

Sistemas de Aprendizaje Automáticos

UT1: Inteligencia Artificial Fuerte y Débil

1. ¿Qué es la Inteligencia Artificial? Introducción.

Existen muchas definiciones válidas de inteligencia artificial, una que se considera sencilla y completa es aquella que afirma que es la disciplina que estudia la forma en la que se puede hacer que las máquinas piensen. En otras palabras, es la disciplina que se dedica a crear máquinas que son capaces de realizar tareas que, cuando las hacen los humanos, implican el uso de su inteligencia.

Una primera definición bastante común que podemos encontrar para la Inteligencia Artificial es: "Habilidad para aprender y resolver problemas, llevada a cabo por una máquina o software" En general, la mayoría de los expertos coinciden en que es la simulación de procesos de inteligencia humana por parte de máquinas, especialmente sistemas informáticos.

Estos procesos incluyen:

- 1.- El aprendizaje a través de la adquisición de información y reglas para el uso de la información.
- 2.- El razonamiento usando las reglas para llegar a conclusiones aproximadas o definitivas.
- 3.- La autocorrección.

Una definición más concreta y consensuada podría ser:

"La inteligencia artificial es la inteligencia llevada a cabo por máquinas. En ciencias de la computación, una máquina «inteligente» ideal es un agente flexible que percibe su entorno y lleva a cabo acciones que maximicen sus posibilidades de éxito en algún objetivo o tarea"

Se suele considerar que los primeros trabajos en el campo de la inteligencia artificial fueron los llevados a cabo por McCulloch y Pitts, quienes en la década de 1940 y desarrollaron un modelo automático de la neurona, modelo de neurona artificial. Estos autores se apoyaron en

los conocimientos que en aquel momento había disponibles de lógica, así como en los trabajos de Turín relativos a computación.

La hipótesis de que las máquinas puedan simular el pensamiento humano se denomina la hipótesis de la inteligencia artificial fuerte. Si se consigue que esas máquinas no sólo simulen la forma de actuar y de comportarse del ser humano, sino que realmente piensen, se habría logrado la inteligencia artificial fuerte.

Si bien hoy en día existen numerosas aplicaciones que realizan tareas que simulan de manera parcial el comportamiento de los seres humanos, y por tanto se podría considerar que se ha alcanzado la inteligencia artificial débil.

En la actualidad la evolución de la Inteligencia Artificial comprende un campo tan amplio, con tantas ramas, que es complicado poder atender a una única clasificación. De hecho, encontramos diferentes clasificaciones según la diferente visión con la que se aborda dicha tecnología. Las hay más filosóficas, más técnicas, y según su aplicación.

Test de Turing.

En un artículo publicado en 1950 el matemático Alan Turing propuso una prueba encaminada a verificar desde un punto de vista operacional la inteligencia de una máquina. En opinión de Turín la pregunta correcta que se debe formular no era si las máquinas pueden pensar, sino que resulta de mucho mayor utilidad preguntarse si las máquinas podrían pasar una prueba, denominada desde entonces test de Turín, y que consiste en hacer que esta mantenga una conversación a través de un texto que se presenta en una pantalla con un interlocutor humano durante cierto intervalo de tiempo.

El interlocutor tiene que discernir si la conversación que está manteniendo es con un programa cómo una persona; el programa se considera que ha pasado la prueba si es capaz de engañar a su interlocutor durante un cierto periodo de tiempo.

Un ordenador o programa que pase la prueba de Turing debe poner de capacidad para el procesamiento del lenguaje natural, capacidad de almacenamiento de razonamiento para responder a preguntas a través de la información que tiene almacenada, capacidad de adaptación al desarrollo de una conversación, etc.

En la actualidad, la inteligencia artificial ha hecho que los humanos se comuniquen con los chatbots, manteniendo conversaciones que hace un par de décadas parecían imposibles.

2. Objetivos y enfoques de la IA

Los objetivos principales de la IA incluyen la *deducción y el razonamiento*, la *representación del conocimiento*, la *planificación*, el *procesamiento del lenguaje natural (NLP)*, el *aprendizaje*, la *percepción* y la *capacidad de manipular y mover objetos*.

Los objetivos a largo plazo incluyen el logro de la Creatividad, la Inteligencia Social y la Inteligencia General (a nivel Humano).

Cuatro enfoques principales:

- Sistemas que se **comportan como humanos**: la idea es desarrollar máquinas capaces de realizar funciones para las cuales se requeriría un humano inteligente.
- Sistemas que **piensan como humanos**: las máquinas piensan como humanos en el sentido más literal; es decir, que tengan capacidades de toma de decisiones, resolución de problemas, aprendizaje, etc.
- Sistemas que **piensan racionalmente**: descubrir los cálculos que hacen posible percibir, razonar y actuar; es decir, encontrar las leyes que rigen el pensamiento racional.
- Sistemas que se **comportan racionalmente**: diseñar agentes inteligentes.

Dentro de las distintas clasificaciones que vamos a ver, hay algunas tipologías que están definidas "en teoría" aunque aún no exista ninguna aplicación IA de esa clase.

Tipos de Inteligencia Artificial según distintas clasificaciones:



Fran Bartolomé - Elaboración propia (CC BY-SA)

3. Inteligencia Artificial débil y fuerte.

Aunque, en un principio, el objetivo del desarrollo de la inteligencia artificial era construir una máquina capaz de seguir las leyes de la lógica del pensamiento a la perfección y sin equivocarse, pronto se vio que esto era algo casi utópico. Tratar de programar todas las reglas necesarias, con sus propias correlaciones y generalizaciones, era una misión inabarcable y que siempre acababa en alguna incoherencia o fallo.

Los sistemas expertos, marcaron la clara separación entre dos tipos de inteligencia artificial. La inteligencia artificial fuerte o general, similar a la mente de los seres humanos, y la inteligencia artificial débil o estrecha, capaz de realizar con gran precisión una tarea concreta, dentro de un campo de especialización.

Toda la evolución que ha habido en el ámbito del aprendizaje automático profundo se ha desarrollado en el plano de la inteligencia artificial débil. Aun así, cada cierto tiempo, se van produciendo hitos y avances que vuelven a rescatar la idea de contar con una inteligencia artificial fuerte como la que aparece en las películas y novelas de ciencia ficción, capas de relacionarse y servir a los humanos en modo multitarea.

Aunque la mayoría de alusiones a las diferencias entre ambas se refieren casi siempre a la capacidad de imitación de la inteligencia humana, esto no es del todo correcto. El campo de la inteligencia artificial actualmente no se centra en emular las capacidades o el comportamiento humanos, a no ser que nos refiramos a robots de tipo humanoide o videojuegos, y ni siquiera sería del todo exacta esta concepción en esos casos. La clave de la inteligencia artificial actual es la evolución y avance de la automatización.

La primera clasificación de la Inteligencia Artificial que vamos a ver se basa en qué tipo de tareas nos ayuda a resolver o ejecutar.

Cuando hablamos de tarea a resolver nos referimos, por ejemplo, a si pretendemos que la IA sea capaz de jugar al ajedrez en un ordenador o si pretendemos que sea capaz de gestionar una cocina sin intervención humana (desde el abastecimiento de alimentos, procesado, decisión de qué cocinar en cada momento, limpieza y mantenimiento de utensilios, resolver imprevistos o accidentes...). Hay tareas sencillas, concretas, y puntuales que son relativamente sencillas de programar, mientras que hay tareas complejas en las que, además influyen muchos factores exteriores (sentimientos, contexto, moral, ética, creencia religiosa...).

Seguro que eres capaz de ir intuyendo que nuestros programas de Inteligencia Artificial actuales son más bien de los que resuelven tareas en un entorno muy concreto y acotado, mientras que hay que irse a las películas o series de ficción para poder hablar de algún caso de Inteligencia Artificial capaz de actuar en escenarios complejos, cambiantes y con "conciencia".

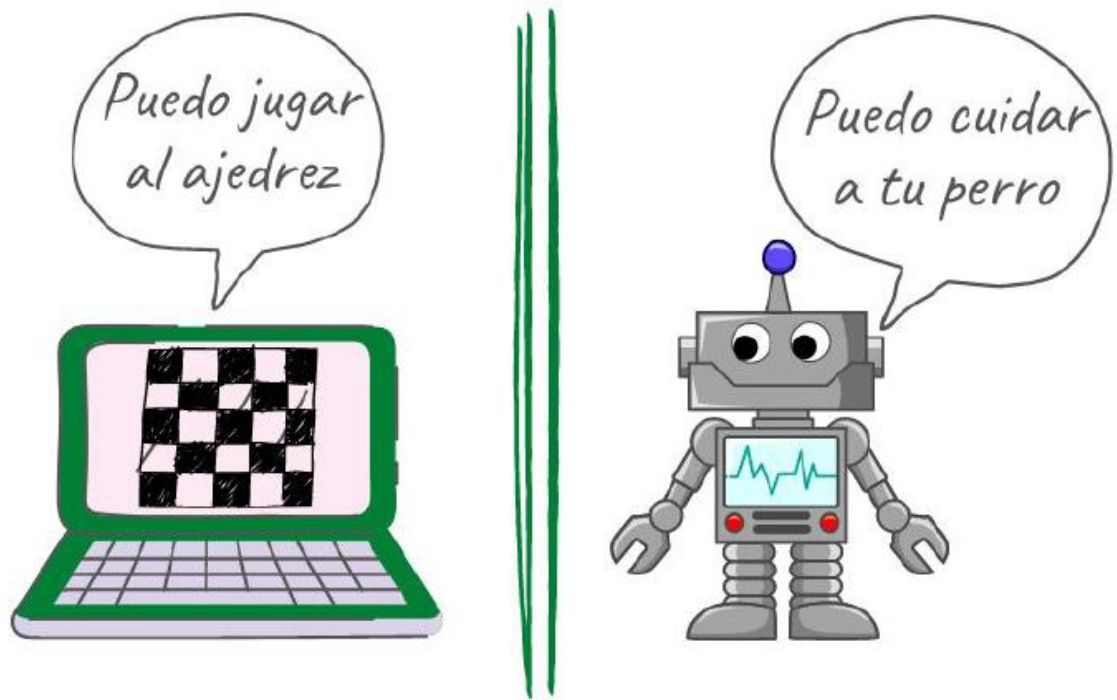
Esta clasificación según tareas es una aproximación bastante simple, con dos opciones:

- ✓ IA Débil.
- ✓ IA Fuerte.

En la actualidad aún no hemos sido capaces de desarrollar ninguna IA Fuerte. Todo lo que conocemos en este momento quedaría dentro de la clasificación de IA Débil.

Vamos a ver qué características tienen (o tendrán) las IA según esta clasificación.

Tipos de IA según las TAREAS que pueden resolver:



Fran Bartolomé - Elaboración propia (CC BY-SA)

IA Débil

La IA **débil** intenta construir sistemas para ejecutar eficientemente ciertas tareas intelectuales que un ser humano puede realizar. Este precepto conduce a que el desarrollo de IA débil no requiere implementar exhaustivamente la inteligencia humana para obtener las funcionalidades del sistema deseado. La IA débil implementa agentes que abordan tareas en un específico entorno.

El desarrollo de la IA ha ido evolucionando por una visión débil de la misma, marcada por hechos clave como la aparición de las redes neuronales o la introducción del término machine learning (aprendizaje automático) en 1959 por Arthur Samuel.

No obstante, la investigación en IA para simular el comportamiento humano ha contribuido a crear nuevas disciplinas como la Computación del Conocimiento focalizada en construir máquinas que tengan habilidades de razonamiento análogas al cerebro humano. La Figura 1.1 sintetiza las relaciones entre estos conceptos. La relación fundamental es que las técnicas de aprendizaje automático constituyen una parte de la IA débil.

La IA débil ha tenido numerosos éxitos que han llegado a la opinión pública. La empresa IBM construyó el ordenador Deep Blue que derrotó al campeón del mundo de ajedrez el 11 de mayo de 1997. Desarrollaron el ordenador Watson que podía procesar lenguaje natural y aprender sin supervisión de documentos.

En febrero de 2011, Watson ganó a dos campeones previos en el concurso televisivo estadounidense Jeopardy, que trata de preguntas de conocimientos sobre una variedad diversa de temas. Este hecho mostró que la capacidad de Watson de entender las preguntas formuladas en lenguaje natural, buscar en bases de datos los hechos relevantes y hallar la respuesta correcta.

Otras empresas como Google emplean redes neuronales profundas para tareas de reconocimiento de imágenes. La red Google Net fue la ganadora del concurso 2014 ImageNet Large-Scale Visual Recognition Challenge. En el año 2021, DeepMind, una empresa comprada por Google, es capaz de predecir con gran exactitud la estructura de las proteínas que forman un ser humano. Esta información puede resultar esencial para el desarrollo de tratamiento de las enfermedades. Esta herramienta de IA va a cambiar radicalmente la investigación biológica entre otros motivos porque acelera el proceso.

Los métodos tradicionales requieren meses o incluso años para dilucidar la forma de una proteína y este sistema lo hace en minutos, aunque con ciertos errores. J. C. R. Licklider fue un pionero de la IA, pronosticando en la década de los años 60 del siglo pasado que la aparición de la IA fuerte no iba a ser inminente y requeriría un periodo interino de entre 10 y 500 años en el cual los humanos y los ordenadores existiría una relación simbiótica en el que las máquinas ayuden de forma efectiva el proceso de pensamiento. Argumentó que trabajarían con los humanos, existiendo como organismos diferentes, pero viviendo juntos en una asociación íntima que mejoraría ciertas capacidades del pensamiento humano. El reto no el de construir máquinas que simulen la mente humana sino entender el proceso de conocimiento humano y desarrollar máquinas que ayuden/realicen aquellos aspectos de la solución de problemas más difíciles o tediosos para el ser humano.

Desde el punto de vista de esta clasificación (por tarea a resolver), las características de la Inteligencia Artificial débil son:

- Ya existen en la vida real: Como hemos comentado, los asistentes virtuales, programas como Watson o Alpha Go que vimos en la unidad anterior.
- Se orientan a resolver problemas muy concretos: El programa que "sabe" jugar al Go, no sabe hacer otra cosa. Ni tiene posibilidades de aprender a jugar a otra cosa, por muy similar que sea.
- Son reactivas: No tienen iniciativa, es necesario que se desencadene la acción que tienen programada para que se inicie su rutina. En el ejemplo del asistente virtual,



- tiene programado responder cuando le preguntas, y por tanto nunca tomará la iniciativa de ofrecerte nada sin que tú lo actives previamente.
- No son flexibles: Colapsan si se encuentran en un caso no previsto en su programación.
 - Quedan limitadas por lo que programa un humano: Es el humano quien programa lo que "tiene que pensar" la máquina. Si el humano no programa deja sin considerar ciertas opciones o posibles situaciones, la IA nunca será capaz de suplirlo o aprenderlo sobre la marcha por sí misma.
 - Se programan con pocas redes neuronales: Hablaremos más adelante sobre las redes neuronales. Por el momento es suficiente entender que el nivel de computación compleja que requieren este tipo de Inteligencias Artificiales es menor que otros casos.
 - No razonan, solo computan: No tienen en cuenta ningún factor moral, contextual, circunstancial, emocional... que a un humano le haría reaccionar de manera diferente. La máquina está programada para alcanzar tal objetivo o funcionar de tal manera, y así lo hará sin "entender" lo que está haciendo. Por tanto: no tiene conciencia.
 - Aprenden a base de ejemplos: Necesita conocer muchos ejemplos de lo que tiene que hacer (la base de datos), con todas las variantes posibles. Por ejemplo, en la máquina que juega al Go, se la "entrenó" con 50 millones de partidas de dicho juego.
 - Son repetitivas: No se cansan nunca, son implacables, siempre la misma rutina.
 - No salen de su marco de trabajo: Y esto supone que pueden ocuparse de tareas mecánicas, repetitivas, "aburridas" para sustituir al humano mejorando rendimiento y precisión. Pero necesitará siempre una supervisión humana que vaya decidiendo cómo adaptar el programa a las cambiantes circunstancias.

Según se mire la Inteligencia Artificial débil podría llegar a ser peligrosa. Porque este tipo de IA, al no tomar en consideración todo un contexto amplio, ni seguir las reglas sociales, éticas, ... ejecuta las tareas para las que se le ha entrenado con **eficacia** y **contundencia**. No evalúa las consecuencias como lo hacemos los humanos, considerando un espectro amplio de efectos y relaciones. Por eso, es una opción incompleta, inestable y peligrosa si no se utiliza con prudencia, o si quien el programa pretende, precisamente, causar mal.

Ventajas de la IA Débil:

- ✓ Actualmente es relativamente sencilla de conseguir y desarrollar.
- ✓ Se controla más fácilmente y no presenta comportamientos demasiado inesperados.
- ✓ Permite arquitecturas modulares, con las que combinar distintos talentos o capacidades según las necesidades del entorno en el que se despliega.

Inconvenientes de la IA Débil:

- ✓ Es un sistema "ciego", que no cuenta con todo el contexto de la situación, por lo que puede mal-interpretar entradas inusuales.
- ✓ Es muy limitada; solo sirve para aplicaciones muy concretas.
- ✓ Suelen ser sistemas que no ofrecen una experiencia de usuario tan cercana o personalizada como la de un ser humano.

IA Fuerte

La IA **fuerte** es aquella inteligencia artificial que iguala o excede la inteligencia humana promedio. Este tipo de IA será capaz de realizar todas las tareas que un ser humano podría hacer. Hay mucha investigación en este campo, pero todavía no ha habido grandes avances.

La IA Fuerte es solo una inteligencia artificial hipotética, no existe de la forma que se plantea tendría que ser. Lo que sí hay es una fuerte investigación al respecto, intentando crear complejos algoritmos matemáticos que serían la base del sistema.

La Inteligencia Artificial fuerte (IAF) o general o (IAG), desde el punto de vista de la tarea a resolver, sería aquella que iguala o excede la inteligencia humana promedio. Sería capaz de realizar con éxito cualquier tarea intelectual del ser humano, teniendo en cuenta todos los factores y matices que pueden intervenir cuando una persona toma decisiones en cada momento mientras realiza una tarea.

No existe todavía ninguna máquina que se pueda considerar completamente IA fuerte. Salvo en algunas películas o series de televisión.

En comparación con la Inteligencia Artificial débil, las características de la fuerte son:

- No existe en la realidad

- Resolverán problemas abiertos: Deberían poder abarcar múltiples posibles tareas, distintas unas de otras (reparar una puerta, ir a recoger a los niños del colegio, regar las plantas, darte conversación...).
- Serán proactivas: En función de la misión u objetivo que tenga, y de las circunstancias, iniciará cualquier tipo de rutina sin esperar a que un humano se lo pida o esté pendiente.
- Serán flexibles: Podrán encontrar similitudes entre algo que conocen y algo que se le parezca un poco. Por ejemplo, aunque inicialmente solo haya sido programada para saber andar, será capaz de aprender a correr sin necesidad de intervenir en su programa.
- Se auto programarán: Serán capaces de detectar sus propios límites y se regularán a sí mismas para no excederlos.
- Usarán muchas redes neuronales: Y además podrán entrar en conflicto entre ellas en algunas ocasiones. Esto quiere decir que necesitarán una capacidad de almacenaje de información y cómputo que aún hoy no hemos llegado a alcanzar.
- Imitarán el comportamiento humano: Serán capaces de razonar, y, por tanto, de alcanzar algún tipo de consciencia.
- Aprenderán como las personas: Podrán recordar datos, observar nuevas situaciones y encontrar relaciones entre diferentes acciones. Esto quiere decir que, si saben jugar al ajedrez y "observan" el juego de las damas, podrán aprender a jugar a las damas basándose en lo que saben sobre jugar al ajedrez.
- Serán capaces de aprender nuevas tareas: Modificarán la tarea o cómo realizan la tarea para adaptarse a las circunstancias.
- Serán capaces de adaptarse a nuevos escenarios: Podrán adaptarse a cambios y nuevas situaciones para seguir cumpliendo su objetivo.

Este tipo de IA es la que sería capaz de analizar cualquier situación y deducir el conjunto de acciones más adecuado para dicha situación y contexto. Lo mismo sabría conducir un coche, que resolver una ecuación matemática o mantener una conversación sobre un tema concreto.

Aunque aún no existe este tipo de IA, todas las empresas e instituciones dedicadas a la investigación y desarrollo de IA están buscando formas de avanzar hacia este tipo de Inteligencia Artificial. De momento, al menos, se está trabajando en conseguir una IAF en el campo de los asistentes virtuales.

Sin duda uno de los ámbitos más ambiciosos de aplicar esta IAF (los asistentes virtuales humanoides) necesitan contar también con otras ramas científicas como son la robótica y mecatrónica.

4. Antecedentes del Machine Learning

La evolución de la inteligencia artificial se ha basado en tratar de ser capaz de actuar como los humanos, en primera instancia, y tratar de predecir el comportamiento más probable o deseable de un sistema en el planteamiento más actual. Para esto, las técnicas más prometedoras son las que utilizan modelos probabilísticos que, precisamente, son capaces de modelizar mejor los datos observados.

Podemos decir, que las técnicas de aprendizaje automático, básicamente, miran el problema desde fuera y configuran sus propias reglas para emular el comportamiento del sistema. De esta forma, se consigue una mayor tasa de acierto. Todo empieza por modelar probabilísticamente una relación entre causa y efecto, estableciendo un aprendizaje a partir de observaciones.

Estadística y probabilidad.

El Machine Learning se basa en las estadísticas y el desarrollo de la informática de los últimos 70 años. Arthur Samuel, pionero estadounidense en el campo de la inteligencia artificial, acuñó el término «Aprendizaje automático» en 1959 durante su trayectoria en IBM.

Como campo de estudio, el Machine Learning comenzó a florecer en la década de 1990. Andrew Ng (profesor asociado del departamento de Ciencias de la Computación y del departamento de Ingeniería Electrónica de la Universidad de Stanford) resume el proceso de desarrollo del Machine Learning en la siguiente premisa:

“La escala (la disponibilidad de datos y la escala computacional) impulsa el progreso del aprendizaje automático”.

Larry Wasserman es profesor del departamento de Estadística y del departamento de Aprendizaje Automático de Carnegie Mellon, una de las principales universidades de estadística y Machine Learning a nivel mundial. Cuando se le preguntó sobre las diferencias entre la Estadística y el Machine Learning, respondió: “Están preocupados por la misma pregunta: ¿cómo aprendemos de los datos?”

Pero concretando más, explica: “Las estadísticas enfatizan la inferencia estadística formal (intervalos de confianza, pruebas de hipótesis, estimadores óptimos) en problemas de baja dimensión (conjuntos de datos más pequeños) y el aprendizaje automático se centra más en hacer predicciones precisas de alta dimensión (grandes conjuntos de datos)”.

Algunos como Robert Tibshirani (estadístico y experto en aprendizaje automático de Stanford) incluso llama al aprendizaje automático «estadísticas glorificadas”.

Ambos métodos se centran en extraer conocimiento o ideas de los datos, pero sus métodos se ven afectados por sus diferencias culturales inherentes. La razón principal de que estos temas sean efectivamente los mismos es que cubren casi exactamente el mismo material y usan casi exactamente las mismas técnicas.

Además, cada paso en un proyecto de Machine Learning requiere el uso de un método estadístico. Tanto para comprender los datos utilizados en el entrenamiento de un modelo de aprendizaje automático como en la interpretación de los resultados obtenidos tras probar diferentes modelos de aprendizaje automático, se requieren métodos estadísticos.

Bibliografía:

Fuente de información: Ministerio de Educación y Formación Profesional

