection/document/75790. |
| [5] | European Union, “Directrices éticas para una IA fiable,” 2018. [Online]. Available: https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1. |
| [6] | Gobierno de Navarra, “Manifiesto ético del dato,” 2020. [Online]. Available: https://gobiernoabierto.navarra.es/sites/default/files/manifiesto\_etico\_dato\_v1.pdf. |
| [7] | Wikipedia, “Leyes de la robótica,” 2021. [Online]. Available: https://es.wikipedia.org/wiki/Leyes\_de\_la\_rob%C3%B3tica. |