



ÉTICA DE LA INTELIGENCIA ARTIFICIAL

En este documento se pretende establecer varios principios y normas que cualquier máquina, sistema con inteligencia artificial debe regir para un uso ético y moralmente aceptable.

En lo documento se presentan de forma resumida los objetivos de la Inteligencia Artificial, así como lo que es capaz de realizar con esta tecnología y para lo que se podría llegar a emplear. Por ello, finalmente se ofrecen ciertos principios y normas que aseguren un uso responsable de esta tecnología, a pesar de ser principios muy superficiales, entendiendo que deberán de detallarse ampliamente para cada aplicación.

Al final del documento, se ofrece una pequeña bibliografía con los enlaces consultados.



1. Objetivos de la Inteligencia Artificial

- Lograr que las máquinas adquirieran una inteligencia similar a la humana, con la capacidad de razonar y tomar decisiones por sí mismas.
- Adquisición de conocimiento especializado o conocimiento de sentido común, tal y como trabaja el cerebro humano.
- Tener un impacto positivo en multitud de sectores como energía, transporte, salud, seguridad, productividad, educación,...

2. Ejemplos de posibles aplicaciones de la Inteligencia Artificial. ¿Tiene límites?

Se presentan ejemplos ordenados de menor a mayor riesgo según el criterio del autor, debido a una mayor responsabilidad bien en la salud, privacidad o discriminación de los seres humanos. Estos ejemplos se presentan con la intención de que el lector se pregunte acerca de la potencialidad de esta tecnologías y sus límites.

- Redes sociales, elección de anuncios.
- Creación de piezas de arte según el estilo de un importante pintor. ¿Quién sería designado como el autor de estas nueva piezas, los ingenieros, el algoritmo, el pintor mismo,... ? ¿Qué hay del plagiarismo y piratería en las que se podrían ver involucradas las máquinas de IA?
- Banco, elección de dar créditos. ¿Discriminación?
- Sistemas de reconocimiento facial. ¿Localizables en cualquier lugar?
- Diseño de nuevos medicamentos y vacunas.
- Ciberseguridad. ¿Saboteo de redes de infraestructuras de países?
- Conducción autónoma. ¿Qué sucede si la máquina tiene que tomar la decisión de salvar la vida de una persona u otra?
- Conflictos armados. ¿Drones de ataque con IA?



3. Principios y normas que toda Inteligencia Artificial debe seguir

Aquí se recogen diversos principios y normas éticos con los que estamos conformes y que ya se recogen en multitud de documentos de empresas y estados con el objetivo de regular la actividad de la IA.

No obstante, aún falta mucha concreción en como deben usarse, y es que para cada aplicación, posiblemente se necesiten unas normas muy concretas. Estas normas se irán detallando conforme se vayan aplicando estas tecnologías y vayan apareciendo nuevos riesgos. Para un uso extendido, deberían pasar por diferentes órganos que avalen su uso en cumplimiento con dichos principios y normas.

De manera superficial, estas son algunos de los principios que toda Inteligencia Artificial debe seguir.

- **Prudencia:** se deben considerar estrictos requisitos en la aplicación de la IA, siendo extremadamente cuidadosos en la elección de los algoritmos y de los parámetros en cada aplicación de uso. Si esto no se cumple, se puede llegar a resultados potencialmente peligrosos.
- **Fiabilidad:** los sistemas de IA deben demostrar ser fiables en las funciones que desempeñen, sobre todo en salud y robots autónomos. Diferencias agencias internacionales deberán testar dichos sistemas.
- **Seguridad:** deben cumplir con los requisitos de seguridad en los diferentes entornos en los que se pueden emplear, así como con los seres humanos, animales y medio ambiente.
- **Rendición de cuentas:** deben ser capaces de argumentar el porqué de haber tomado ciertas decisiones. Ésta es hasta ahora una de las grandes limitaciones de la IA.
- **Legibilidad:** deben ser transparentes e legibles para los seres humanos, sobretodo en aquellas que tomen decisiones.
- **Responsabilidad:** debe poder localizarse en todo momento a los responsables de las máquinas de IA.
- **Autonomía de los sistemas de IA:** se deberán establecer reglas que limiten la capacidad de toma de decisiones de los sistemas autónomos. Además, habrá que clarificar quién es el responsable en caso de algún tipo de accidente o malfuncionamiento.
- **Sesgos:** toda máquina de IA necesita de un humano para ser programada, y por tanto sufren de ciertos sesgos humanos difícilmente de eliminar. Además, las máquinas de IA son entrenadas con datos históricos, por lo que esto puede provocar ciertas discriminaciones actuales por circunstancias pasadas. Por ende, no podemos aplicar sistemas de IA como sistemas de justicia perfectos.



- **Privacidad:** los sistemas de IA no deberán aplicarse en invasiones de la privacidad de los seres humanos, ya que éste es un derecho fundamental.
- **Opacidad:** los modelos de IA son prácticamente modelos de caja negra donde se suele conocer bien las entradas y salidas pero no lo que sucede en el interior. El conocimiento almacenado en las capas intermedias de la red neuronal difícilmente es conocido. Esto provoca que no sepamos con certeza el conocimiento adquirido por la red y si estará preparado para afrontar cualquiera de los infinitos escenarios posibles. De alguna forma, se deberá exigir un acercamiento a lo que conocemos como caja negra para evitar problemas de seguridad, discriminaciones, invasiones de la privacidad, etc.

-



Bibliografía

<https://www.iiia.csic.es/barcelonadeclaration/>

<https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/el-multiverso-cuntico-711/tica-en-la-inteligencia-artificial-15492>

<https://www.tendencias.kpmg.es/claves-decada-2020-2030/robots-etica-inteligencia-artificial/>

La Inteligencia Artificial y su Implicancia en los Conflictos Armados Contemporáneos [Pablo Agustín Mastragostino, 2020]

<https://en.unesco.org/artificial-intelligence/ethics/cases>

<https://plato.stanford.edu/entries/ethics-ai/>

https://blogs.elconfidencial.com/tecnologia/tribuna/2016-12-06/caja-negra-software-big-data-ai-informatica_1299837/