

# Trabajo Práctico Nro 12 - Inteligencia Artificial

## Ejercicio 1

1. Defina Inteligencia e Inteligencia Artificial.

La **inteligencia** se puede definir como la capacidad de un individuo para aprender, razonar, resolver problemas, comprender ideas complejas, adaptarse a nuevas situaciones y utilizar el conocimiento para manipular su entorno.

La **inteligencia artificial** es un campo de la informática que se centra en la creación de sistemas y programas capaces de realizar tareas que normalmente requieren inteligencia humana. Esto incluye habilidades como el aprendizaje, el razonamiento, la percepción y la toma de decisiones.

2. De ejemplos de sistemas de IA actuales.

Como ejemplos podemos ver *Asistentes Virtuales* y *Sistemas de Procesamiento de Lenguaje Natural* como pueden ser: Siri, Alexa, Chatbots, entre otros.

3. Explique la diferencia entre IA débil e IA fuerte.

Característica	IA Débil	IA Fuerte
Definición	Sistemas de inteligencia artificial diseñados y entrenados para realizar una <b>tarea específica y limitada</b> . Estos sistemas operan dentro de un conjunto predefinido de reglas y datos, y su "inteligencia" se manifiesta únicamente en la ejecución eficiente de esa tarea particular.	Tipo de inteligencia artificial que posee la capacidad de <b>entender, aprender y aplicar inteligencia a cualquier tarea intelectual que un ser humano pueda realizar</b> . Implica una comprensión genuina, conciencia y la capacidad de razonar, planificar, resolver problemas y aprender de la experiencia en una amplia gama de dominios.
Características Claves	<ul style="list-style-type: none"><li>- <b>Especialización</b>: Son expertos en una única función o dominio.</li><li>- <b>Ausencia de Conciencia o Comprensión Genuina</b>: No poseen autoconciencia, emociones, intenciones o una comprensión real del mundo. Simplemente procesan</li></ul>	<ul style="list-style-type: none"><li>- <b>Generalización</b>: Puede realizar cualquier tarea cognitiva, no solo una específica.</li><li>- <b>Conciencia y Autoconciencia (Teórica)</b>: Se postula que una IA fuerte tendría conciencia, autoconciencia, emociones y la capacidad de experimentar el mundo de manera</li></ul>

Característica	IA Débil	IA Fuerte
	<p>información y ejecutan algoritmos.</p> <ul style="list-style-type: none"> <li>- <i>No Generalizable</i>: Su conocimiento y habilidades no pueden transferirse fácilmente a otras tareas o dominios diferentes.</li> <li>- <i>Basada en Reglas o Aprendizaje de Patrones</i>: Su funcionamiento se basa en algoritmos que siguen reglas explícitas o en modelos que aprenden patrones a partir de grandes volúmenes de datos.</li> </ul>	<p>similar a los humanos.</p> <ul style="list-style-type: none"> <li>- <i>Comprensión Genuina</i>: No solo procesa información, sino que la "entiende" en un sentido profundo y contextual.</li> <li>- <i>Transferencia de Conocimiento</i>: Puede aplicar el conocimiento adquirido en un dominio a problemas en otros dominios completamente diferentes.</li> <li>- <i>Creatividad y Sentido Común</i>: Sería capaz de generar ideas novedosas, mostrar creatividad y aplicar el sentido común en situaciones ambiguas o no estructuradas.</li> </ul>
Actualidad	Actualmente es <b>prácticamente realizable</b> .	Actualmente es un <b>concepto teórico y un objetivo a largo plazo</b> .

4. Explique la diferencia entre un sistema de IA y un sistema de software complejo (como el que calcula la trayectoria de un cohete a Marte).

Aunque tanto los sistemas de IA como los sistemas de software complejos implican algoritmos sofisticados y grandes volúmenes de procesamiento de datos, su **naturaleza fundamental, su propósito y su forma de operar** son distintos. La clave reside en la **adaptabilidad, el aprendizaje y la autonomía** que caracterizan a la IA.

- Un **sistema de software complejo** es determinista, se basa en reglas explícitas, busca la optimización y precisión, tiene una entrada y salida definidas, no aprende ni se adapta, entre otras cosas.
- Un **sistema de IA** es probabilístico, se basa en el aprendizaje, busca la adaptabilidad y generalización, maneja incertidumbre y ambigüedad, tiene un aprendizaje continuo, entre otras cosas.

5. Defina IA simbólica e IA no-simbólica.

Característica	IA Simbólica	IA No-Simbólica
Enfoque	Se basa en la idea de que la inteligencia puede ser modelada mediante la <b>manipulación de símbolos y reglas lógicas</b> . Se busca representar el conocimiento humano de forma	Se basa en la idea de que la inteligencia emerge de <b>patrones y conexiones aprendidas a partir de grandes volúmenes de datos</b> , sin la necesidad de una

Característica	IA Simbólica	IA No-Simbólica
	explícita y estructurada, utilizando lenguajes formales y sistemas de lógica.	representación explícita y simbólica del conocimiento.
Principios Clave	<ul style="list-style-type: none"> <li>- <i>Representación del Conocimiento</i>: El conocimiento se codifica como símbolos (conceptos, objetos, propiedades) y relaciones entre ellos (reglas, hechos, lógica).</li> <li>- <i>Razonamiento Lógico</i>: La inteligencia se logra mediante la aplicación de reglas de inferencia lógica sobre estas representaciones simbólicas para deducir nuevas conclusiones o tomar decisiones.</li> <li>- <i>Programación Explícita</i>: Los sistemas se construyen programando explícitamente el conocimiento y las reglas de razonamiento.</li> <li>- <i>Transparencia</i>: En teoría, el proceso de razonamiento es interpretable y se puede seguir la cadena de inferencias que llevó a una conclusión.</li> </ul>	<ul style="list-style-type: none"> <li>- <i>Aprendizaje a partir de Datos</i>: Los sistemas aprenden a identificar patrones complejos y a hacer predicciones o clasificaciones directamente de los datos.</li> <li>- <i>Representaciones Distribuidas</i>: El conocimiento no se almacena en símbolos discretos, sino en las ponderaciones y conexiones de una red (como una red neuronal).</li> <li>- <i>Tolerancia a la Ambigüedad</i>: Son más robustos ante datos ruidosos o incompletos, ya que aprenden a generalizar a partir de ejemplos.</li> <li>- <i>Énfasis en la Percepción y el Reconocimiento</i>: Excelentes para tareas donde el conocimiento es difícil de formalizar simbólicamente.</li> </ul>
Ventajas	<ul style="list-style-type: none"> <li>- <i>Interpretabilidad</i>: Facilita la comprensión de cómo el sistema llega a sus conclusiones.</li> <li>- <i>Precisión Lógica</i>: Ideal para problemas bien definidos con reglas claras.</li> <li>- <i>Facilidad para Incorporar Conocimiento Experto</i>: Permite a los expertos en un dominio codificar su conocimiento directamente en el sistema.</li> </ul>	<ul style="list-style-type: none"> <li>- <i>Robustez</i>: Manejan bien la incertidumbre y los datos ruidosos.</li> <li>- <i>Escalabilidad</i>: Pueden procesar y aprender de cantidades masivas de datos.</li> <li>- <i>Descubrimiento de Patrones</i>: Capaces de encontrar patrones ocultos en los datos que no serían evidentes para los humanos.</li> </ul>
Desventajas	<ul style="list-style-type: none"> <li>- <i>Problema del Sentido Común</i>: Dificultad para codificar la vasta cantidad de conocimiento de sentido común que los humanos adquieren de forma natural.</li> <li>- <i>Fragilidad</i>: Los sistemas son frágiles ante situaciones no previstas o datos ambiguos, ya que dependen de reglas explícitas.</li> <li>- <i>Escalabilidad</i>: Se vuelve muy complejo y costoso mantener y</li> </ul>	<ul style="list-style-type: none"> <li>- <i>Falta de Interpretabilidad ("Caja Negra")</i>: A menudo es difícil entender por qué un sistema no-simbólico tomó una decisión particular.</li> <li>- <i>Dependencia de Datos</i>: Requieren grandes cantidades de datos de entrenamiento de alta calidad.</li> <li>- <i>Dificultad con el Razonamiento Lógico Explícito</i>: No son</li> </ul>

Característica	IA Simbólica	IA No-Simbólica
	expandir bases de conocimiento muy grandes.	inherentemente buenos para el razonamiento lógico complejo o la planificación a largo plazo sin técnicas híbridas. - <i>Vulnerabilidad a Sesgos</i> : Pueden aprender y amplificar sesgos presentes en los datos de entrenamiento.

## 6. Defina Machine Learning (ML). ¿ML está incluido en IA?

Machine Learning es una rama de la Inteligencia Artificial que se enfoca en el desarrollo de algoritmos y modelos que permiten a los sistemas **aprender de los datos sin ser programados explícitamente** para cada tarea. En lugar de seguir instrucciones paso a paso para resolver un problema, los sistemas de ML identifican patrones y relaciones en grandes conjuntos de datos, y utilizan esos patrones para hacer predicciones, clasificaciones o tomar decisiones sobre datos nuevos y no vistos.

Es una de las **técnicas o metodologías principales** que se utilizan para lograr la IA. Es la forma en que muchas de las capacidades inteligentes se implementan en los sistemas de IA actuales.

El **proceso fundamental del ML** implica:

- *Datos de entrenamiento.*
- *Aprendizaje de Patrones.*
- *Construcción de un Modelo.*
- *Predicción o Decisión.*

## 7. ¿Qué es una Red neuronal artificial (ANN)? Dibuje un ejemplo señalando cada una de sus partes.

Una **Red neuronal artificial (ANN)** es un modelo computacional inspirado en la estructura y el funcionamiento del cerebro humano. Está diseñada para reconocer patrones, clasificar datos y hacer predicciones, aprendiendo de grandes volúmenes de información.

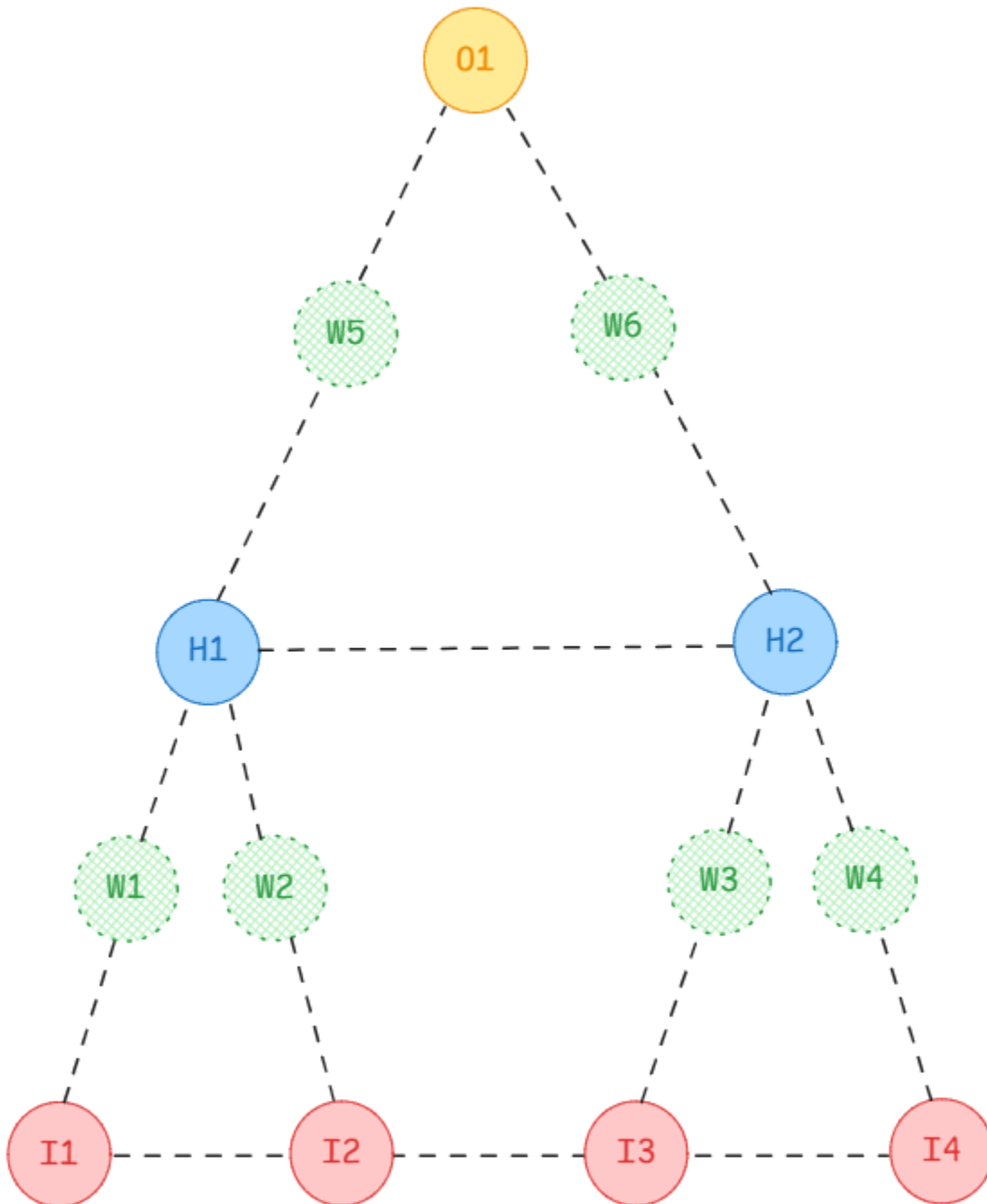
Están compuestas por un conjunto de **nodos interconectados** (llamados "neuronas" o "unidades") organizados en **capas**. Cada conexión entre neuronas tiene un **"peso" asociado**, que representa la fuerza o importancia de esa conexión. Durante el **proceso de "entrenamiento"**, la red ajusta estos pesos para minimizar la diferencia entre sus predicciones y los resultados reales, aprendiendo así a realizar una tarea específica.

## Funcionamiento Básico:

1. **Entrada:** Las neuronas de la capa de entrada reciben datos (características o atributos).
2. **Propagación:** Estos datos se propagan a través de las capas ocultas. En cada neurona, se realiza una suma ponderada de las entradas (cada entrada multiplicada por el peso de su conexión) y se le añade un "sesgo" (bias).
3. **Función de Activación:** El resultado de esta suma ponderada se pasa a través de una "función de activación" (no lineal) que decide si la neurona debe "activarse" y pasar información a la siguiente capa.
4. **Salida:** Finalmente, las neuronas de la capa de salida producen el resultado de la red (una predicción, una clasificación, etc.).
5. **Aprendizaje (Backpropagation):** Durante el entrenamiento, si la salida de la red es incorrecta, el error se propaga hacia atrás a través de la red (algoritmo de retro-propagación o *backpropagation*), ajustando los pesos de las conexiones para reducir el error en futuras predicciones.

## EJEMPLO

# CAPA DE SALIDA



# CAPA DE ENTRADA

## Capa de Entrada:

- *I1, I2, I3, I4:* Son las neuronas de entrada. Cada una de ellas recibe un valor de una característica o atributo del conjunto de datos. Por ejemplo, si la red está diseñada para predecir el precio de una casa, I1 podría ser el número de habitaciones, I2 el tamaño en

metros cuadrados, la antigüedad, etc. No realizan cálculos, solo distribuyen los datos de entrada.

### Capas Ocultas:

- **H1, H2:** Son las neuronas de la capa oculta. Una red puede tener una o varias capas ocultas. Estas capas son donde se realizan la mayoría de los cálculos complejos y donde la red aprende a extraer características y patrones de los datos de entrada. Cada neurona en una capa oculta recibe entradas de todas las neuronas de la capa anterior, realiza una suma ponderada de estas entradas, le añade un sesgo.

### Capa de Salida:

- **O1:** Es la neurona de salida. Produce el resultado final de la red. Por ejemplo, si la red predice el precio de una casa, O1 sería el precio estimado.

### Conexiones:

- Las líneas que conectan las neuronas entre capas. Representan el flujo de información.

### Pesos:

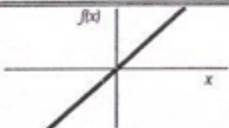
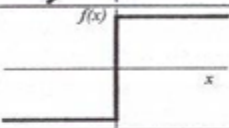
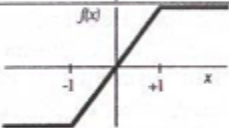
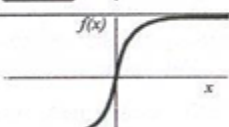
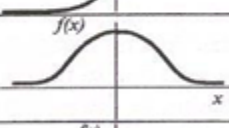
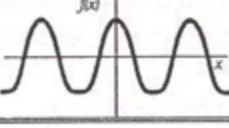
- **W1, W2, W3, W4, W5, W6:** Son los valores numéricos asociados a cada conexión. Indican la fuerza o importancia de la conexión entre dos neuronas. Durante el entrenamiento, estos pesos son ajustados por el algoritmo de aprendizaje para que la red pueda hacer predicciones más precisas. Un peso alto significa que la entrada de esa conexión es muy influyente en la neurona receptora, mientras que un peso bajo significa lo contrario.

### Sesgos:

- Aunque no están dibujados como elementos separados en el diagrama, cada neurona (excepto las de entrada) tiene un valor de sesgo asociado. El sesgo es un término constante que se añade a la suma ponderada de las entradas antes de aplicar la función de activación. Permite que la función de activación se desplace, lo que ayuda a la red a modelar relaciones más complejas.

### Funciones de Activación:

- No se muestran explícitamente como un componente gráfico, pero son una parte crucial de cada neurona en las capas ocultas y de salida (excepto las de entrada). Son funciones matemáticas no lineales que introducen no linealidad en la red, permitiéndole aprender patrones complejos y no lineales en los datos. Sin funciones de activación no lineales, una red neuronal, sin importar cuántas capas tenga, se comportaría como un modelo lineal simple.

	Función	Rango	Gráfica
Identidad	$y = x$	$[-\infty, +\infty]$	
Escalón	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
Lineal a tramos	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } -l \leq x \leq +l \\ +1, & \text{si } x > +l \end{cases}$	$[-1, +1]$	
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
Gaussiana	$y = Ae^{-Bx^2}$	$[0, +1]$	
Sinusoidal	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

## Ejercicio 2

Lea el artículo original de Turing sobre IA (Turing 1950). En él se comenta algunas objeciones potenciales a su propuesta y a su prueba de inteligencia. ¿Cuáles de estas objeciones tiene todavía validez? ¿Son válidas sus refutaciones? ¿Se le ocurren nuevas objeciones a esta propuesta teniendo en cuenta los desarrollos realizados desde que se escribió el artículo?

En el artículo, Turing predijo que para el año 2000 sería probable que una computadora tuviera un 30 por ciento de posibilidades de superar una Prueba de Turing dirigida por un evaluador inexperto con una duración de cinco minutos. ¿Consideras razonable lo anterior en el mundo actual?

### Contexto del Artículo >

El artículo de Alan Turing de 1950, "**Computing Machinery and Intelligence**", es un texto fundacional en el campo de la inteligencia artificial. En él, Turing propone el "Juego de la Imitación" (conocido popularmente como la Prueba de Turing) como un criterio para determinar si una máquina puede "pensar", y luego aborda una serie de objeciones potenciales a su propuesta.



## Objeciones potenciales a la Propuesta de Turing y su validez actual

- *Objeción Teológica:*

- Su **argumento** es que el pensamiento es una función del alma inmortal del hombre, dada por Dios, y las máquinas no tienen alma.
- La **refutación de Turing** argumenta que esto limita la omnipotencia de Dios, sugiriendo que Dios podría, si quisiera, conferir un alma a una máquina. Además, señala que los argumentos teológicos han demostrado ser insatisfactorios en el pasado.
- Esta objeción sigue siendo **válida en el ámbito teológico o filosófico-religioso**, ya que se basa en una creencia metafísica que no puede ser refutada por la ciencia o la tecnología. Sin embargo, **carece de validez científica o empírica** para la discusión sobre la capacidad de las máquinas para simular o exhibir inteligencia.

- *Objeción de la "Cabeza en la Arena":*

- Su **argumento** es que las consecuencias de que las máquinas pensaran serían demasiado terribles; por lo tanto, esperamos y creemos que no pueden hacerlo.
- La **refutación de Turing** no es lógica, sino que la descarta como una objeción emocional, sugiriendo que es un sentimiento de superioridad humana que no requiere refutación.
- Esta objeción sigue siendo **muy válida como preocupación social y ética**.

- *Objeción Matemática:*

- Su **argumento** se basa en teoremas como el de Gödel y los resultados de Church, Kleene, Rosser y Turing mismo, que demuestran limitaciones inherentes a cualquier sistema formal o máquina de estado discreto (incluyendo computadoras digitales). Se argumenta que hay preguntas que una máquina no podría responder correctamente o no podría responder en absoluto, a diferencia del intelecto humano.
- La **refutación de Turing** reconoce la validez de las limitaciones matemáticas para **cualquier máquina particular**, pero argumenta que no se ha demostrado que estas limitaciones se apliquen al intelecto humano. Además, señala que los humanos también cometen errores y que la superioridad percibida es ilusoria o limitada a casos específicos.
- La objeción **sigue siendo matemáticamente válida** para cualquier sistema computacional formal. Sin embargo, la **refutación de Turing sigue siendo fuerte**.

- *Argumento de la Conciencia:*

- Su **argumento** se basa en que una máquina no puede pensar a menos que tenga conciencia, sentimientos, emociones, etc.
- La **refutación de Turing** dice que la única forma de saber si algo tiene conciencia es ser ese algo (solipsismo), lo cual dificulta la comunicación de ideas. Propone que la Prueba de Turing es una forma práctica de evaluar la inteligencia sin resolver el misterio de la

conciencia. Si una máquina puede responder de manera convincente a preguntas sobre emociones o creatividad, ¿Cómo podríamos saber que no las "siente"?

- Esta objeción **sigue siendo profundamente válida y es uno de los debates centrales en la filosofía de la mente y la IA.**
- *Argumento de Diversas Incapacidades:*
  - Su **argumento** se basa en que las máquinas nunca podrán hacer X (ser amables, ingeniosas, cometer errores, enamorarse, aprender de la experiencia, etc.).
  - La **refutación de Turing** atribuye estas objeciones a la inducción científica defectuosa (basada en máquinas primitivas) y a la falta de imaginación.
  - La mayoría de estas objeciones han sido **refutadas por los avances de la IA.**
- *Objeción de Lady Lovelace:*
  - Su **argumento** se basa en que la máquina analítica "no pretende crear nada. Puede hacer cualquier cosa que sepamos ordenarle que haga". Es decir, las máquinas solo hacen lo que se les programa, no pueden ser originales o "sorprendernos".
  - La **refutación de Turing** argumenta que la máquina puede sorprendernos si no hemos calculado todas las consecuencias de su programación.
  - Esta objeción **ha sido en gran medida refutada por el Machine Learning y el Deep Learning.**
- *Argumento de la Continuidad del Sistema Nervioso*
  - Su **argumento** se basa en que el sistema nervioso es continuo, no discreto, y pequeños errores en la entrada pueden tener grandes efectos en la salida. Por lo tanto, una máquina discreta no puede imitarlo.
  - La **refutación de Turing** argumenta que una máquina discreta puede simular una máquina continua con suficiente precisión. Además, la Prueba de Turing se basa en la indistinguibilidad del comportamiento, no en la replicación exacta de la fisiología.
  - La objeción **tiene cierta validez teórica** en cuanto a la naturaleza fundamental de los sistemas. Sin embargo, la **refutación de Turing es práctica y efectiva.**
- *Argumento de la Informalidad del Comportamiento*
  - Su **argumento** nos dice que no es posible producir un conjunto de reglas que describa lo que una persona debe hacer en cada circunstancia concebible (ej. semáforos defectuosos). Por lo tanto, los humanos no pueden ser máquinas.
  - La **refutación de Turing** distingue entre "reglas de conducta" (preceptos conscientes) y "leyes de comportamiento" (leyes naturales). Argumenta que la ausencia de reglas de conducta explícitas no significa la ausencia de leyes de comportamiento subyacentes.
  - La objeción **sigue siendo relevante para la IA simbólica pura**, que lucha con la informalidad y el sentido común. Sin embargo, la **refutación de Turing es muy fuerte y ha sido validada por el ML.**
- *Argumento de la Percepción Extrasensorial (PES)*

- Su **argumento** nos dice que si la PES existe (telepatía, clarividencia, etc.), entonces el pensamiento no se rige solo por las leyes físicas conocidas, y una máquina física no podría replicarlo.
- La **refutación de Turing** reconoce la solidez del argumento si la PES es real y relevante para el pensamiento. Sugiere que, si la PES es un factor, la Prueba de Turing debería realizarse en una "habitación a prueba de telepatía".
- Esta objeción **depende de la existencia y relevancia de la PES**. Dado que la PES no es un fenómeno científicamente aceptado o replicable, la objeción **carece de validez empírica en el contexto científico actual**.

## Nuevas Objeciones a la Propuesta de Turing

- *El Problema de la "Caja Negra" y la Explicabilidad:*
  - **Objeción:** Los sistemas de IA modernos son "cajas negras". Pueden pasar la Prueba de Turing, pero no podemos entender **cómo** llegaron a sus respuestas o **por qué** tomaron ciertas decisiones. Esto contrasta con la comprensión humana y plantea problemas de confianza, sesgo y depuración.
  - **Relevancia para Turing:** Turing no anticipó la opacidad de los modelos de aprendizaje profundo. Su prueba se centra en el **comportamiento externo**, no en el **proceso interno**.
- *El Problema de la "Alucinación" y la Veracidad:*
  - **Objeción:** Los modelos generativos de IA pueden "alucinar", es decir, generar información falsa o sin sentido con gran fluidez y confianza, lo que dificulta distinguirlos de un humano que miente o se equivoca. Esto complica la Prueba de Turing, ya que un examinador podría confundir una alucinación con un error humano.
  - **Relevancia para Turing:** Turing consideró que una máquina podría cometer errores deliberadamente. Sin embargo, la "alucinación" es diferente; no es un error deliberado, sino una limitación inherente al modelo que genera contenido plausible pero incorrecto.

## Predicción de Turing para el Año 2000 y su Razonabilidad Actual

- Turing predijo que para el año 2000, una computadora tendría un **30% de posibilidades de superar una Prueba de Turing dirigida por un evaluador inexperto con una duración de cinco minutos**.
- Considero que esta predicción fue notablemente razonable y, en cierto sentido, se ha superado en el mundo actual.
- **En el año 2000**, aunque no hubo un "ganador" oficial de la Prueba de Turing en ese año, los chatbots y sistemas de IA de la época (como ELIZA o A.L.I.C.E.) ya podían engañar a algunos usuarios durante conversaciones cortas, especialmente si el evaluador era inexperto y la conversación se mantenía en un dominio limitado. El 30% de probabilidad para un evaluador inexperto en 5 minutos parece plausible para esa época.

- **En el mundo actual (2023-2024) con LLMs como ChatGPT** creo que los modelos son capaces de mantener conversaciones mucho más coherentes, fluidas y contextualmente relevantes que cualquier IA del 2000. Esto hace que muy probablemente un modelo actual pueda engañar a un evaluador inexperto durante una conversación de cinco minutos.