

# Problemas éticos y morales de la Inteligencia Artificial. Evolución a lo largo de la historia

Historia de las matemáticas - Universidad de Granada

Antonio Martín Ruiz  
Laura Gómez Garrido  
Fernando de la Hoz Moreno

11 de diciembre de 2019

# Contenido

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar

# Introducción. ¿Qué es la IA?

Cuatro visiones:

- Sistemas que piensan como humanos
- Sistemas que actúan como humanos
- Sistemas que piensan racionalmente
- Sistemas que actúan racionalmente

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA**
- 3 Objeciones de Turing
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar

- Génesis de la Inteligencia Artificial (1943-1955)
- Nacimiento de la Inteligencia Artificial (1956)
- Entusiasmo inicial, grandes esperanzas (1952-1969)
- Una dosis de realidad (1966-1973)

Fernando

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing**
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar



## Objeción Teológica:

*El pensamiento es una función del alma inmortal del hombre. Dios ha proporcionado un alma inmortal a todos los hombres y mujeres, pero no así a ningún otro animal, ni tampoco a las máquinas. Por consiguiente, ningún animal o máquina puede pensar.*

## El argumento de la percepción extrasensorial.

Si estas objeciones fueran ciertas, ya no podríamos considerar que nuestros cuerpos se mueven de acuerdo a las leyes físicas conocidas ni por las que aún están por descubrir.

## La objeción de la *cabeza en la arena*:

*Las consecuencias de que las máquinas pensaran serían demasiado terribles. Esperemos y creamos que no pueden hacerlo.*

**Argumento de la conciencia:** Se basa en que no podemos considerar que una máquina iguala al cerebro humano hasta que no sea realmente consciente de lo que ella misma esta creando y pueda tener sentimientos y emociones.

## La objeción matemática:

*Existen muchos resultados de lógica matemática que pueden utilizarse para demostrar que hay limitaciones al potencial de las máquinas de estado discreto. [...] Este es el resultado matemático: se afirma que prueba que las máquinas adolecen de una incapacidad a la que no se encuentra sujeto el intelecto humano.*

## Argumentos sobre diversas incapacidades:

Son de la forma: *"Acepto que puedas hacer que las máquinas hagan todo lo que hasta ahora has mencionado, pero nunca podrás hacer que una de ellas haga X"*

## La objeción de Lady Lovelace

*La máquina no pretende crear nada. Puede hacer lo que sea que sepamos ordenarle.* - Ada Lovelace, 1842.

## **El argumento de la continuidad del sistema nervioso**

El sistema nervioso no es una máquina de estado discreto y, por ello, no deberíamos de poder ser capaces de imitar su comportamiento de forma discreta.

## El argumento de la informalidad del comportamiento

*No es posible producir un conjunto de reglas que pretenda describir lo que una persona debe hacer en cada grupo de circunstancias concedible. Podría, por ejemplo, haber una regla que dictara que debemos detenernos al ver la luz roja de un semáforo y avanzar cuando la luz cambie a verde. Entonces, ¿qué sucedería si por algún desperfecto ambas aparecieran al mismo tiempo? Tal vez se decidiría que lo más seguro sería detenerse. No obstante, más adelante podría surgir otra dificultad a raíz de esta decisión. Intentar proporcionar reglas de conducta que cubran cualquier eventualidad, incluso las que surjan a partir de las luces de los semáforos, parecería imposible.*

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing
- 4 Singularidad tecnológica**
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar

## Contenido



- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA**
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar

## Contenido

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA**
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar

Características que hacen peligrosa a la IA:

- Posible utilización destructiva
- Eficientes y escalables
- Aumentan anonimato y distancia psicológica
- Rápida difusión
- Vulnerabilidades

Los sistemas distribuidos favorecen la automatización de ataques.

En la actualidad solo se conocen ataques llevados a cabo por sombreros blancos.

Actual utilización de sistemas de IA en ciberseguridad.

Algunos casos concretos:

- Automatización de ataques de ingeniería social.
- Automatización del descubrimiento de vulnerabilidades.
- Ataques DDoS imitando el comportamiento humano.

Aplicaciones armamentísticas.

Aumento de la diferencia entre potencial ofensivo y defensivo.

- Reutilización terrorista de sistemas comerciales.
- Aumento de la capacidad ofensiva.
- Aumento de la escala de los ataques.

Avance tecnológico e inestabilidad política están ligados.

Producción y detección de información manipulativa.

Control de la población.

Ataques contra la seguridad política:

- Plataformas de vigilancia automatizada.
- Informes de noticias falsa.
- Campañas de desinformación.

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar



## Contenido

- 1 Introducción. ¿Qué es la IA?
- 2 Historia de la IA
- 3 Objeciones de Turing
- 4 Singularidad tecnológica
- 5 Peligros no intencionados. Seguridad en IA
- 6 Peligros intencionados. Uso malicioso de la IA
- 7 Consecuencias de la IA sobre el empleo
- 8 Para ampliar

*Artificial Intelligence: A Modern Approach* Stuart J. Russell y Peter Norvig

*Concrete Problems in AI Safety* Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané

*The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* Miles Brundage, Shahar Avin et al.