

Ética e Inteligencia Artificial

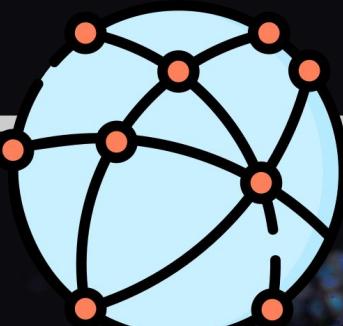
Inteligencia artificial en el mundo



Sistema de IA global (jerárquico)



Personas
Culturas
Sociedades



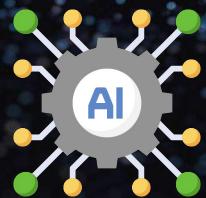
Instituciones
Estados
Empresas



Sistemas (con y sin IA)

Inteligencia artificial y software

- Software
 - Colección de funciones
 - Forman un programa
- Inteligencia artificial
 - Función **especial** dentro de un programa
 - **Difícil caracterizar**
 - Ejemplos
 - Jugar al Ta-Te-Ti
 - tateti(Estado) → Acción
 - Manejar un auto automáticamente
 - manejar(Estado) → Acción



Modelos de IA basados en *datos*

- Modelos que *aprenden* o se *entrenan*
- Datos → Ejemplos



Ejemplos



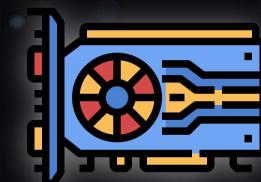
Algoritmo de
Aprendizaje



Modelo

Inteligencia artificial: cambio de paradigma

- 1950 - 2010: Sistemas de IA **custom**
 - Basados en conocimiento experto
 - Lógica y búsqueda
- 2010 - Hoy: Sistemas de IA como APIs
 - Modelos basados en **datos masivos**
 - Crecimiento de información disponible
 - Modelos conexionistas o no simbólicos
 - Redes Neuronales o Deep Learning
 - Aumento del poder de cómputo
 - GPUs, TPUs, dispositivos especiales



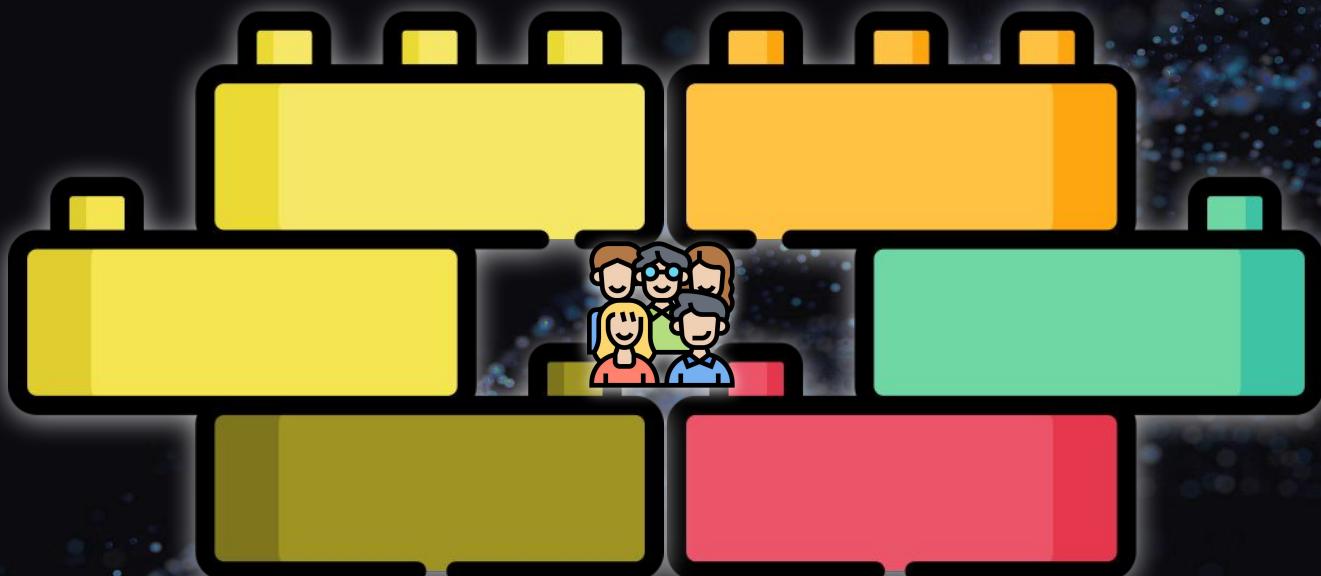
Impacto de sistemas de AI: Ejemplo de GPT

- **GPT**: Modelo de generación de texto
- Dada una entrada, completa la salida. Ejemplos:
 - Entrada: *Te hacías un café con azúcar, pero no tenías cuchara así que revolviste con una lapisera. Resulta que fue una mala idea porque ...*
 - Salida: **la lapisera se derritió.**
- Alta calidad, posibilita nuevas aplicaciones
 - Chatbots, asistencia para escribir/programar, etc
 - **Generar Spam o contenido falso**
 - Requiere análisis de impacto
 - Release Strategies and the Social Impacts of Language Models

Aspectos a evaluar

- **Desempeño**
 - ¿El modelo realiza la tarea correctamente?
- **Interpretabilidad**
 - ¿Cómo afecta a la salida un cambio en la entrada?
 - ¿Se pueden comprender los pasos para generar la salida?
 - Caja blanca vs Caja Negra
- **Sesgo de datos**
 - Datos usados para generar un modelo
 - ¿reflejan la realidad? ¿deben reflejarla? ¿la de quienes?
- **Interacción con el mundo real**
 - ¿Cómo afectará el sistema al mundo real?

Agregar o modificar sistemas de IA



Considerar el impacto

Problemáticas asociadas a modelos de IA

- Estás desarrollando alguno de estos sistemas
 - a. Reconocedor de habla para celulares/computadoras.
 - b. Generador de imágenes artísticas para un software de diseño.
 - c. Predictor del crecimiento económico de un país para evaluación de riesgos financieros.
- ¿Qué impactos podés preveer a largo plazo si el sistema es exitoso?
- ¿Qué medidas podés tomar para mitigar los impactos negativos?