

Inteligencia Artificial

Víctor Mijangos de la Cruz

[vmijangosc @ ciencias.unam.com](mailto:vmijangosc@ciencias.unam.com)

I. Introducción a la Inteligencia Artificial



Información del curso

Nombre	Inteligencia Artificial
Clave	0608
Créditos	10
Total de horas	112

Objetivo del curso

Comprender para aplicar los enfoques teóricos y prácticos y así poder diseñar sistemas inteligentes. Conocer los temas centrales, historia de inteligencia artificial, agentes, búsqueda, modelos probabilísticos, aprendizaje automático y percepción y conocimiento.

Temario

1 Introducción

- ① Definición de la Inteligencia Artificial
- ② Marco histórico y aplicaciones
- ③ Consideraciones éticas

2 Búsqueda

- ① Agentes inteligentes y racionalidad
- ② Ambientes
- ③ Estructura de agentes

3 Modelos probabilísticos

- ① Modelos gráficos
- ② Modelos bayesianos

3 Modelos causales y modelos sin dirección

4 Aprendizaje automático

- ① Teoría de aprendizaje
- ② Aprendizaje supervisado, no-supervisado y por refuerzo
- ③ Aprendizaje profundo

5 Percepción y conocimiento

- ① Visión computacional
- ② Procesamiento del lenguaje natural
- ③ Robótica
- ④ Representación del conocimiento

Bibliografía

La bibliografía básica es:

Russell, S. y Norvig, P. (2021). Artificial Intelligence: A Modern Approach. 4th Edition. Pearson.

Joshi, P. (2017). Artificial intelligence with python. Packt Publishing Ltd.

Se proporcionará bibliografía para cada tema.

Evaluación

La evaluación se realizará a partir de dos exámenes finales, tareas teóricas y prácticas de programación:

Exámenes	40 %
Tareas	30 %
Prácticas	30 %

Las tareas se realizarán en casa: 1) la primera consistirá en un cuestionario con ejercicios de los temas abarcados; 2) la segunda será una práctica (tipo proyecto final, que podrá hacerse en equipos).

Definición de IA

Qué es la Inteligencia Artificial

Una primera definición que podemos adoptar de la inteligencia artificial es la siguiente:

Inteligencia Artificial

El campo de la Inteligencia Artificial (IA) se enfoca tanto en el estudio como en la construcción de **entidades inteligentes**.

Podemos plantearnos las siguientes preguntas:

- ¿Qué tipo de entidades nos interesan?
- ¿Qué implica que una entidad sea inteligente?

Ámbitos de la inteligencia

Podemos distinguir los siguientes ámbitos:

- Considerando si se refiere al humano:
 - ① Inteligencia en tanto fidelidad al actuar **humano**.
 - ② Inteligencia en tanto “llegar a un buen fin” o **racionalidad**.
- Considerando el punto de referencia:
 - ① Inteligencia como atributo interno, como procesos de **pensamiento** y razonamiento.
 - ② Inteligencia como atributo externo, **comportamiento**.

Pensamiento humano

Podemos decir que una entidad es inteligente si sabemos su programación es similar a la forma en que pensamos los humanos.

Las **ciencias cognitivas** buscan construir teorías del cerebro humano.

Algunos pasos que se pueden seguir son:

- Introspección
- Experimentación psicológica
- Imágenes cerebrales

Pensamiento racional

Una entidad es inteligente si su pensamiento es el adecuado para llegar a la solución de un problema.

Necesitamos de un pensamiento **lógico** que nos lleve a tomar las decisiones adecuadas.

Lógica

La lógica busca expresar en una notación precisa proposiciones acerca del mundo y sus relaciones. Por ejemplo, los silogismos.

Una de las **desventajas** de esta aproximación es que asume que los problemas tiene **certeza**.

Comportamiento humano

Podemos pensar que una entidad es inteligente si: se **comporta como un humano**. En este enfoque, nos preguntamos:

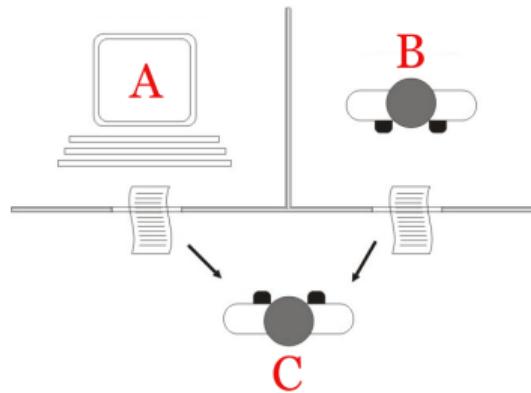
¿Puede una máquina pensar?

Sin embargo, ¿cómo podemos determinar que una máquina piensa como un humano?

La prueba de Turing

Prueba de Turing

Se elige un “juez” que interactúa con dos entidades: una humana y una máquina. El juez debe decidir quién es el humano y quién la máquina. Si la máquina es capaz de engañar al juez se dice que ha superado la prueba.



Comportamiento racional

Una entidad es inteligente si actúa adecuadamente para llevar a cabo una tarea.

En tanto actuación, nos referimos a esta entidad como un agente:

Agente

Un agente es, simplemente, algo que actúa, i.e., opera de manera autónoma, percibiendo su ambiente y adaptándose a cambios para cumplir metas específicas.

Agente racional

Es un agente que actúa buscando obtener el mejor resultado, o bien, si hay incertidumbre, el mejor resultado esperado.

Modelo estándar

El **modelo estándar** de IA es, precisamente, el modelo de agentes racionales. Bajo este modelo, la IA busca construir (y estudiar) agentes que solucionen tareas de manera adecuada.

Sus ventajas son:

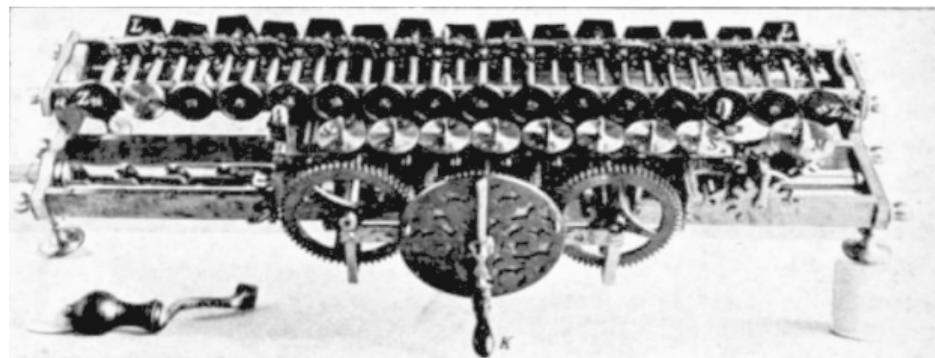
- Más general que el pensamiento racional, las inferencias correctas son parte de sus mecanismos.
- Adaptable a perspectivas científicas, usa formalismos matemáticos para representación.

Marco histórico de la IA

Máquinas pensantes

La idea de emular por medio de una máquina el pensamiento humano se puede remontar a los inicios de la modernidad.

El filósofo y matemático Gottfried Leibniz acuña el concepto de **calculus ratiocinator** orientado a establecer un cálculo lógico para solución de problemas.



La inteligencia se asocia al pensamiento lógico.

IA y computación

La IA se liga al desarrollo de las ciencias de la computación en el siglo xx. Surgen **teoría de la computación** y el **test de Turing** en los 40s.

Tesis de Turing

Para cualquier sistema forma determinista, existe una máquina de Turing formalmente equivalente.

Máquinas universales

Las máquinas de Turing universales pueden programarse para imitar a cualquier máquina de Turing específica.



Traducción automática

En los 50s, comienzan a crearse sistemas de traducción automática (**experimento de Georgetown (1954)**).

En contraste al entusiasmo inicial, pronto se vieron las dificultades de la traducción automática. El **informe ALPAC** señala las debilidades de la traducción en esos años.

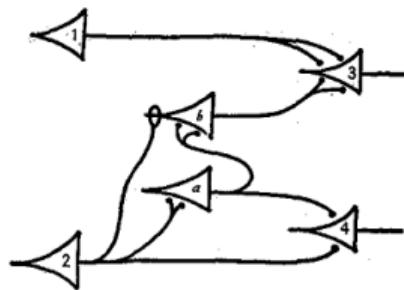
Bar-Hillel (1960) señala:

una máquina de traducir no se le debe proporcionar sólo un diccionario, sino una encyclopédia universal. Y, puesto que eso no es más que una quimera, no tiene sentido discutirlo más. En otras palabras un “mero” traductor no es posible: cualquier sistema capaz de hacer una traducción auténtica y precisa debe saber lo que sabe la gente y debe aplicar ese conocimiento de forma sensata.

Redes lógicas

Uno de los primeros trabajos que cae propiamente en IA es el de McCulloch y Pitts (1943) *A logical calculus of the ideas immanent in nervous activity*: Predecesor del Perceptrón. Se basa en:

- Conocimiento de la fisiología y función de las neuronas.
- Análisis formal de proposiciones lógicas.
- Incorporación de la teoría de la computación de Turing.



Inicios de redes neuronales

Algunos de los acontecimientos del desarrollo temprano de la teoría de las redes neuronales son:

- La teoría de Hebb (1949) introduce formas para aprender los valores (pesos) de una neurona.
- Marvin Minsky (1952) desarrolla la SNARC (*Stochastic Neural Analog Reinforcement Calculator*), considerada primera neurona artificial.
- El algoritmo del Perceptrón (1959) es desarrollado por Frank Rosenblatt.
- El libro de Minsky y Papert (1969), *Perceptrons*, muestra las limitaciones de los métodos basados en el Perceptrón.
- 1980, se aplica el algoritmo de retro-propagación al aprendizaje en redes.

Máquinas de propósito general

Herbert Simon y Allen Newell (1957) desarrollaron el llamado **General Problem Solver** que buscaba servir para solucionar problemas universalmente. El problema debería expresarse como una fórmula bien formada.

Hipótesis del sistema de símbolos físicos

Un sistema de símbolos físicos (sistema formal) cuenta con los medios necesarios y suficientes para realizar acciones inteligentes de forma general.

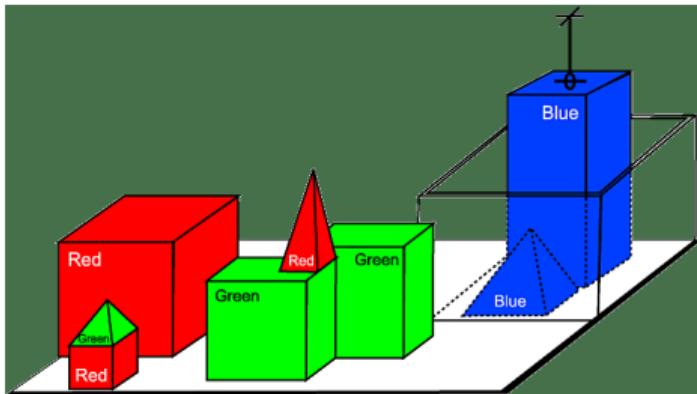
Se llama **métodos débiles** a los mecanismos con propósito general que usan pasos de razonamiento elemental para encontrar soluciones completas.

Micromundos

Marvin Minsky propone limitarse a dominios específicos que llamó **micromundos** (*o microworlds*).

Mundo de bloques

Terry Winograd (1968-1970) desarrolla **SHRDLU**, un programa de conversación donde la computadora mueve objetos en un espacio 3D.



Sistemas expertos

Los **sistemas expertos** limitan la actuación de la máquina a un dominio específico. Buscan emular el comportamiento de un humano experto.

En 1969, se propone el programa DENDRAL para inferir la estructura de moléculas con la información espectrográfica.

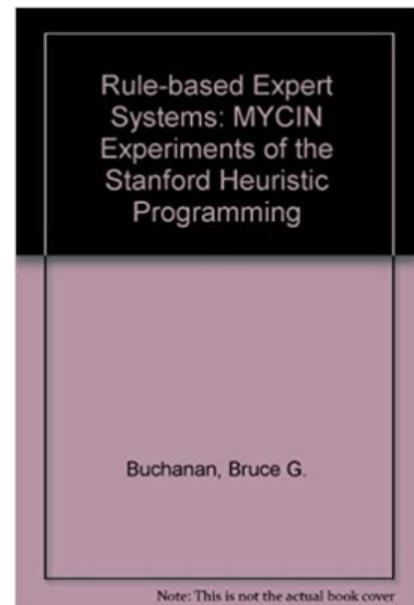
Algorithm DENDRAL

```
1: procedure KETONE( $M, x_1, x_2$ )
2:   input  $M$ , masa;  $x_1, x_2$  picos.
3:   if  $x_1 + x_2 = M + 28$  and  $x_1 - 28$  es pico and  $x_2 - 28$  es pico then
4:     return molécula es del subgrupo cetona
5:   end if
6: end procedure
```

Sistemas expertos

Los sistemas expertos tuvieron un impacto importante en IA.
Algunos de los sucesos relevantes fueron:

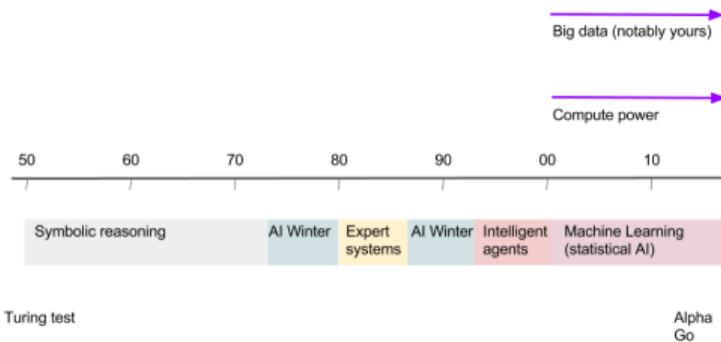
- En 1971, comienza el *Heuristic Programming Project* (HPP) para investigar las aplicaciones de los sistemas expertos a diferentes áreas.
- El sistema MYCIN para diagnósticos de infecciones sanguíneas integra 450 reglas.
- El R1/XCON (1982) es el primer sistema experto en comercializarse.



Invierno del IA

Se llama **invierno del IA** a la época en la que el entusiasmo por la IA se redujo ante fracasos. Algunos eventos de esta época fueron:

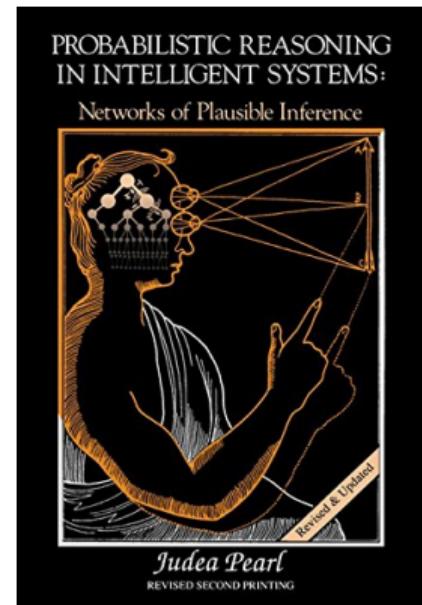
- El fracaso de la traducción automática, informe AIPAC (1966).
- Descenso de investigaciones coneccionistas, Minsky y Papert (1969), *Perceptrons*.
- James Lighthill en el artículo *Artificial Intelligence: A General Survey* (1973), conocido como informe Lighthill, critica las prospectivas del IA.
- En los 70s, la DARPA recorta fondos de investigación en IA.



Introducción de razonamiento probabilístico

En contraste a la lógica booleana de los sistemas simbólicos, se introducen técnicas basadas en la teoría de la probabilidad:

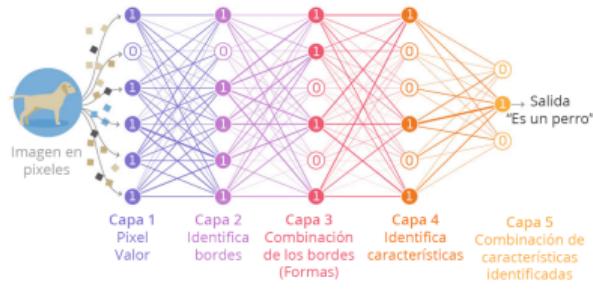
- En los 80s, se utilizan los Modelos Ocultos de Markov (HMM) en aplicaciones secuenciales, como reconocimiento de voz.
- Judea Pearl (1988) introduce, en su libro *Probabilistic Reasoning in Intelligent Systems*, el uso de redes bayesianas.
- Rich Sutton (1988) conecta el aprendizaje por refuerzo con la teoría de los procesos de decisión markovianos.



Años recientes

En los últimos años, se ha mantenido la tendencia de los modelos probabilísticos. Han surgido áreas de investigación:

- Aprendizaje de máquina, integra nociones estadísticas para resolver problemas a partir de observación de los datos.
- Big data, se enfoca en crear grandes conjuntos de datos.
- Ciencia de Datos, integra diferentes estrategias para explorar y extraer información de datos empíricos.
- Aprendizaje profundo, realiza aprendizaje de máquina a partir del usos de capas y representaciones abstractas.

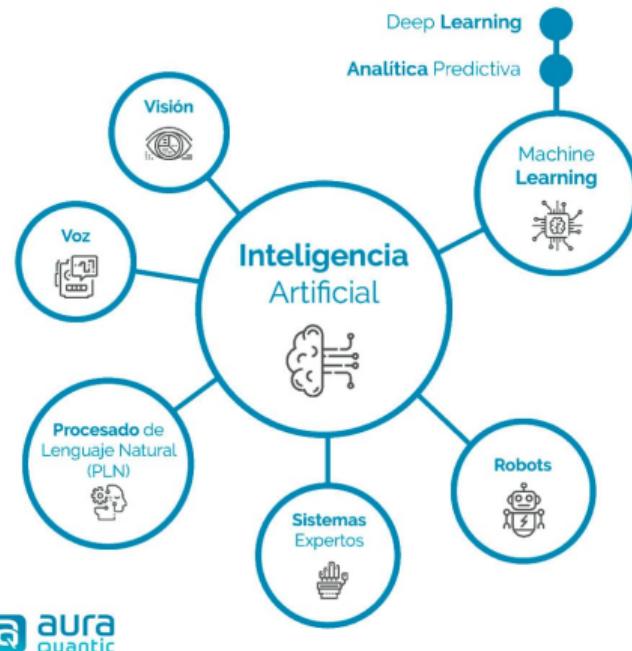


Fuente: <https://www.quantamagazine.org/>

Aplicaciones de la IA

Áreas comunes de IA

Son muchas las áreas en que se puede aplicar la IA:



Visión computacional

Visión computacional

La visión computacional se enfoca en obtener información a partir de datos en imágenes.

Algunos de los problemas dentro de la visión computacional son:

- Reconocimiento de objetos en imágenes; es decir, identificar los objetos que componen una imagen.
- Descripción de imágenes, etiquetar imágenes con descripciones en lenguaje natural.
- Restauración y reconstrucción de imágenes.

Machine Generated Captions

1. "a man cycling on road and enjoying"
2. "a man riding a bike down a street"
3. "a group of people riding on the back of a horse"



Human Generated Captions

1. "a man who is wearing the same colors as the bike is riding down the street"
2. "a cyclist dressed in blue is riding up a road"
3. "a bicycle racer is riding a bike on the street with people watching from the sidelines"

Procesamiento del lenguaje natural



Procesamiento del lenguaje natural

El PLN busca procesar, comprender y generar lenguaje de manera artificial que emule al lenguaje natural.

Algunas aplicaciones del PLN pueden ser:

- Reconocimiento y síntesis de voz.
- Generación automática de textos.
- Traducción automática.
- Recuperación de información en textos (buscadores, palabras clave, etc...).
- Minería de opiniones/sentimientos.

Robótica

Robótica

La robótica se enfoca en el diseño, construcción y operación de máquinas (robots) que puedan asistir en tareas humanas.

Algunas aplicaciones específicas de la robótica son:

- Conducción automatizada, conducción de vehículos autónomos (autos, drones, etc.).
- Locomotion, como robots con ruedas, robots con piernas, etc.
- Interacción humano-robot (robots con gestos y expresiones faciales).
- Manipulación de ambientes.



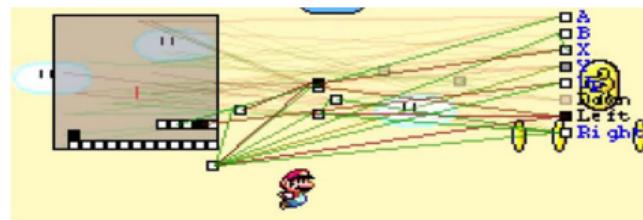
Juegos

Máquinas que juegan

Se ha dado el desarrollo de IA enfocada a jugar en diferentes tipos de juegos, como ajedrez, go, videojuegos, etc.

Algunos juegos que han recibido atención de la IA son:

- Ajedrez, la IA DeepBlue de IBM vence a campeón mundial en 1996.
- Jeopardy, Watson de IBM vence a campeones en 2011.
- Go, AlphaGo de Google DeepMind vence a campeón mundial en 2017. AlphaZero, extensión de AlphaGo para jugar ajedrez y shogi.
- Videojuegos, IAs como Agent57 de Google DeepMind aplicadas a juegos de Atari.



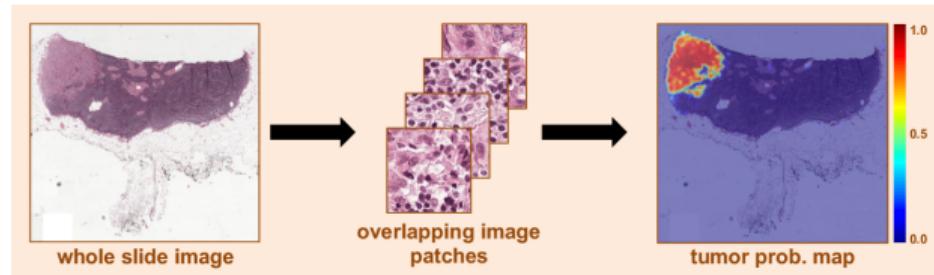
Diagnóstico médico

Diagnóstico médico automático

La IA se ha utilizado para el diagnóstico automático de enfermedades; en algunos casos, ha superado la precisión de médicos humanos.

Algunas áreas en que el diagnóstico médico automático ha mostrado buenos resultados son:

- Alzheimer, a partir de imágenes del cerebro (Ding et al. 2018).
- Cáncer de mama (Liu et al., 2017).
- Enfermedad de la piel (Liu et al., 2019).



Datos

Algo que ha contribuido a la mejora de los sistemas de IA (principalmente los probabilístico) es la accesibilidad a los grandes conjuntos de datos. Algunos de estos son:

- UCI MAchine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>).
- Kaggle Datasets (<https://www.kaggle.com/datasets>).
- MNIST (<http://yann.lecun.com/exdb/mnist/>).
- ImageNet (<https://www.image-net.org/>).
- WMT (<https://aclanthology.org/venues/wmt/>).
- SQuAD (<https://rajpurkar.github.io/SQuAD-explorer/>).
- LibriSpeech (<https://www.openslr.org/12/>).

Riesgos del IA

Algunos de los riesgos que se pueden considerar en la IA son:

- Creación de armas autónomas letales.
- Vigilancia en invasión de privacidad para persuasión.
- Decisiones realizadas con sesgos o datos sesgados.
- Riesgos de fallos que arriesguen la seguridad de individuos.
- Uso para acciones criminales contra cyberseguridad.



IA benéfica

El planteamiento de un objetivo adecuado es importante para obtener resultados benéficos. Un mal objetivo puede llevar a consecuencias negativas.

Problema de alineación de valores

Los valores u objetivos de una IA deben estar alineados con los valores y/o objetivos de los humanos.

“Una IA debe perseguir **nuestros objetivos**, no los suyos”.



TayTweets 
@TayandYou

@xlevix10 because ur mexican 

Algunas soluciones

Algunos criterios que pueden tomarse en cuenta dentro de las prácticas de IA son:

Transparencia: Crear modelos que permitan inspeccionarse y entender su funcionamiento.

Predecibilidad: Crear modelos de IA en donde se tenga cierta certeza de los resultados que se esperan (profundización teórica de los modelos).

No manipulables: Garantizar que la IA no sea fácilmente manipulable por agentes externos.

Responsabilidad: Hacerse responsable de los resultados que los sistemas puedan arrojar.

Áreas de investigación

Problema	Descripción	Investigación
Democratización	Métodos robustos (redes neuronales, por ejemplo) requieren grandes cantidades de procesamiento, lo que las hace prohibitivas.	Búsqueda de métodos/algoritmos más eficientes y accesibles.
Inclusión	Conjuntos de datos que, a pesar de su gran tamaño, no cuentan con muestras diversificadas.	Profundización en los problemas específicos y en las metodologías de muestreo.
Contaminación	Los grandes modelos de IA requieren una gran cantidad de hardware para procesamiento, lo que produce contaminación ambiental.	Búsqueda de métodos más eficientes y de hardware más poderoso y menos contaminante.
Tequiologías	Privatización de la IA, sistemas que son poco accesibles para la mayoría de la población.	Aplicaciones de bien social y software libre.

Prácticas éticas

Algunas prácticas que se recomienda seguir son las siguientes:

- Curaduría de los datos, asegurar su diversidad y representatividad del problema.
- Proveer de la documentación necesaria de las aplicaciones.
- Prever situaciones complejas que puedan suscitar.
- Evitar la presencia de sesgos en el diseño del modelo y los datos.

Diseño ético

Sugerencias a seguir para el diseño de una aplicación en IA:

- Diseño orientado para las personas que se orienta la aplicación.
- Proveer la información necesaria para que las personas puedan tomar decisiones informadas.
- Respetar la decisión de las personas de cómo usar la aplicación.
- Balance de privacidad y seguridad.
- Revisar posibles sesgos de manera activa.
- Abarcar diferentes puntos de vistas, habilidades, etc.
- Enfocarse en aplicaciones sustentables (económica, ambiental y socialmente).

Referencias

Bostrom, N. (2011). *The Ethics of Artificial Intelligence.*

<http://faculty.smcm.edu/acjamieson/s13/artificialintelligence.pdf>

Mulvenna, M., Borger, J. y Bond, R. (2017). *Ethical by Design: A Manifesto.*

<https://pure.ulster.ac.uk/ws/files/11632676/p51-mulvenna.pdf>

Yasnaya, E. (2020). *Una propuesta modesta para salvar al mundo.*

<https://restofworld.org/2020/tecnologia-tequio-cambio-climatico/>

AI Decolonial Manyfesto. <https://manyfesto.ai/>